Identification of cell signaling pathways based on biochemical reaction kinetics repositories

Gustavo Estrela

May 2019

Instituto de Matemática e Estatística
Centro de Toxinas, Resposta-imune e Sinalização Celular (CeTICS)
Laboratório Especial de Ciclo Celular, Instituto Butantan

Introduction

Cell Signaling

Cell signaling allows cells to respond to signals that come from its environment changing its behaviour accordingly.

Cell Signaling

Cell signaling allows cells to respond to signals that come from its environment changing its behaviour accordingly.

This mechanism is essential for many cell functions, including reproduction, growth and death.

Cell Signaling

Cell signaling allows cells to respond to signals that come from its environment changing its behaviour accordingly.

This mechanism is essential for many cell functions, including reproduction, growth and death.

Understanding the functioning of cell signaling is important in many biological areas.

Cell Signaling

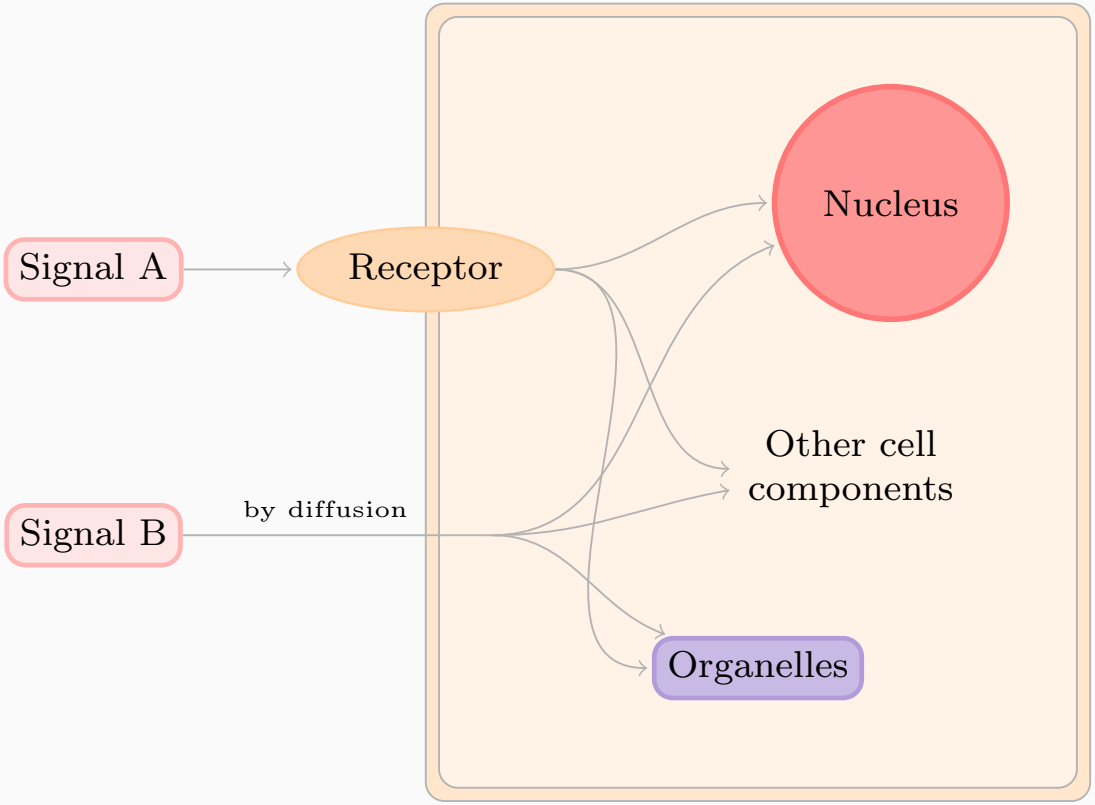

Figura 1: A general cell signaling mechanism.

Cell Signaling Pathways

A cell signaling network can be characterized by a sequence of chemical reactions that allows the presence of a signal to modify the state or behaviour of a cell.

