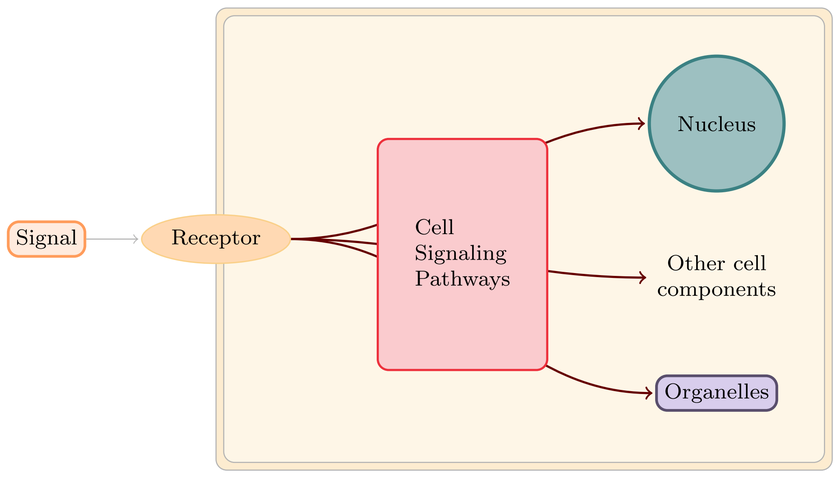
Identification of cell signaling pathways based on biochemical reaction kinetics repositories

Gustavo Estrela de Matos[a, b, c]
Hugo A. Armelin[b, c], Marcelo S. Reis[b, c]

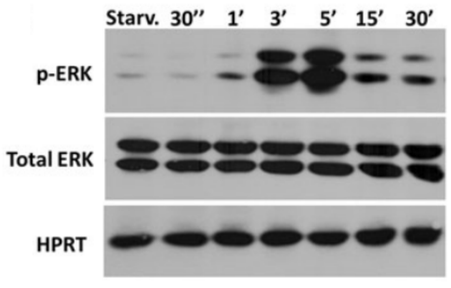
[a]Institute of Mathematics and Statistics, University of São Paulo, Brazil
[b]Center of Toxins, Immune-response and Cell Signaling (CeTICS), Instituto Butantan, Brazil
[c]Special Laboratory of Cell Cycle, Instituto Butantan, Brazil

Cell Signaling Pathways

Cell Signaling is a mechanism that allows the cell to change its behaviour according to the environment.



A signal flows in a cell thorugh a cell signaling pathway, which can be characterized by a sequence of chemical reactions.



We can summarize the state of a cell signaling pathway by measuring the concentration over time of some chemical species that are present on the pathway, yielding experimental data D.

Identification of Signaling Pathways

What is the structure of a cell signaling pathway, given a set of concentration measurements? We answer this question with a computational model, created for a set of chemical reactions, that can reproduce the concentration dynamics observed experimentally. These models are created using the laws of mass action, deriving a system of ordinary differential equations.

As an example, we can model the following equation:

$$\mathrm{A} + \mathrm{B} \longrightarrow \mathrm{C} \implies \frac{d[C]}{dt} = k[A][B]$$

Where k is a reaction rate constant.

However, to derive the model, we still need to determine what is the set of chemical reactions of the signaling pathway.

Bayesian Ranking of Models

Given some experimental data D and a model M (composed by a a set o reactions), we use an estimative of $p(D|M)$ as a quality measure of model M. To create this estimative, we need to take samples from the posterior distribution of model parameters (reactions rate constants), denoted by $\theta|D, M$.

$$\underbrace{p(\theta|M, D)}_{\text{posterior}} \propto \underbrace{p(D|M, \theta)}_{\text{likelihood}} \underbrace{p(\theta|M)}_{\text{prior}}$$

This ranking score was implemented as a Python package called SigNetMS. This is an open source software and it is available on GitHub: github.com/gustavoem/SigNetMS.

