Biological Pathways Representation by Petri Nets and extensions

Andrea Marin

December 6, 2006

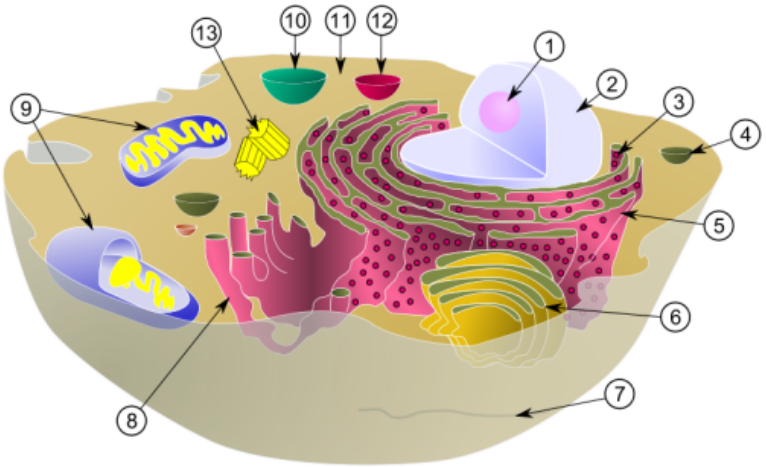
The cell (1)

The cell is the structural and functional unit of all living organisms.
It is the *building block* of life.
Two types of cells:

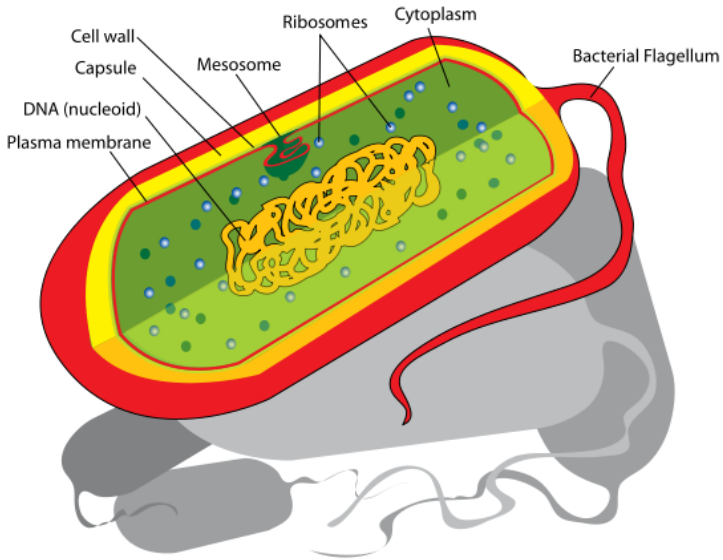
- Prokaryotic cells: small, lack of nuclear membrane, usually singletons;
- Eukaryotic cells: 10 time bigger than Prokaryotic (1000 in volume), contain membrane-bound compartments in which specific metabolic activities take place, usually found in multi-cellular organisms;

The cell (2)

Eukaryotic Cell



Prokaryotic Cell



Cell Activities

All cells share the following activities:

- Reproduction by cell division
- Use and production of enzymes and other proteins coded by DNA genes
- Response to external and internal stimuli such as changes in temperature, pH or nutrient levels
- Metabolism: take in raw materials, building cell components, converting energy, molecules.

Biological Pathways

Biological Pathways

A Biological Pathway is a molecular interaction network in biological processes.

We consider three types of biological pathways:

- Metabolic pathways
- Message Passing Regulatory pathways
- Gene Regulatory Pathways

Metabolic Pathways

Metabolic Pathways

A metabolic pathway is a series of chemical reactions occurring within a cell, catalyzed by enzymes, resulting in either the formation of a metabolic product to be used or stored by the cell, or the initiation of another metabolic pathway.

The functioning of a cell depends upon its ability to extract and use chemical energy stored in organic molecules. This energy is derived from metabolic pathways to ATP. Example of metabolic pathways are:

- glycolysis which generates high-energy molecules, ATP and NADH
- Pentose phosphate pathway
- ...

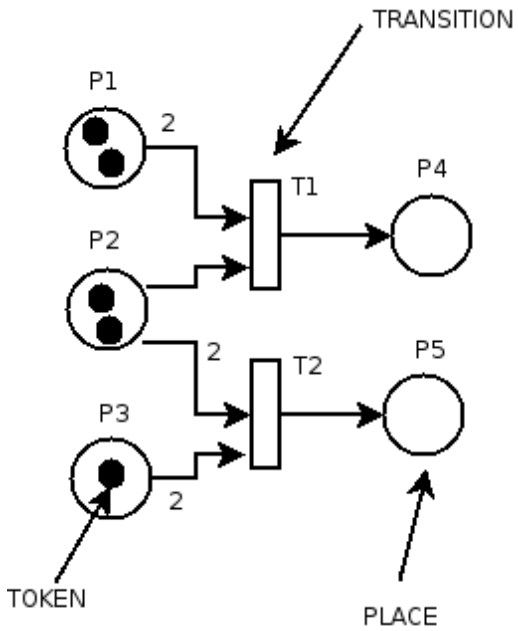
Regulatory Pathways

Regulatory Pathways

Regulatory Pathways can be classified in:

- Genetic information processing: Transcription, Translation, Sorting and Degradation. Replication and Repair
- Environmental information processing: Membrane transport, Signal transduction, Ligand receptor interaction
- Cellular processes: Cell motility, Cell growth and death, Cell communication, Development, Behaviour

Standard Petri Net (definition)



A Petri Net is a 5-tuple $PN = \{\mathcal{P}, \mathcal{T}, \mathcal{A}, \mathcal{W}, \mathbf{m_0}\}$ where:

- $\mathcal{P}$ is the set of places. E.g. $\mathcal{P} = \{P_1, P_2, P_3, P_4, P_5\}$
- $\mathcal{T}$ is the set of transitions. E.g. $\mathcal{T} = \{T_1, T_2\}$
- $\mathcal{A} \subseteq T \times P \cup P \times T$ is the set of arcs. E.g. $\mathcal{A} = \{(P_1, T_1), (P_2, T_1), (P_2, T_2), (P_3, T_2)\} \cup \{(T_1, P_4), (T_2, P_5)\}$
- $\mathcal{W} : \mathcal{A} \rightarrow \mathbb{N}$ is a function which assigns to each arc a weight
- $\mathbf{m_0}$ is the initial marking vector. E.g. $\mathbf{m_0} = (2, 2, 1, 0, 0)$.

