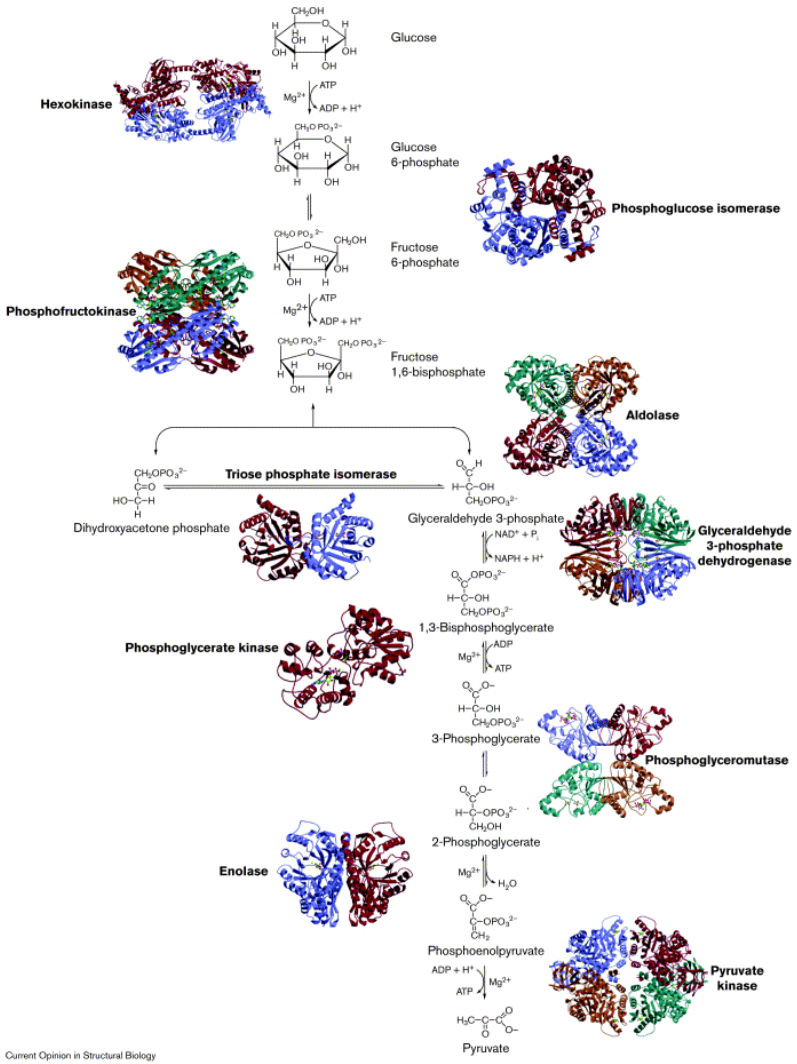
Unravelling the biochemical reaction kinetics from time-series data

Santiago Schnell
Indiana University School of Informatics and Biocomplexity Institute
Email: schnell@indiana.edu
WWW: http://www.informatics.indiana.edu/schnell

Achievements of the biomedical sciences

- The identification and structural characterisation of molecules.
- The determination of rate constant for large number of biochemical reactions and physiological interactions.
- The design, construction and synthesis of molecules. The study of their physiological effects.

Identification of the reaction mechanism



However, it is not until recently that some progress has been made on dissecting a complex biochemical/physiological mechanism into its key components.

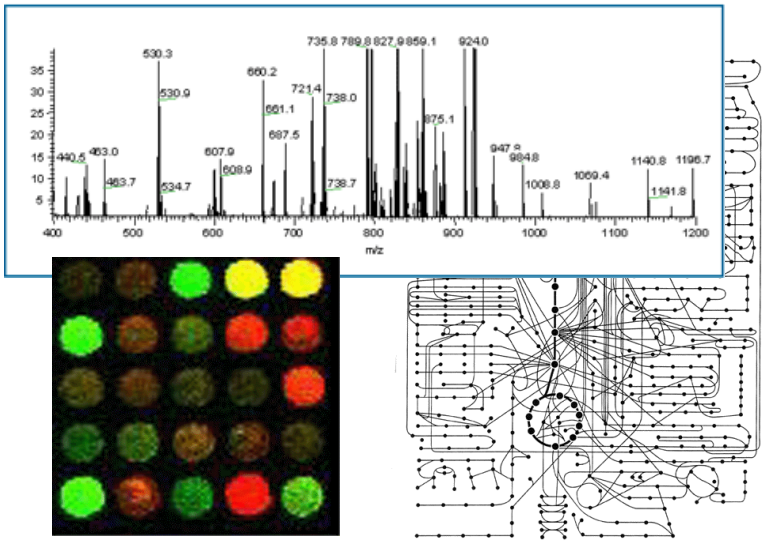
Classical approach for determining the reaction mechanism