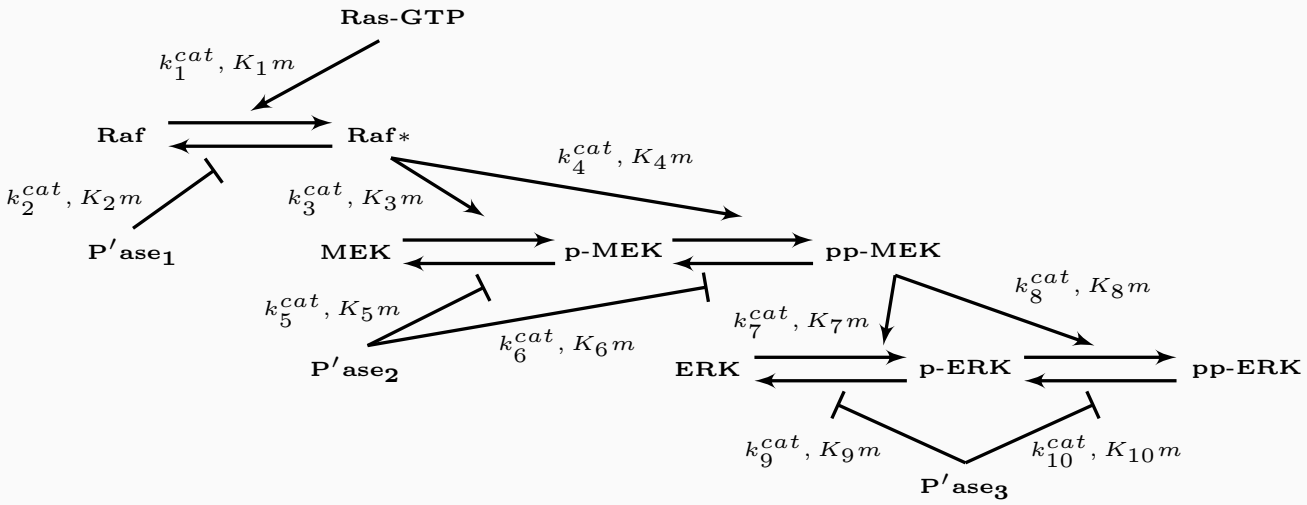

Figura 2: An example of a signaling pathway.

Mathematical Models of Signaling Networks

In many applications, we summarize the state of the cell with a measurement based on the concentration of chemical species of the cell.

Mathematical Models of Signaling Networks

In many applications, we summarize the state of the cell with a measurement based on the concentration of chemical species of the cell.

Using biochemical and enzymatic kinetics, we can write equations that represent the rate of change of concentration for a chemical species.

As an example, for the chemical reaction

$$\mathrm{A} \longrightarrow \mathrm{B},$$

As an example, for the chemical reaction

$$\mathrm{A} \longrightarrow \mathrm{B},$$

we can write the following equation:

$$\frac{d[\mathrm{A}]}{dt} = k_1[\mathrm{A}];$$

where k_1 is a reaction rate constant.

Mathematical Models of Signaling Networks

As an example, for the chemical reaction

$$\mathrm{A} \longrightarrow \mathrm{B},$$

we can write the following equation:

$$\frac{d[\mathrm{A}]}{dt} = k_1[\mathrm{A}];$$

where k_1 is a reaction rate constant.

Repeating this procedure for all reactions of a pathway allows us to derive a system of ordinary differential equations that can model the signaling pathway.

Identification of Cell Signaling Pathways

The problem of identification of cell signaling pathways is the problem of finding the components of a signaling pathway and how they interact given a set of experimental measurement.

Identification of Cell Signaling Pathways

The problem of identification of cell signaling pathways is the problem of finding the components of a signaling pathway and how they interact given a set of experimental measurement.

As the input, a description of a biological experiment and a set of experimental measurements are given.

Identification of Cell Signaling Pathways

The problem of identification of cell signaling pathways is the problem of finding the components of a signaling pathway and how they interact given a set of experimental measurement.

As the input, a description of a biological experiment and a set of experimental measurements are given. A possible output to the problem is composed by:

Identification of Cell Signaling Pathways

The problem of identification of cell signaling pathways is the problem of finding the components of a signaling pathway and how they interact given a set of experimental measurement.

As the input, a description of a biological experiment and a set of experimental measurements are given. A possible output to the problem is composed by:

- a model composed by a set of chemical reactions that are relevant for the biological experiment;

Identification of Cell Signaling Pathways

The problem of identification of cell signaling pathways is the problem of finding the components of a signaling pathway and how they interact given a set of experimental measurement.

As the input, a description of a biological experiment and a set of experimental measurements are given. A possible output to the problem is composed by:

- a model composed by a set of chemical reactions that are relevant for the biological experiment;
- information about the reaction rate constants of the model.

Identification of Cell Signaling Pathways

One can search for the set of chemical reactions relevant for a biological experiment in repositories like the Kyoto Encyclopedia of Genes and Genomes (KEGG).

Identification of Cell Signaling Pathways

One can search for the set of chemical reactions relevant for a biological experiment in repositories like the Kyoto Encyclopedia of Genes and Genomes (KEGG). However, the pathway maps from KEGG may be incomplete or have impertinent reactions for the biological experiment of interest.

