Quantitative Evaluation of Biological Systems

Jane Hillston.
LFCS, University of Edinburgh

11th September 2006

Outline

Rough Timetable

- Introduction to Systems Biology (40 min)
- Biological models (45 min)
- Questions and Discussion (10 min)
- **Coffee Break 10:30 – 11:00**
- Performance Techniques applied to Systems Biology
  - Stochastic Activity Networks (20 min)
  - PEPA (40 min)
  - PRISM and Biological Reasoning (20 min)
- Summary, other work and conclusions (10 min)

Outline

Systems Biology

- Biological advances mean that much more is now known about the components of cells and the interactions between them.

Systems Biology

- Biological advances mean that much more is now known about the components of cells and the interactions between them.
- Systems biology aims to develop a better understanding of the processes involved.

Systems Biology

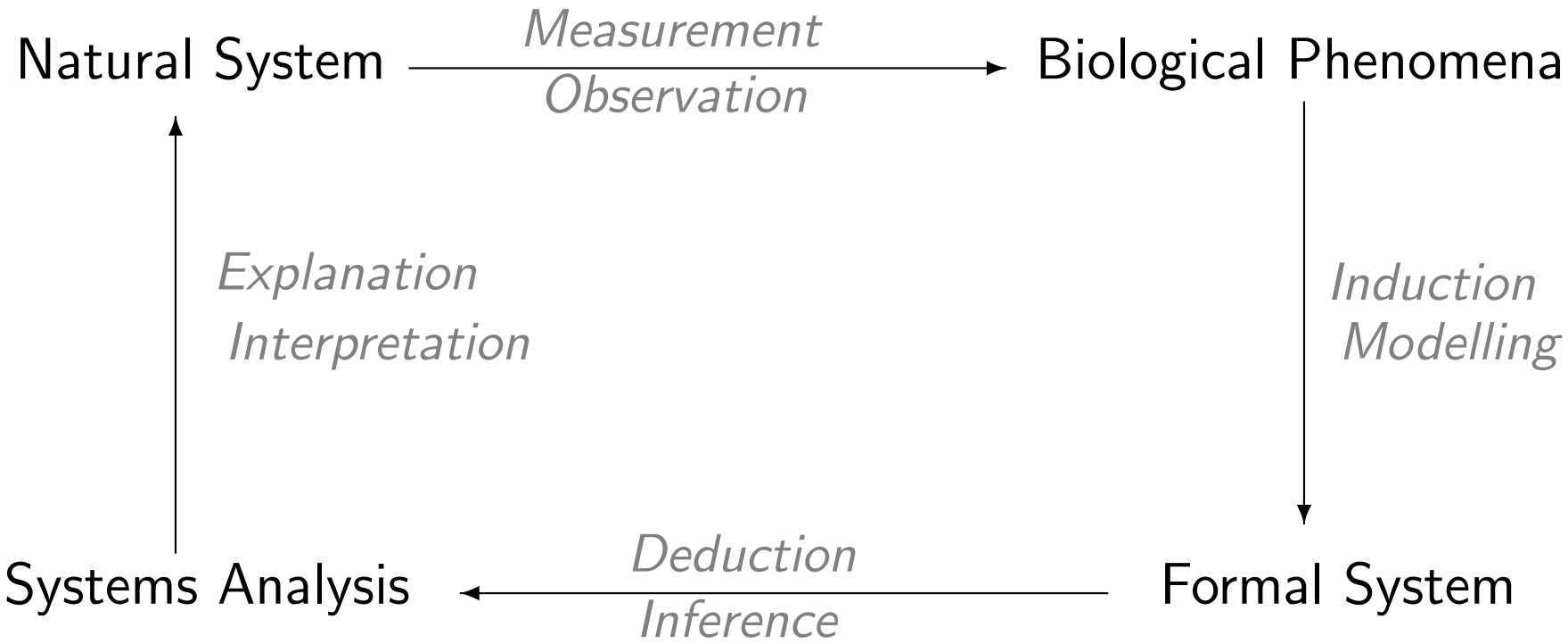
- Biological advances mean that much more is now known about the components of cells and the interactions between them.
- Systems biology aims to develop a better understanding of the processes involved.
- Formalisms from theoretical computer science have found a new role in developing models for systems biology, allowing biologists to test hypotheses and prioritise experiments.

What is Systems Biology?

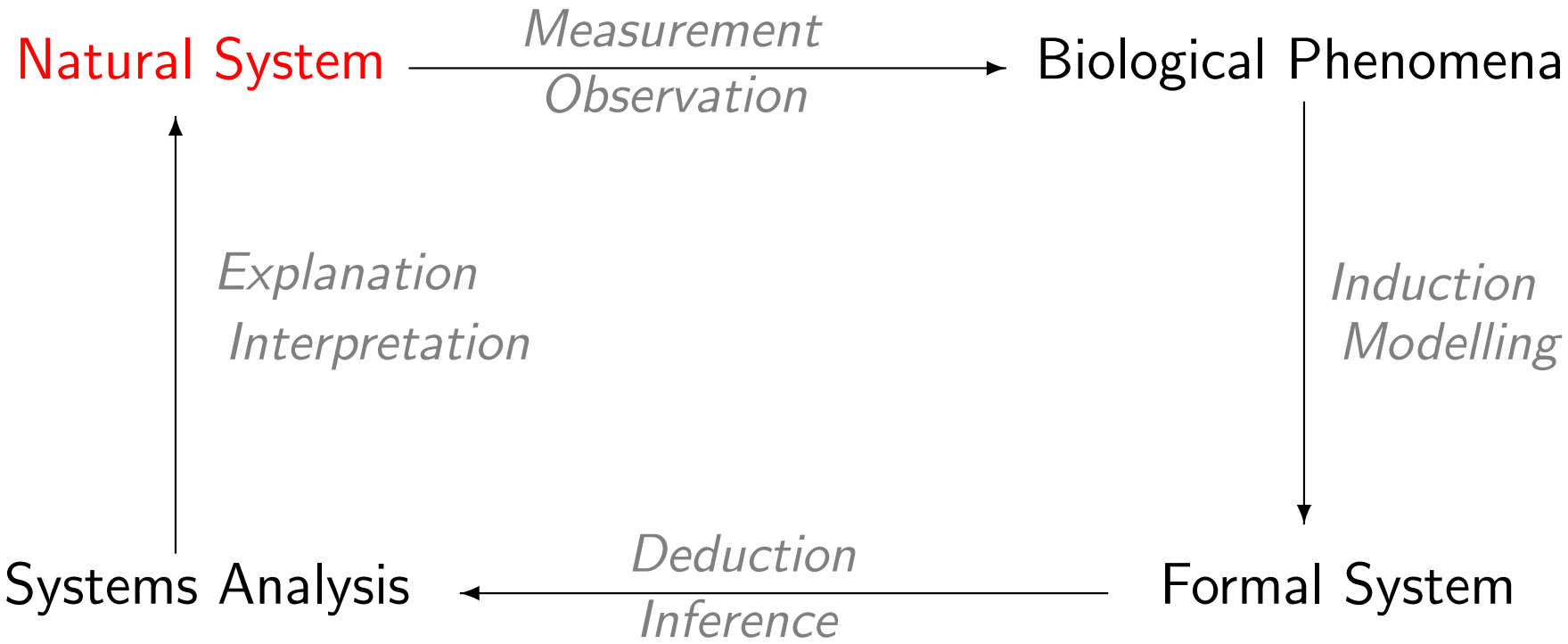
What is Systems Biology?

"The principal aim of systems biology is to provide both a conceptual basis and working methodologies for the scientific explanation of biological phenomena" – Olaf Wolkenhauer