- Identify the primary stoichiometry of the overall reaction.
- Compile a list of chemically plausible species.
- Break down the overall reaction into likely elementary steps.
- Assemble all relevant experimental data available.
- Put all the thermodynamically plausible steps and the corresponding kinetics data together to form a trial mechanism.
- Use numerical methods to simulate the experimental results.
- Continue to refine and improve the mechanism, testing it against all new experimental results.

Changes in biomedical research

- Traditional labour intensive! Hypothesis testing

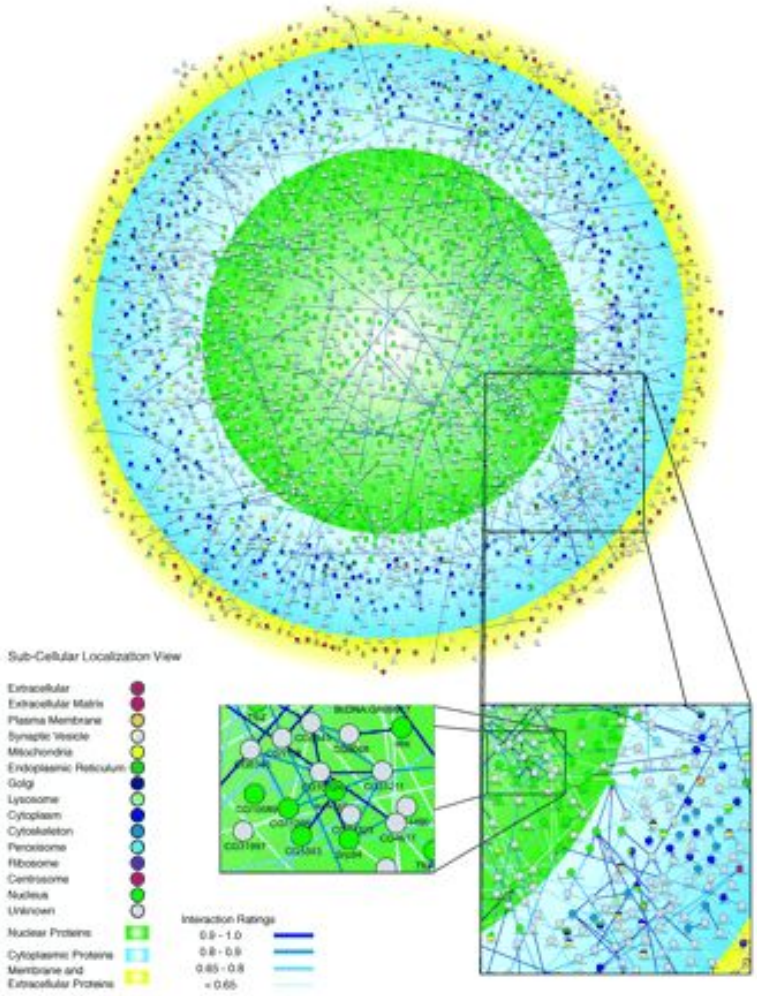
- High throughput measurements, *-omic* technologies, databases: data-driven modelling