Identification of Cell Signaling Pathways

One can search for the set of chemical reactions relevant for a biological experiment in repositories like the Kyoto Encyclopedia of Genes and Genomes (KEGG). However, the pathway maps from KEGG may be incomplete or have impertinent reactions for the biological experiment of interest.

Hence, it is desirable to construct a method that can systematically modify these models and choose the one that better represents the experiment.

Identification of Cell Signaling Pathways

Lulu Wu (2015) presented in her master dissertation a methodology that proposes to systematically modify models of signaling network in order to better represent experiments.

Identification of Cell Signaling Pathways

Lulu Wu (2015) presented in her master dissertation a methodology that proposes to systematically modify models of signaling network in order to better represent experiments.

On her work, the problem of identification of cell signaling pathways is treated as a feature selection problem.

Feature Selection Problem

The feature selection problem is a combinatorial optimization problem:

> Given a set of features S and a cost function c, find subset $X \in \mathcal{P}(S)$, with minimum cost $c(X)$.

Feature Selection Problem

The feature selection problem is a combinatorial optimization problem:

> Given a set of features S and a cost function c, find subset $X \in \mathcal{P}(S)$, with minimum cost $c(X)$.

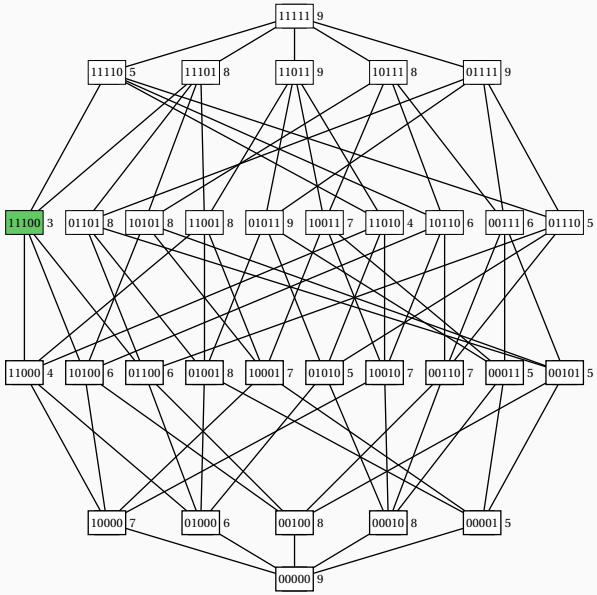

Figura 3: An example of feature selection search space with 5 features.

Feature Selection for Identification of Signaling Pathways

The methodology proposed by Wu defines the set of features as a set of chemical reactions that can be added to a starting model.

Feature Selection for Identification of Signaling Pathways

The methodology proposed by Wu defines the set of features as a set of chemical reactions that can be added to a starting model.This set of chemical reactions is fetched from KEGG.

Wu's Search Algorithm for Feature Selection

The search algorithm used by Wu is the Sequential Forward Selection (SFS).

Wu's Cost Function for Feature Selection

Wu defines the cost function as the minimum distance between experimental and simulated data.

Wu's Cost Function for Feature Selection

Wu defines the cost function as the minimum distance between experimental and simulated data. The minimum distance is found using a Simmulated Annealing that traverses the parameter space.

Wu's Cost Function for Feature Selection

If we define that:

Wu's Cost Function for Feature Selection

If we define that:

- the experimental measurement is D;

Wu's Cost Function for Feature Selection

If we define that:

- the experimental measurement is D;
- the simulated measure is $\phi(M, \theta)$;

Wu's Cost Function for Feature Selection

If we define that:

- the experimental measurement is D;
- the simulated measure is $\phi(M, \theta)$;
- all elements visited on the Simmulated Annealing process define Θ'.

Wu's Cost Function for Feature Selection

If we define that:

- the experimental measurement is D;
- the simulated measure is $\phi(M, \theta)$;
- all elements visited on the Simmulated Annealing process define Θ'.

Then, the cost function used on Wu's work is:

$$c(M) = min_{\{\theta \in \Theta' \subset \Theta\}} dist(\phi(M, \theta), D) + R(M),$$

Wu's Cost Function for Feature Selection

If we define that:

- the experimental measurement is D;
- the simulated measure is $\phi(M, \theta)$;
- all elements visited on the Simmulated Annealing process define Θ'.

Then, the cost function used on Wu's work is:

$$c(M) = min_{\{\theta \in \Theta' \subset \Theta\}} dist(\phi(M, \theta), D) + R(M),$$

where $R(M)$ is a regularization function.

Fundamental Concepts

Model Selection

Experiments on Model Selection

Next Steps