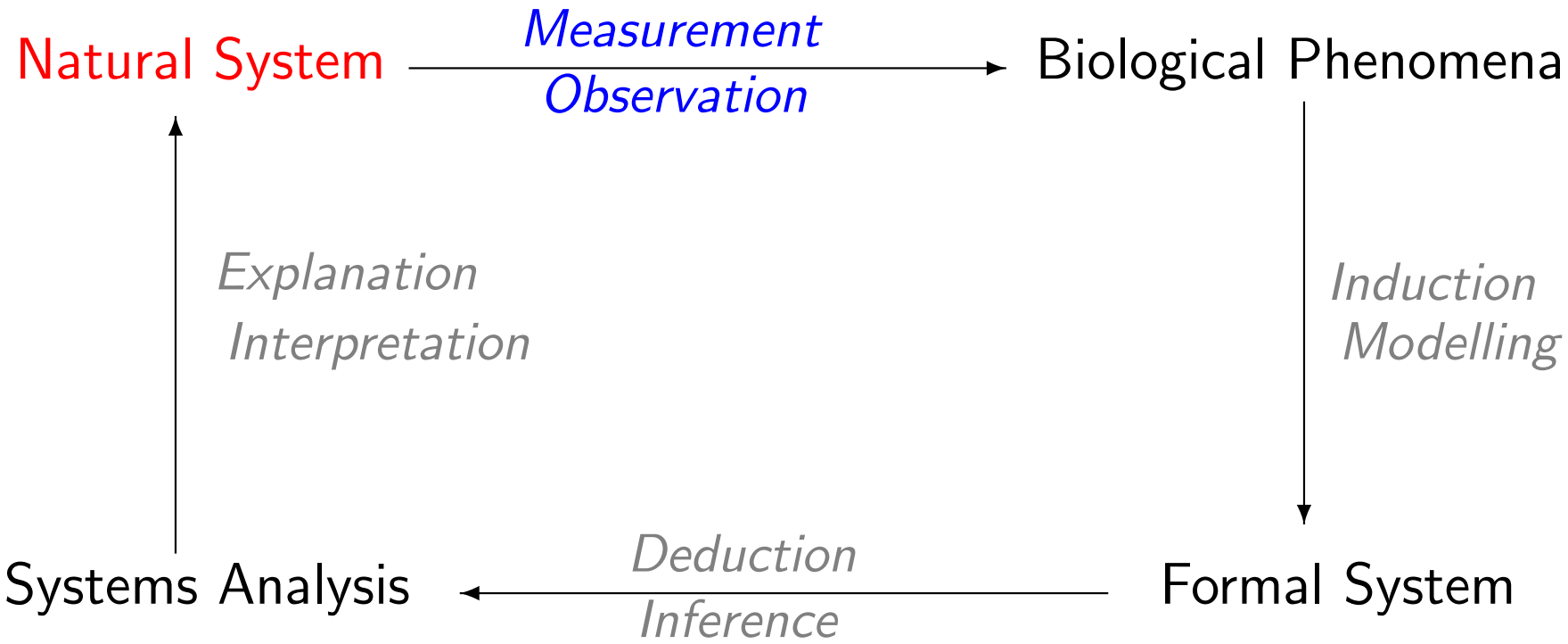
Systems Biology Methodology



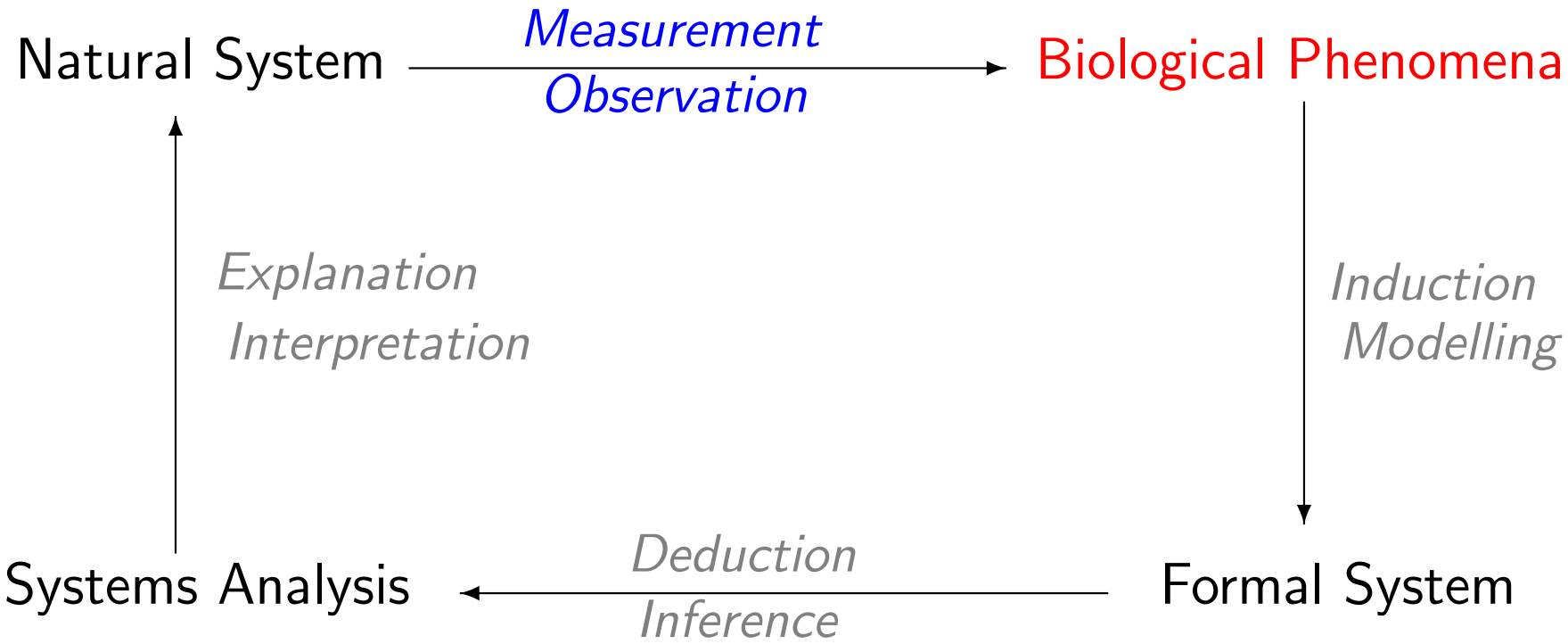
Systems Biology Methodology



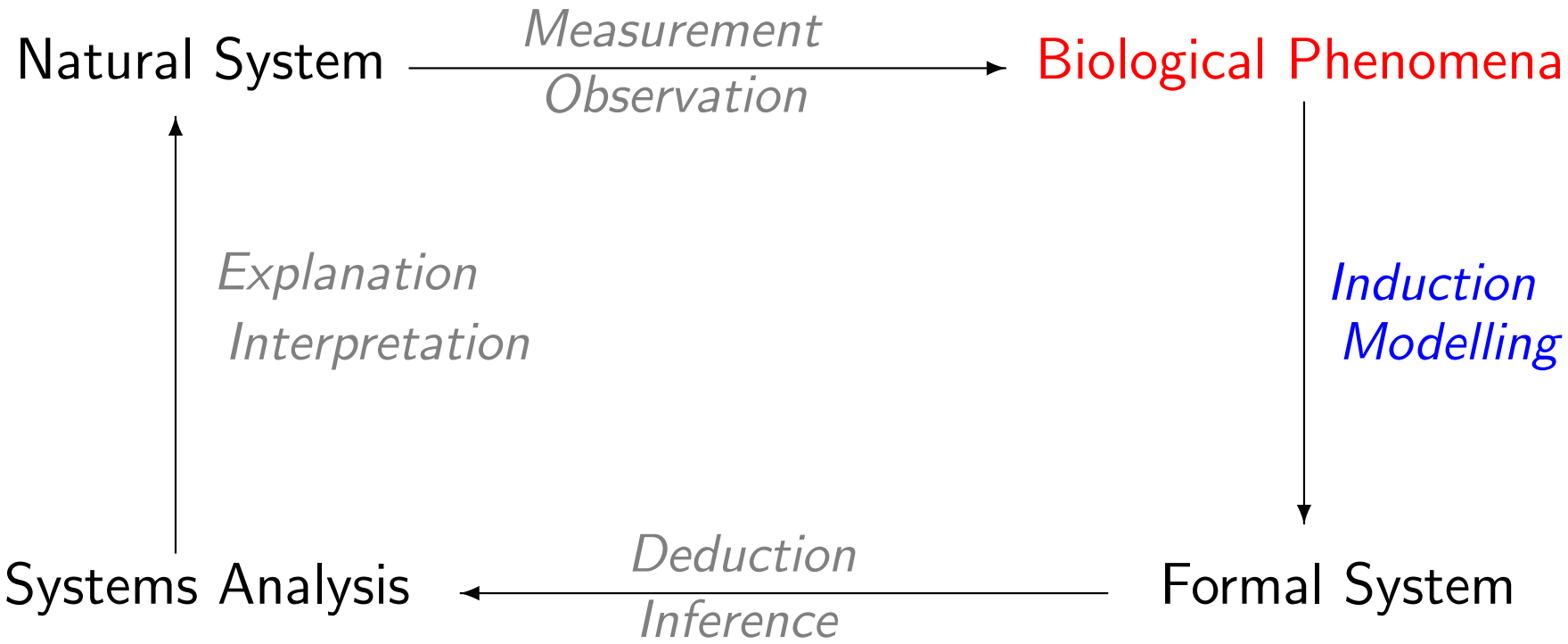
Systems Biology Methodology



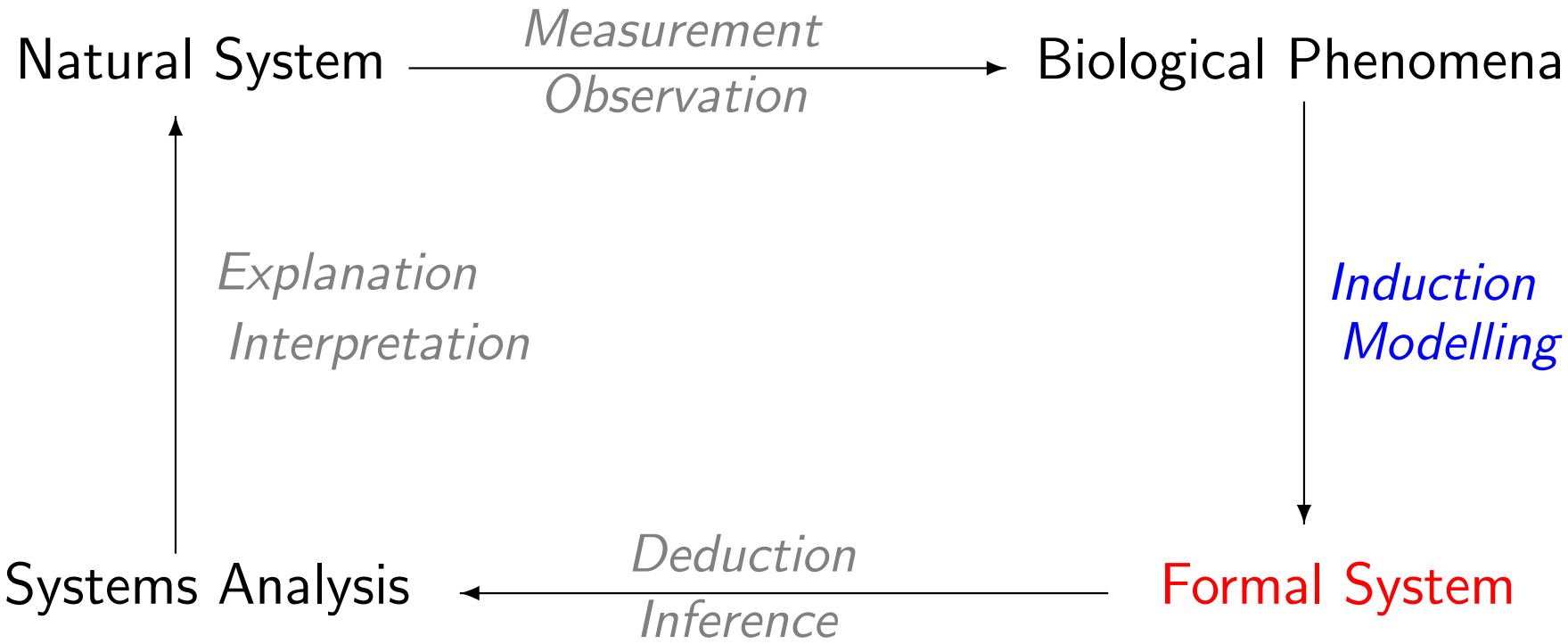
Systems Biology Methodology



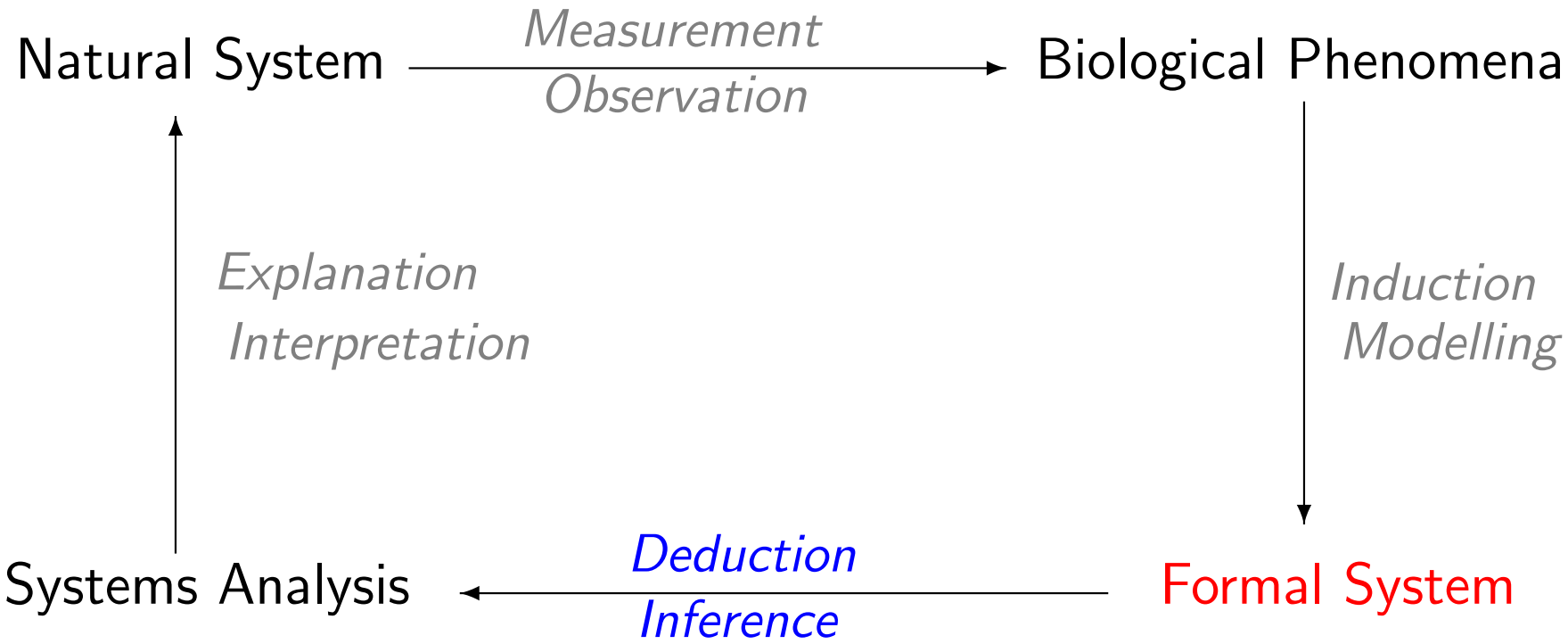
Systems Biology Methodology



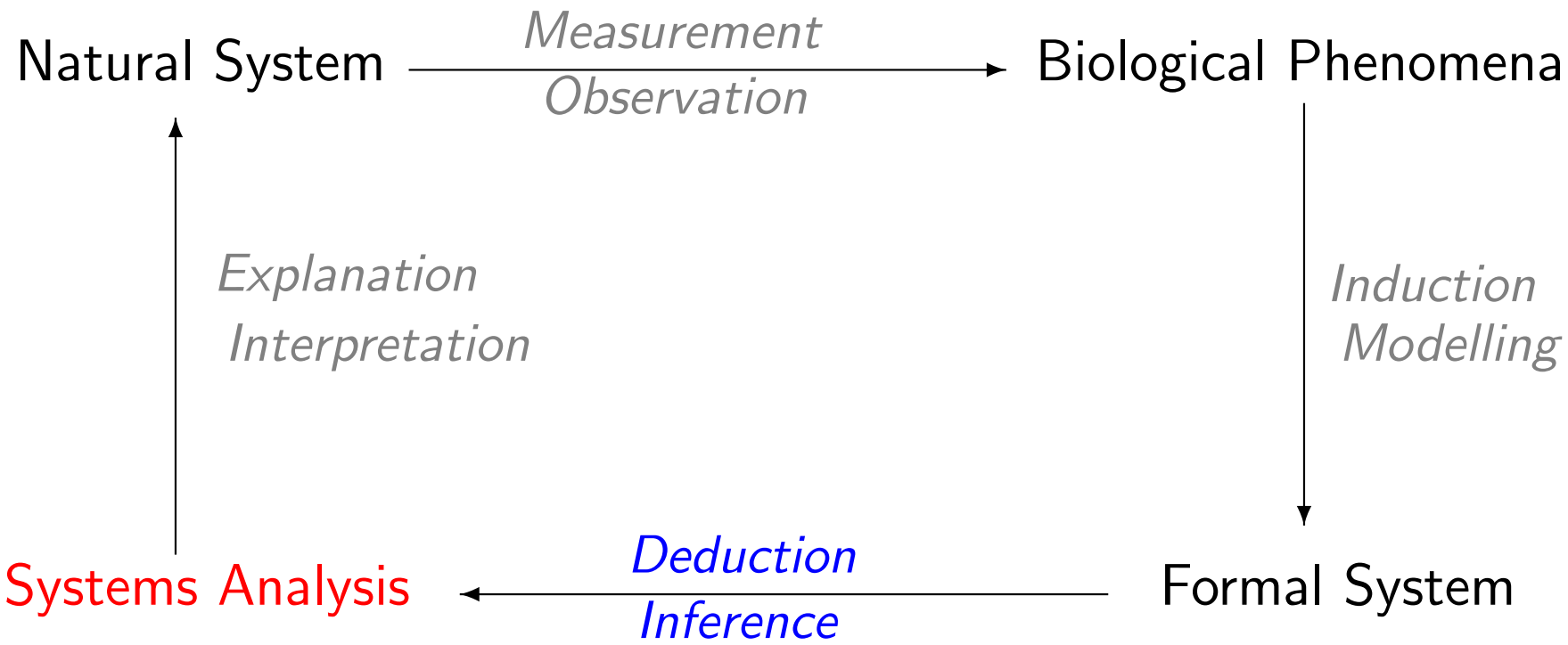
Systems Biology Methodology



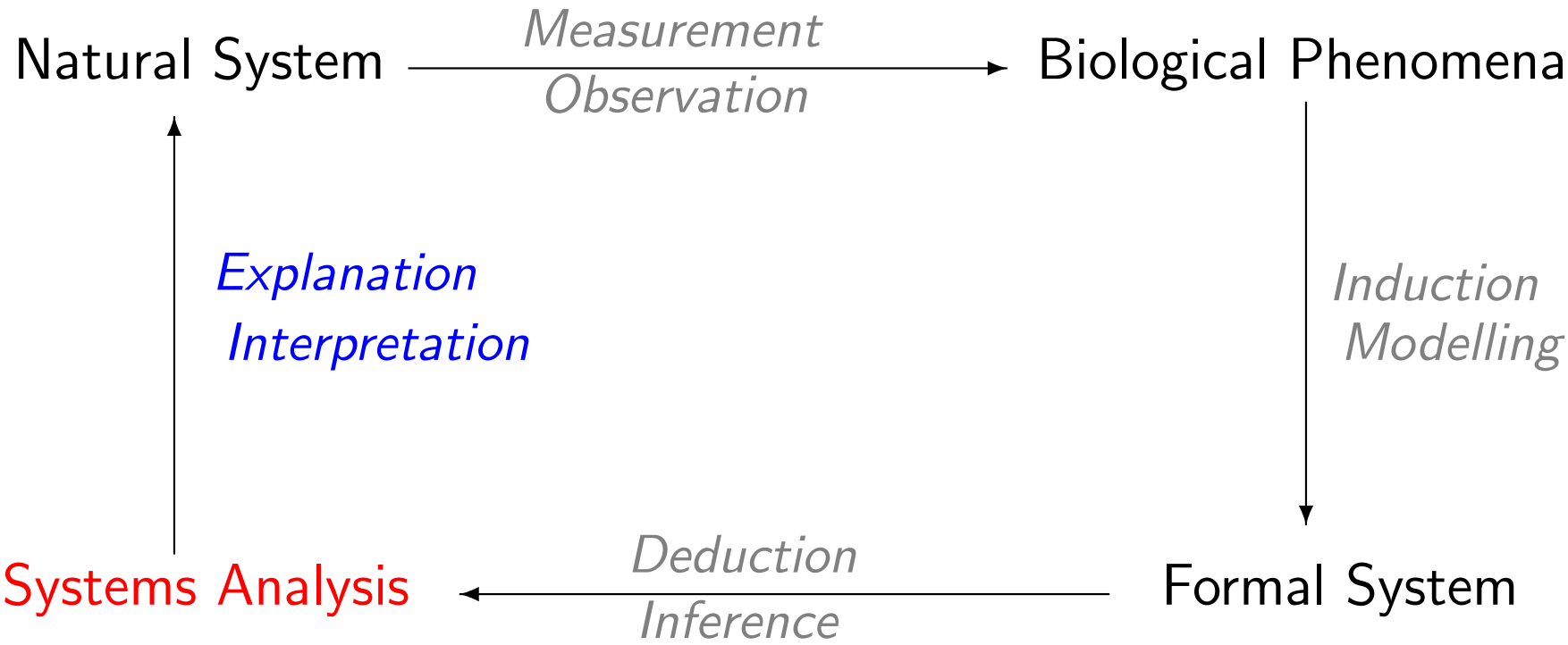
Systems Biology Methodology



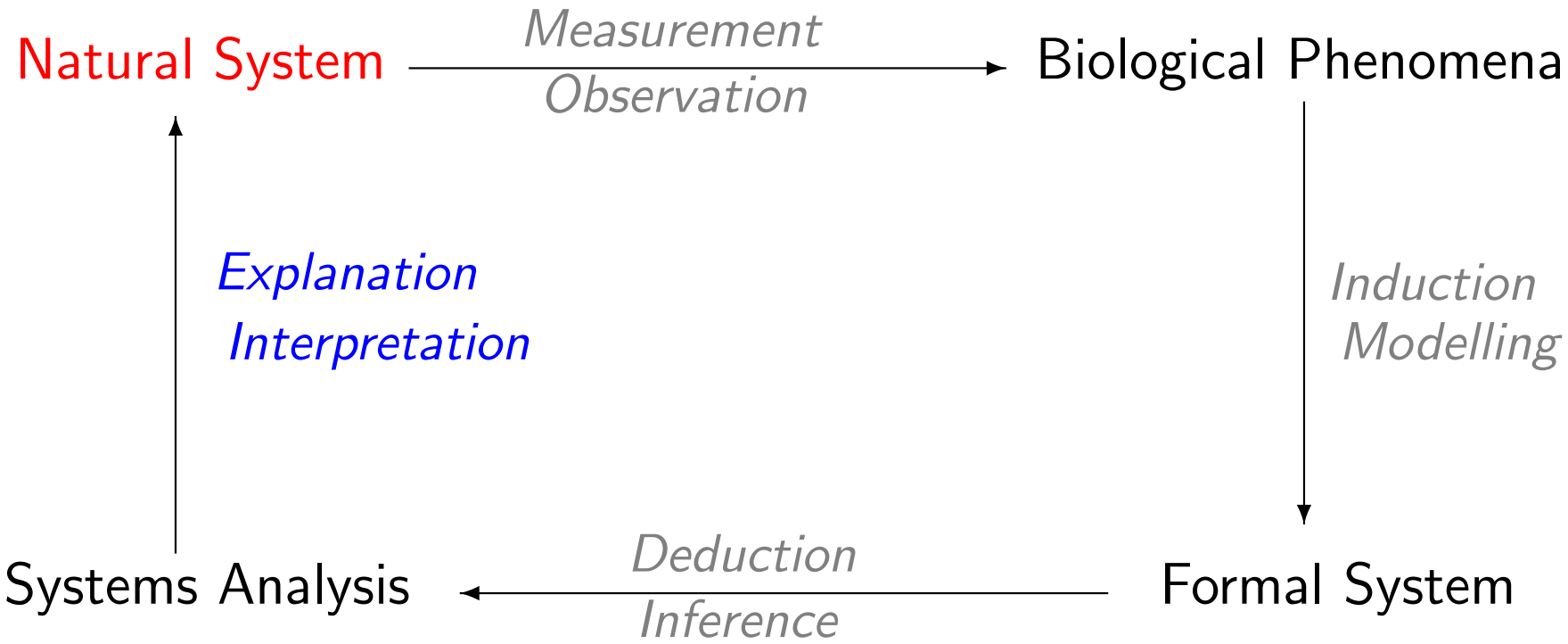
Systems Biology Methodology



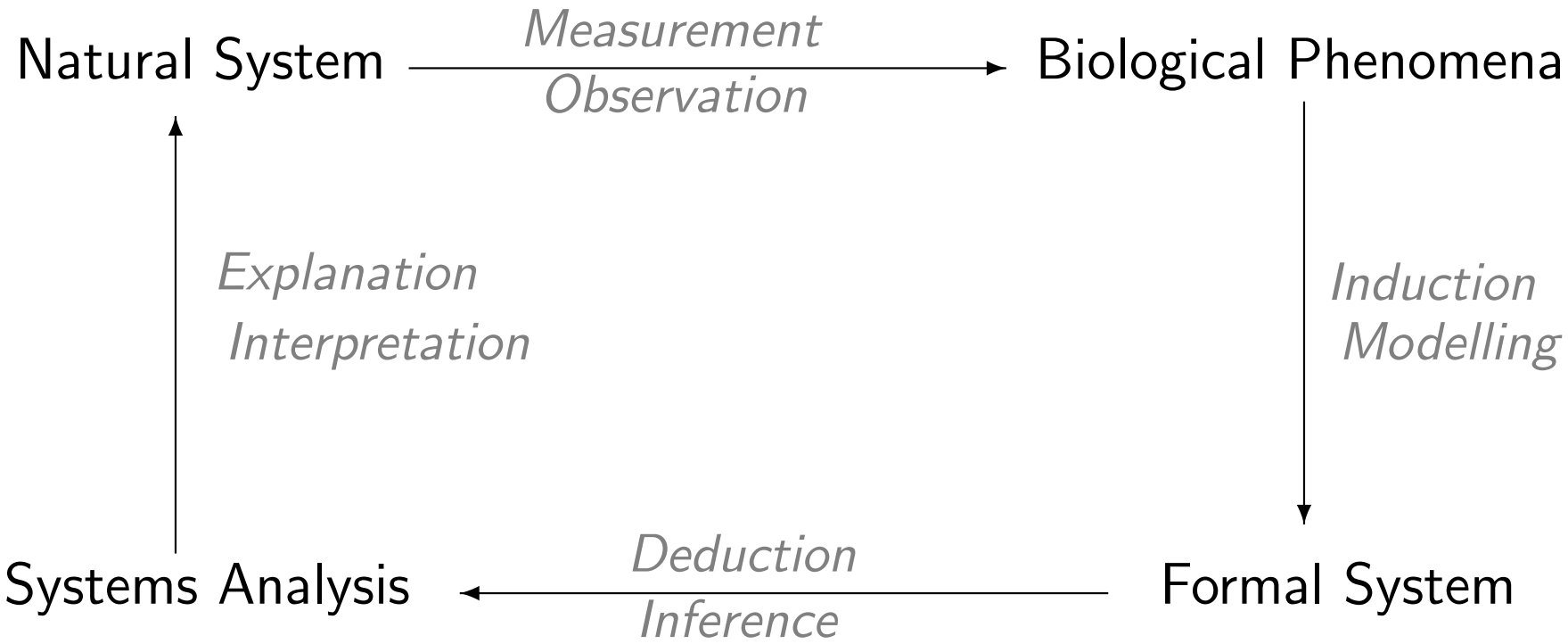
Systems Biology Methodology



Systems Biology Methodology



Systems Biology Methodology



Measurement, Observation and Induction

- Robot Scientist project — Kell, King, Muggleton *et al.*

Measurement, Observation and Induction

- Robot Scientist project — Kell, King, Muggleton *et al.*
- Combination of machine learning for hypothesis generation and genetic algorithms for automatic experimental tuning.