Enabled Transitions, Input and Output Vector

Definition

A transition T_i is enabled by the marking $\mathbf{m}$ when $\forall(P_j, T_i) \in \mathcal{T}, \mathcal{W}(P_j, T_i) \leq m_j$. It is disabled otherwise.

In the previous example T_1 is enabled and T_2 is disabled.

Definition

The Input vector $\mathbf{I}(T_i)$ is a vector whose k-th component is $\mathcal{W}(P_k, T_i)$ if $(P_k, T_i) \in \mathcal{A}$, 0 otherwise. The Output vector $\mathbf{O}(T_i)$ is a vector whose k-th component is $\mathcal{W}(T_i, P_k)$ if $(T_i, P_k) \in \mathcal{A}$, 0 otherwise.

In the previous example $\mathbf{I}(T_1) = (2, 1, 0, 0, 0)$ and $\mathbf{O}(T_1) = (0, 0, 0, 1, 0)$.

Incidence Matrix

Definition

The Incidence Matrix of a PN is a matrix with a row for each transition and a column for each place. The k-th matrix row is the vector $\mathbf{O}(T_k) - \mathbf{I}(T_k)$ and it represents the marking change when T_k fires.

The incidence matrix of the previous example is:

$$\mathbf{A} = \begin{bmatrix} -2 & -1 & 0 & 1 & 0 \\ 0 & -2 & -2 & 0 & 1 \end{bmatrix}$$

Dynamic

Definition

When a transition is enabled it *can* fire. The firing of the transition T_i changes the marking of the net from $\mathbf{m}$ to $\mathbf{m} - \mathbf{I}(T_i) + \mathbf{O}(T_i)$.

Definition

The reachability set $RS(\mathbf{m_0})$ of a Petri Net is the set of all possbile markings reachable from the initial marking $\mathbf{m_0}$.

Note that:

- In the general case, the number of marking of the RS grows exponentially with the number of places of the net and the number of tokens of the initial marking,
- The problem of deciding if a marking is reachable is NP-hard.

S-invariant

Definition

A S-Invariant for a PN with $\mathbf{A}$ as incidence matrix is a vector such that:

$$\mathbf{A} \cdot \mathbf{S} = \mathbf{0}.$$

The existence of a no-zero components S-invariant tells us that the weight sum of the tokens in the net is costant.

T-invariant

Definition

A T-Invariant for a PN with **A** as incidence matrix is a vector such that:

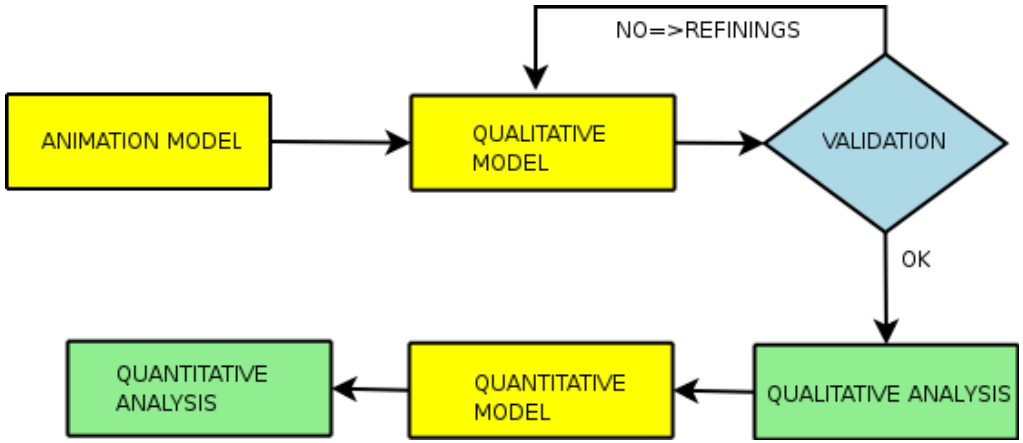
$$\mathbf{A}^{\mathbf{T}} \cdot \mathbf{T} = \mathbf{0}.$$

Suppose that a PN has 4 transitions $\mathcal{T} = \{T_1, \ldots, T_4\}$ and let $T = (2, 0, 1, 0)$ be a T-invariant. Then we know that, starting from any marking **m**, **if** T_1 fires twice and T_3 fires once (in any order), and no other transition fires, the final marking is again **m**.

What should a suitable model allow?

1. readability $\Longrightarrow$ supports understanding both for computer scientists and biologists
2. executability $\Longrightarrow$ allows even no-experts to get familiar with the model
3. validations techniques $\Longrightarrow$ consistency checks, does the model respect natural laws?
4. analysis techniques $\Longrightarrow$ qualitative and/or quantitative analysis.

A possible path for models



- The animation step gives the idea on the possible simplifications, e.g. ignoring reactions which seem not important
- The qualitative model is a simplified model which allows one to compare pathways, or point out cycles or steady states
- The quantitative model is an extended model which consider compound concentrations and reaction kinetic.

Catalyzed chemical reaction

A chemical reaction takes its substrate and give a product. For example the glycosulfatase reaction has as substrate D-glucose 6-sulfate + H20 and for product D-glucose and sulfate:

$$\text{D-glucose 6-sulfate} + \text{H2O} = \text{D-glucose} + \text{sulfate}$$

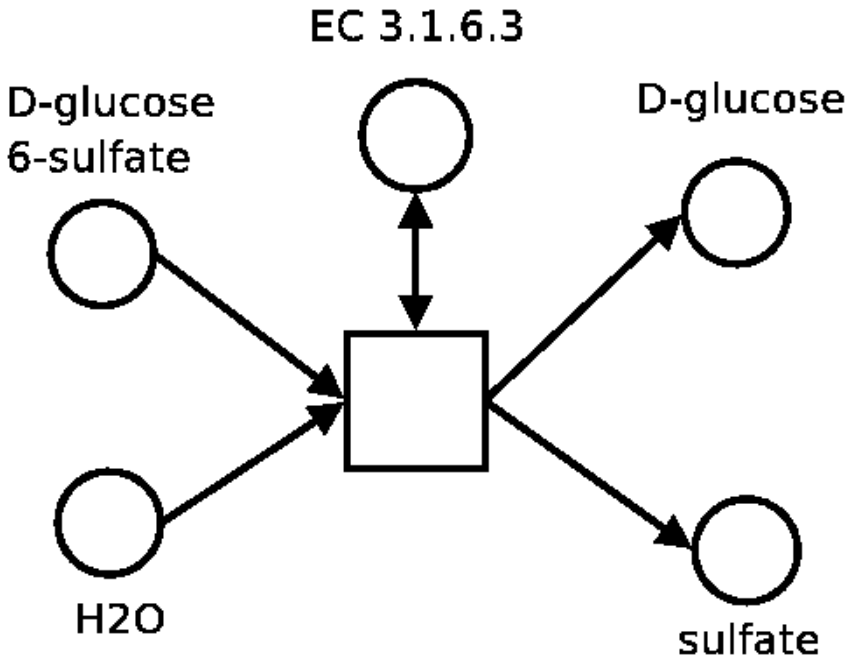
Most of the reactions in cell are catalyzed by enzymes. Note that:

- enzymes are not consumed by the reaction
- enzymes speed up the reaction

In the glycolysis the glycosulfatase is catalyzed by the **EC 3.1.6.3** enzyme.