- Experimentation is now non-hypothesis driven

Changes in biomedical research

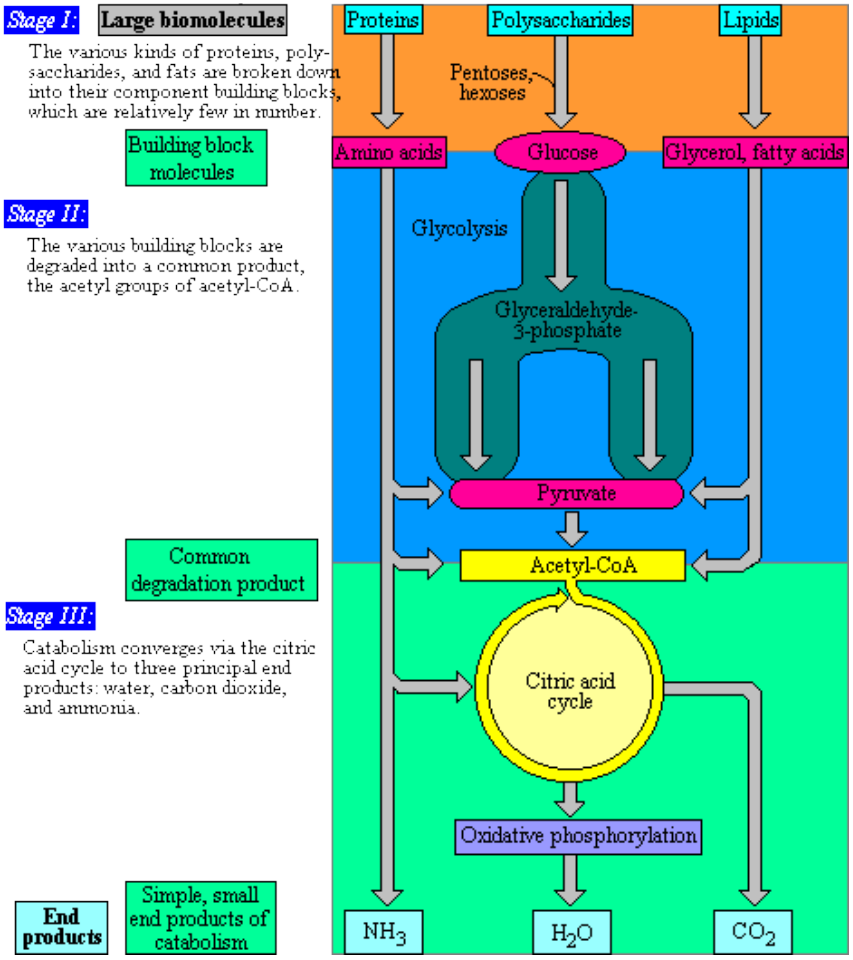
Oimics Science



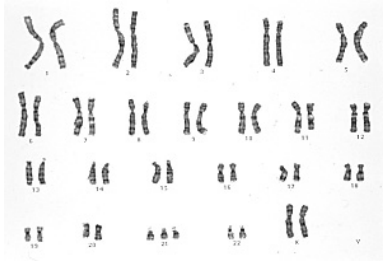
While there is a wealth of genomics, proteomics and microarray data available today, we still have more questions than answers when it comes to understand the functions of genes and proteins and their downstream effects on the behaviour of a cell or an organism.

Limitation of –oimics data

While the high-throughput experimental assays would have been unimaginable in the era preceding genome sequencing, an protein interaction map doesn't get us close to one key aspect of biology: dynamism.



Fundamental problems in biomedical sciences

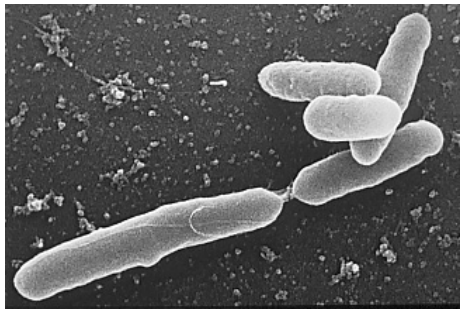


One of the most profound surprises arising from the sequencing of different species' genomes is the degree of similarity among sequences:

We see this similarity not only in

- the numbers of protein-coding genes, and
- the degree of homology between genes belonging to different species, but also in
- the organization of genes and in their regulation.