Measurement, Observation and Induction

- Robot Scientist project — Kell, King, Muggleton *et al.*
- Combination of machine learning for hypothesis generation and genetic algorithms for automatic experimental tuning.
- Experiments are carried out by a robot.

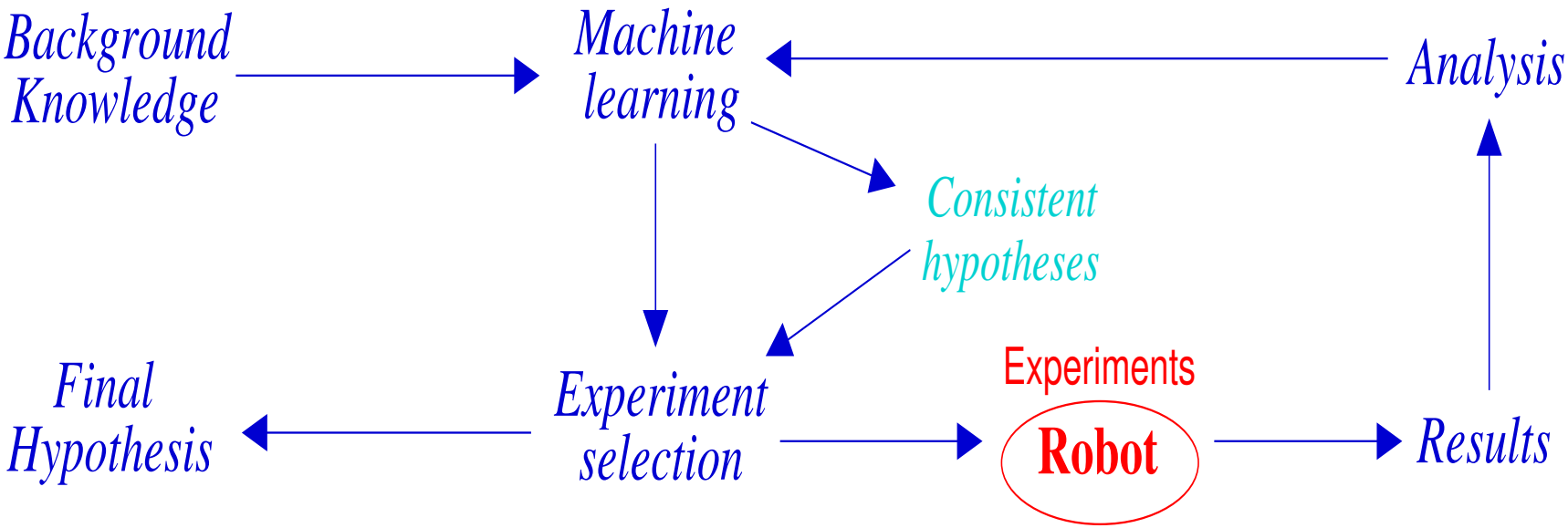
Measurement, Observation and Induction

- Robot Scientist project — Kell, King, Muggleton *et al.*
- Combination of machine learning for hypothesis generation and genetic algorithms for automatic experimental tuning.
- Experiments are carried out by a robot.
- Data is generated at rates which exceed what is possible when there are humans in the loop.

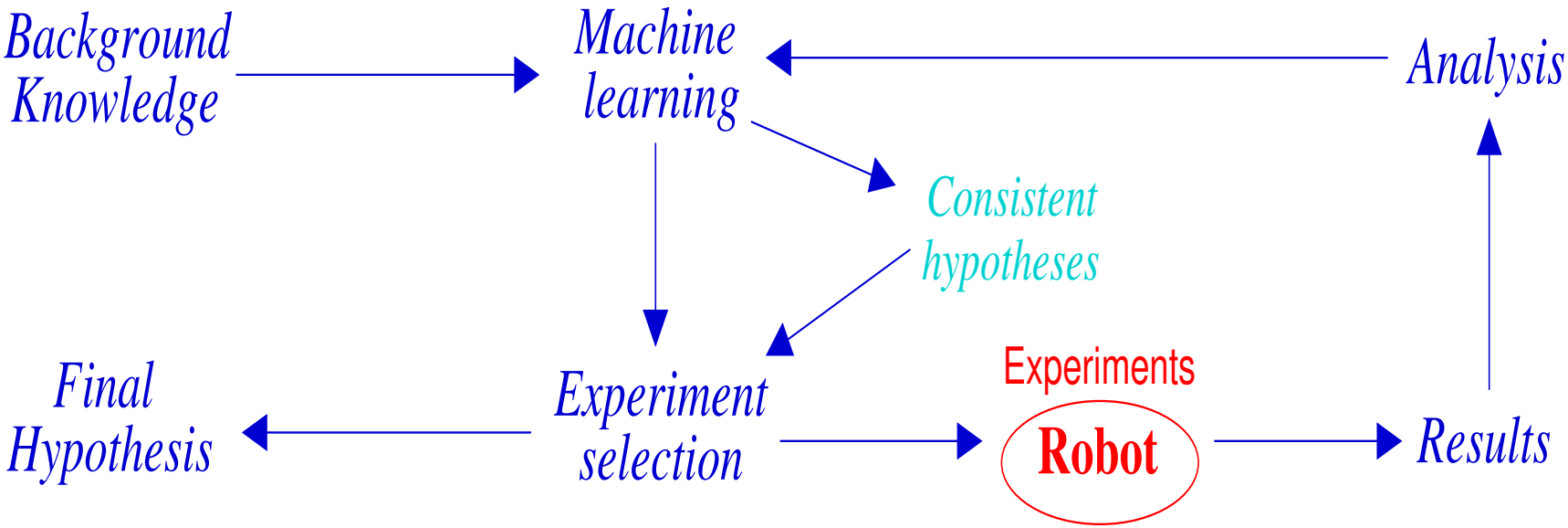
Measurement, Observation and Induction

- Robot Scientist project — Kell, King, Muggleton *et al.*
- Combination of machine learning for hypothesis generation and genetic algorithms for automatic experimental tuning.
- Experiments are carried out by a robot.
- Data is generated at rates which exceed what is possible when there are humans in the loop.
- Moreover the intelligent experiment selection strategy is competitive with (good) human strategies, and significantly outperforms *cheapest* and *random* selection strategies.

The Robot Scientist

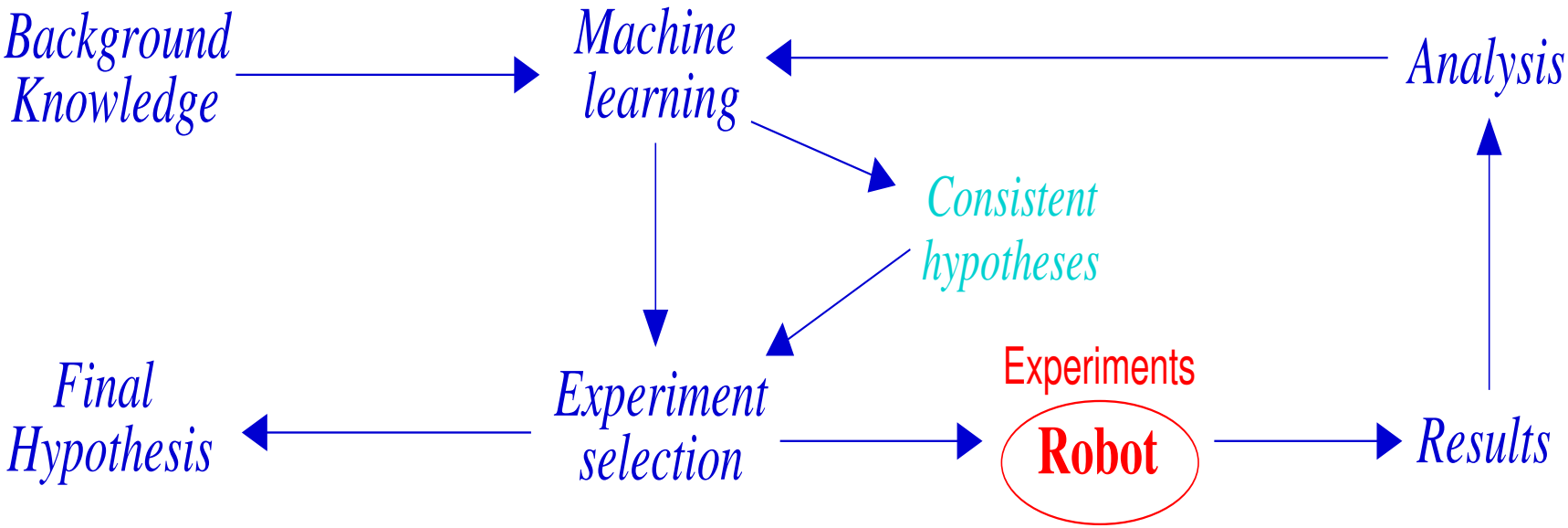


The Robot Scientist



- No human intellectual input in the design of experiments or the interpretation of data.

The Robot Scientist



- No human intellectual input in the design of experiments or the interpretation of data.
- Integrates scientific discovery software with laboratory robotics.

Case Study: Circadian Rhythms

J.C.W. Locke, M.M. Southern, L. Kozma-Bognár, V. Hibberd, P.E. Brown, M.S. Turner and A.J. Millar.

Extension of a genetic network model by iterative experimentation and mathematical analysis.

Molecular Systems Biology, msb4100018-E2, 2005.

D. Forger, M. Drapeau, B. Collins and J. Blau.

A new model for circadian clock research?

Molecular Systems Biology, msb4100019-E1, 2005.

Case Study: Circadian Rhythms – Overview

The study by Locke *et al.* focuses on the circadian rhythms in plants, combining mathematical models and molecular biology.

Case Study: Circadian Rhythms – Overview

The study by Locke *et al.* focuses on the circadian rhythms in plants, combining mathematical models and molecular biology.

Their objective is to identify the genes (and proteins) responsible for maintaining the daily rhythms observed in the plants.

Case Study: Circadian Rhythms – Overview

The study by Locke *et al.* focuses on the circadian rhythms in plants, combining mathematical models and molecular biology.

Their objective is to identify the genes (and proteins) responsible for maintaining the daily rhythms observed in the plants.

The research exploits an interplay between mathematical models, experiments in the laboratory and literature search.

Case Study: Circadian Rhythms – Overview

The study by Locke *et al.* focuses on the circadian rhythms in plants, combining mathematical models and molecular biology.

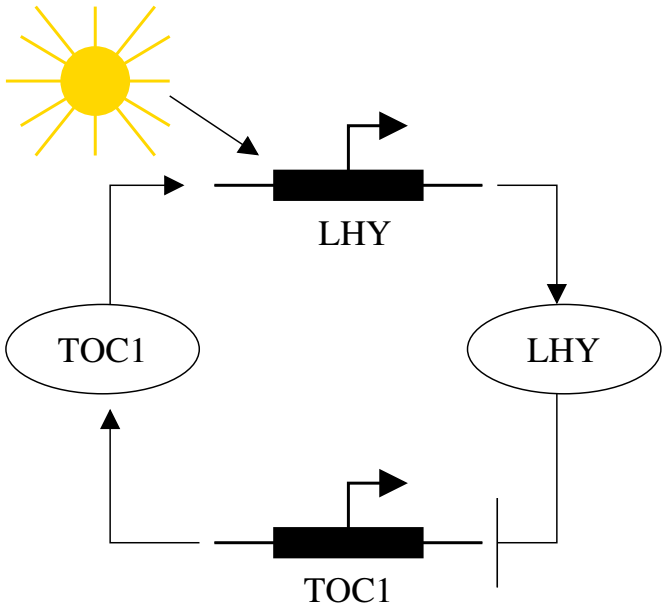
Their objective is to identify the genes (and proteins) responsible for maintaining the daily rhythms observed in the plants.

The research exploits an interplay between mathematical models, experiments in the laboratory and literature search.

It is held up as an exemplar of what systems biology is trying to achieve, and the breakthroughs that it can bring about when it is successful.

Case Study: Circadian Rhythms – Initial Model

From initial experiments Locke *et al.* identified a two genes and two proteins which appeared to operate in a simple loop:



An initial mathematical model (ODEs) was constructed to capture this model.

Case Study: Circadian Rhythms – Role of Mathematics

Initial simulations with the mathematical model showed good agreement with the experimental data for some of the observed phenomena but significant discrepancies for others.

Case Study: Circadian Rhythms – Role of Mathematics

Initial simulations with the mathematical model showed good agreement with the experimental data for some of the observed phenomena but significant discrepancies for others.

Experiments were then undertaken with the mathematical model to find an alternative model which was biologically plausible but produced a better fit.

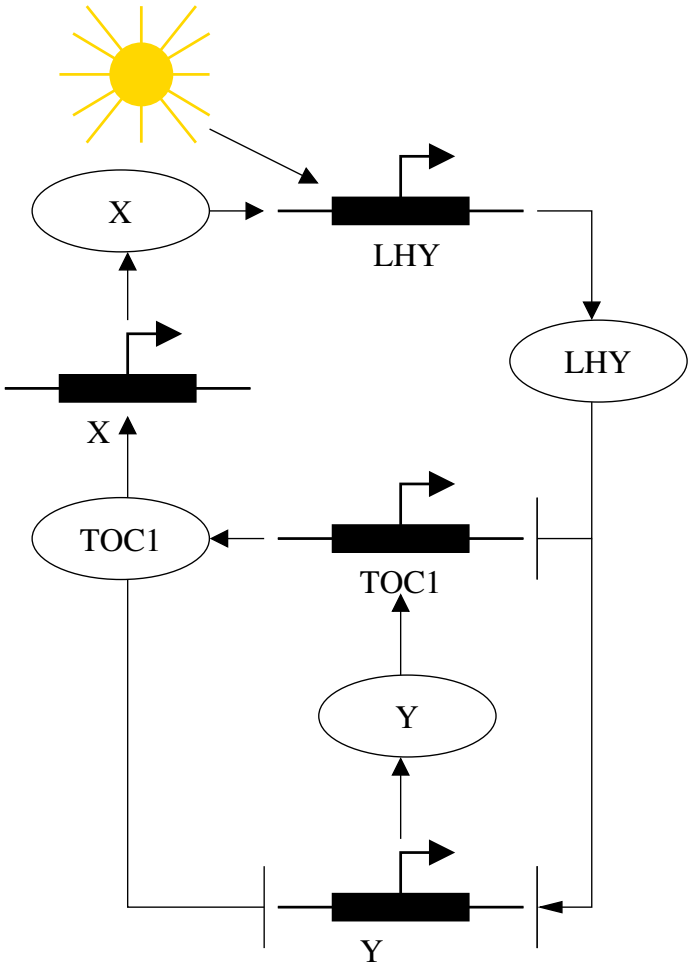
Case Study: Circadian Rhythms – Role of Mathematics

Initial simulations with the mathematical model showed good agreement with the experimental data for some of the observed phenomena but significant discrepancies for others.

Experiments were then undertaken with the mathematical model to find an alternative model which was biologically plausible but produced a better fit.

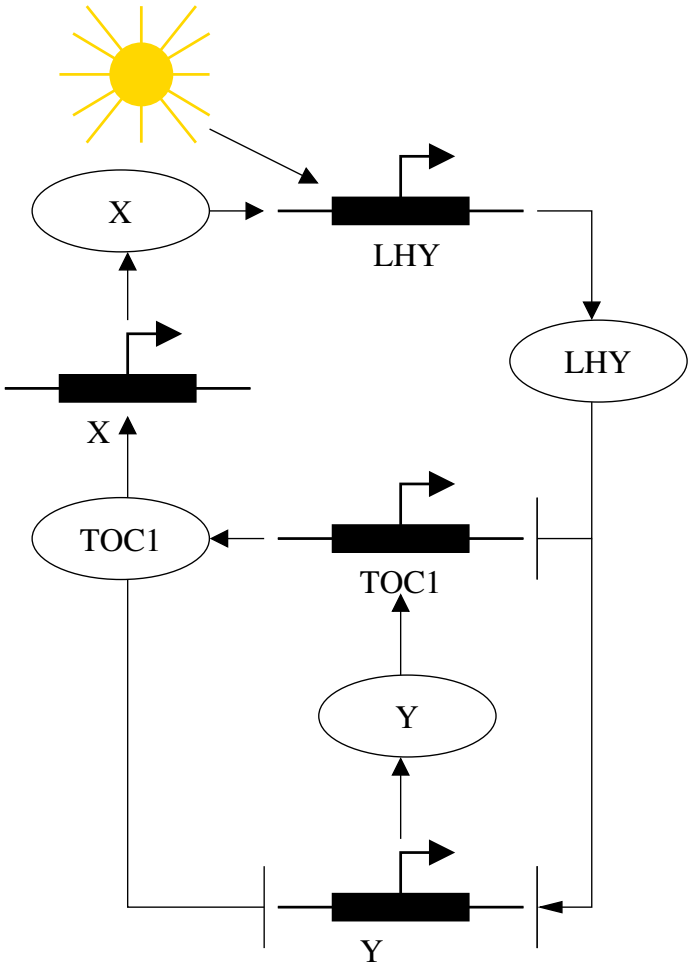
These mathematical experiments conjectured a network with two interacting loops.

Case Study: Circadian Rhythms – Elaborated Model



Two "new" genes were introduced to the model which now has interlocking loops and more complex feedback.

Case Study: Circadian Rhythms – Elaborated Model



Two "new" genes were introduced to the model which now has interlocking loops and more complex feedback.

The simulation results from this model showed much better agreement with the observed data.

Case Study: Circadian Rhythms – Validating the Model

The researchers then sought to identify the "new" genes X and Y.

Searching the literature elicited several candidate genes which previous experimental studies had suggested were implicated in the circadian rhythm.

In particular, "knockout" data for one, GIGANTEA (GI), coincided with the pattern from simulation experiments of the original model with a single loop.

Subsequent wet lab experiments have reinforced this impression that GI is gene Y.

Networks in cells

We can distinguish three distinct types of links or networks in cells

Networks in cells

We can distinguish three distinct types of links or networks in cells

Gene networks: Genes control the production of proteins but are themselves regulated by the same or different proteins.

Networks in cells

We can distinguish three distinct types of links or networks in cells

Gene networks: Genes control the production of proteins but are themselves regulated by the same or different proteins.