Petri Net for a simple chemical reaction

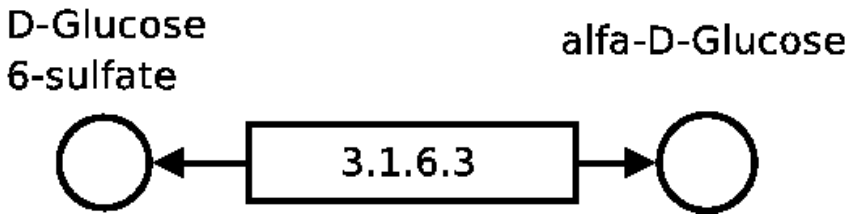
- The basic idea is to associate to each substrate or product compound a place, to the enzyme another place and set the arcs to destroy the substrate and return the product:



- EASY!

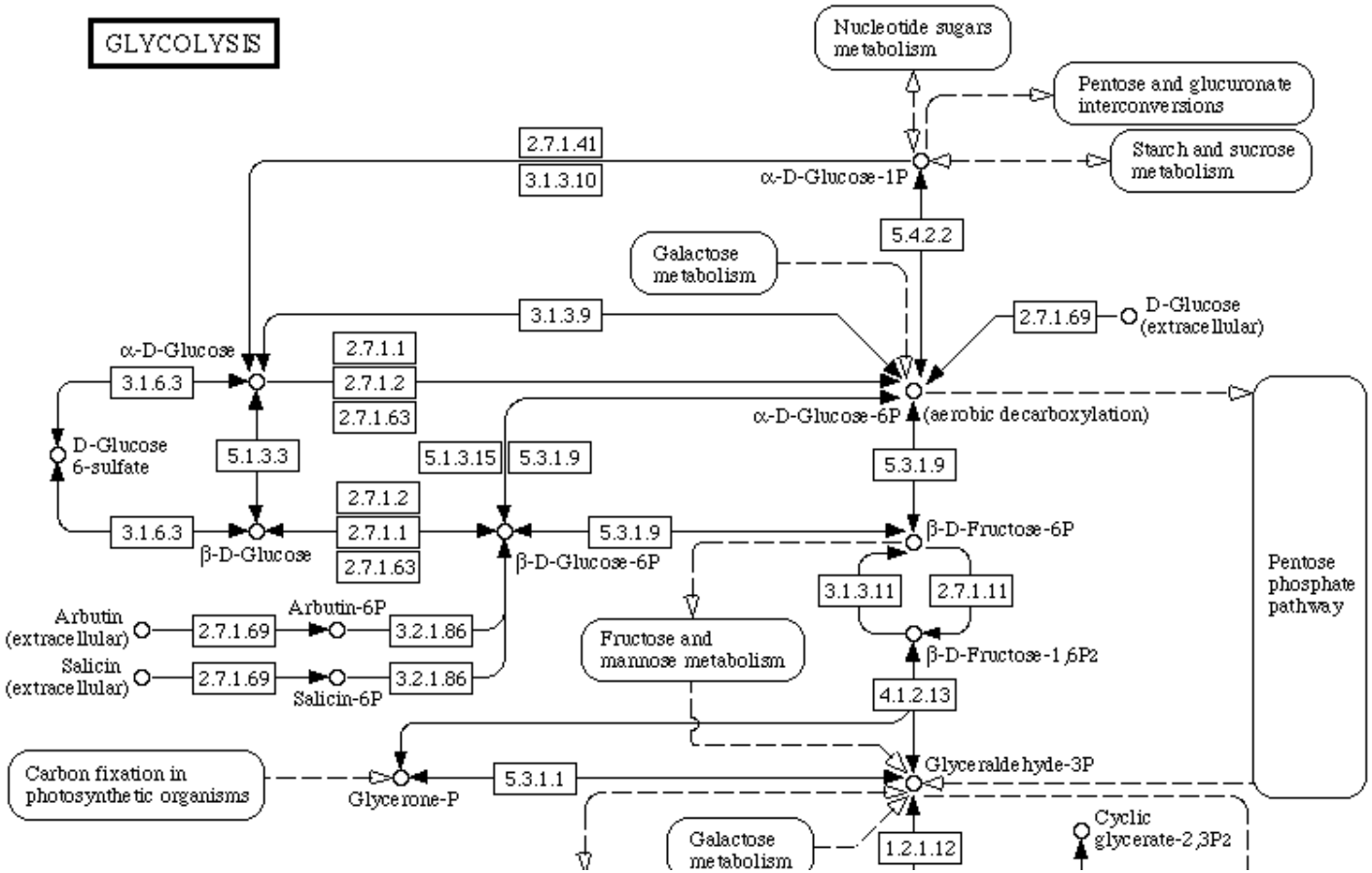
The reaction in KEGG

- KEGG (Kyoto Encyclopedia of Genes and Genomes) is a large database which stores 43,471 pathways. Address: http://www.genome.jp/kegg
- How is the glycosulfatase represented?



- Note that H20 and sulfate molecules are not present. They are considered not relevant.

Example: Part of the glycolysis pathway



Problem 1: reversibility

Reversibility

A reaction is reversible when it can work either taking the substrate and giving the product or vice versa. The actual verse is determined by the type of the reaction and the chemical state of the cell.

- Most of the reactions in a pathway are reversible.
- Representing a reaction and its inverse by a Petri Net could give a possible dynamic behaviour where the cell keeps looping in producing and destroying the same substance. This is biologically meaningless.
- When most of the reactions are reversible, most of the T-Invariants of the net are not meaningful.
- Can reversibility be ignored?

On the reversibility

Glycolysis features excellent examples of the problem of reaction reversibility:

1. As glucose enters a cell it is immediately phosphorylated by ATP to glucose 6-phosphate in the irreversible first step. This is to prevent the glucose leaving the cell.
2. In times of excess lipid or protein energy sources glycolysis may run in reverse (gluconeogenesis) in order to produce glucose 6-phosphate for storage as glycogen or starch.

