Accurate parameter optimization leads to predictive dynamical models for systems biology

Angiotensin signaling model: Data Fitting, Convergence and Identifiability issues

T. Bourquard[1], M. A. Ayoub[1], D. Heitzler[1], N. Hansen[2], J. Azé[3], R. Yvinec[1], P. Crépieux[1], E. Reiter[1] & A. Poupon[1]

[1] BIOS group, UMR7247, INRA Tours
[2] LRI, UMR8623, CNRS, Orsay
[3]LIRMM, UMR5506, CNRS, Montpellier

Signaling Biology

Modeling and data fitting framework

Optimization for parameter estimation

Results on the angiontensin signaling pathway

Outline

Signaling Biology

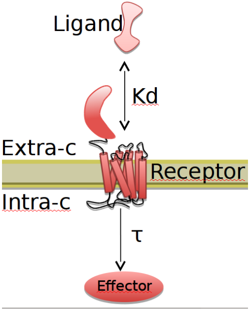
Modeling and data fitting framework

Optimization for parameter estimation

Results on the angiontensin signaling pathway

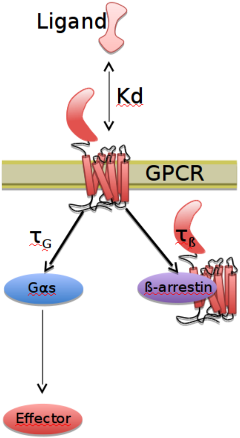
General issues

- A **Ligand** binds a **receptor** in the cell surface and leads to a signal.
- The bound receptor-ligand complex leads to a **cascade of reactions** (enzymatic catalysis, phosphorylation,...) up to some effector molecule that leads to a cellular response.



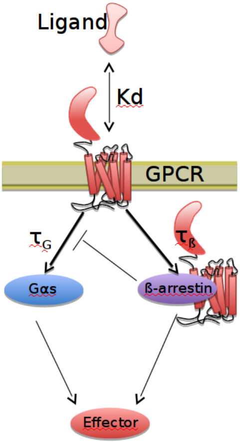
General issues

- A Ligand binds a receptor and signals.
- The bound receptor-ligand complex leads to a cascade of reactions.
- **G Protein Coupled Receptor** (GPRC) : Family of receptor, widely targeted by drugs.
- Two main pathways : **G protein** pathway and β-**arrestin** (signal vs internalization)



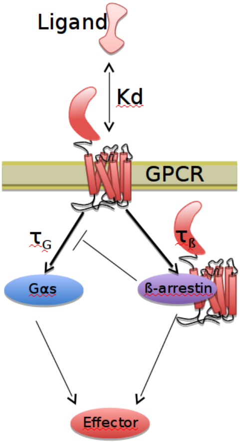
General issues

- A Ligand binds a receptor and signals.
- The bound receptor-ligand complex leads to a cascade of reactions.
- G Protein Coupled Receptor (GPRC)
- Two pathways : G protein and β-arrestin.
- More complex issues : β-arrestin induced pathway leads to a **different signal on the same effector**.



General issues

- A Ligand binds a receptor and signals.
- The bound receptor-ligand complex leads to a cascade of reactions.
- G Protein Coupled Receptor (GPRC)
- Complex interactions between G protein and β-arrestin pathways.

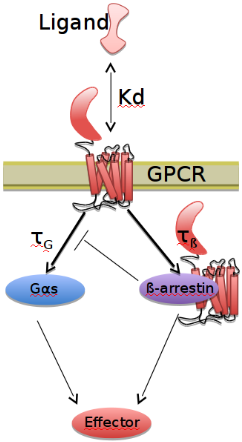


Drug discovery

- Signaling through one pathway and not another one : **Bias** (synthetic hormone, mutant receptor, small molecules...)

General issues

- A Ligand binds a receptor and signals.
- The bound receptor-ligand complex leads to a cascade of reactions.
- G Protein Coupled Receptor (GPRC)
- Complex interactions between G protein and β-arrestin pathways.



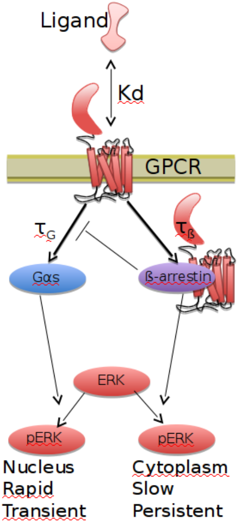
Computational Modeling

- Help **deciphering** the intricate effect of each pathway.
- **Quantify** the precise effect of a specific couple Ligand-Receptor.

GPCR signaling through ERK phosphorylation

The extracellular signal-regulated kinase ERK is activated both by the G protein and the β-arrestin pathway but (*Ahn et al. J Biol Chem (2004)*) :

- The **spatial distribution** are distinct.
- The **kinetics** are distinct.

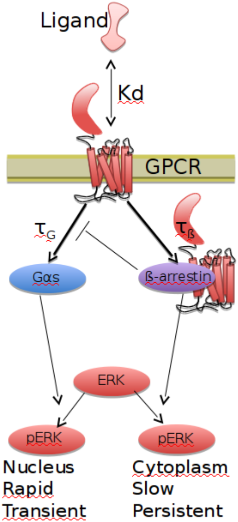


GPCR signaling through ERK phosphorylation

The extracellular signal-regulated kinase ERK is activated both by the G protein and the β-arrestin pathway but (*Ahn et al. J Biol Chem (2004)*) :

- The spatial distribution are distinct.
- The kinetics are distinct.

Transient and sustained ERK activation have been shown to regulate cell fates such as **growth** and **differentiation**.(*Sasagawa et al. Nat Cell Biol (2005)*)

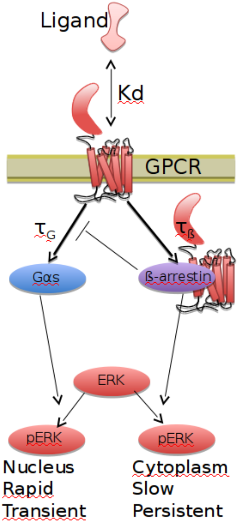


GPCR signaling through ERK phosphorylation

The extracellular signal-regulated kinase ERK is activated both by the G protein and the β-arrestin pathway but (*Ahn et al. J Biol Chem (2004)*) :

- The spatial distribution are distinct.
- The kinetics are distinct.