Signal transduction networks: External stimuli initiate messages that are carried through a cell via a cascade of biochemical reactions.

Networks in cells

We can distinguish three distinct types of links or networks in cells

Gene networks: Genes control the production of proteins but are themselves regulated by the same or different proteins.

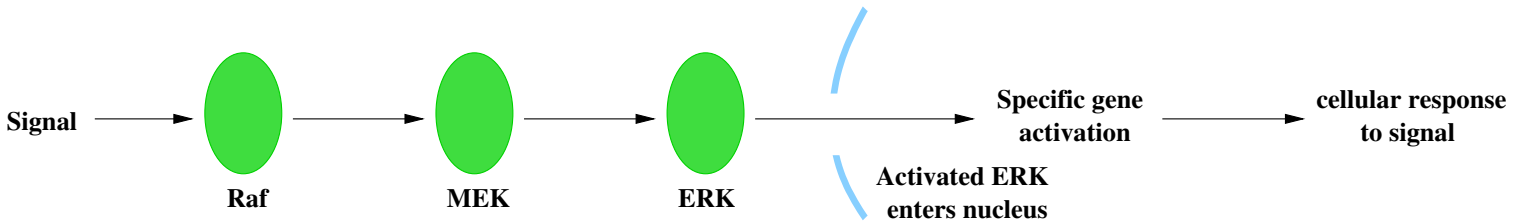
Signal transduction networks: External stimuli initiate messages that are carried through a cell via a cascade of biochemical reactions.

Metabolic pathways: The survival of the cell depends on its ability to transform nutrients into energy.

Extracellular signalling

Extracellular signalling — communication between cells.

- Signalling molecules released by one cell migrate to another;
- These molecules enter the cell and instigate a pathway, or series of reactions, which carries the information from the membrane to the nucleus;
- For example, the Ras/Raf-1/MEK/ERK pathway conveys differentiation signals to the nucleus of a cell.



Cell signalling

- All signalling is biochemical:

Cell signalling

- All signalling is biochemical:
- Increasing protein concentration broadcasts the information about an event; for example, that a gene promoter is “on”.

Cell signalling

- All signalling is biochemical:
- Increasing protein concentration broadcasts the information about an event; for example, that a gene promoter is "on".
- The message is "received" by a concentration dependent response at the protein signal's site of action.

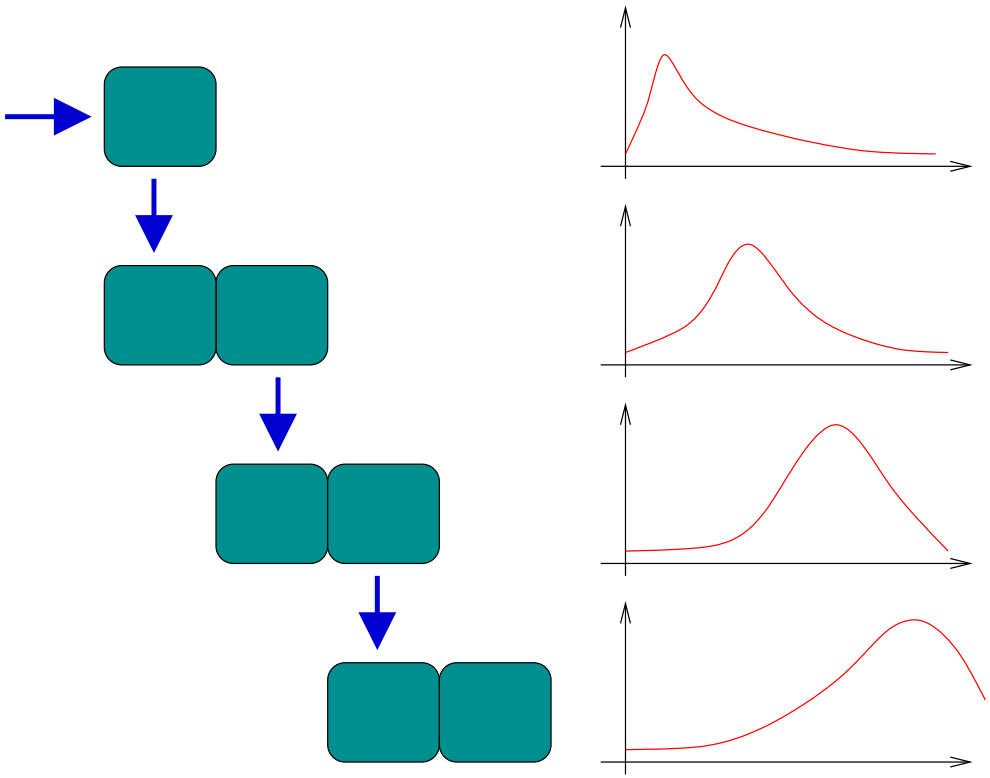
Cell signalling

- All signalling is biochemical:
- Increasing protein concentration broadcasts the information about an event; for example, that a gene promoter is "on".
- The message is "received" by a concentration dependent response at the protein signal's site of action.
- This stimulates a response at the signalling protein's site of action.

Cell signalling

- All signalling is biochemical:
- Increasing protein concentration broadcasts the information about an event; for example, that a gene promoter is "on".
- The message is "received" by a concentration dependent response at the protein signal's site of action.
- This stimulates a response at the signalling protein's site of action.
- Signals propagate through a series of protein accumulations.

Signal transduction pathways



A series of biochemical reactions serve to pass a message from the cell membrane to the nucleus.

Gene expression pathways

- Genetic activity is controlled by molecular signals that determine when and how often a given gene is transcribed.

Gene expression pathways

- Genetic activity is controlled by molecular signals that determine when and how often a given gene is transcribed.
- The product encoded by one gene often regulates the expression of other genes.

Gene expression pathways

- Genetic activity is controlled by molecular signals that determine when and how often a given gene is transcribed.
- The product encoded by one gene often regulates the expression of other genes.
- Moreover two or more proteins may act together to activate or repress a gene.

Gene expression pathways

- Genetic activity is controlled by molecular signals that determine when and how often a given gene is transcribed.
- The product encoded by one gene often regulates the expression of other genes.
- Moreover two or more proteins may act together to activate or repress a gene.
- For appropriate combinations of input signals transcription is initiated and protein product accumulates when production exceeds degradation.

Gene expression pathways

- Genetic activity is controlled by molecular signals that determine when and how often a given gene is transcribed.
- The product encoded by one gene often regulates the expression of other genes.
- Moreover two or more proteins may act together to activate or repress a gene.
- For appropriate combinations of input signals transcription is initiated and protein product accumulates when production exceeds degradation.
- Links are established between genes when the product of one regulates the expression of another.

Gene expression pathways

- Genetic activity is controlled by molecular signals that determine when and how often a given gene is transcribed.
- The product encoded by one gene often regulates the expression of other genes.
- Moreover two or more proteins may act together to activate or repress a gene.
- For appropriate combinations of input signals transcription is initiated and protein product accumulates when production exceeds degradation.
- Links are established between genes when the product of one regulates the expression of another.
- Thus networks of interaction can be deduced and these may be quite complex.

Dynamic issues

- In biochemical regulatory networks, the delay between events are determined by the delay while signal molecule concentrations accumulate or decline sufficiently.

Dynamic issues

- In biochemical regulatory networks, the delay between events are determined by the delay while signal molecule concentrations accumulate or decline sufficiently.
- For example, delay from the activation of a gene promoter until reaching an effective level to control the next promoter depends on the rate of protein accumulation.

Dynamic issues

- In biochemical regulatory networks, the delay between events are determined by the delay while signal molecule concentrations accumulate or decline sufficiently.
- For example, delay from the activation of a gene promoter until reaching an effective level to control the next promoter depends on the rate of protein accumulation.
- The accumulation of protein is a stochastic process affected by several factors in the cell (temperature, pH, etc.).

Dynamic issues

- In biochemical regulatory networks, the delay between events are determined by the delay while signal molecule concentrations accumulate or decline sufficiently.
- For example, delay from the activation of a gene promoter until reaching an effective level to control the next promoter depends on the rate of protein accumulation.
- The accumulation of protein is a stochastic process affected by several factors in the cell (temperature, pH, etc.).
- Thus the "switching delay" is a distribution rather than a deterministic time, and this can account for some of the cellular phenomena which can be observed across a cell population.

Stochastic behaviour

- The stochastic reaction rate of a chemical reaction is a function of only those molecular species involved as reactants or catalysts, and a stochastic rate constant c.
- The stochastic rate constant takes into account volume, temperature, pH and other environmental factors.
- The stoichiometry of the reaction — how many molecules of each reactant species are required — also has an impact.

Stochastic behaviour

- The stochastic reaction rate of a chemical reaction is a function of only those molecular species involved as reactants or catalysts, and a stochastic rate constant c.
- The stochastic rate constant takes into account volume, temperature, pH and other environmental factors.
- The stoichiometry of the reaction — how many molecules of each reactant species are required — also has an impact.

Stochastic behaviour

- The stochastic reaction rate of a chemical reaction is a function of only those molecular species involved as reactants or catalysts, and a stochastic rate constant c.
- The stochastic rate constant takes into account volume, temperature, pH and other environmental factors.
- The stoichiometry of the reaction — how many molecules of each reactant species are required — also has an impact.

Outline

Background

- The modelling of chemical reactions using deterministic rate laws has proven extremely successful in both chemistry and biochemistry for many years.

Background

- The modelling of chemical reactions using deterministic rate laws has proven extremely successful in both chemistry and biochemistry for many years.
- This deterministic approach has at its core the law of mass action, an empirical law giving a simple relation between reaction rates and molecular component concentrations.

Background

- The modelling of chemical reactions using deterministic rate laws has proven extremely successful in both chemistry and biochemistry for many years.
- This deterministic approach has at its core the law of mass action, an empirical law giving a simple relation between reaction rates and molecular component concentrations.
- Given knowledge of initial molecular concentrations, the law of mass action provides a complete picture of the component concentrations at all future time points.

Background: Law of Mass Action

- The law of mass action considers chemical reactions to be macroscopic under convective or diffusive stirring, continuous and deterministic.

Background: Law of Mass Action

- The law of mass action considers chemical reactions to be macroscopic under convective or diffusive stirring, continuous and deterministic.
- These are evidently simplifications, as it is well understood that chemical reactions involve discrete, random collisions between individual molecules.

Background: Law of Mass Action

- The law of mass action considers chemical reactions to be macroscopic under convective or diffusive stirring, continuous and deterministic.
- These are evidently simplifications, as it is well understood that chemical reactions involve discrete, random collisions between individual molecules.
- As we consider smaller and smaller systems, the validity of a continuous approach becomes ever more tenuous.

Background: Law of Mass Action

- The law of mass action considers chemical reactions to be macroscopic under convective or diffusive stirring, continuous and deterministic.
- These are evidently simplifications, as it is well understood that chemical reactions involve discrete, random collisions between individual molecules.
- As we consider smaller and smaller systems, the validity of a continuous approach becomes ever more tenuous.
- As such, the adequacy of the law of mass action has been questioned for describing intracellular reactions.

Background: Application of Stochastic Models

Arguments for the application of stochastic models for chemical reactions come from at least three directions, since the models:

1. take into consideration the discrete character of the quantity of components and the inherently random character of the phenomena;

Background: Application of Stochastic Models

Arguments for the application of stochastic models for chemical reactions come from at least three directions, since the models:

1. take into consideration the discrete character of the quantity of components and the inherently random character of the phenomena;
2. are in accordance with the theories of thermodynamics and stochastic processes; and

Background: Application of Stochastic Models

Arguments for the application of stochastic models for chemical reactions come from at least three directions, since the models:

1. take into consideration the discrete character of the quantity of components and the inherently random character of the phenomena;
2. are in accordance with the theories of thermodynamics and stochastic processes; and
3. are appropriate to describe "small systems" and instability phenomena.

Deterministic: The law of mass action

The fundamental empirical law governing reaction rates in biochemistry is the law of mass action.

Deterministic: The law of mass action

The fundamental empirical law governing reaction rates in biochemistry is the law of mass action.

This states that for a reaction in a homogeneous, free medium, the reaction rate will be proportional to the concentrations of the individual reactants involved.

Deterministic: Michaelis-Menten kinetics

Consider the simple Michaelis-Menten reaction

$$S + E \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} C \xrightarrow{k_2} E + P$$

Deterministic: Michaelis-Menten kinetics

Consider the simple Michaelis-Menten reaction

$$S + E \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} C \xrightarrow{k_2} E + P$$

For example, we have

$$\frac{\mathrm{d}C}{\mathrm{d}t} = k_1 SE - (k_{-1} + k_2)C$$

Deterministic: Michaelis-Menten kinetics

Consider the simple Michaelis-Menten reaction

$$S + E \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} C \xrightarrow{k_2} E + P$$

For example, we have

$$\frac{\mathrm{d}C}{\mathrm{d}t} = k_1 SE - (k_{-1} + k_2)C$$

Hence, we can express any chemical system as a collection of coupled non-linear first order differential equations.

Stochastic: Random processes

- Whereas the deterministic approach outlined above is essentially an empirical law, derived from *in vitro* experiments, the stochastic approach is far more physically rigorous.
- Fundamental to the principle of stochastic modelling is the idea that molecular reactions are essentially random processes; it is impossible to say with complete certainty the time at which the next reaction within a volume will occur.

Stochastic: Random processes

- Whereas the deterministic approach outlined above is essentially an empirical law, derived from *in vitro* experiments, the stochastic approach is far more physically rigorous.
- Fundamental to the principle of stochastic modelling is the idea that molecular reactions are essentially random processes; it is impossible to say with complete certainty the time at which the next reaction within a volume will occur.

Stochastic: Random processes

- Whereas the deterministic approach outlined above is essentially an empirical law, derived from *in vitro* experiments, the stochastic approach is far more physically rigorous.
- Fundamental to the principle of stochastic modelling is the idea that molecular reactions are essentially random processes; it is impossible to say with complete certainty the time at which the next reaction within a volume will occur.

Stochastic: Predictability of macroscopic states

- In macroscopic systems, with a large number of interacting molecules, the randomness of this behaviour averages out so that the overall macroscopic state of the system becomes highly predictable.
- It is this property of large scale random systems that enables a deterministic approach to be adopted; however, the validity of this assumption becomes strained in *in vivo* conditions as we examine small-scale cellular reaction environments with limited reactant populations.

Stochastic: Predictability of macroscopic states

- In macroscopic systems, with a large number of interacting molecules, the randomness of this behaviour averages out so that the overall macroscopic state of the system becomes highly predictable.
- It is this property of large scale random systems that enables a deterministic approach to be adopted; however, the validity of this assumption becomes strained in *in vivo* conditions as we examine small-scale cellular reaction environments with limited reactant populations.

Stochastic: Predictability of macroscopic states

- In macroscopic systems, with a large number of interacting molecules, the randomness of this behaviour averages out so that the overall macroscopic state of the system becomes highly predictable.
- It is this property of large scale random systems that enables a deterministic approach to be adopted; however, the validity of this assumption becomes strained in *in vivo* conditions as we examine small-scale cellular reaction environments with limited reactant populations.

Stochastic: Propensity function

As explicitly derived by Gillespie, the stochastic model uses basic Newtonian physics and thermodynamics to arrive at a form often termed the propensity function that gives the probability a_μ of reaction μ occurring in time interval $(t, t + \mathrm{d}t)$.

$$a_\mu \mathrm{d}t = h_\mu c_\mu \mathrm{d}t$$

where the M reaction mechanisms are given an arbitrary index μ $(1 \leq \mu \leq M)$, h_μ denotes the number of possible combinations of reactant molecules involved in reaction μ, and c_μ is a stochastic rate constant.

Stochastic: Fundamental hypothesis

The rate constant c_μ is dependent on the radii of the molecules involved in the reaction, and their average relative velocities – a property that is itself a direct function of the temperature of the system and the individual molecular masses.

Stochastic: Fundamental hypothesis

The rate constant c_μ is dependent on the radii of the molecules involved in the reaction, and their average relative velocities – a property that is itself a direct function of the temperature of the system and the individual molecular masses.

These quantities are basic chemical properties which for most systems are either well known or easily measurable. Thus, for a given chemical system, the propensity functions, a_μ can be easily determined.

Stochastic: Grand probability function

The stochastic formulation proceeds by considering the grand probability function $\Pr(\mathbf{X}; t) \equiv$ probability that there will be present in the volume V at time t, X_i of species S_i, where $\mathbf{X} \equiv (X_1, X_2, \ldots X_N)$ is a vector of molecular species populations.

Stochastic: Grand probability function

The stochastic formulation proceeds by considering the grand probability function $\Pr(\mathbf{X}; t) \equiv$ probability that there will be present in the volume V at time t, X_i of species S_i, where $\mathbf{X} \equiv (X_1, X_2, \ldots X_N)$ is a vector of molecular species populations.

Evidently, knowledge of this function provides a complete understanding of the probability distribution of all possible states at all times.

Stochastic: Infinitesimal time interval

By considering a discrete infinitesimal time interval $(t, t + \mathrm{d}t)$ in which either 0 or 1 reactions occur we see that there exist only $M + 1$ distinct configurations at time t that can lead to the state $\mathbf{X}$ at time $t + \mathrm{d}t$.

Stochastic: Infinitesimal time interval

By considering a discrete infinitesimal time interval $(t, t + \mathrm{d}t)$ in which either 0 or 1 reactions occur we see that there exist only $M + 1$ distinct configurations at time t that can lead to the state $\mathbf{X}$ at time $t + \mathrm{d}t$.

$$\Pr(\mathbf{X}; t + \mathrm{d}t)$$

$$= \Pr(\mathbf{X}; t) \Pr(\text{no state change over } \mathrm{d}t)$$

Stochastic: Infinitesimal time interval

By considering a discrete infinitesimal time interval $(t, t + \mathrm{d}t)$ in which either 0 or 1 reactions occur we see that there exist only $M + 1$ distinct configurations at time t that can lead to the state $\mathbf{X}$ at time $t + \mathrm{d}t$.

$$\begin{aligned}
&\Pr(\mathbf{X}; t + \mathrm{d}t) \\
&\quad = \Pr(\mathbf{X}; t) \Pr(\text{no state change over } \mathrm{d}t) \\
&\quad + \textstyle\sum_{\mu=1}^{M} \Pr(\mathbf{X} - \mathbf{v}_\mu; t) \Pr(\text{state change to } \mathbf{X} \text{ over } \mathrm{d}t)
\end{aligned}$$

Stochastic: Infinitesimal time interval

By considering a discrete infinitesimal time interval $(t, t + \mathrm{d}t)$ in which either 0 or 1 reactions occur we see that there exist only $M + 1$ distinct configurations at time t that can lead to the state $\mathbf{X}$ at time $t + \mathrm{d}t$.

$$\begin{aligned}
&\Pr(\mathbf{X}; t + \mathrm{d}t) \\
&\quad = \Pr(\mathbf{X}; t) \Pr(\text{no state change over } \mathrm{d}t) \\
&\quad + \textstyle\sum_{\mu=1}^{M} \Pr(\mathbf{X} - \mathbf{v}_\mu; t) \Pr(\text{state change to } \mathbf{X} \text{ over } \mathrm{d}t)
\end{aligned}$$

where $\mathbf{v}_\mu$ is a stoichiometric vector defining the result of reaction μ on state vector $\mathbf{X}$, i.e. $\mathbf{X} \rightarrow \mathbf{X} + \mathbf{v}_\mu$ after an occurrence of reaction μ.