Fundamental problems in biomedical sciences



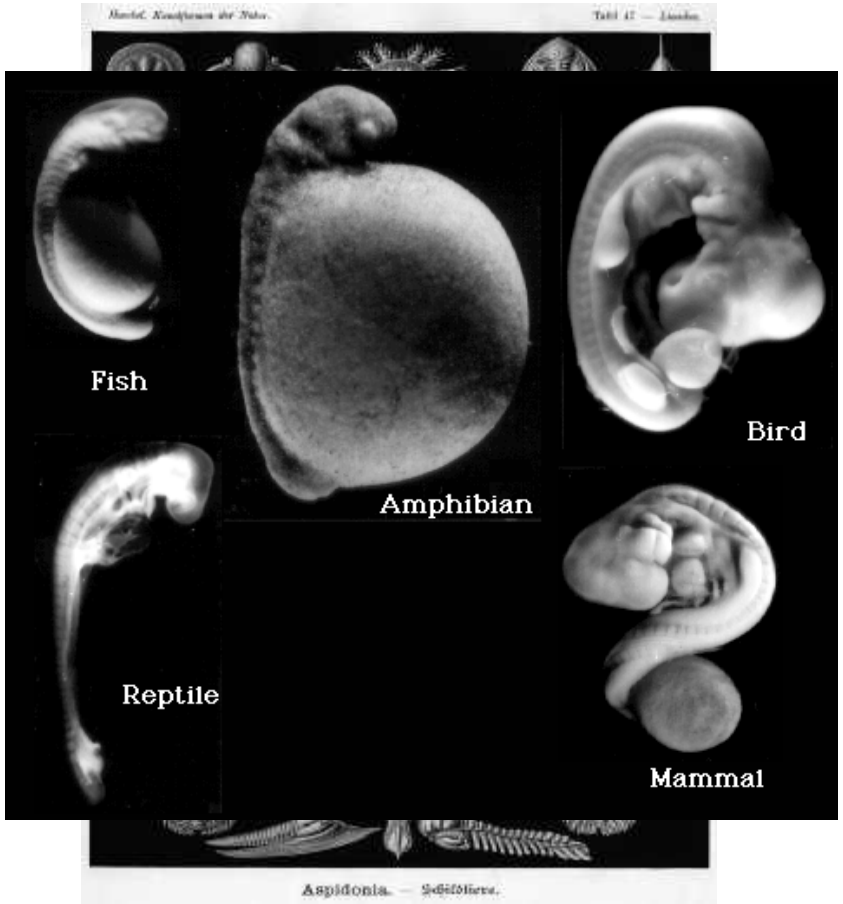
Even when full genomic sequences for an organism are available, the functions and interactions of only a small number of gene components are clear.

- For humans, a function is known for 3-5% of the genes.

- For the well-studied E. coli, more than 60% of the genes have a known function.

As we understand the gene functions, we can better predict and control their responded to perturbations.

Possible outcomes of functional genomics



- Similar sets of genes in different developmental programs
- Different organisms hook up in different ways
- Kinetic properties of biological pathways may be important determinants in differentiating among the products of different genomes.

Biochemical pathway inference methodologies under consideration

Models to investigate reaction mechanisms from time course data:

- **Impulse-response method**
- **Distribution-delay method**
- **Time-series analysis**

Unravelling the reaction
Why kinetic models?

- Biological processes are time dependent
- Sequence of events (causality)
- To understand regulatory mechanisms
- Over 25 years of kinetic modelling and experimental work on biochemical reaction networks

Principles for biochemical kinetics pathway modelling

Identify the flow of mass between variables.

- Irreversible and reversible processes

$$E + S \leftrightarrow C \rightarrow P + E$$

- Open and closed reactions

$$\rightarrow S_1 \leftrightarrow S_2 \leftrightarrow S_3 \rightarrow \qquad\qquad S_1 \leftrightarrow S_2 \leftrightarrow S_3$$

- Stoichiometry

$$2\,X \rightarrow A + P \qquad\qquad X + Y \rightarrow 2\,P$$

Principles for biochemical kinetics pathway modelling

Identify stoichiometry in the pathway.