β-arrestin 2 dependent ERK pathway can be activated **independently** of G proteins with a mutant receptor (*Wei et al. PNAS (2004)*).



Case study : Angiotensin receptor

- **Angiotensin** II type 1A receptor (AT1AR) transfected in cultured human embryonic kidney (HEK 293 cells).
- **ERK phosphorylation** data : Phosphorylated ERK in immunoblots, quantified by densitometry (*Kim et al. PNAS 2005*)
- **DAG** accumulation and **PKC** activity data, measured in real time by FRET sensors.
- **Four perturbed** conditions in addition to control :
  - β-arrestin 2 siRNA
  - G protein-coupled receptor kinases (GRK2/3 and GRK5/6) siRNA
  - PKC inhibitor.

Outline

Signaling Biology

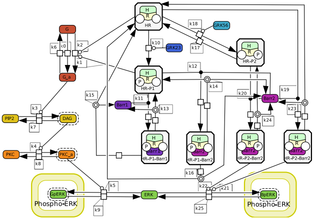
Modeling and data fitting framework

Optimization for parameter estimation

Results on the angiontensin signaling pathway

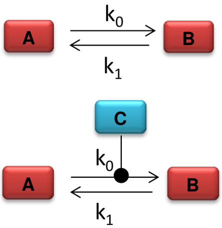
Overview of the methodology

- Starting point : **graph** of interaction of molecules (based on biological knowledge, literature)



Overview of the methodology

- Graph of interaction of molecules
- Law of mass-action : Ordinary Differential Equations (ODE) produce time-dependent trajectories, that depend on **parameters** (kinetic rate, initial condition)

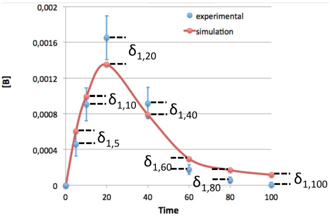


$$\frac{d[B]}{dt} = k_0[A] - k_1[B]\,.$$

$$\frac{d[B]}{dt} = k_0[A][C] - k_1[B]\,.$$

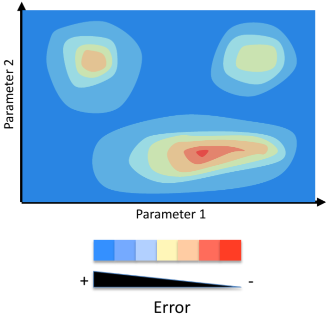
Overview of the methodology

- Graph of interaction of molecules.
- Law of mass-action : ODE.
- Quality of the model based on the introduction of a **cost function** (based on statistical error model, or heuristic arguments).



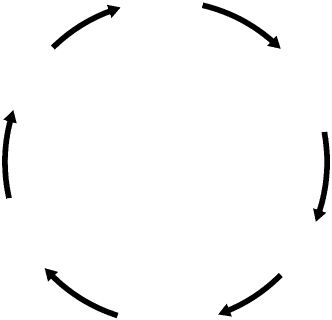
Overview of the methodology

- Graph of interaction of molecules.
- Law of mass-action : ODE.
- Cost function.
- **Optimization** of the cost function (Frequentist / Bayesian approach). Numerical search.



Overview of the methodology

- Graph of interaction of molecules.
- Law of mass-action : ODE.
- Cost function.
- Optimization.
- Validation data, prediction and experimental design...



Outline

Signaling Biology

Modeling and data fitting framework

Optimization for parameter estimation

Results on the angiontensin signaling pathway

The major difficulties are due to

- **Large dimension** of parameter space $(10 - 100)$ and state space (> 10).
- **Few** molecule concentrations measured, and not in absolute numbers.
- Large ODE's may be numerically costly to simulate if they are **stiff**.
- Parameters can be **non-identifiable** (non-convexity, presence of many local minima)

The major difficulties are due to

- **Large dimension** of parameter space $(10 - 100)$ and state space (> 10).
- **Few** molecule concentrations measured, and not in absolute numbers.
- Large ODE's may be numerically costly to simulate if they are **stiff**.
- Parameters can be **non-identifiable** (non-convexity, presence of many local minima)

A variety of methods can be employed : local/global methods and deterministic/stochastic methods, hybrid method.

- **Gradient descent** methods with many random initial start (D2D, *Raue A., et al. Bioinformatics (2015)*).
- **Hybrid** local and global method, based on heuristics (HYPE, T. Bourquard & A. Poupon)

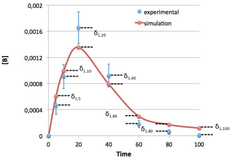
H.Y.P.E

Random set of parameters

Genetic algorithm

CMA-ES

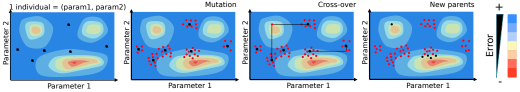
Final set of parameters



Error function

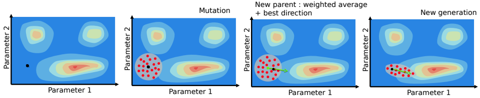
GENETIC ALGORITHM:

- A random set of *n* *parameter sets* is chosen, called **parents.**
- **Mutation** and **cross-over** produce new sets of parameters, as **offsprings**.
- The *n* *best* parameter sets giving the lowest errors are **selected**, and become the parents of next generation.



CMA-ES:

- Start with one set of parameter, called **parent**.
- Local **mutations** (gaussian) produce **offsprings.**
- The next parent is the **average** of the children weighted by their errors.
- Successive steps have **memory**, mutation are made along the direction of the previous steps. (Hansen *et al.* 2003)



A random set of parameters is optimized using Genetic algorithm. The resulting parameter set is optimized using CMA-ES. This operation is repeated until a sufficient number of parameter sets giving small errors is obtained.