Stochastic: State change probabilities

Pr(no state change over $\mathrm{d}t$)

$$1 - \sum_{\mu=1}^{M} a_\mu(\mathbf{X})\mathrm{d}t$$

Stochastic: State change probabilities

Pr(no state change over dt)

$$1 - \sum_{\mu=1}^{M} a_\mu(\mathbf{X})\mathrm{d}t$$

Pr(state change to $\mathbf{X}$ over dt)

$$\sum_{\mu=1}^{M} \Pr(\mathbf{X} - \mathbf{v}_\mu; t) a_\mu(\mathbf{X} - \mathbf{v}_\mu)\mathrm{d}t$$

Stochastic: Partial derivatives

We are considering the behaviour of the system in the limit as $\mathrm{d}t$ tends to zero. This leads us to consider *partial derivatives*, which are defined thus:

$$\frac{\partial \Pr(\mathbf{X}; t)}{\partial t} = \lim_{\mathrm{d}t \to 0} \frac{\Pr(\mathbf{X}; t + \mathrm{d}t) - \Pr(\mathbf{X}; t)}{\mathrm{d}t}$$

Stochastic: Chemical Master Equation

Applying this, and re-arranging the former, leads us to an important *partial differential equation* (PDE) known as the Chemical Master Equation (CME).

$$\frac{\partial \Pr(\mathbf{X}; t)}{\partial t} = \sum_{\mu=1}^{M} a_\mu(\mathbf{X} - \mathbf{v}_\mu) \Pr(\mathbf{X} - \mathbf{v}_\mu; t) - a_\mu(\mathbf{X}) \Pr(\mathbf{X}; t)$$

Stochastic: Chemical Master Equation

Applying this, and re-arranging the former, leads us to an important *partial differential equation* (PDE) known as the Chemical Master Equation (CME).

$$\frac{\partial \Pr(\mathbf{X}; t)}{\partial t} = \sum_{\mu=1}^{M} a_\mu(\mathbf{X} - \mathbf{v}_\mu) \Pr(\mathbf{X} - \mathbf{v}_\mu; t) - a_\mu(\mathbf{X}) \Pr(\mathbf{X}; t)$$

Stochastic: Chemical Master Equation

Applying this, and re-arranging the former, leads us to an important *partial differential equation* (PDE) known as the Chemical Master Equation (CME).

$$\frac{\partial \Pr(\mathbf{X}; t)}{\partial t} = \sum_{\mu=1}^{M} a_\mu(\mathbf{X} - \mathbf{v}_\mu) \Pr(\mathbf{X} - \mathbf{v}_\mu; t) - a_\mu(\mathbf{X}) \Pr(\mathbf{X}; t)$$

Stochastic: Chemical Master Equation

Applying this, and re-arranging the former, leads us to an important *partial differential equation* (PDE) known as the Chemical Master Equation (CME).

$$\frac{\partial \Pr(\mathbf{X}; t)}{\partial t} = \sum_{\mu=1}^{M} a_\mu(\mathbf{X} - \mathbf{v}_\mu) \Pr(\mathbf{X} - \mathbf{v}_\mu; t) - a_\mu(\mathbf{X}) \Pr(\mathbf{X}; t)$$

The problem with the Chemical Master Equation

- The CME is really a set of nearly as many coupled ordinary differential equations as there are combinations of molecules that can exist in the system!
- The CME can be solved analytically for only a very few very simple systems, and numerical solutions are usually prohibitively difficult.

D. Gillespie and L. Petzold.

chapter *Numerical Simulation for Biochemical Kinetics*, in *System Modelling in Cellular Biology*, editors Z. Szallasi, J. Stelling and V. Periwal.

MIT Press, 2006.

Stochastic simulation algorithms

Gillespie's Stochastic Simulation Algorithm (SSA) is essentially an exact procedure for numerically simulating the time evolution of a well-stirred chemically reacting system by taking proper account of the randomness inherent in such a system.

Stochastic simulation algorithms

Gillespie's Stochastic Simulation Algorithm (SSA) is essentially an exact procedure for numerically simulating the time evolution of a well-stirred chemically reacting system by taking proper account of the randomness inherent in such a system.

It is rigorously based on the same microphysical premise that underlies the chemical master equation and gives a more realistic representation of a system's evolution than the deterministic reaction rate equation (RRE) represented mathematically by ODEs.

Stochastic simulation algorithms

Gillespie's Stochastic Simulation Algorithm (SSA) is essentially an exact procedure for numerically simulating the time evolution of a well-stirred chemically reacting system by taking proper account of the randomness inherent in such a system.

It is rigorously based on the same microphysical premise that underlies the chemical master equation and gives a more realistic representation of a system's evolution than the deterministic reaction rate equation (RRE) represented mathematically by ODEs.

As with the chemical master equation, the SSA converges, in the limit of large numbers of reactants, to the same solution as the law of mass action.

Gillespie's exact SSA (1977)

- The algorithm takes time steps of variable length, based on the rate constants and population size of each chemical species.

Gillespie's exact SSA (1977)

- The algorithm takes time steps of variable length, based on the rate constants and population size of each chemical species.
- The probability of one reaction occurring relative to another is dictated by their relative propensity functions.

Gillespie's exact SSA (1977)

- The algorithm takes time steps of variable length, based on the rate constants and population size of each chemical species.
- The probability of one reaction occurring relative to another is dictated by their relative propensity functions.
- According to the correct probability distribution derived from the statistical thermodynamics theory, a random variable is then used to choose which reaction will occur, and another random variable determines how long the step will last.

Gillespie's exact SSA (1977)

- The algorithm takes time steps of variable length, based on the rate constants and population size of each chemical species.
- The probability of one reaction occurring relative to another is dictated by their relative propensity functions.
- According to the correct probability distribution derived from the statistical thermodynamics theory, a random variable is then used to choose which reaction will occur, and another random variable determines how long the step will last.
- The chemical populations are altered according to the stoichiometry of the reaction and the process is repeated.

Stochastic simulation: realisations and ensembles

The SSA computes one realisation of a dynamic trajectory of a chemically reacting system. Often an ensemble of trajectories is computed, to obtain an estimate of the probability density function of the system.

The dynamic evolution of the probability density function is given by the Chemical Master Equation.

Gillespie's SSA is a Monte Carlo Markov Chain simulation

The SSA is a Monte Carlo type method. With the SSA one may approximate any variable of interest by generating many trajectories and observing the statistics of the values of the variable. Since many trajectories are needed to obtain a reasonable approximation, the efficiency of the SSA is of critical importance.

Computational cost of Gillespie's exact algorithm

The cost of this detailed stochastic simulation algorithm is the likely large amounts of computing time.

The key issue is that the time step for the next reaction can be very small indeed if we are to guarantee that only one reaction can take place in a given time interval.

Computational cost of Gillespie's exact algorithm

The cost of this detailed stochastic simulation algorithm is the likely large amounts of computing time.

The key issue is that the time step for the next reaction can be very small indeed if we are to guarantee that only one reaction can take place in a given time interval.

Increasing the molecular population or number of reaction mechanisms necessarily requires a corresponding decrease in the time interval. The SSA can be very computationally inefficient especially when there are large numbers of molecules or the propensity functions are large.

Gibson and Bruck (2000)

Gibson and Bruck refined the first reaction SSA of Gillespie by reducing the number of random variables that need to be simulated.

This can be effective for systems in which some reactions occur much more frequently than others.

M.A. Gibson and J. Bruck.

Efficient exact stochastic simulation of chemical systems with many species and many channels.

J. Comp. Phys., 104:1876–1889, 2000.

Circadian clock

To adapt to natural periodicity, such as the alternation of day and night, most living organisms have developed the capability of generating oscillating expressions of proteins in their cells with a period close to 24 hours (circadian rhythm).

Circadian clock

To adapt to natural periodicity, such as the alternation of day and night, most living organisms have developed the capability of generating oscillating expressions of proteins in their cells with a period close to 24 hours (circadian rhythm).

The Vilar-Kueh-Barkai-Leibler (VKBL in short) description of the circadian oscillator incorporates an abstraction of a minimal set of essential, experimentally determined mechanisms for the circadian system.

Circadian clock

- The VKBL model involves two genes, an activator A and a repressor R, which are transcribed into $mRNA$ and subsequently translated into proteins.

Circadian clock

- The VKBL model involves two genes, an activator A and a repressor R, which are transcribed into $mRNA$ and subsequently translated into proteins.

Circadian clock

- The VKBL model involves two genes, an activator A and a repressor R, which are transcribed into $mRNA$ and subsequently translated into proteins.
- The activator A binds to the A and R promoters and increases their expression rate.

Circadian clock

- The VKBL model involves two genes, an activator A and a repressor R, which are transcribed into $mRNA$ and subsequently translated into proteins.
- The activator A binds to the A and R promoters and increases their expression rate.
- Therefore, A implements a positive loop acting on its own transcription.

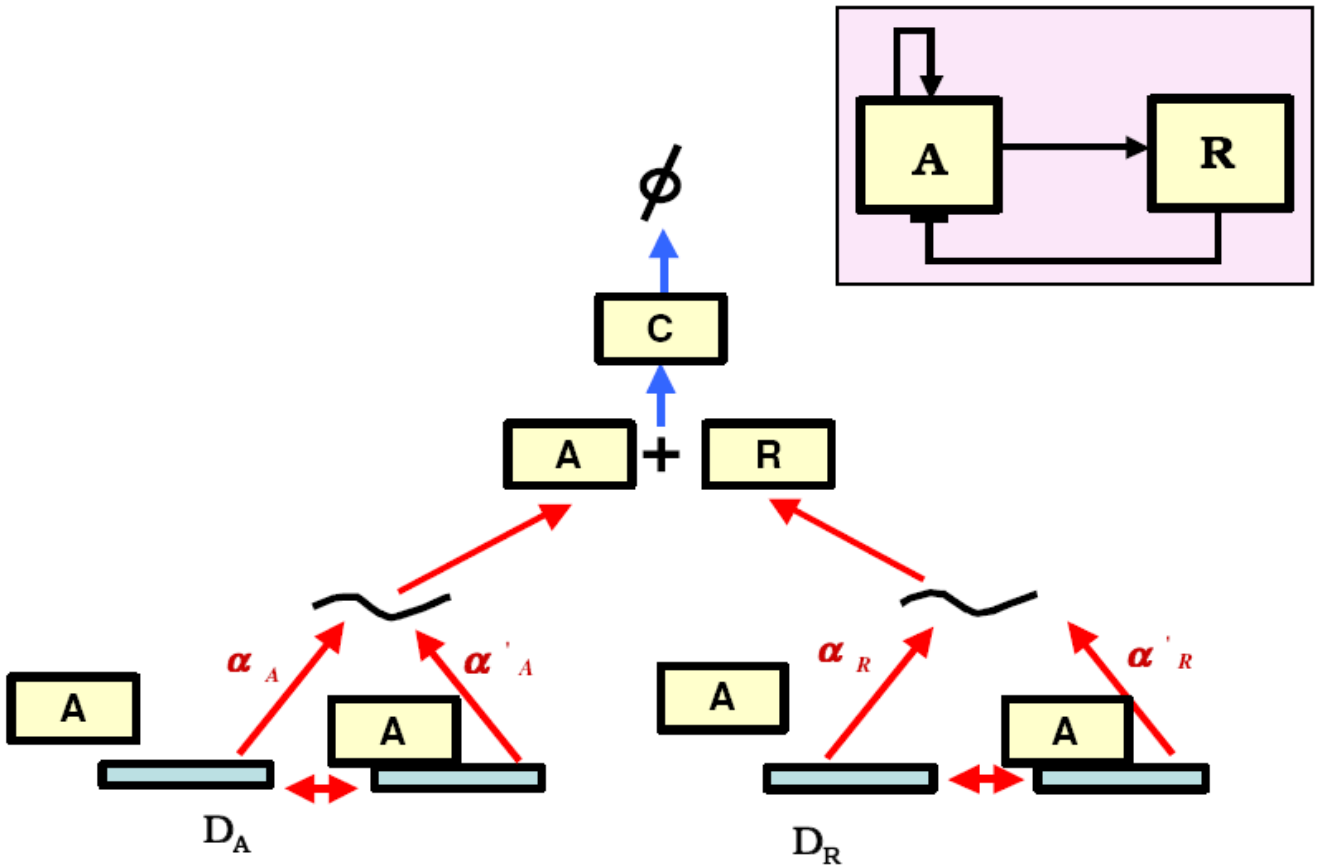
Circadian clock

- The VKBL model involves two genes, an activator A and a repressor R, which are transcribed into $mRNA$ and subsequently translated into proteins.
- The activator A binds to the A and R promoters and increases their expression rate.
- Therefore, A implements a positive loop acting on its own transcription.
- At the same time, R sequesters A to form a complex C, therefore inhibiting it from binding to the gene promoter and acting as a negative feedback loop.

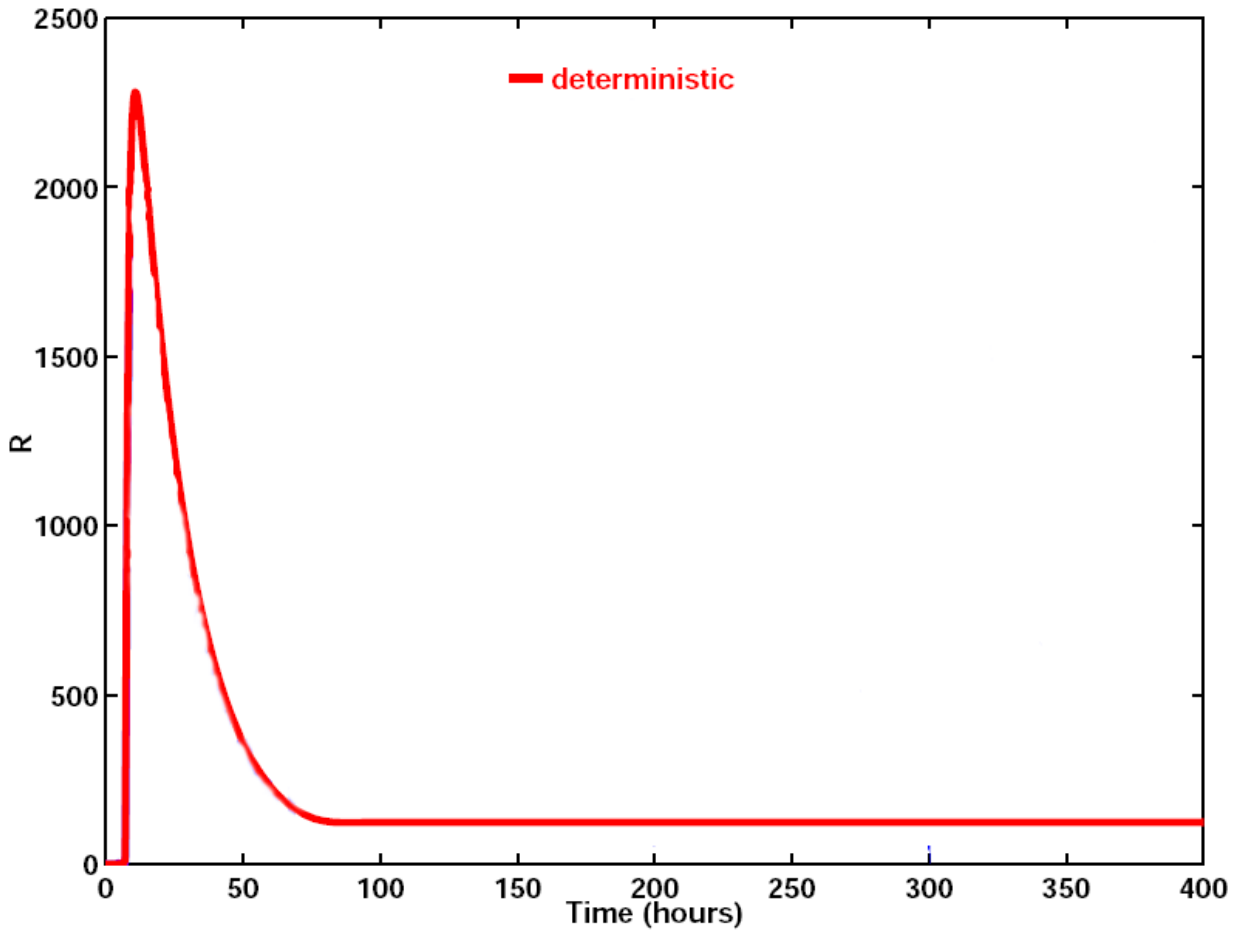
Circadian clock

- The VKBL model involves two genes, an activator A and a repressor R, which are transcribed into $mRNA$ and subsequently translated into proteins.
- The activator A binds to the A and R promoters and increases their expression rate.
- Therefore, A implements a positive loop acting on its own transcription.
- At the same time, R sequesters A to form a complex C, therefore inhibiting it from binding to the gene promoter and acting as a negative feedback loop.