Stoichiometry of the reactions

$$A \rightarrow C + D$$

$$A + B \rightarrow C + D$$

$$A + B + M \rightarrow C + D$$

Principles for biochemical kinetics pathway modelling

For each variable X_i that changes over time, define an equation that relates its change over time to influxes and effluxes.

Change in X_i = Fluxes into X_i - Fluxes out of X_i

The change is equivalent to the derivative of the variable X_i with respect to time: dX_i/dt

Principles for biochemical kinetics pathway modelling

Writing the rate equations.

In general, we can write

$$X_1 \xrightarrow{v_1} X_2 \xrightarrow{v_2} X_3$$

$$\frac{\mathrm{d}X_2}{\mathrm{d}t} = v_1 - v_2$$

where v_1 and v_2 depend on the chemical kinetics of the reaction.

Principles for biochemical kinetics pathway modelling

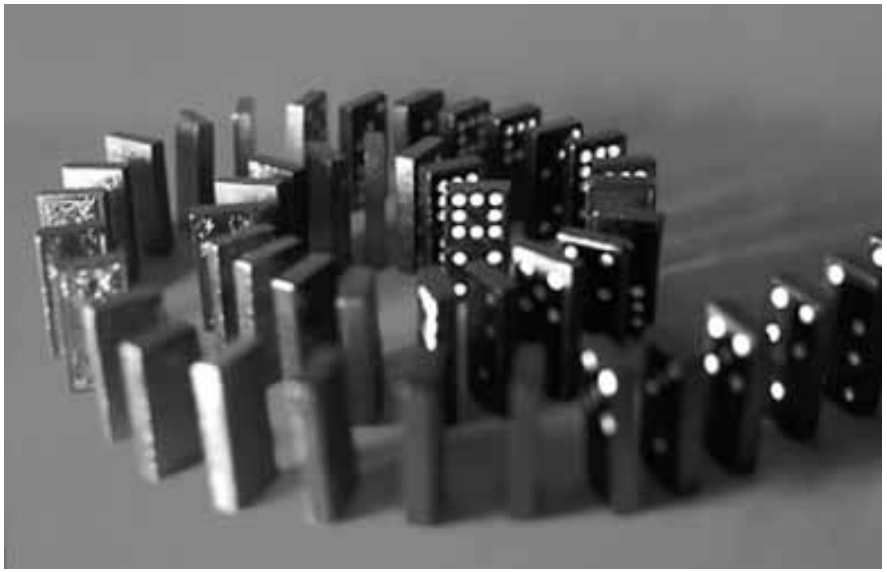
Law of mass action

$$X_1 + A \xrightarrow{v_1} X_2 \xrightarrow{v_2} X_3$$

$$\frac{\mathrm{d}X_2}{\mathrm{d}t} = \underbrace{k_1 X_1 A}_{v_1} - \underbrace{k_2 X_2}_{v_2}$$

where k_1 and k_2 are rate constants.