Problem 2: the kinetic factor

Kinetic

In nature the fact that something can happen does not mean that it is relevant for the behaviour of a system. Kinetic represents the speed of a reaction. If we consider a single irreversible reaction, speed is not important. When reactions compete for the substrate, the speed is relevant. It is not important the absolute speeds of the reactions, but their relative speeds.

Example:

- R is a reversible reaction,
- the speed of the reverse reaction, under certain conditions, is much lower than the direct one,
- the effects of the reversibility are unimportant.

Problem 3: locality

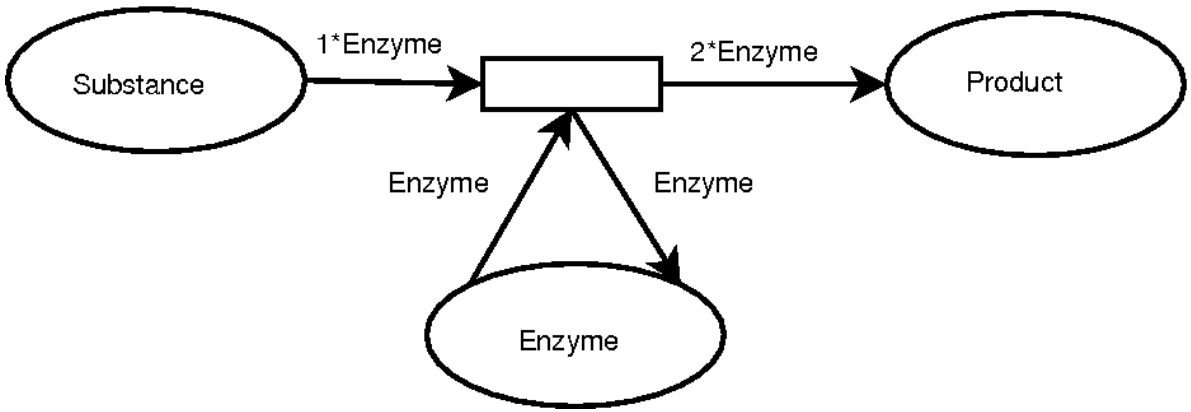
Locality

Is the fact that a substance is present in a cell enough to ensure that it can be used as substrate by a reaction?

- The answer is yes for most of the cases;
- Eucaryotes are big cells so the substance position can influence the reaction;
- Eucaryotes have some membranes inside which isolate parts of the cell. Substances inside and outside the membrane edge can not be considered for the same reactions;
- Petri Nets can not represent locality.

Self-modified Petri net

- Used for quantitative modelling by Hofestädt and Thelen
- main idea: assign to each arc of the Petri Net a function depending on the marking
- Example:



- Possible structural analysis?

Timed Petri Net

- Each transition firing requires a time
- Firing time can be a random variable (Stochastic Petri Net)
- Complex analysis if firing time are *not* modelled by exponential random variables (which is our case...) $\Rightarrow$ Infinite and dense reachability set.
- Easy simulation
- Most of reactions follow the Michaelis-Menten scheme:

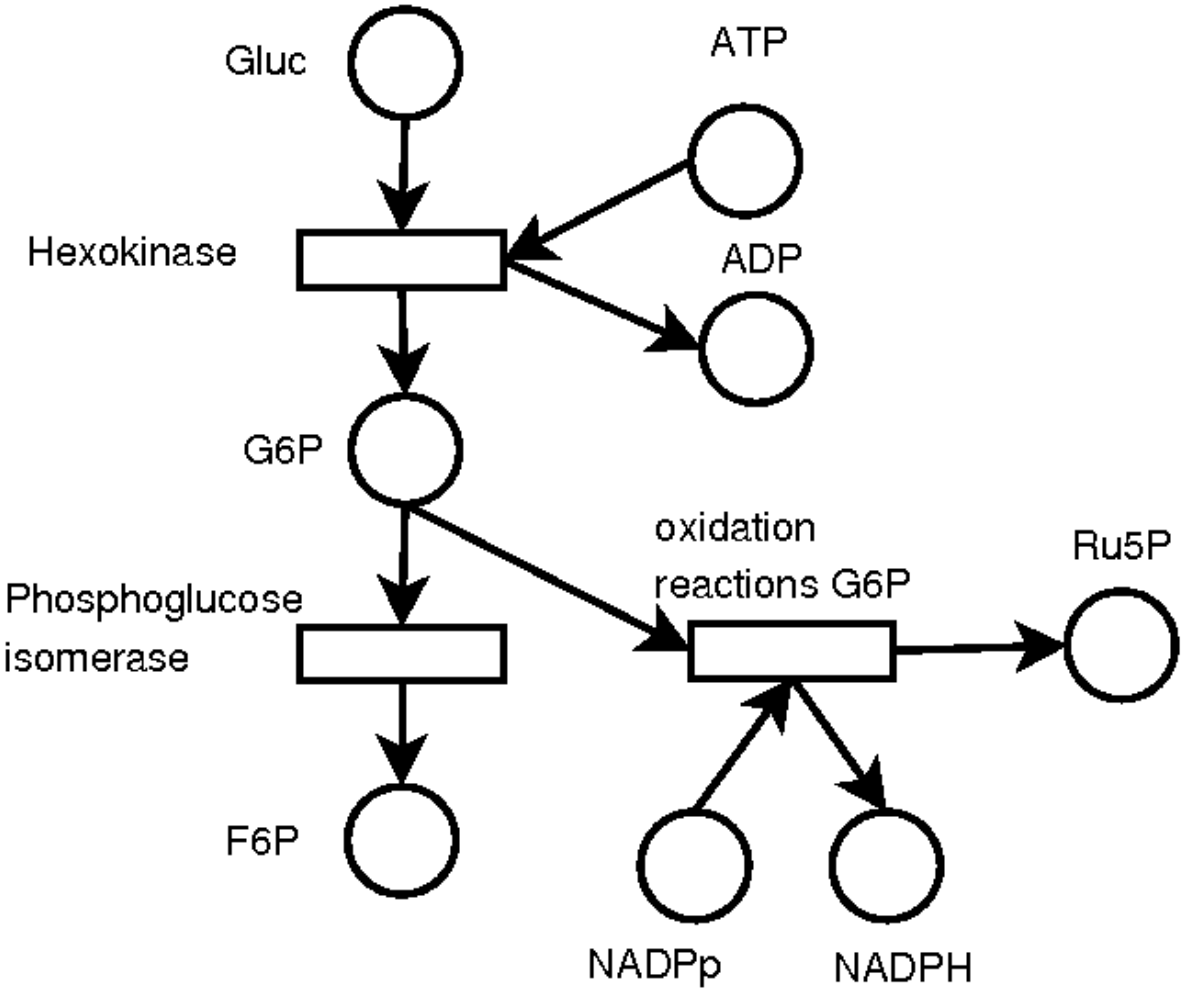
$$V = \frac{V_{max} \cdot S}{S + K_m}$$

where V_{max} is the maximum reaction rate, S the substrate concentration and K_m characterises the interaction Enzyme/Substrate.