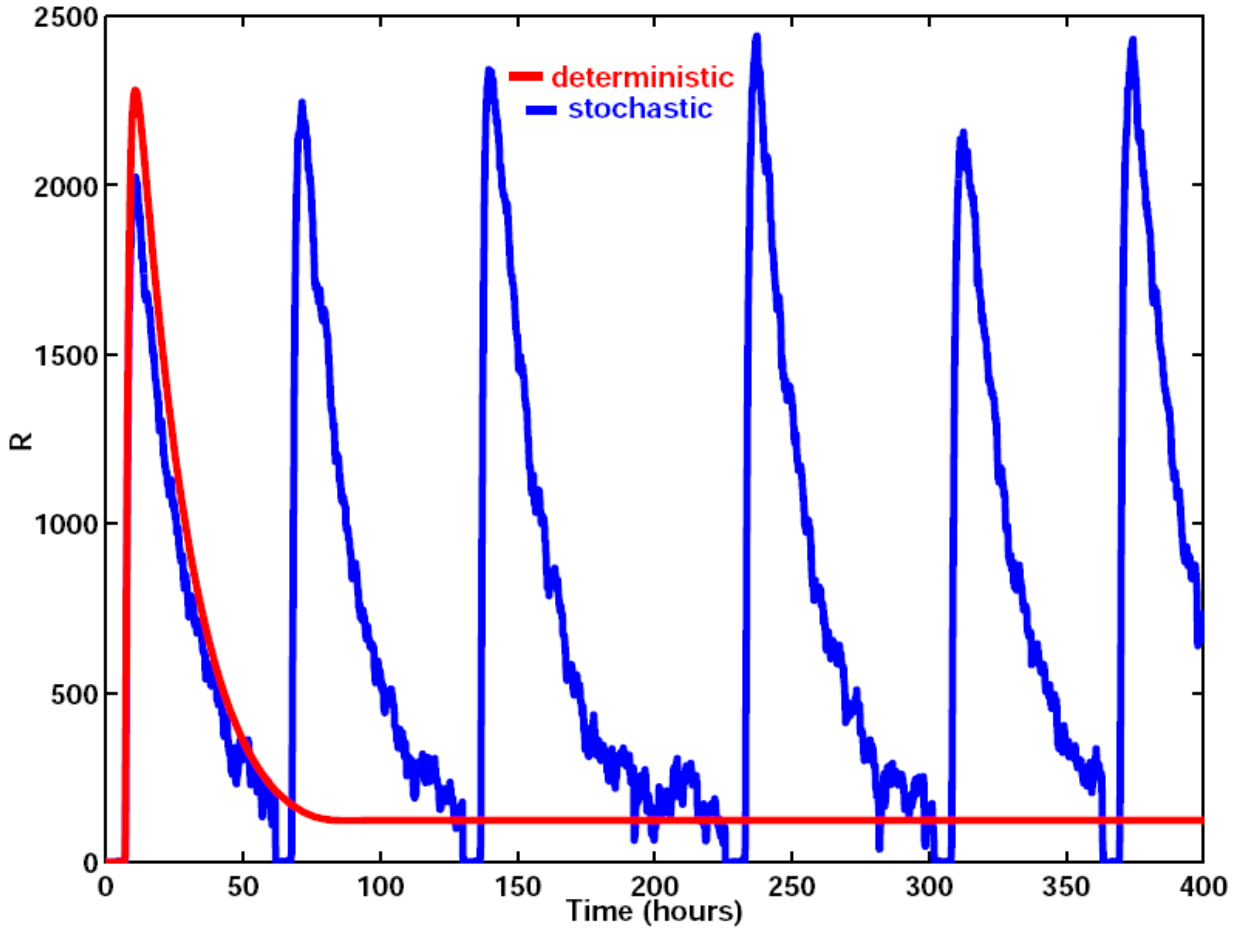
Circadian clock (cartoon)



Circadian clock (deterministically . . .)



Circadian clock (. . . and stochastically)



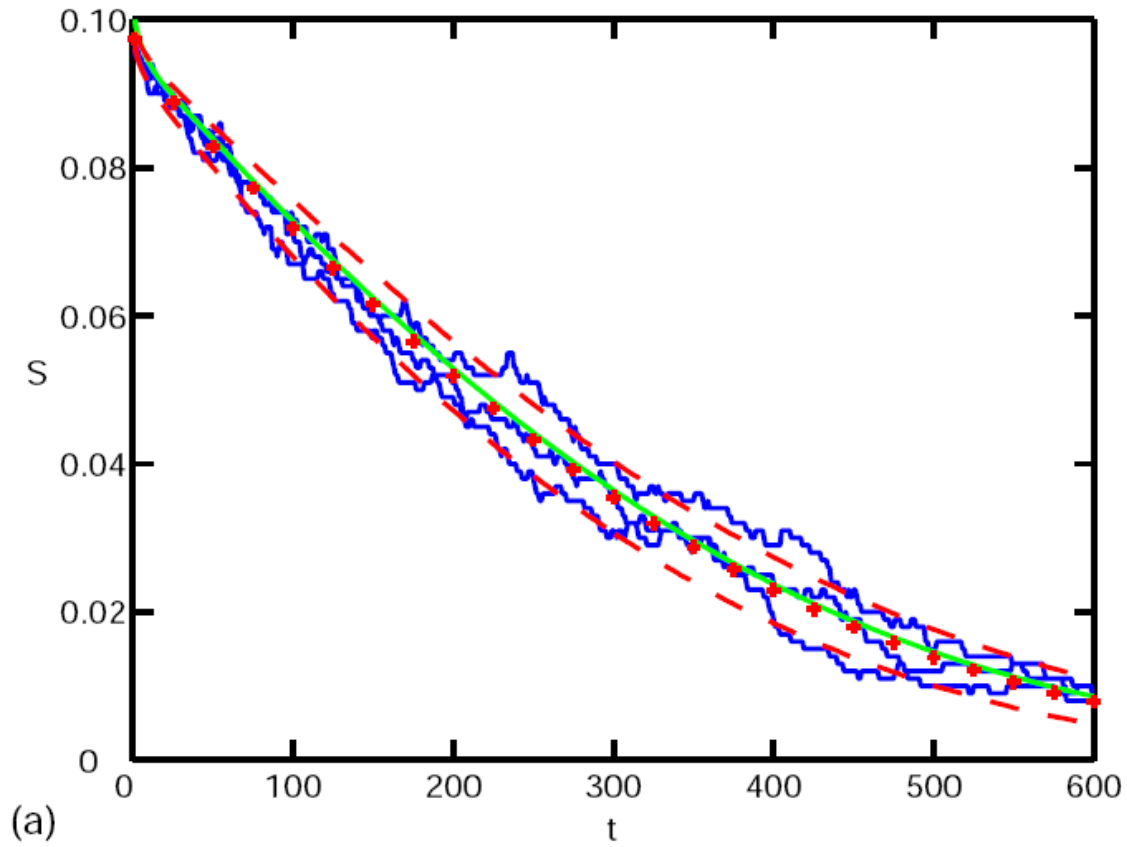
Circadian clock (Conclusions)

- For some parameter values a differential equation model exhibits autonomous oscillations.
- These oscillations disappear from the deterministic model as the degradation rate of the repressor δ_R is decreased.
- The system of ODEs undergoes a bifurcation at this point and the unique deterministic equilibrium of the system becomes stable.
- However, if the effects of molecular noise are incorporated the oscillations in the stochastic system pertain.
- This phenomenon is a manifestation of coherence resonance, and illustrates the crucial interplay between noise and dynamics.

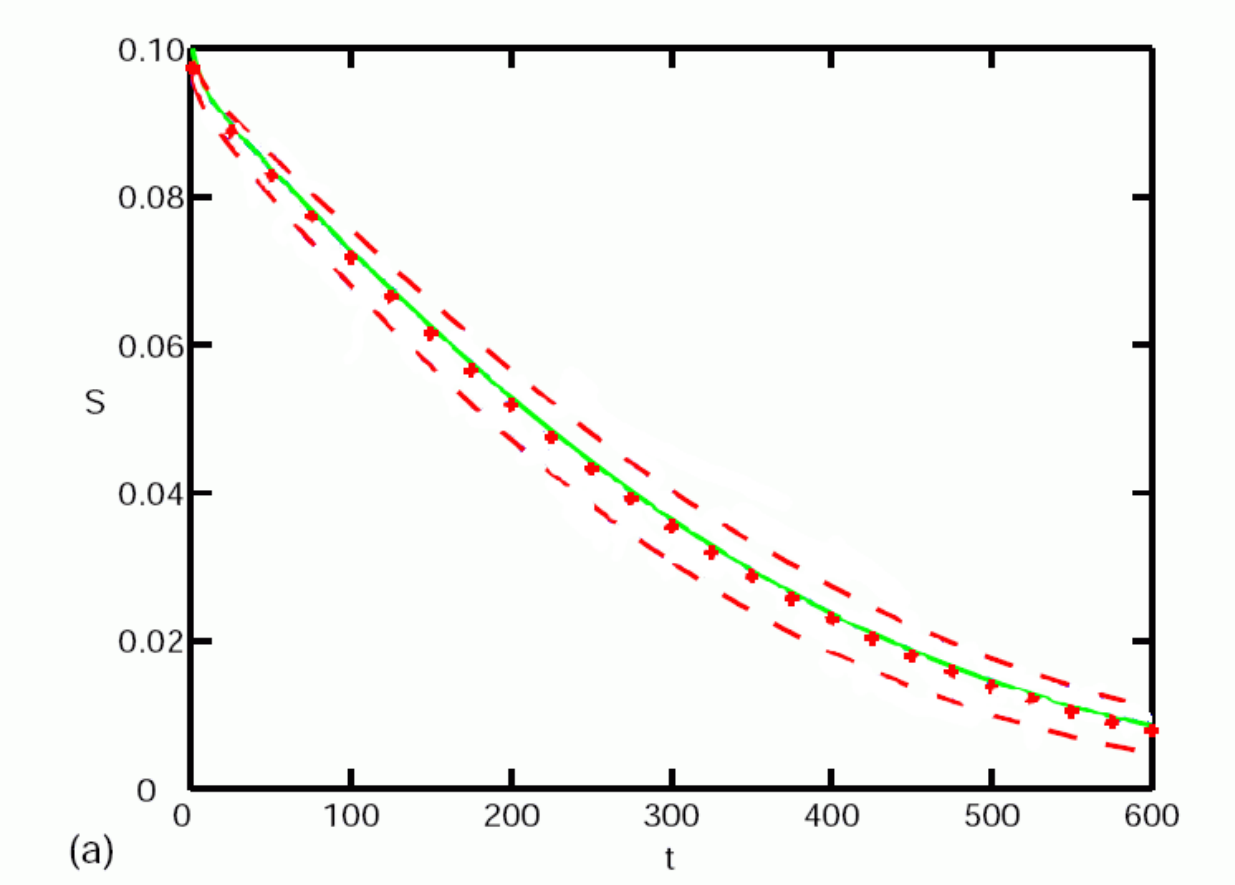
Comparing stochastic simulation and ODEs

It is relatively straightforward to contrast the results of the two methods. We compare the results of 2000 runs of the stochastic algorithm simulating a system with initial molecular populations $S_0 = 100, E_0 = 10, C_0 = 0, P_0 = 0$ and a volume of 1000 units.

Results for $S_0 = 100, E_0 = 10, C_0 = 0, P_0 = 0$ (vol 1000)



Results for $S_0 = 100, E_0 = 10, C_0 = 0, P_0 = 0$ (vol 1000)



Comparing stochastic simulation and ODEs

It is clear that there is a close correspondence between the predictions of the deterministic approach and the stochastic approach, with the deterministic curve falling well within one standard deviation (S.D.) of the stochastic mean.

Comparing stochastic simulation and ODEs

It is clear that there is a close correspondence between the predictions of the deterministic approach and the stochastic approach, with the deterministic curve falling well within one standard deviation (S.D.) of the stochastic mean.

This is a very close match, especially considering our stochastic simulation is modelling a system containing just 110 molecules—well within what we might consider to be the microscopic domain.

The variance of the stochastic approach

However, it is worth bearing in mind that an actual *in vivo* biochemical reaction would follow just one of the many random curves that average together producing the closely fitting mean. This curve may deviate significantly from that of the deterministic approach, and thus call into question its validity.

The variance of the stochastic approach

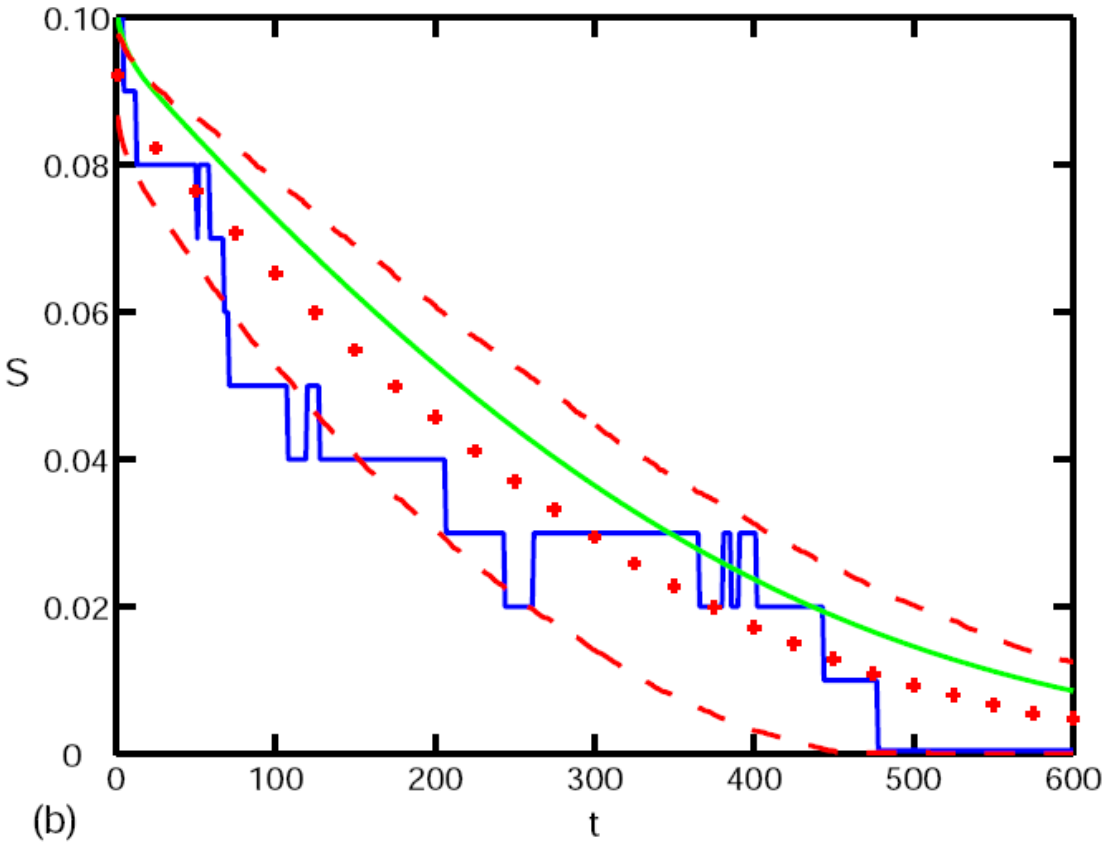
However, it is worth bearing in mind that an actual *in vivo* biochemical reaction would follow just one of the many random curves that average together producing the closely fitting mean. This curve may deviate significantly from that of the deterministic approach, and thus call into question its validity.

Hence, it is perhaps most important to consider the variance of the stochastic approach—with a larger variance indicating a greater deviation from the mean and hence from the deterministic curve.

Comparing results at lower population sizes

Consider exactly the same simulation setup, except this time we are modelling a system consisting of just 11 molecules within a volume of 100 units [thus the molecular *concentrations* are equal to those earlier].

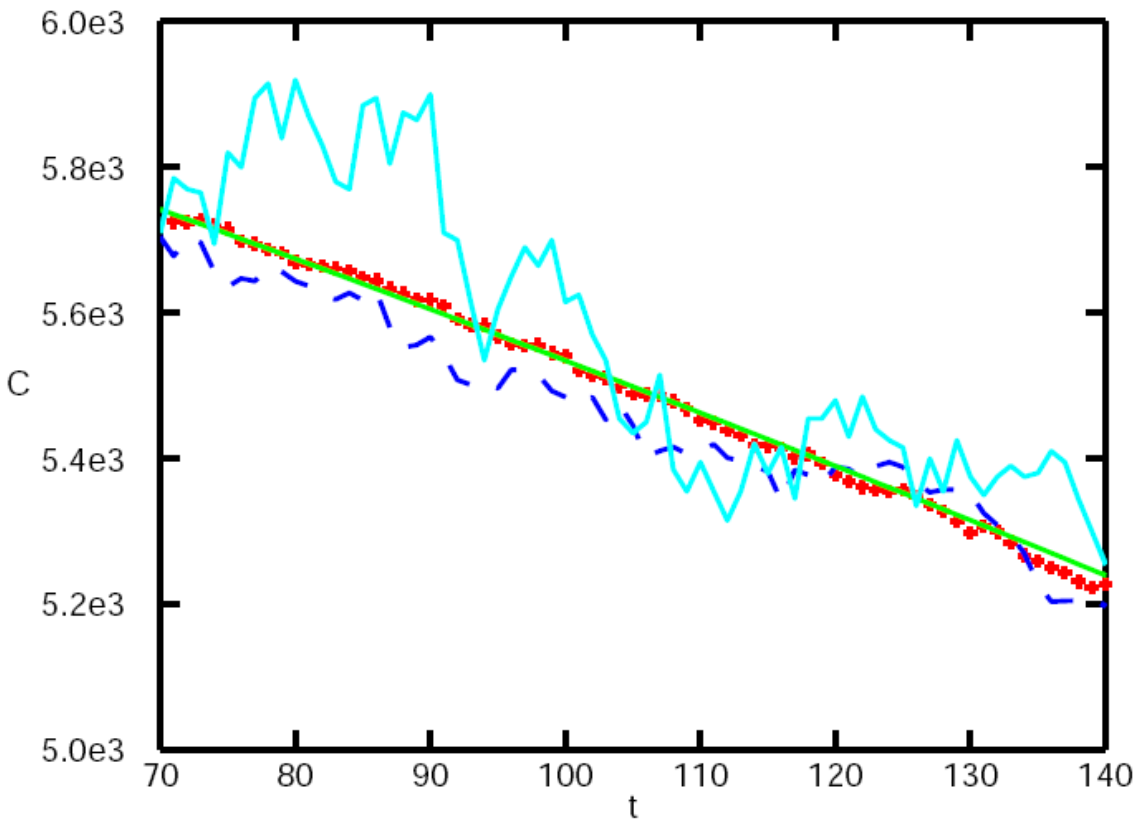
Results for $S_0 = 10, E_0 = 1, C_0 = 0, P_0 = 0$ (vol 100)



Compatibility of the two approaches

On average, the stochastic approach tends to the same solution as the deterministic approach as the number of molecules in the system increases, and we hence move from the microscopic to the macroscopic domain.