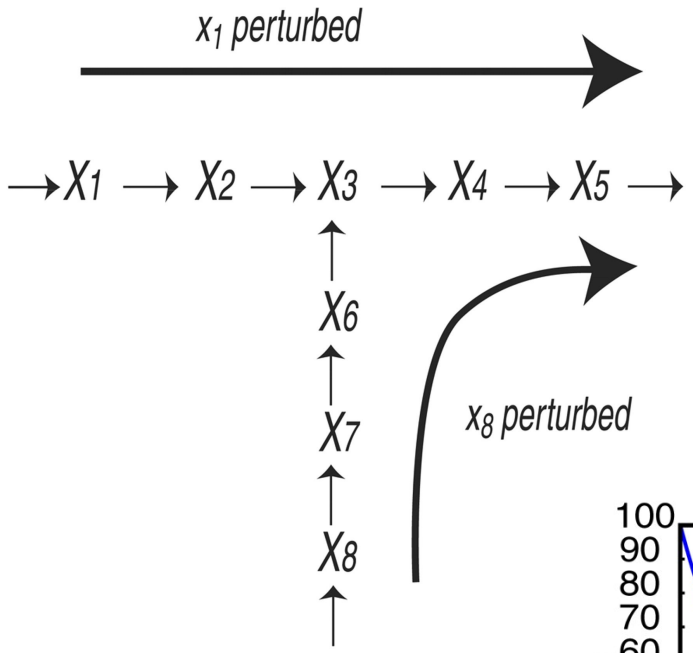
Impulse-response method



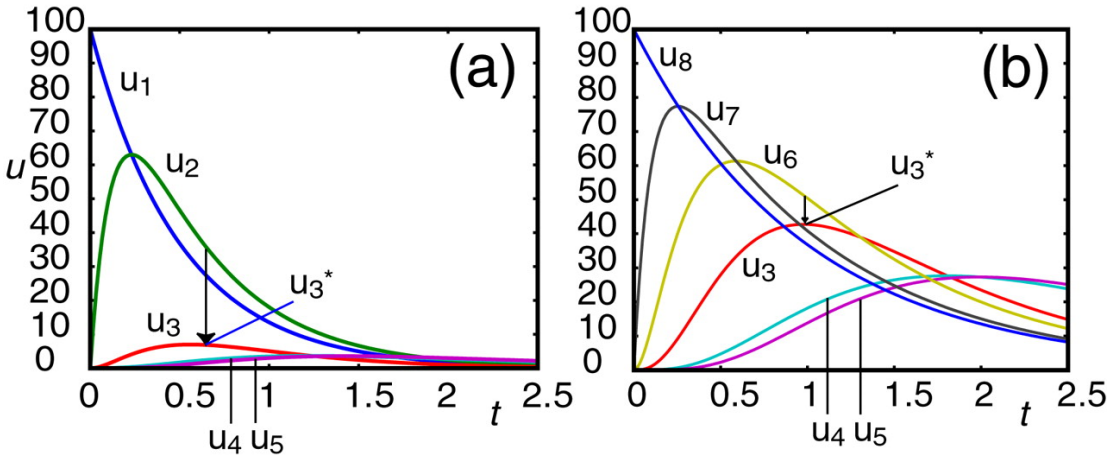
Domino-effect analysis

A qualitative form of impulse response analysis can be used to gather information on the connectivity of a biochemical reaction pathway, which can be pieced together to determine the network connectivity, and in some cases the entire wiring diagram.

Impulse-response method



Peaks can be quantified determining the response timescale of the species.



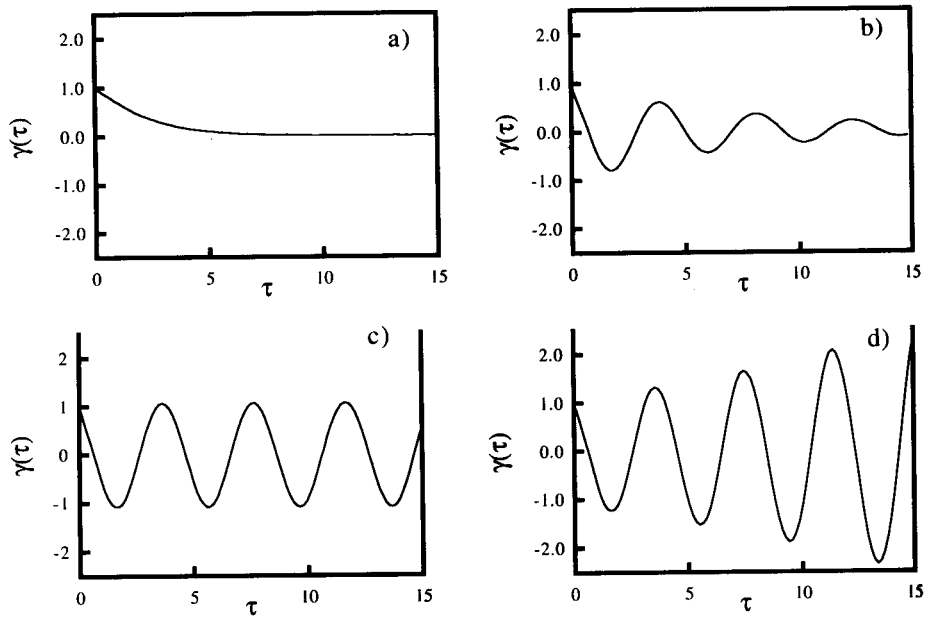
Limitation of the impulse-response method

- For complicated networks, the responses are harder to interpret and the network more difficult to reconstruct.
- The impulse-response diffuses with the "distance" from the perturbed species.
- The method requires to known all the species involved in the reaction.