Critical assessment of methods

How to judge different method ? How to asses the quality of a fit ?

- **Toy models** with *in silico* simulated data / Benchmark models.
- Absolute value of the **error function**.
- **Speed** of the algorithm.
- **Convergence curve** (number of best error function value over number of runs/function evaluation.
- Robustness of the minima.

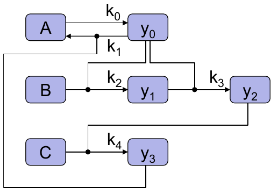
Critical assessment of methods

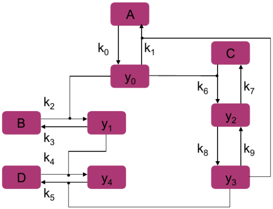
How to judge different method ? How to asses the quality of a fit ?

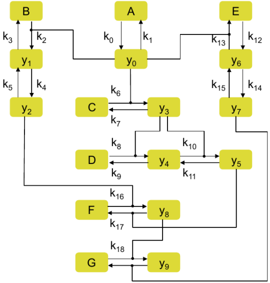
Why/How to deal with Non-Identifiability (NI) ?

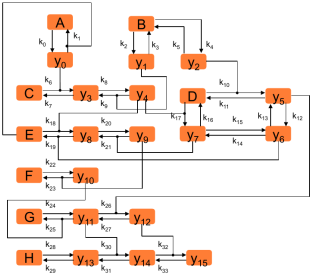
- It slows down the numerical search and leads to **unreliable results**.
- **Theoretical** NI : reduction / algebraic relations.
- **Numerical** NI : distinguish between **structural** and **practical** NI. Calculate sensitivity and one-dimensional profile likelihood.

Toy models

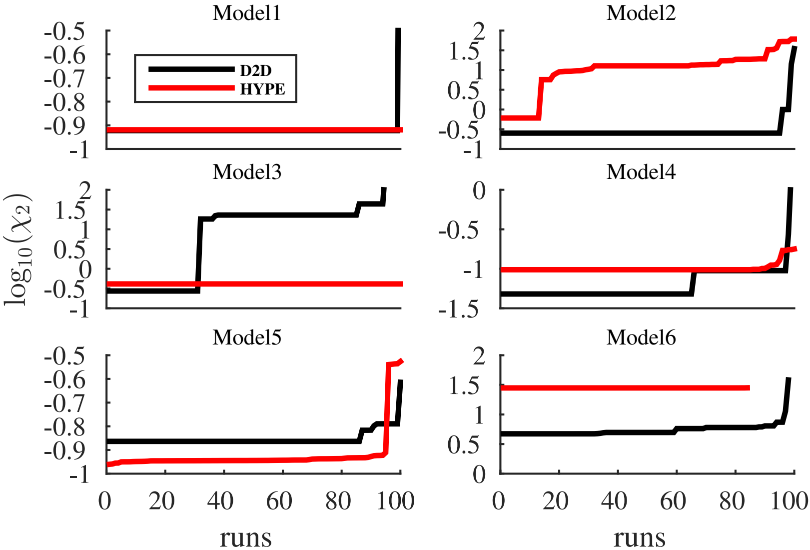

A
k_0
y_0
k_1
B
k_2
y_1
k_3
y_2
C
k_4
y_3


A
k_0
k_1
C
y_0
k_6
k_7
k_2
B
y_1
y_2
k_3
k_4
k_8
k_9
D
y_4
y_3
k_5


B
A
E
k_3
k_2
k_0
k_1
k_{13}
k_{12}
y_1
y_0
y_6
k_5
k_4
k_6
k_{15}
k_{14}
y_2
C
y_3
y_7
k_7
k_8
k_{10}
D
y_4
y_5
k_9
k_{11}
k_{16}
F
y_8
k_{17}
k_{18}
G
y_9


A
B
k_0
k_1
k_2
k_3
k_5
k_4
y_0
y_1
y_2
k_6
k_8
k_{10}
C
y_3
y_4
D
y_5
k_7
k_9
k_{11}
k_{18}
k_{20}
k_{17}
k_{16}
k_{15}
k_{13}
k_{12}
E
y_8
y_9
y_7
y_6
k_{19}
k_{21}
k_{14}
k_{22}
F
y_{10}
k_{23}
k_{24}
k_{26}
G
y_{11}
y_{12}
k_{25}
k_{27}
k_{28}
k_{30}
k_{32}
H
y_{13}
y_{14}
y_{15}
k_{29}
k_{31}
k_{33}