Mean results for 11, 110 and 1100 molecules



From the microscopic to the macroscopic domain

Each specific run is individually in closer and closer agreement with the deterministic approach as the number of molecules in the system increases.

This is a direct effect of the inherent averaging of macroscopic properties of a system of many particles.

Conclusions from the comparison

1. These results provide clear verification of the compatibility of the deterministic and stochastic approaches.
2. They also illustrate the validity of the deterministic approach in systems containing as few as 100 copies of components.

Modelling challenges: stiffness

A problem for modelling temporal evolution is *stiffness*. Some reactions are much faster than others and quickly reach a stable state. The dynamics of the system is driven by the slow reactions.

Modelling challenges: stiffness

A problem for modelling temporal evolution is *stiffness*. Some reactions are much faster than others and quickly reach a stable state. The dynamics of the system is driven by the slow reactions.

Most chemical systems, whether considered at a scale appropriate to stochastic or to deterministic simulation, involve several widely varying time scales, so such systems are nearly always stiff.

Modelling challenges: stiffness

A problem for modelling temporal evolution is *stiffness*. Some reactions are much faster than others and quickly reach a stable state. The dynamics of the system is driven by the slow reactions.

Most chemical systems, whether considered at a scale appropriate to stochastic or to deterministic simulation, involve several widely varying time scales, so such systems are nearly always stiff.

Modelling challenges: multiscale populations

The *multiscale population* problem arises when some species are present in relatively small quantities and should be modelled by a discrete stochastic process, whereas other species are present in larger quantities and are more efficiently modelled by a deterministic ordinary differential equation (or at some scale in between). SSA treats all of the species as discrete stochastic processes.

Gillespie's multiscale SSA methods (2005)

SSA is used for slow reactions or species with small populations. The multiscale SSA method generalizes this idea to the case in which species with small population are involved in fast reactions.

Gillespie's slow-scale SSA methods (2005)

The setting for Gillespie's slow-scale SSA method is

$$S + E \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} C \xrightarrow{k_2} E + P$$

where

$$k_{-1} \gg k_2$$

Gillespie's slow-scale SSA methods (2005)

The setting for Gillespie's slow-scale SSA method is

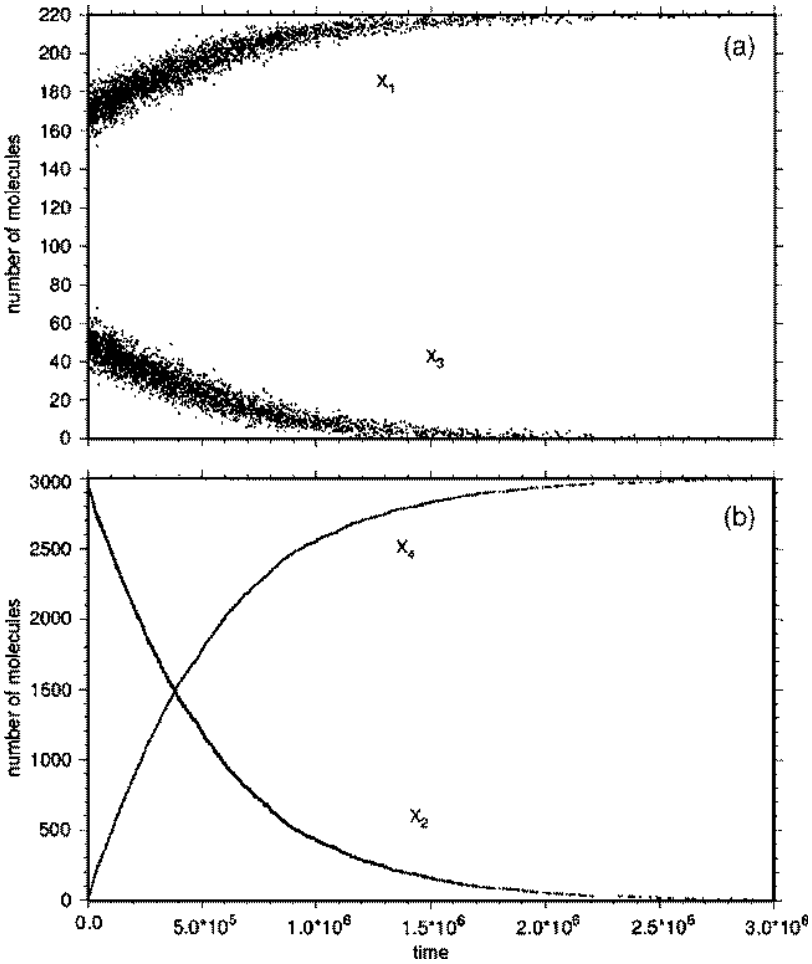
$$S + E \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} C \xrightarrow{k_2} E + P$$

where

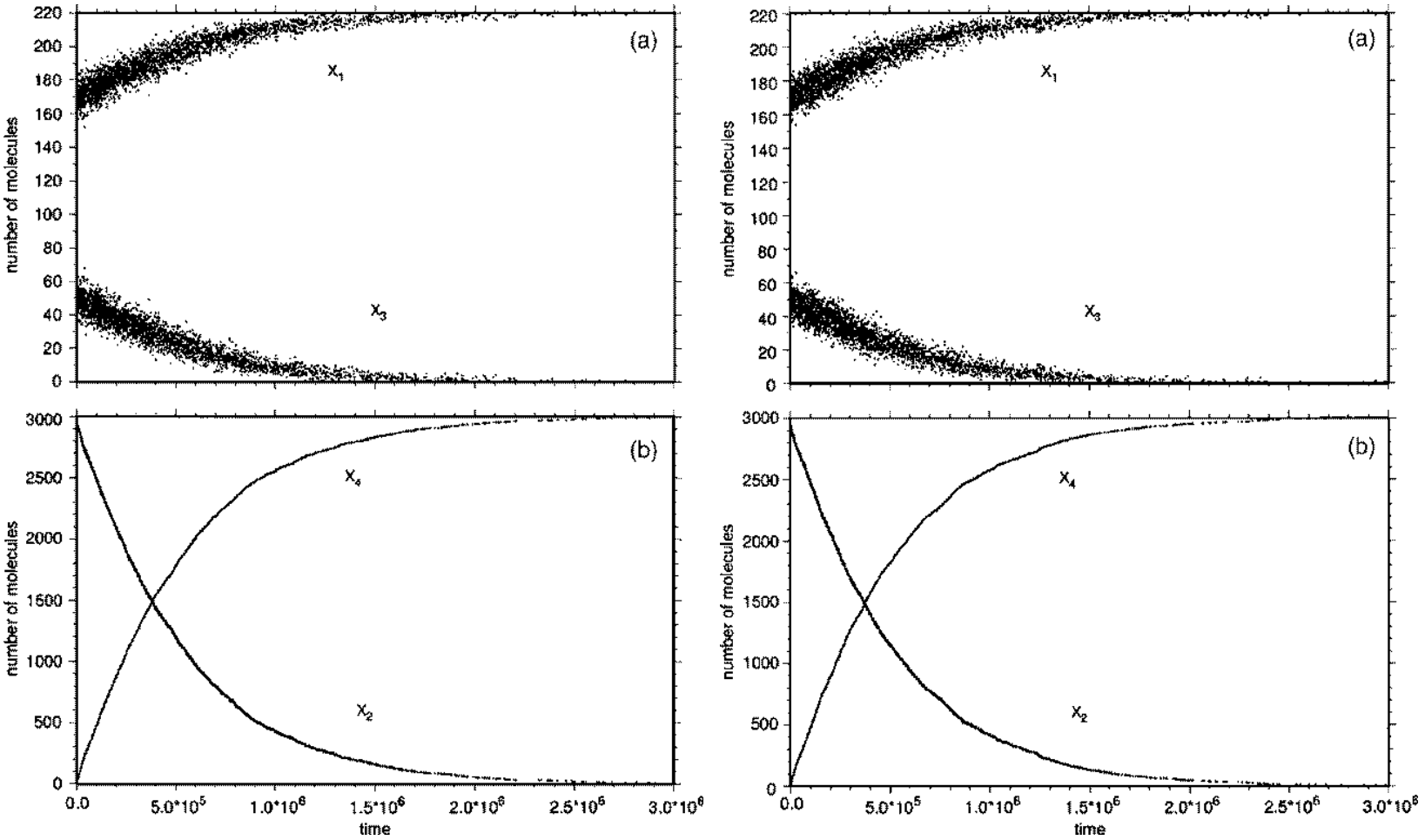
$$k_{-1} \gg k_2$$

Slow-scale SSA explicitly simulates only the relatively rare conversion reactions, skipping over occurrences of the other two less interesting but much more frequent reactions.

Comparing SSA and Slow-Scale SSA results



Comparing SSA and Slow-Scale SSA results



Conclusions

- Stochastic simulation is a well-founded method for simulating *in vivo* reactions.
- Gillespie's SSA can be more accurate than ODEs at low molecular numbers; compatible with them at large molecular numbers.
- Recent explosion of interest in the subject with many new variants of the SSA algorithm.

Excellent introductory papers

T.E. Turner, S. Schnell, and K. Burrage.

Stochastic approaches for modelling in vivo reactions.

Computational Biology and Chemistry, 28:165–178, 2004.

D. Gillespie and L. Petzold.

System Modelling in Cellular Biology, chapter Numerical Simulation for Biochemical Kinetics,.

MIT Press, 2006.

Stochastic simulation software

S. Ramsey, D. Orrell, and H. Bolouri.

Dizzy: stochastic simulation of large-scale genetic regulatory networks.

J. Bioinf. Comp. Biol., 3(2):415–436, 2005.

http://magnet.systemsbiology.net/software/Dizzy

Outline

Current Hypothesis

Some of the techniques we have developed over the last thirty years for modelling complex software systems can be beneficially applied to the modelling aspects of systems biology.

Current Hypothesis

Some of the techniques we have developed over the last thirty years for modelling complex software systems can be beneficially applied to the modelling aspects of systems biology.

In particular formalisms which encompass support for

- Abstraction
- Modularity and
- Reasoning

have a key role to play.

Current Hypothesis

Some of the techniques we have developed over the last thirty years for modelling complex software systems can be beneficially applied to the modelling aspects of systems biology.

In particular formalisms which encompass support for

- Abstraction
- Modularity and
- Reasoning

have a key role to play.

Stochastic mechanisms are crucial for the dynamic analysis of many phenomena so system descriptions should also capture such mechanisms.

SPN for Systems Biology

P.J.E. Goss and J. Peccoud

Quantitative modeling of stochastic systems in molecular biology by using stochastic Petri nets

Proceedings of National Academy of Science, USA, Volume 95(12), pp. 6750–6755, June 1998 (Biochemistry)

D. Gilbert and M. Heiner

From Petri Nets to Differential Equations — an Integrative Approach for Biochemical Network Analysis

Proceedings of the 27th International Conference on Application and Theory of Petri Nets, LNCS Volume 4024, pp. 181–200, June 2006. (Biochemistry)

Stochastic Petri Nets

Stochastic Petri nets (SPN) emerged as a modelling formalism for performance analysis in the early 1980s.

They are based on untimed Petri nets which were developed in the 1960s for modelling and analysing causality, concurrency and conflict within scheduling systems.

Molloy established that the reachability graph of a SPN can be regarded as the state transition diagram of an underlying continuous time Markov process.

Subsequently SPN were generalised in several ways e.g. immediate transitions, inhibitor arcs, colours, arc functions, marking dependent firings, input and output gates.

Molecular biology using SPN

Goss and Peccoud, writing for biochemists, explain the stochastic Petri net (SPN) formalism and illustrate it through a number of examples.

They highlight that the stochastic process resulting from the SPN representation is equivalent to the chemical master equation.

They use the standard UltraSAN simulation tools to simulate their models rather than an implementation of Gillespie's algorithm.

Molecular biology using SPN

Goss and Peccoud, writing for biochemists, explain the stochastic Petri net (SPN) formalism and illustrate it through a number of examples.

They highlight that the stochastic process resulting from the SPN representation is equivalent to the chemical master equation.

They use the standard UltraSAN simulation tools to simulate their models rather than an implementation of Gillespie's algorithm.

Biochemical mapping

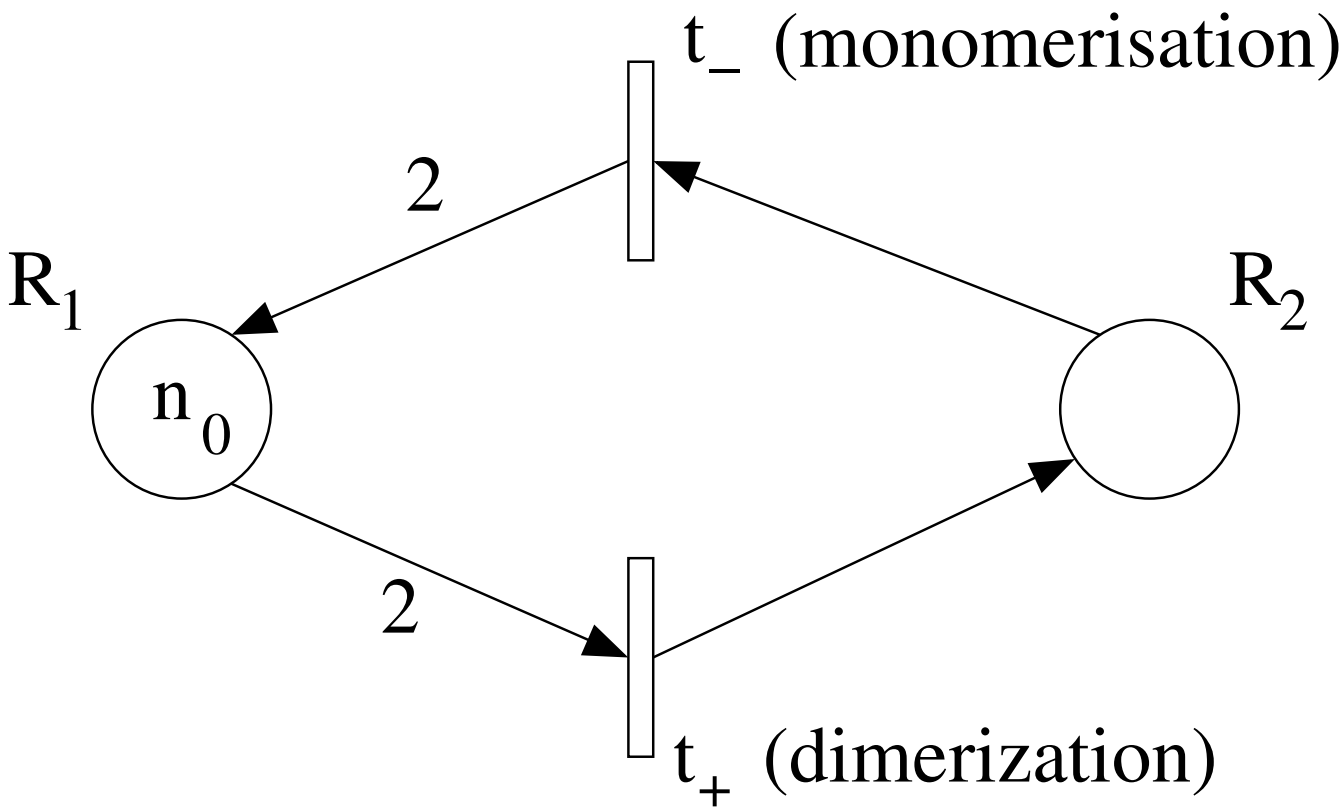
SPN entity	*Molecular interpretation*
Place	Molecular species
Token	Molecule
Marking	Number of molecules
Transition	Reaction
Input place	Reactant
Output place	Product
Weight function	Stoichiometry
To be enable	A possible reaction
To fire	A reaction occurs

Simple example: $2R \leftrightharpoons R_2$

Dimerisation is the process of two molecules of the same species binding to form a single molecule of the dimer species.

Monomerisation is the reverse process when a single dimer disassociates into two individual molecules of the constituent species.

Simple example: $2R \leftrightharpoons R_2$



A remark about kinetics

The rate of a transition in the SPN is the stochastic rate constant.

A remark about kinetics

The rate of a transition in the SPN is the stochastic rate constant.

This takes into account the volume, temperature, pH etc which affect the rate at which the reaction takes place.

A remark about kinetics

The rate of a transition in the SPN is the stochastic rate constant.

This takes into account the volume, temperature, pH etc which affect the rate at which the reaction takes place.

The rate of the reaction will be the stochastic rate constant c multiplied by the number of the ways in which the reaction can be formed from the current state.

A remark about kinetics

The rate of a transition in the SPN is the stochastic rate constant.

This takes into account the volume, temperature, pH etc which affect the rate at which the reaction takes place.

The rate of the reaction will be the stochastic rate constant c multiplied by the number of the ways in which the reaction can be formed from the current state.

For example, for dimerisation the rate will be $c \times N_{monomer} \times (N_{monomer} - 1)$.

A remark about kinetics

The rate of a transition in the SPN is the stochastic rate constant.

This takes into account the volume, temperature, pH etc which affect the rate at which the reaction takes place.

The rate of the reaction will be the stochastic rate constant c multiplied by the number of the ways in which the reaction can be formed from the current state.