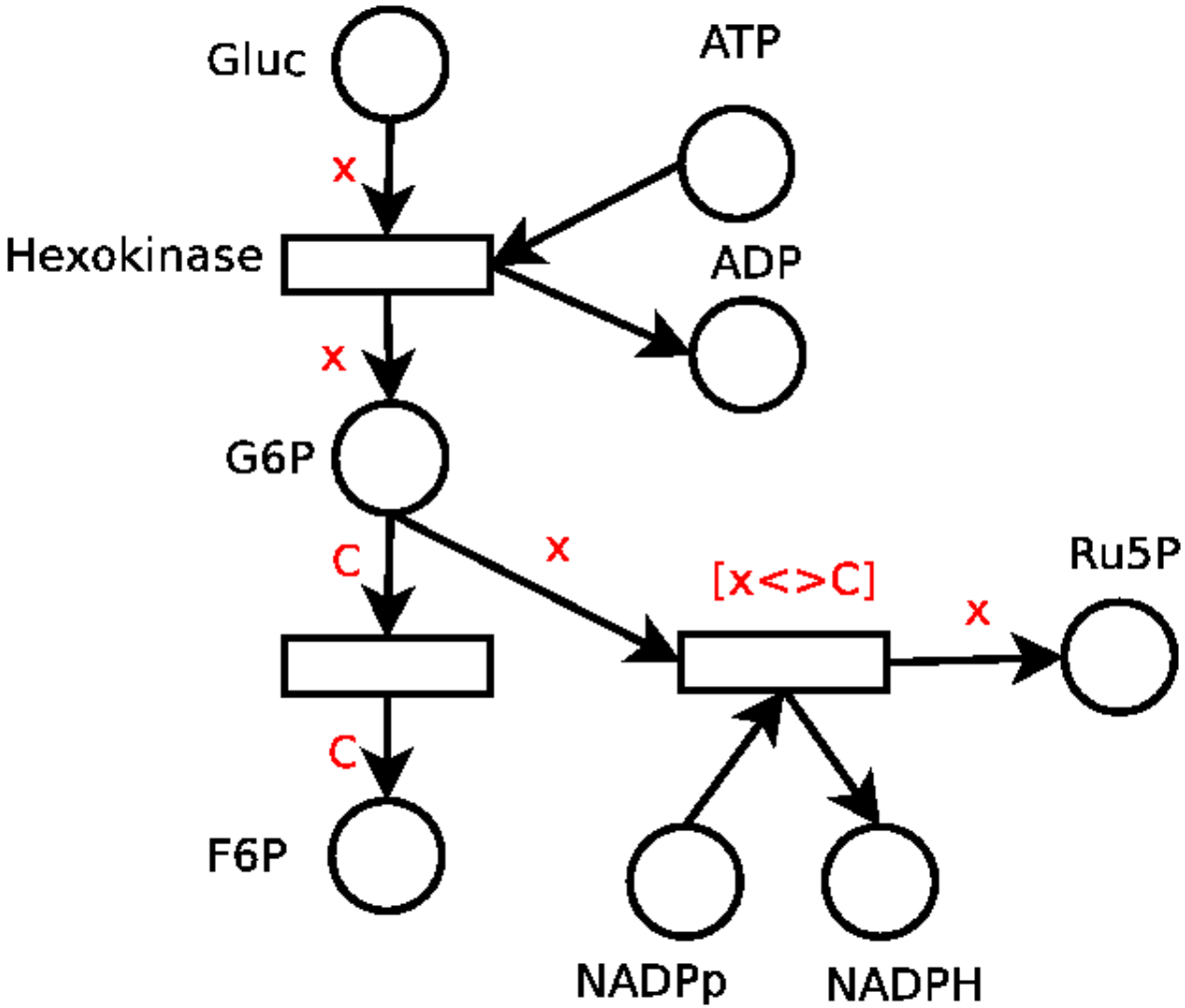
High Level Petri Net

- Used for qualitative analysis and simulation by Koch, Heiner, Voss
- Main idea: each token has a colour, and transition can be labelled with tests on token colours
- For example we can have a transition which fires just with blue token in a place and not with green ones
- See examples in later slides...

Conflict in glycolysis and pentose pathways



HLPN for the conflict



Remarks on HLPN

- The model is not intuitive. It somehow requires to think that the cell has a will.
- It does not seem possible to obtain a HLPN starting from a KEGG diagram algorithmically due to the required simplifications.
- Is there an algorithm to assign colours to the tokens and guards to the transitions?
- It allows better invariant analysis than standard Petri Nets by the decomposition of the net.
- Guards can solve the problems of reversible reactions.
- It does not require a simulator tool.

Qualitative or Quantitative modelling?

The answer to this question is not easy for these reasons:

- Kinetic and quantitative factors determine if an even is important or not, while qualitative models just say that it can happen.
- Biochemists do not agree on which simplifications of the models are possible. It depends much on the results we aim to get.
- Self-modified PN and Timed PN (better a combination of both) are flexible models but they can not be structurally studied (for what we know) for invariant or steady state by algebraical methods. They are excellent models for simulation.
- Qualitative analysis require simplifications on the pathway. It is not clear which are reasonable and which are not.

Example of simplifications

Taken from "Analysis and Simulation of Steady States in Metabolic Pathway with PN" by Heiner, Koch et al

- A great number of reactions can be treated as *irreversible* under normal conditions. [...] In principle each enzymatic reaction is *reversible*.
- For a given reactant concentrations, the higher the concentrations of the product gets, the slower the reaction will occur in the proffered direction [...]. The approach described in this paper concentrates on the mere structure of the pathways.
- Those reactions with high reactant concentration are preferred to those with low. [...] Molecules of the same substance are chemically indistinguishable although their roles may differ depending on the reaction environment in which they appear.