Distribution delay method

In this approach, we consider a reaction network as follows:

$$\text{Reactants} \xrightarrow{\text{delay}} \text{Products}$$



Distribution delay method

From the mathematical point of view, a reaction with delay can be written:

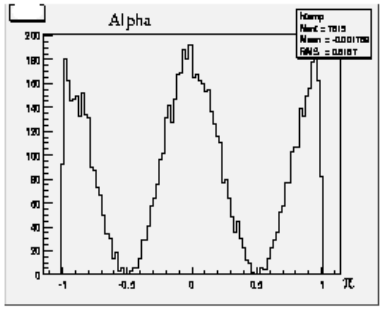
$$X_1 \xrightarrow{v_1} X_2 \xrightarrow{v_2} X_3$$

$$\frac{\mathrm{d}X_2(t)}{\mathrm{d}t} = \underbrace{k_1 X_1(t-\tau)}_{\text{synthesis}} - \underbrace{k_2 X_2(t)}_{\text{decay}}$$

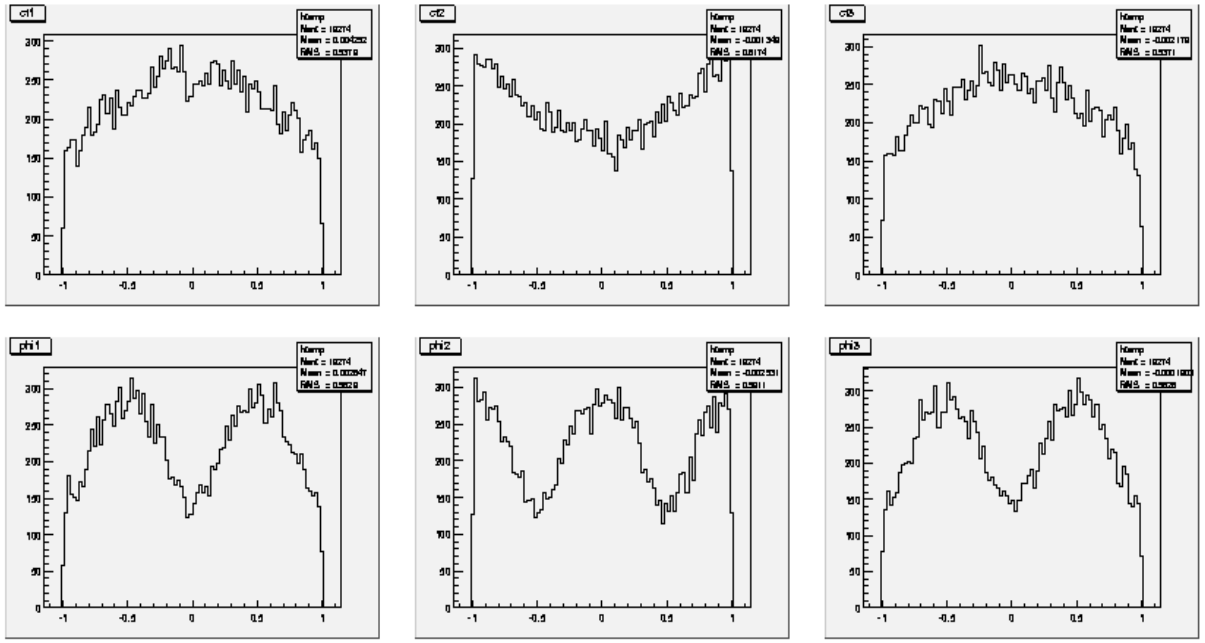
where t is the delay in the reaction.

Distribution delay method

The distribution of delays can tell us about the number of intermediates and the size of rates constants.



The distribution of delays can also tell us about the topology of the network.