Toy models : HYPE gives comparable results to high quality optimization methods



Outline

Signaling Biology

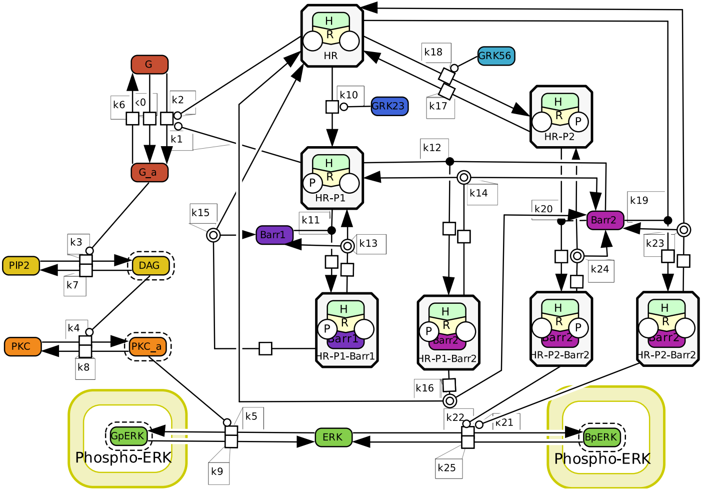
Modeling and data fitting framework

Optimization for parameter estimation

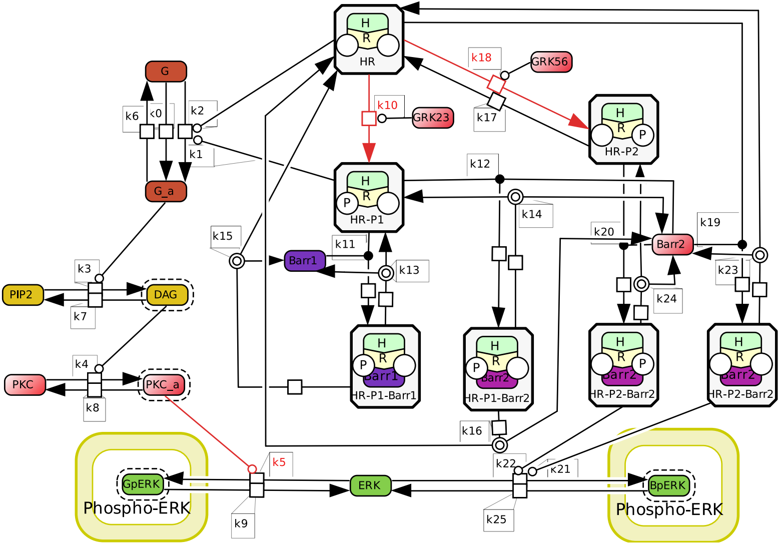
Results on the angiontensin signaling pathway

Full model (Heitzler el al. MSB 2012) : 3 pathways

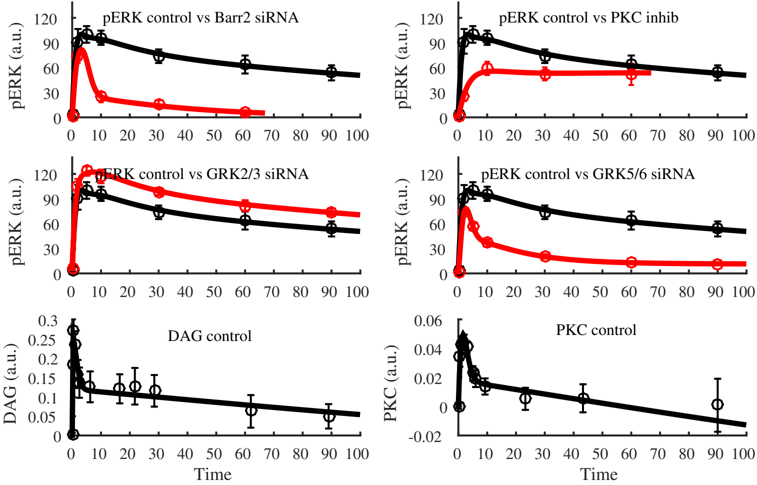
A model that fit the data : role of **G protein-coupled receptor kinases** (GRK) in the balance of signals.



Full model (Heitzler el al. MSB 2012) : Control + 4 pert.



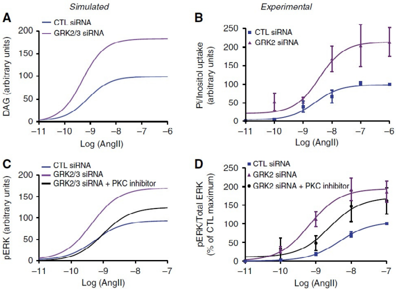
The model fits the data...

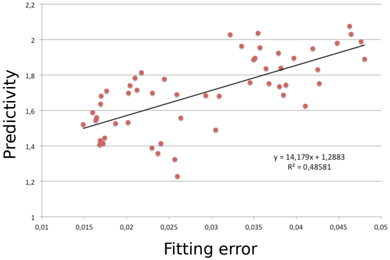


In black : control experiments.

In Red : perturbated experiments.

The model correctly predicts some validation data...

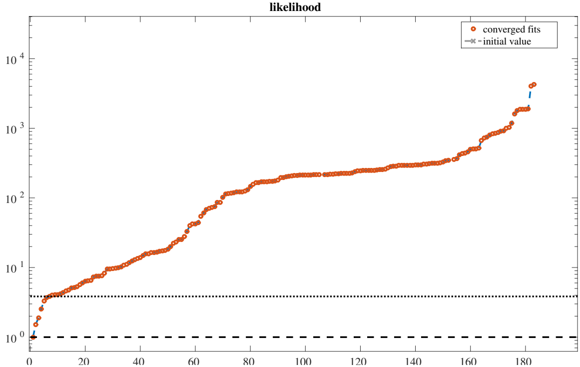




Remark

Good correlation between error and prediction.

…but the convergence is poor, and the identifiability a serious issue !

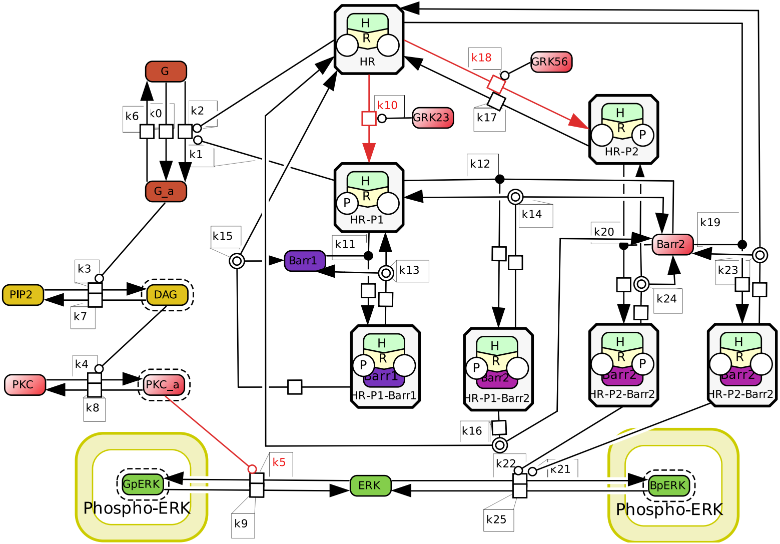


Critical assessments :

- Convergence curve. Identifiability of parameters.
- Parsimonious use of parameters. Model selection.