Regulatory pathways

Environmental information processing

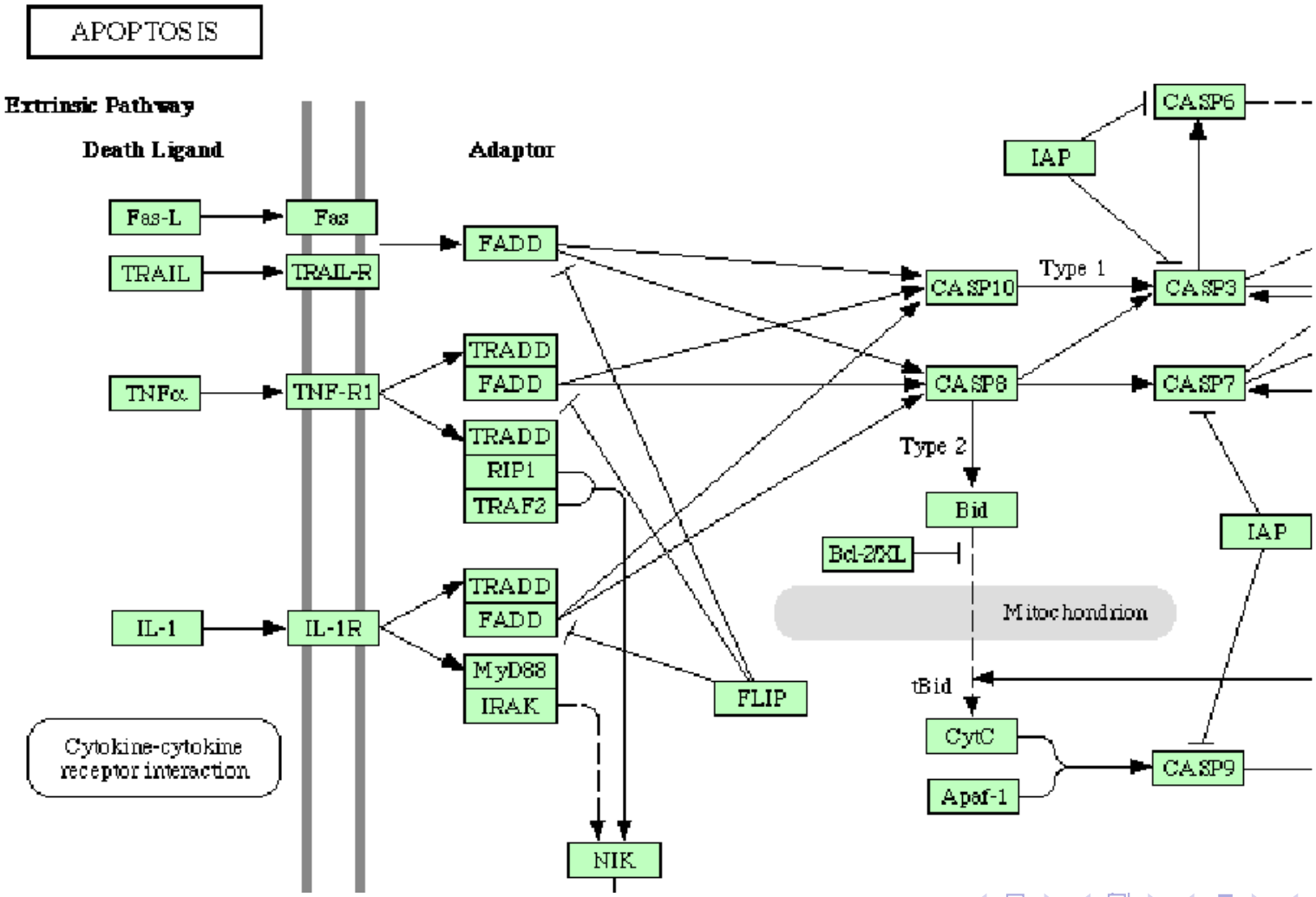
Types: Membrane transport, Signal transduction, Ligand receptor interaction

Cellular processes

Types: Cell motility, Cell growth and death, Cell communication, Development, Behaviour

- They are formed by cascades of activated/deactivated proteins or proteins complexes
- They detect, amplify and integrate external signals to generate responses such as changes in enzyme activity, gene expression, or ion-channel activity

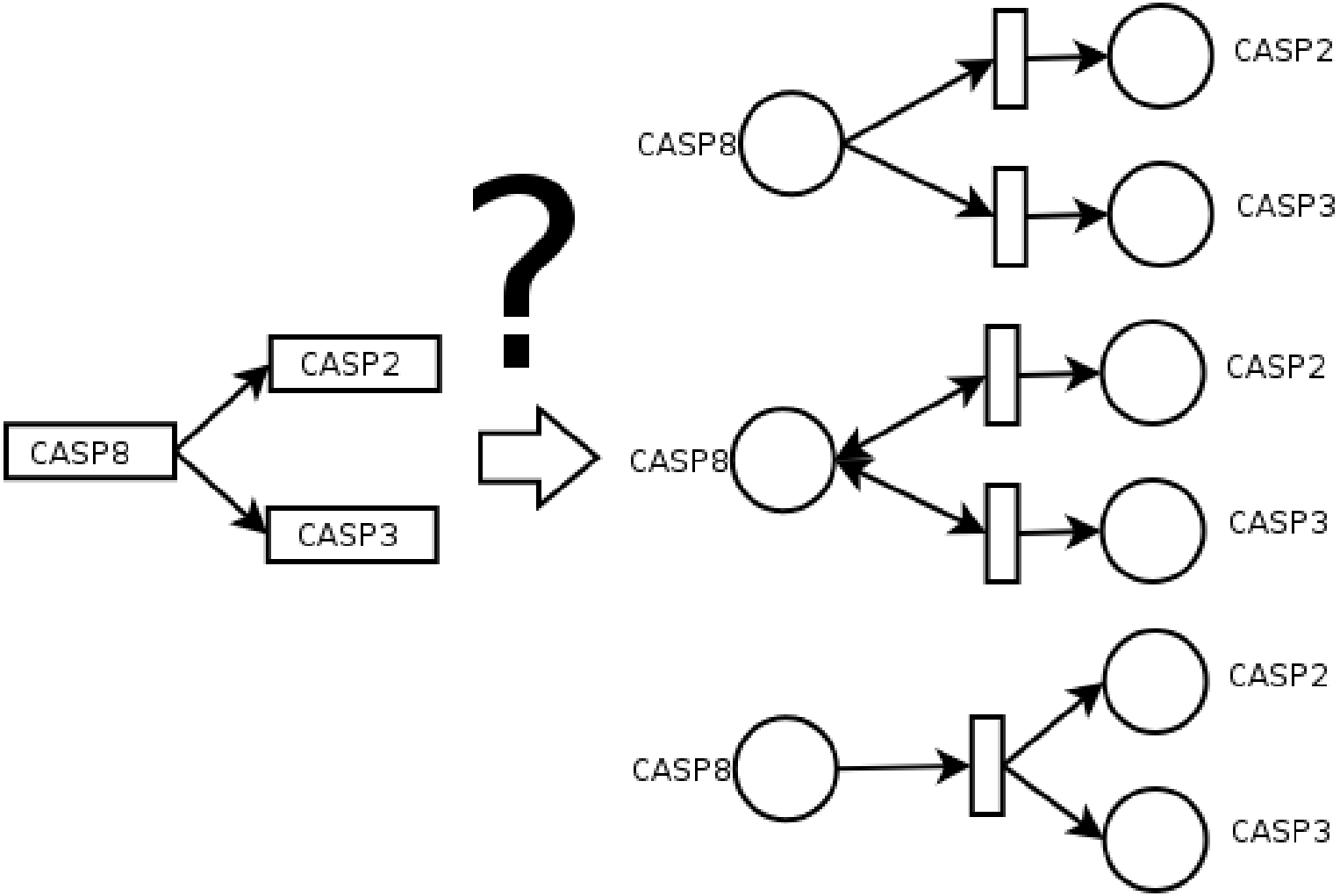
Example of signal transduction pathway. Apoptosis