Limitation of the distribution delay method

- We found that the form of the distribution delay is characteristic for each reaction mechanism.
- However, a particular distribution of delays does not give a unique reaction mechanism.
- Distribution of delays can be employed to determine the conditions under which a number of reaction mechanism are equivalent.

A potential difficulty: Indistinguishable reactions

Example

In 1902, Victor Henri proposed two mechanisms for the enzyme action:

I $E + S \leftrightarrow C \rightarrow E + P$

II $C \leftrightarrow E + S \rightarrow E + P$

The reduce system describing these reaction is:

$$\frac{d[S]}{dt} = -\alpha([E_0]-[C])[S] + \beta[C]$$

$$\frac{d[C]}{dt} = \gamma([E_0]-[C])[S] - \delta[C]$$

where [illegible],[illegible],[illegible] and [illegible] are cons

Time series analysis

In this type of analysis, we try to solve the reverse problem: from the time course data, we would like to fit an expression for the reaction kinetics.

Let us suppose that reaction system governing equation of the chemical species is represented by a polynomial model structure of order *p*.

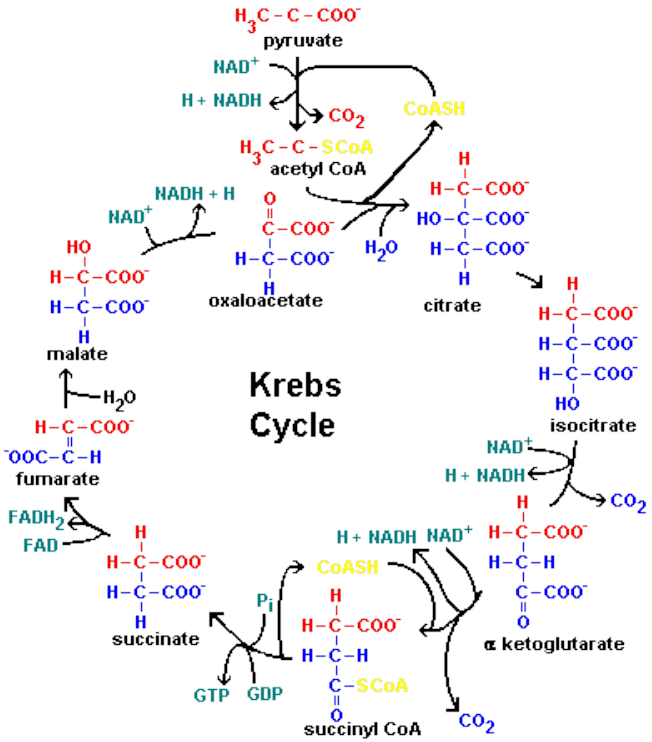
$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = F_i(\mathbf{x}, a_i) = a_i + \sum_{j=1}^{K} b_{ij} x_j + \sum_{j=1}^{K} \sum_{k=1}^{K} c_{ijk} x_j x_k$$

Then, we apply an interative approach to model selection.

Simple-to-General, or General-to-Specific?

How the researchers are divided!

Simple-to-General



General-to-simple

Setting the cost function: minimum description length

- Danger of over fitting
- Avoid by choosing a Cost function to assess "goodness of model" which penalises use of many terms
- Consistent with Sparseness ideas for some biological networks

Dealing with biological complexity: Using prior knowledge

For time series analysis approaches the key is to adapt them to incorporate biological knowledge. This can be done by

- Starting with a partial model
- Favourably weighting suspected reactions and interactions
- Fixing known interactions

The pros and cons

- Simple-to-general approach model building is a divergent branching process, this favours general-to-specific approach.

- The general-to-specific approach is very slow as the number of species increases, particularly for polynomial type expressions.

Limitation of the time series analysis

- Most of the methods proposed require to measurement of the concentrations of all the species. This conditions cannot be met in practice.
- Most of the techniques involve monitoring the concentration relaxation after a perturbation very close to the equilibrium state. The mechanisms deduced by these methods follow pseudo-first-order kinetics.
- The determination of the mechanism by time series analysis does not afford unique solutions.
- There are reaction mechanisms that can be kinetically indistinguishable by time series analysis!

FIN