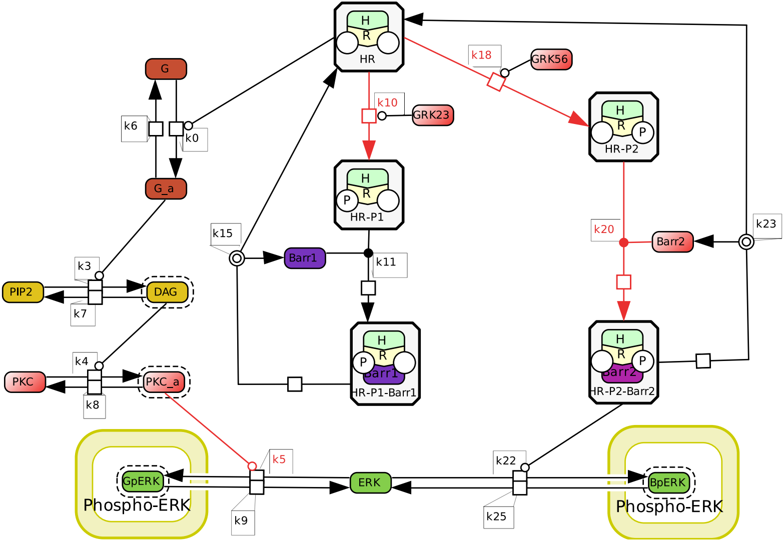
Model reduction

From **50** parameters...



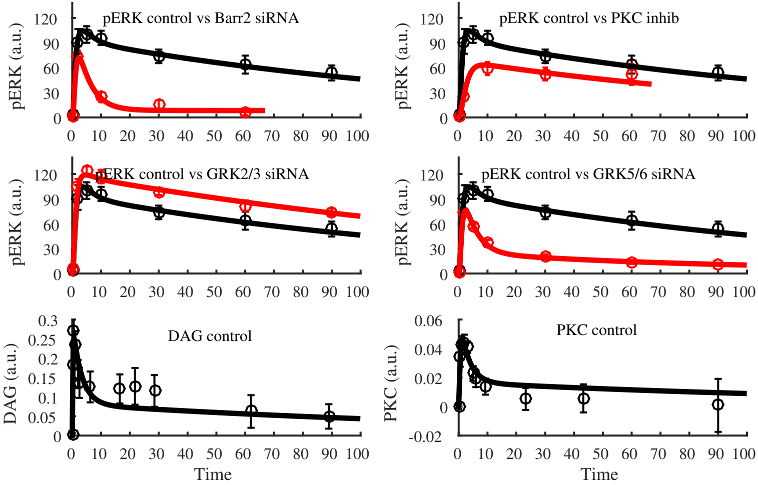
Model reduction

...to **22** parameters !



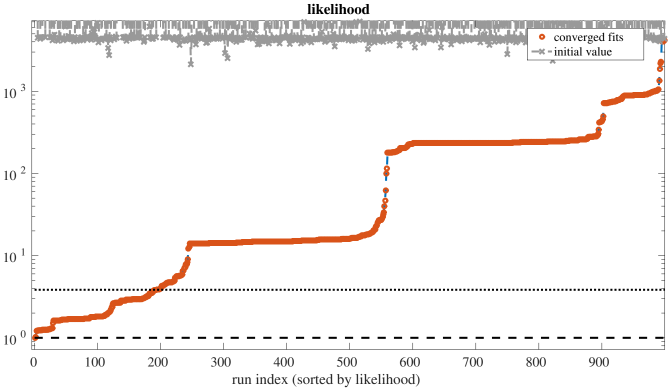
Model reduction

- The reduced model still **fit** (reasonably) well...



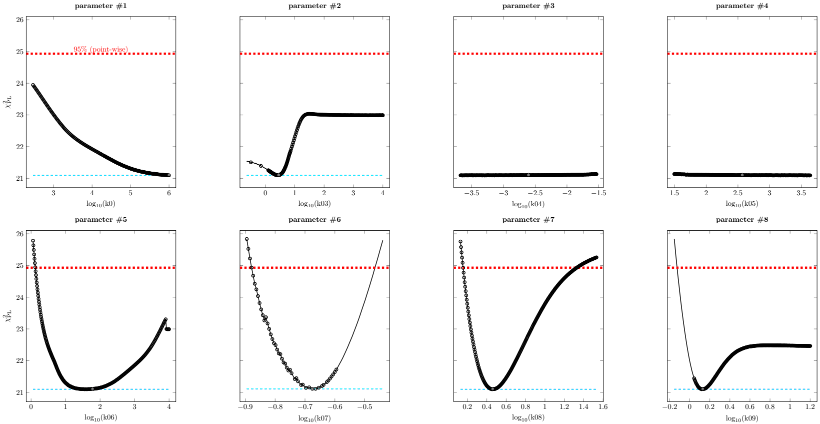
Model reduction

- The reduced model still fit well...
- and the **convergence** is **better**...



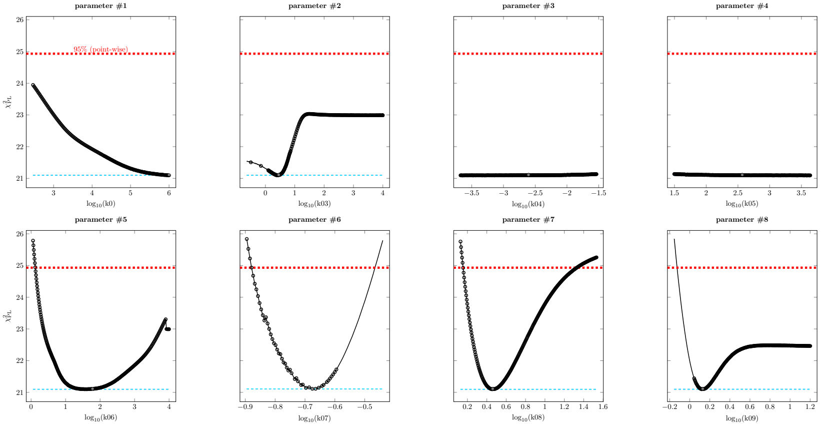
Model reduction

- The reduced model still fit well...
- and the convergence is better...
- but the **identifiability** is still **poor** !