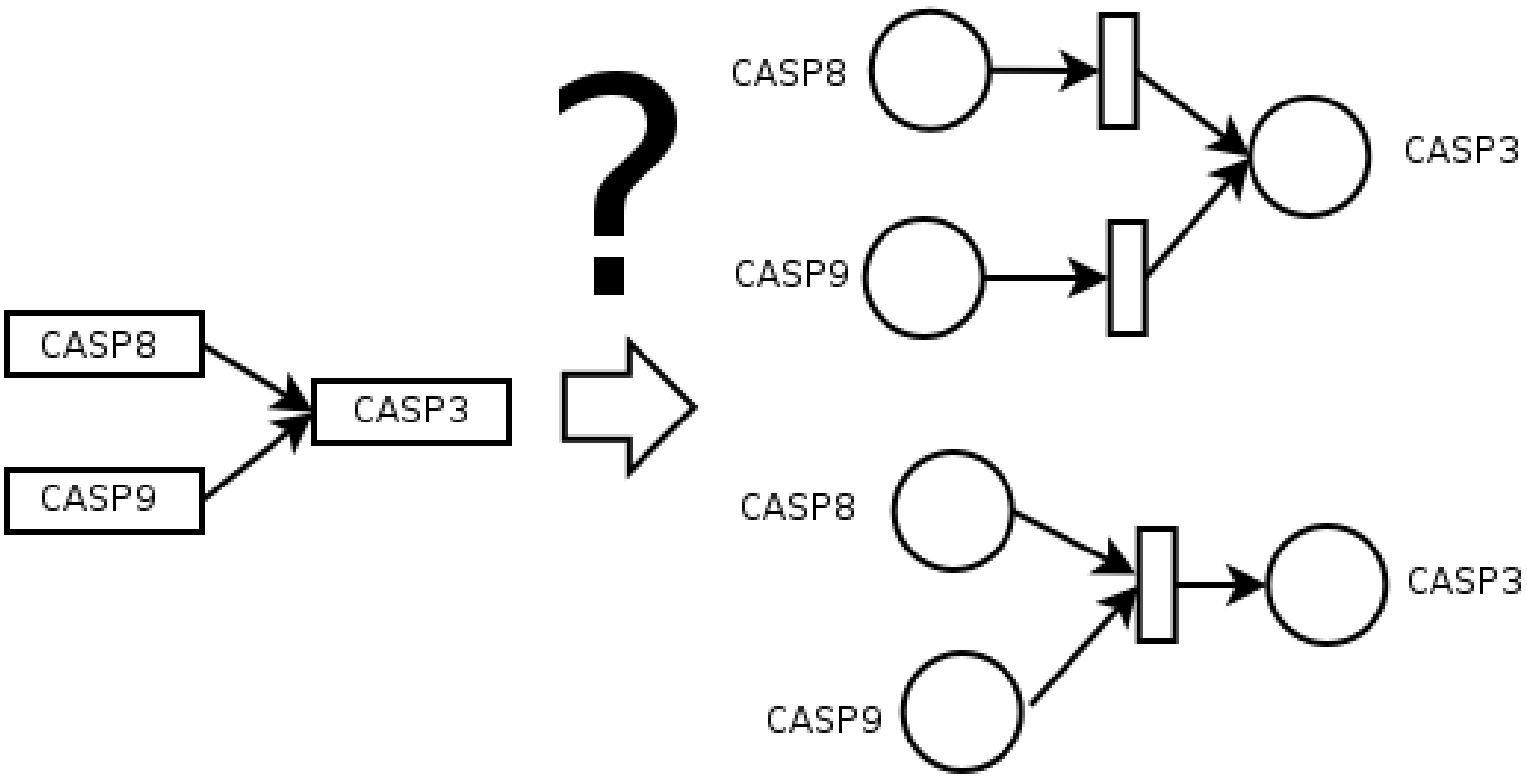
Yet another PN extension

- The easier way to represent the inhibition of a substance is to use inhibitor arcs in PN.
- If the inhibitor arc (p_i, t_j) is labelled with the natural number l then a necessary condition for the marking $\mathbf{m}$ to enable t_j is than $m_i < l$.
- The reachability problem for a PN with inhibitor arcs is equivalent to the halting problem for the Turing Machine.

A semantic ambiguity in KEGG



Another semantic ambiguity in KEGG



Again qualitative or quantitative approach?

- Take the protein (complex) α which is activated by β and inhibited by γ. In Petri Net theory, the presence of β and γ does not enable the transition.
- In biology the state of the protein (complex) α depends on the concentration of the activator and the inhibitor.
- How can we determine the validity range of a qualitative analysis?
- Self-modified Petri Nets is a good model for quantitative analysis.

Regulation of gene expression

Definition

Regulation of gene expression (gene regulation) is the cellular control of the amount and timing of appearance of the functional product of a gene.

Some notes:

- A functional gene product may be an RNA or a protein
- Most of the well-known mechanisms regulate the expression of protein coding genes
- We work on eucatyotes. In this context we call operon a set of genes which are regulated together
- 4 types of regulations: negative inducible operons, negative repressible operons, positive inducible operons, positive repressible operons

A (simplified) example of positive repressible operon: LAC of Escherichia Coli (1)

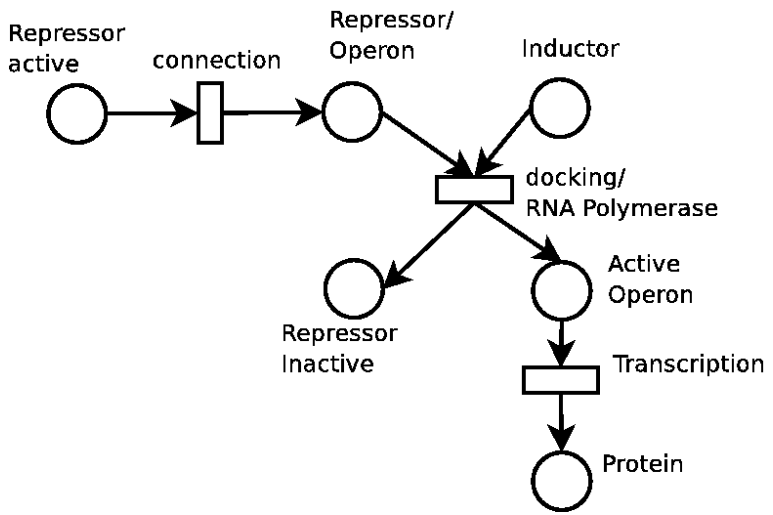
The E.C. can obtain carbon from lactose. This process needs specific enzymes which must be produced only when strictly needed.