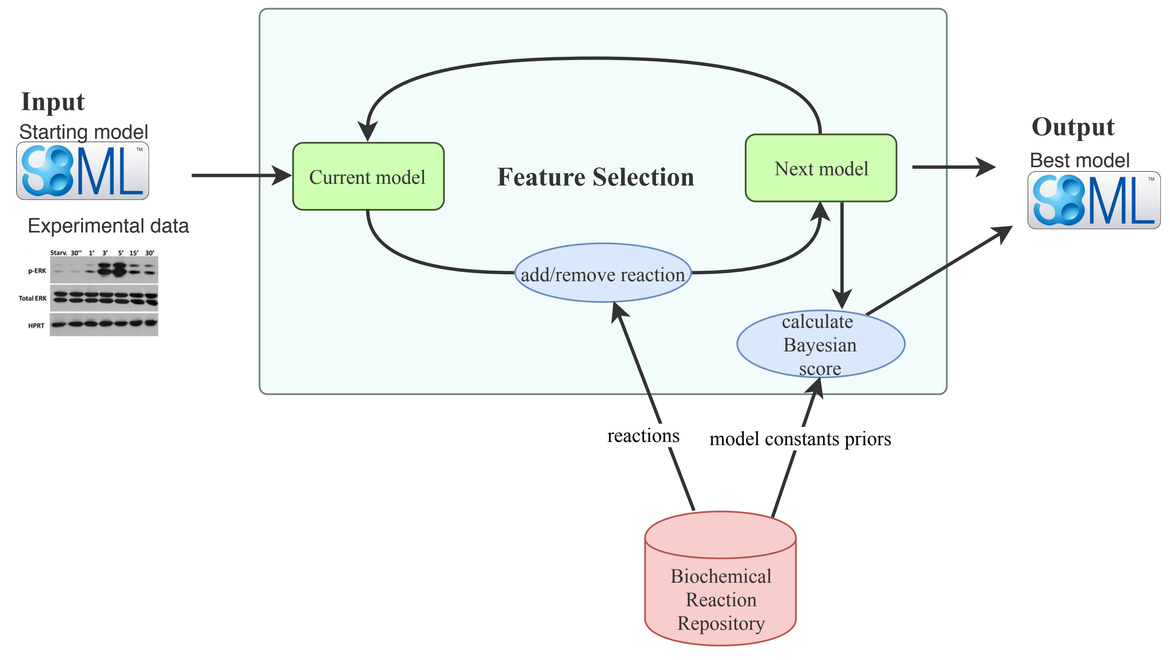
Acknowledgements

FAPESP

IME-USP

The Proposed Methodology

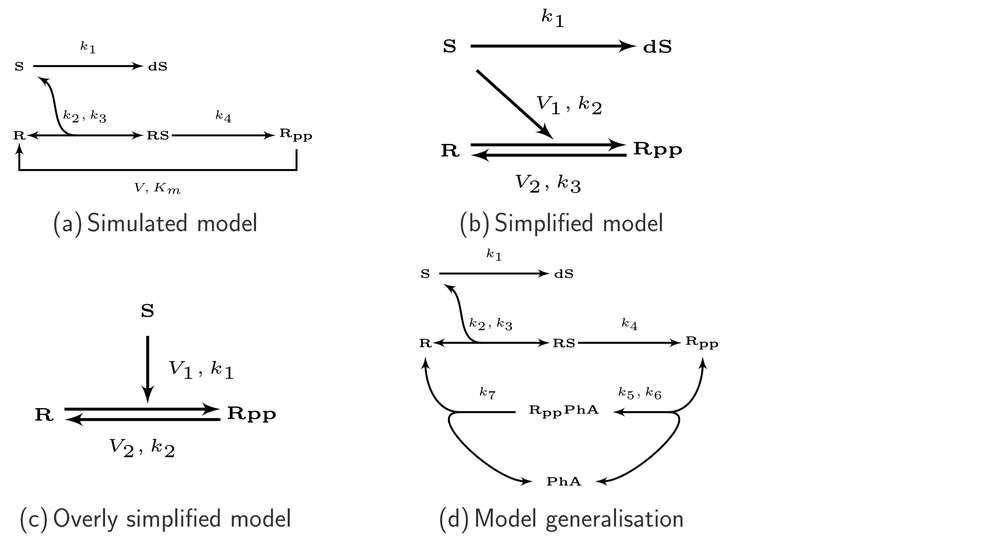
We proposed to solve the identification of cell signaling pathways as a feature selection problem. This problem consists of: given a set of features (candidate reactions) and a score for each subset (estimative of $p(D|M)$), what is the best subset?



The set of candidate reactions is stored in a database, which also stores information about model parameters, namely, the reactions rate constants. The information in the database is extracted from other public databases.

Experiments

We tested our ranking method on an experiment in which we create artificial experimental data and try to compare the found ranking for four models.

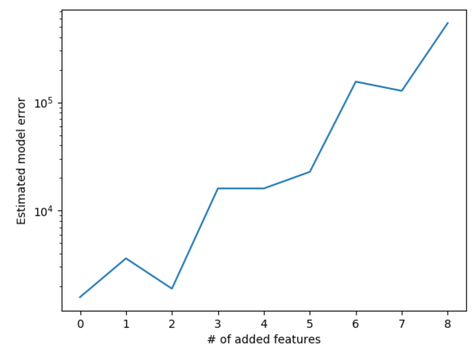


(a) Simulated model (b) Simplified model

(c) Overly simplified model (d) Model generalisation

On this experiment we got the ranking: a > b > d > c. This showed not only that we ranked the correct model first but also that the complex models were penalized.

We also performed an experiment in which we add reactions iteratively to a model. We could have a glance of the surface of the score metric over our search space. We could also notice that the model error does not increase monotonically as we add more reactions.



Conclusion and Future Work

In this work we were able to create a score metric that is able to rank models in a Bayesian approach and embed this function to a feature selection procedure. The next tasks we aim to accomplish are:

- Continue to study the surface of the search space;
- Apply the methodology in ERK signaling pathways of cell lines Y1 and HEK293.

{marcelo.reis, gustavo.matos }@butantan.gov.br