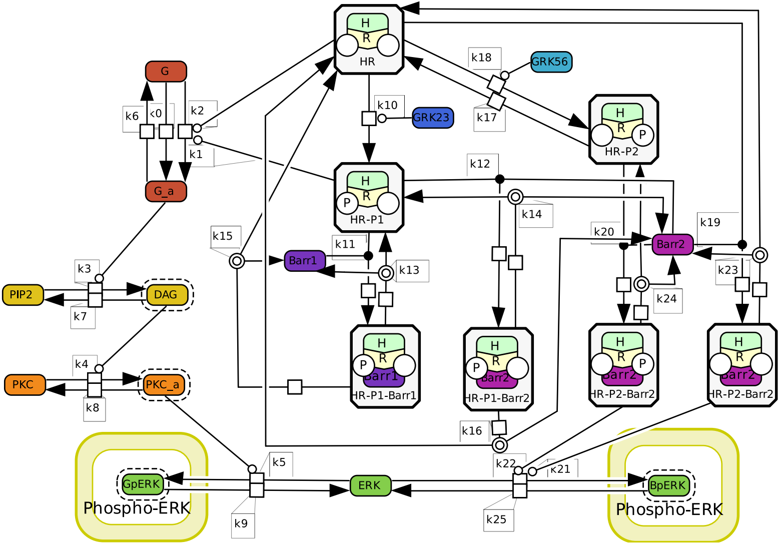


Model reduction

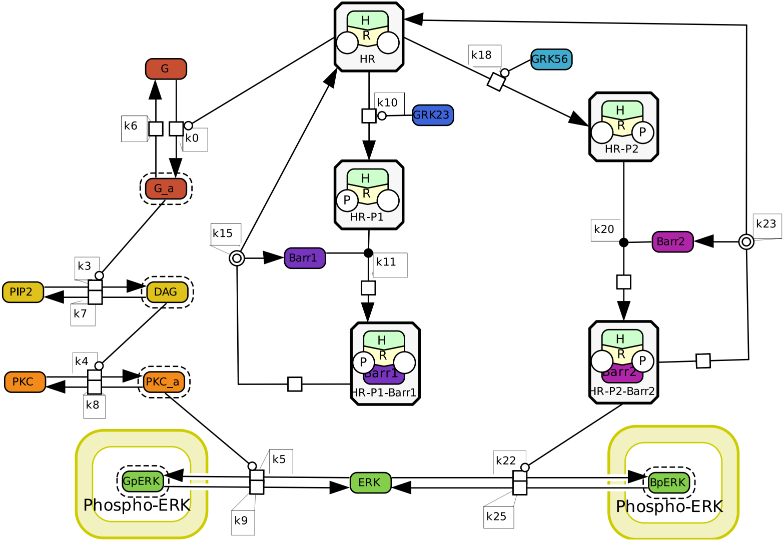
- The reduced model still fit well...
- and the convergence is better...
- but the identifiability is still poor ! Let's **reduced further** ?