- When there is no Lactose in the environment, the cell produces a protein called *lactose repressor*. This protein binds to the *lac-operator* and the transcription of the LAC Enzyme occurs at a very slow rate.
- When cells are grown in the presence of lactose, a lactose metabolite called allolactose binds to the lactose repressor. The repressor is unable to bind to the operator, allowing RNA Polymerase to transcribe the lac genes.
- Extension: what happens when glucose is present?

PN for negative inducible operons

Description

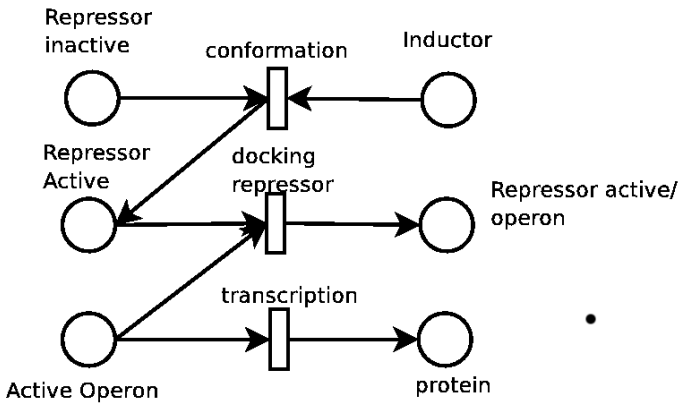
A regulatory repressor protein is normally bound to the operator and it prevents the transcription of the genes on the operon. If an inducer molecule is present, it binds to repressor and changes its conformation so that it is unable to bind to the operator. This allows for the transcription of the genes on the operator.



PN for negative repressible operons

Description

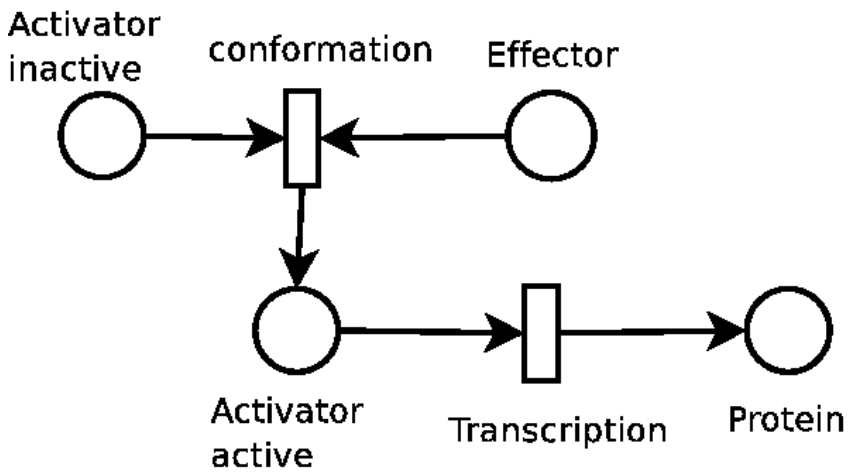
The transcription of the genes on the operon normally takes place. Repressor proteins are produced by a regulator gene but they are unable to bind to the operator in their normal conformation. Corepressors can bind to the repressor protein and change its conformation so that it can bind to the operator. The activated repressor protein binds to the operator and prevents transcription.



PN for positive inducible operons

Description

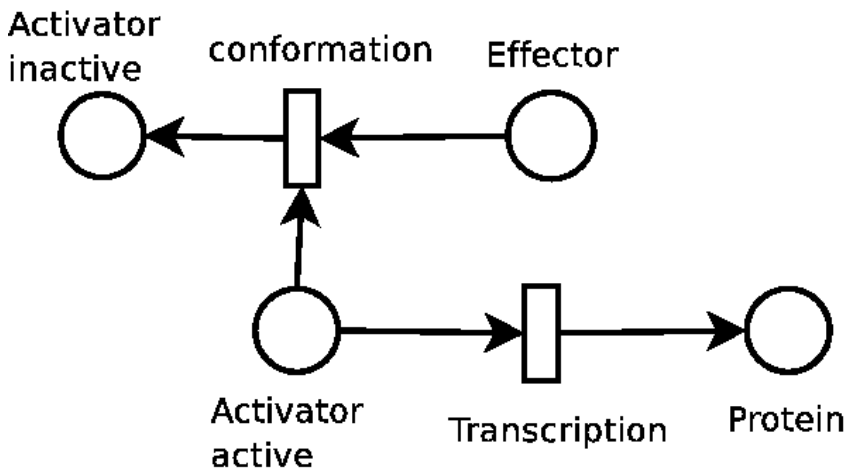
The activator proteins are normally unable to bind to the pertinent DNA. However, certain substrate molecules can bind to the activator proteins and change their conformations so that they can bind to the DNA and enable transcription to take place.



PN for positive repressible operons

Description

The activator proteins are normally bound to the pertinent DNA segment. However, certain molecules can bind to the activator and prevent it from binding to DNA. This prevents transcription.



Bibliography (1)

- The introduction section is mainly taken documents of bioinformatic section of Intelligent Software Lab., Postech. Images are taken from Wikipedia.
- Elementary biological notions are taken for various biology textbooks.
- The application of High Level Petri Nets to metabolic pathways are taken from Koch et al. works, e.g. *Analysis and simulation of Steady States in Metabolic Pathways with Petri Nets.*
- Examples of ambiguity of KEGG are taken from the Koch et al. paper: *Model validation of biological pathways using Petri nets - demonstrated for apoptosis.*

Bibliography (2)

- PN for gene regulatory network as well considerations on self-modified PN are taken from Hofestädt and Thelen paper: *Quantitative modeling of Biochemical Networks.*
- A complete bibliography as well deeper analysis on formalisms can be found in the technical report *Biological Pathways Representation by Petri Nets and extensions* downloadable at http://www.dsi.unive.it/ marin/
- For any mistake or comment please write to marin@dsi.unive.it