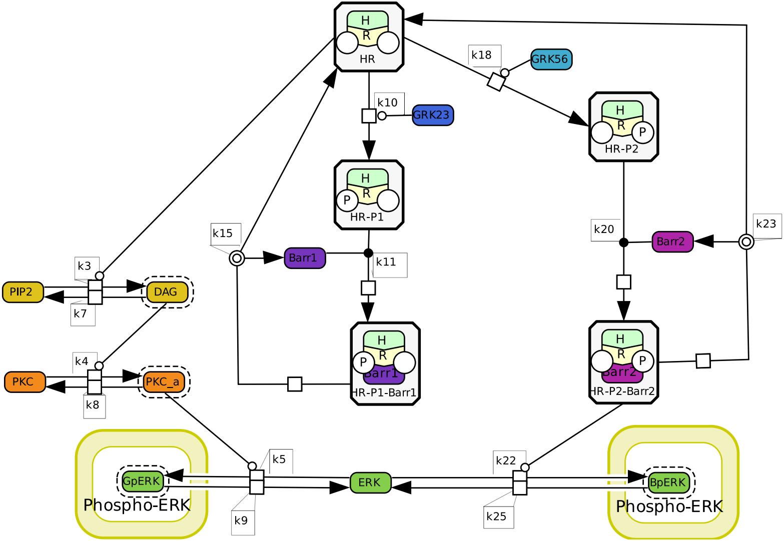
Systematic reduction



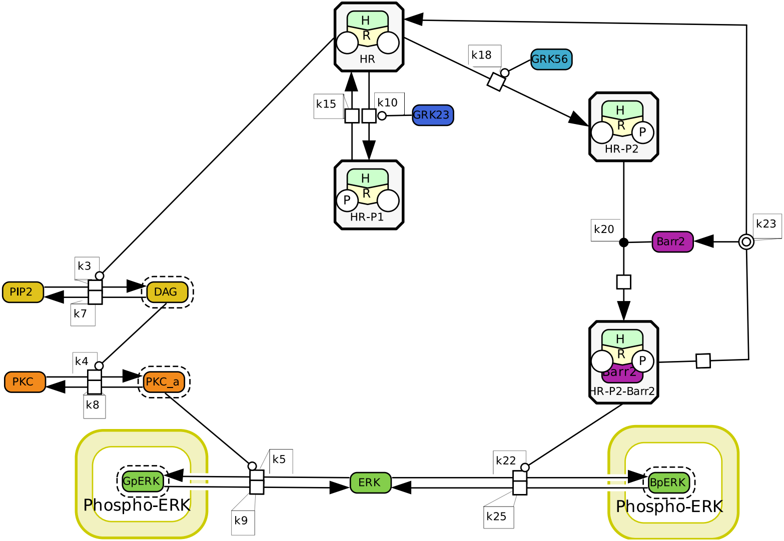
Systematic reduction



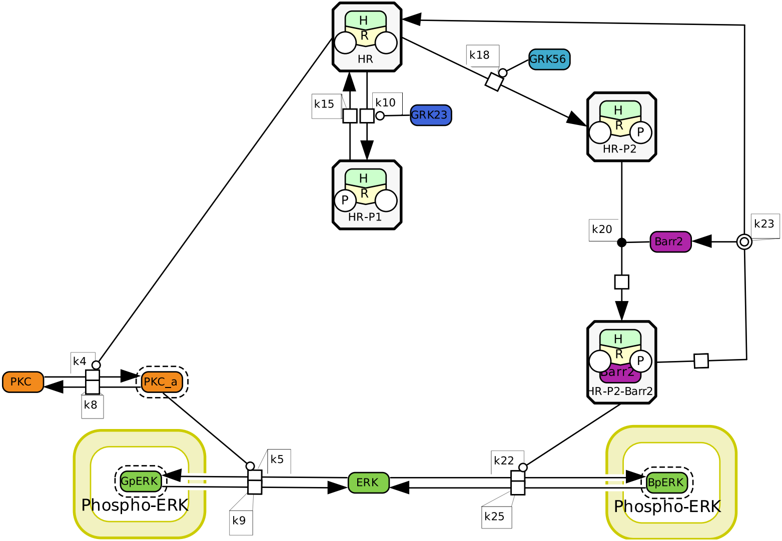
Systematic reduction



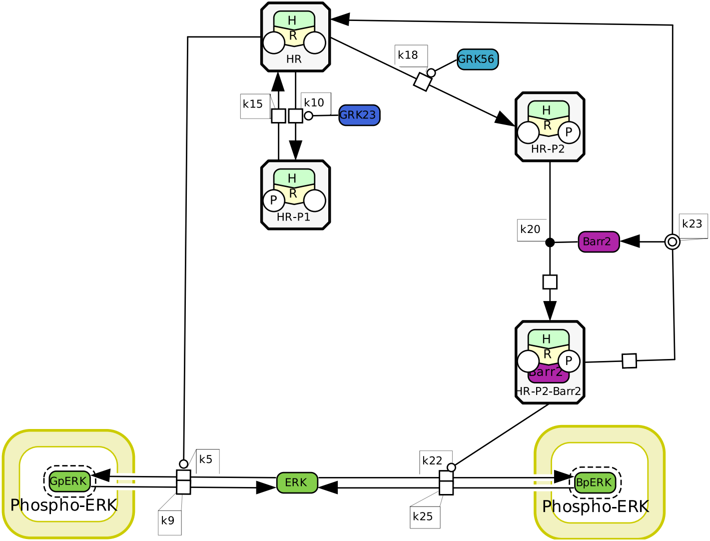
Systematic reduction



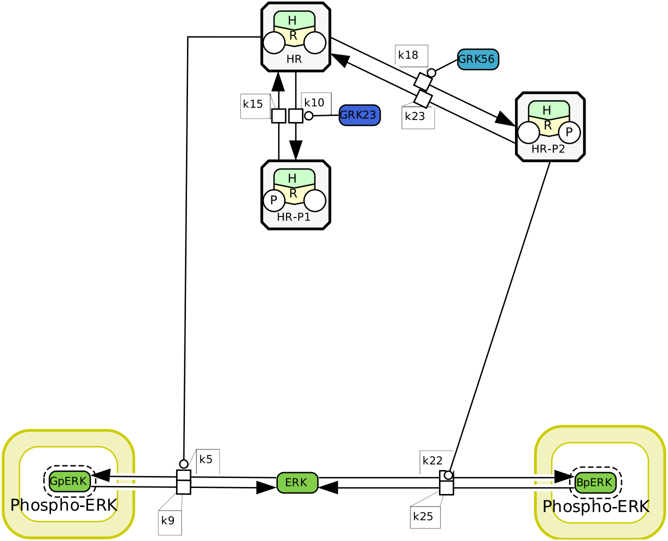
Systematic reduction



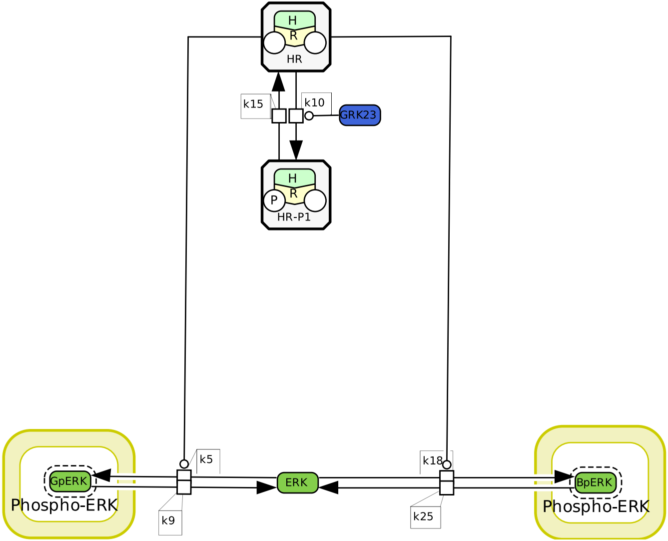
Systematic reduction



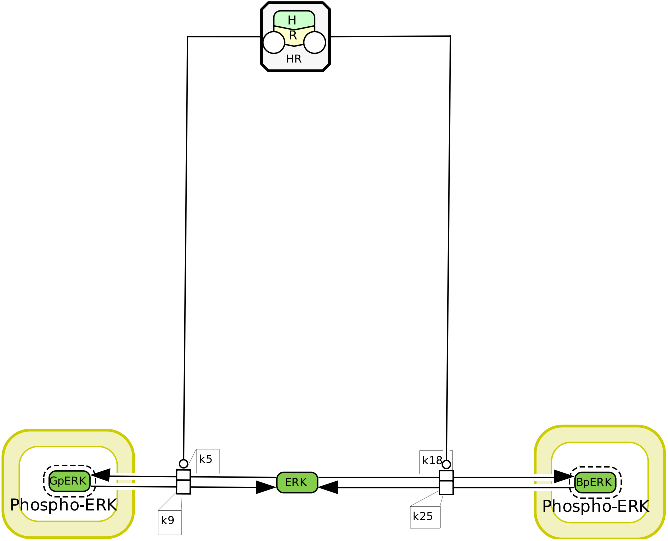
Systematic reduction



Systematic reduction



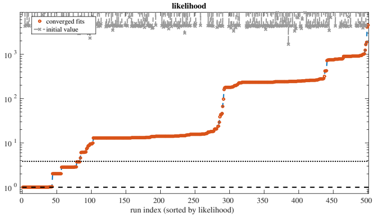
Systematic reduction

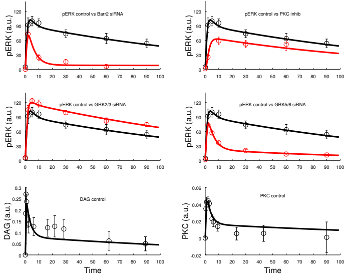


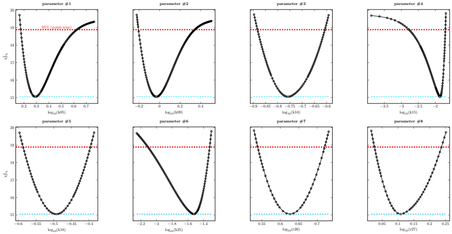
Results

With model selection criteria :

- a model with **19** parameters
- **good convergence** properties (10% of runs reached the optima)
- most parameters are **identifiable**

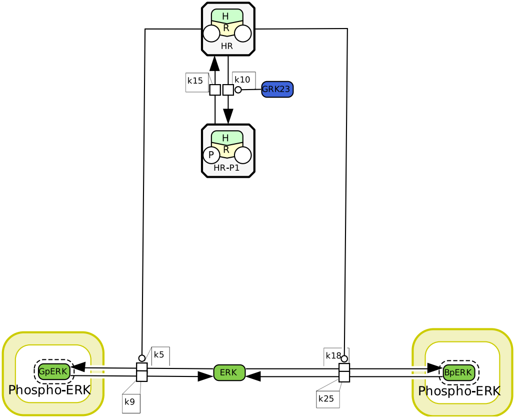






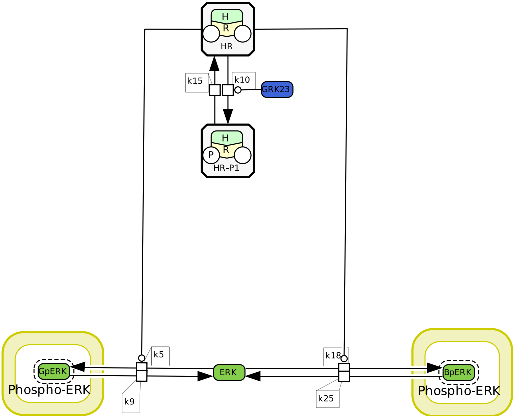
What do we learn (so far) ?

- The **three pathways** are a **necessary** condition to reproduce the data.



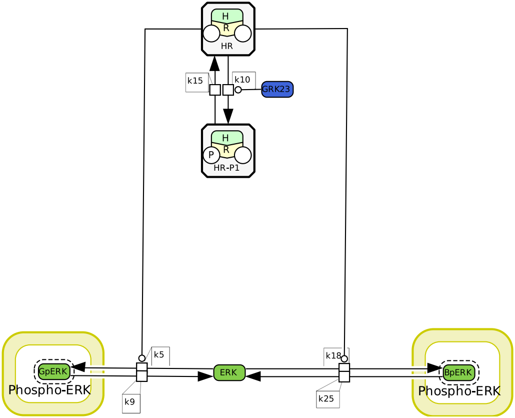
What do we learn (so far) ?

- Three pathways are necessary.
- An internalization pathway, **independent** of the β-arrestin signaling pathway, is mandatory.



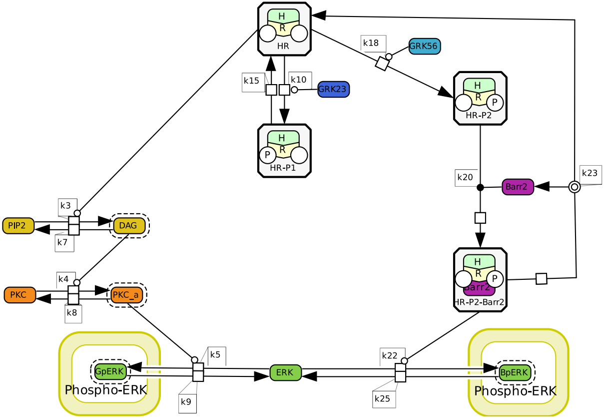
What do we learn (so far) ?

- Three pathways are necessary.
- An independent internalization pathway is mandatory.
- A **minimal** model with three reversible pathway with 10 parameter is able to fit the **phospho ERK data** and its parameter are **all identifiable**.



What do we learn (so far) ?

- Three pathways are necessary.
- An independent internalization pathway is mandatory.
- A minimal model can fit the **phospho ERK** data and is identifiable.
- The best model able to fit **all data** present non-identifiability ⇒ **Experimental design**.



Conclusion

- A full model able to fit the data (*Heitzler el al. MSB 2012*).
- Accurate parameter estimation leads to **accurate prediction**.
- Further improvements with **model reduction/selection**.
- Parameter **identifiability** with a good fit can be achieved.
- We have shed light on the importance of **three pathways** in GPCR signaling, and its regulations.

Conclusion

- A full model able to fit the data (*Heitzler el al. MSB 2012*).
- Accurate parameter estimation leads to accurate prediction.
- Further improvements with model reduction/selection.
- Parameter identifiability with a good fit can be achieved.
- We have shed light on the importance of three pathways in GPCR signaling, and its regulations.

Thanks for your attention !

validation data