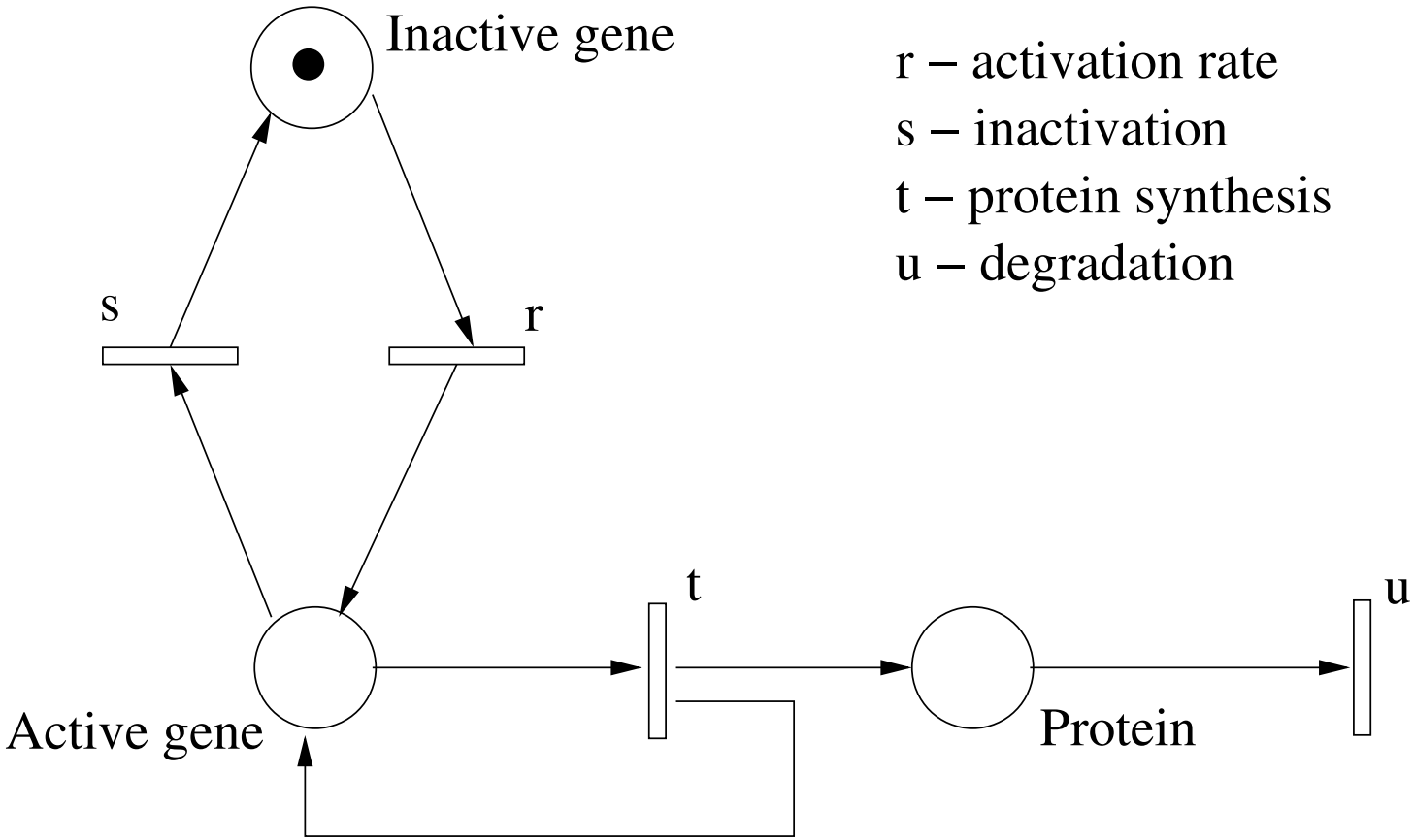
For example, for dimerisation the rate will be $c \times N_{monomer} \times (N_{monomer} - 1)$.

Thus, marking dependent rates are used in the SPN.

Protein synthesis example

- A single gene is represented which is initially inactive, but may later be activated.
- When the gene is activated protein may be produced.
- Once protein is produced it may be degraded.

Protein synthesis example



r – activation rate
s – inactivation
t – protein synthesis
u – degradation

Model analysis

The authors discuss three approaches to analysis of the model:

Model analysis

The authors discuss three approaches to analysis of the model:

- structural analysis;

Model analysis

The authors discuss three approaches to analysis of the model:

- structural analysis;
- numerical analysis;

Model analysis

The authors discuss three approaches to analysis of the model:

- structural analysis;
- numerical analysis;
- simulation.

Model analysis

The authors discuss three approaches to analysis of the model:

- structural analysis;
- numerical analysis;
- simulation.

They consider numerical analysis and simulation for this example but point out that in general state spaces are so large that numerical analysis is precluded.

Numerical analysis

The authors use the input gates of SANs to limit the number of protein molecules to 100.

Numerical analysis

The authors use the input gates of SANs to limit the number of protein molecules to 100.

Rewards are associated with places and transitions in order to calculate the measures of interest — in this case the number of protein molecules.

Numerical analysis

The authors use the input gates of SANs to limit the number of protein molecules to 100.

Rewards are associated with places and transitions in order to calculate the measures of interest — in this case the number of protein molecules.

Transient and steady state analysis are conducted in UltraSAN using both the numerical solver and simulation.

Numerical analysis

The authors use the input gates of SANs to limit the number of protein molecules to 100.

Rewards are associated with places and transitions in order to calculate the measures of interest — in this case the number of protein molecules.

Transient and steady state analysis are conducted in UltraSAN using both the numerical solver and simulation.

These are both compared against a symbolic solution of the Kolmogorov equations for the same system previously derived by Peccoud and Ycart.

Analysis results: number of protein molecules

	Mean	*Variance*
Transient analysis (time $t = 10$)		
Symbolic solution	1.488	1.858
Numerical solution	1.488	1.858
Simulation	1.481 +- 0.004	1.852 +- 0.011
Transient analysis (time $t = 100$)		
Symbolic solution	7.202	8.334
Numerical solution	7.202	8.334
Simulation	7.171 +- 0.009	8.315 +- 0.039
Steady state analysis		
Symbolic solution	8.333	9.487
Numerical solution	8.333	9.487
Simulation	8.333 +- 0.031	9.487 +- 0.100

Conclusions

The authors found advantages in using a high-level stochastic language (SPN) and supporting tool (UltraSAN):

- Allows the biologist to focus on the content of the model rather than its implementation;
- Standard format facilitates replication, extension and exchange of models between researchers;
- Existing solution engines produce results that can be related to biological phenomena.

Process Algebras for Systems Biology

Process algebras have several attractive features which could be useful for modelling and understanding biological systems:

Process Algebras for Systems Biology

Process algebras have several attractive features which could be useful for modelling and understanding biological systems:

- Process algebraic formulations are compositional and make interactions/constraints explicit — not the case with classical ordinary differential equation models.

Process Algebras for Systems Biology

Process algebras have several attractive features which could be useful for modelling and understanding biological systems:

- Process algebraic formulations are compositional and make interactions/constraints explicit — not the case with classical ordinary differential equation models.
- Structure can also be apparent.

Process Algebras for Systems Biology

Process algebras have several attractive features which could be useful for modelling and understanding biological systems:

- Process algebraic formulations are compositional and make interactions/constraints explicit — not the case with classical ordinary differential equation models.
- Structure can also be apparent.
- Equivalence relations allow formal comparison of high-level descriptions.

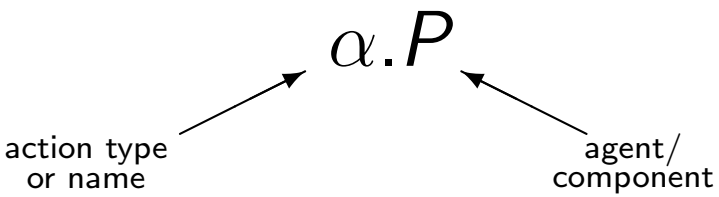
Process Algebras for Systems Biology

Process algebras have several attractive features which could be useful for modelling and understanding biological systems:

- Process algebraic formulations are compositional and make interactions/constraints explicit — not the case with classical ordinary differential equation models.
- Structure can also be apparent.
- Equivalence relations allow formal comparison of high-level descriptions.
- There are well-established techniques for reasoning about the behaviours and properties of models, supported by software. These include qualitative and quantitative analysis, and model checking.

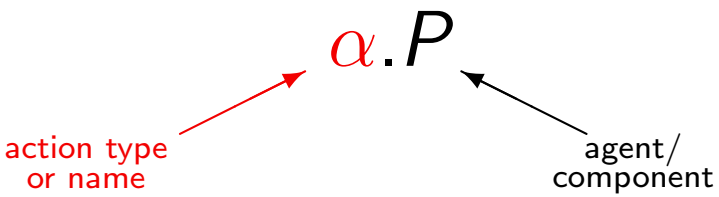
Process Algebra

- Models consist of agents which engage in actions.



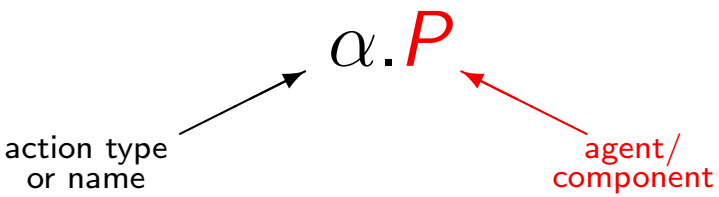
Process Algebra

- Models consist of agents which engage in actions.



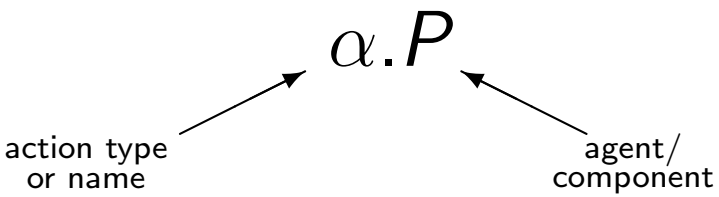
Process Algebra

- Models consist of agents which engage in actions.



Process Algebra

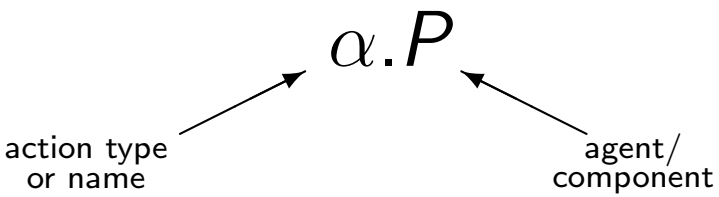
- Models consist of agents which engage in actions.



- The structured operational (interleaving) semantics of the language is used to generate a labelled transition system.

Process Algebra

- Models consist of agents which engage in actions.

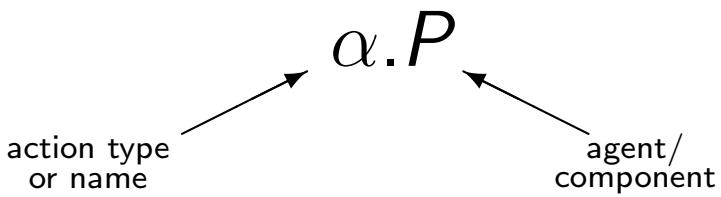


- The structured operational (interleaving) semantics of the language is used to generate a labelled transition system.

Process algebra model

Process Algebra

- Models consist of agents which engage in actions.

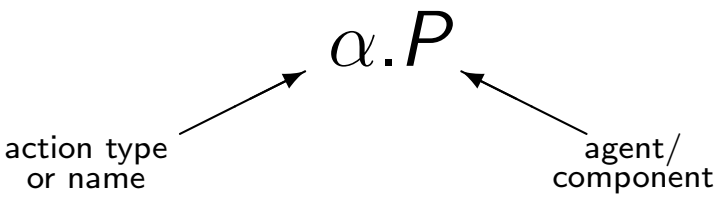


- The structured operational (interleaving) semantics of the language is used to generate a labelled transition system.

Process algebra model $\xrightarrow{\text{SOS rules}}$

Process Algebra

- Models consist of agents which engage in actions.



- The structured operational (interleaving) semantics of the language is used to generate a labelled transition system.

Process algebra model $\xrightarrow{\text{SOS rules}}$ Labelled transition system

Calculus of Communicating Systems (CCS)

Introduced to capture the behaviour of concurrent programs, CCS first appeared in the last 1970s.

Calculus of Communicating Systems (CCS)

Introduced to capture the behaviour of concurrent programs, CCS first appeared in the last 1970s.

$\alpha.P$ Prefix

Calculus of Communicating Systems (CCS)

Introduced to capture the behaviour of concurrent programs, CCS first appeared in the last 1970s.

$\alpha.P$	Prefix
$P_1 + P_2$	Choice

Calculus of Communicating Systems (CCS)

Introduced to capture the behaviour of concurrent programs, CCS first appeared in the last 1970s.

$\alpha.P$	Prefix
$P_1 + P_2$	Choice
$P_1 \mid P_2$	Composition

Calculus of Communicating Systems (CCS)

Introduced to capture the behaviour of concurrent programs, CCS first appeared in the last 1970s.

$\alpha.P$	Prefix
$P_1 + P_2$	Choice
$P_1 \mid P_2$	Composition
$P \backslash L$	Restriction

Calculus of Communicating Systems (CCS)

Introduced to capture the behaviour of concurrent programs, CCS first appeared in the last 1970s.

$\alpha.P$	Prefix
$P_1 + P_2$	Choice
$P_1 \mid P_2$	Composition
$P \backslash L$	Restriction
$P[f]$	Relabelling

Calculus of Communicating Systems (CCS)

Introduced to capture the behaviour of concurrent programs, CCS first appeared in the last 1970s.

$\alpha.P$	Prefix	
$P_1 + P_2$	Choice	
$P_1 \mid P_2$	Composition	
$P \backslash L$	Restriction	
$P[f]$	Relabelling	
$A \stackrel{def}{=} P$	Constant	*or*

Calculus of Communicating Systems (CCS)

Introduced to capture the behaviour of concurrent programs, CCS first appeared in the last 1970s.

$\alpha.P$	Prefix	
$P_1 + P_2$	Choice	
$P_1 \mid P_2$	Composition	
$P \backslash L$	Restriction	
$P[f]$	Relabelling	
$A \stackrel{def}{=} P$	Constant	*or*
$\mathbf{fix}(X = E)$	Recursion	

Dynamic behaviour

- The behaviour of a model is dictated by the semantic rules governing the combinators of the language.

Dynamic behaviour

- The behaviour of a model is dictated by the semantic rules governing the combinators of the language.
- The possible evolutions of a model are captured by applying these rules exhaustively, generating a labelled transition system.

Dynamic behaviour

- The behaviour of a model is dictated by the semantic rules governing the combinators of the language.
- The possible evolutions of a model are captured by applying these rules exhaustively, generating a labelled transition system.
- This can be viewed as a graph in which each node is a state of the model (comprised of the local states of each of the components) and the arcs represent the actions which can cause the move from one state to another.

Dynamic behaviour

- The behaviour of a model is dictated by the semantic rules governing the combinators of the language.
- The possible evolutions of a model are captured by applying these rules exhaustively, generating a labelled transition system.
- This can be viewed as a graph in which each node is a state of the model (comprised of the local states of each of the components) and the arcs represent the actions which can cause the move from one state to another.
- The language is also equipped with observational equivalence which can be used to compare models.

Dynamic behaviour

$$Browser \stackrel{def}{=} display.(cache.Browser + get.download.rel.Browser)$$

Dynamic behaviour

$$Browser \stackrel{def}{=} display.(cache.Browser + get.download.rel.Browser)$$

$$\frac{}{\alpha.P \xrightarrow{\alpha} P}$$

$$\frac{P \xrightarrow{\alpha} P'}{P + Q \xrightarrow{\alpha} P'}$$

$$\frac{Q \xrightarrow{\alpha} Q'}{P + Q \xrightarrow{\alpha} Q'}$$

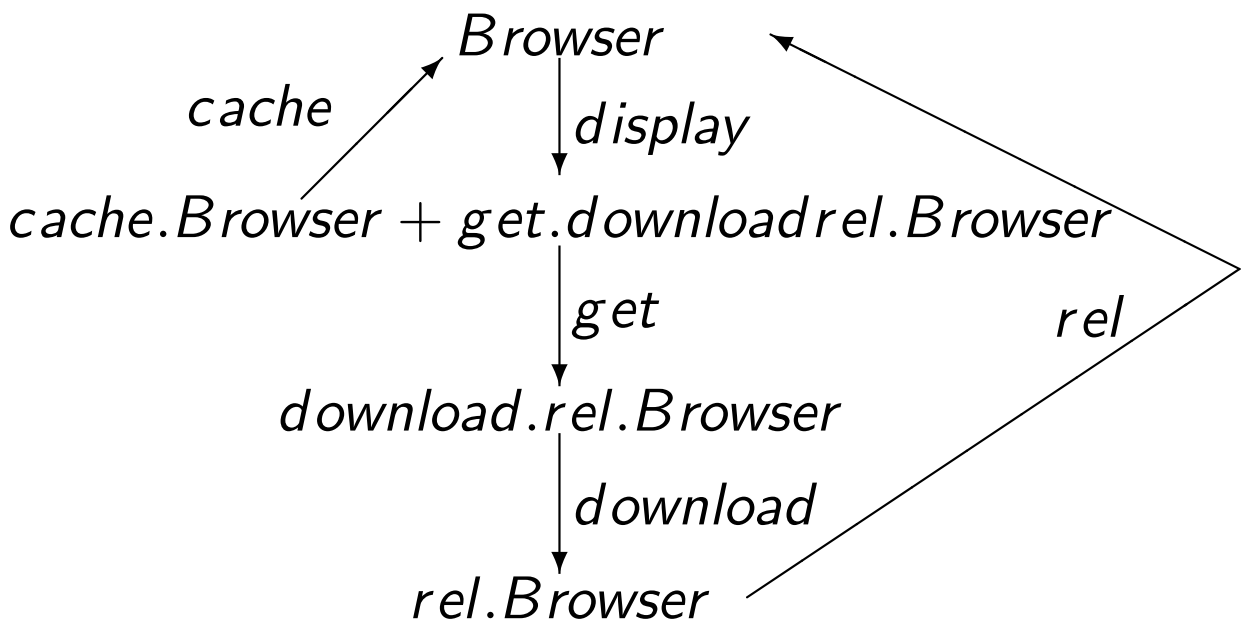
Dynamic behaviour

$$Browser \stackrel{def}{=} display.(cache.Browser + get.download.rel.Browser)$$

$$\frac{}{\alpha.P \xrightarrow{\alpha} P}$$

$$\frac{P \xrightarrow{\alpha} P'}{P + Q \xrightarrow{\alpha} P'}$$

$$\frac{Q \xrightarrow{\alpha} Q'}{P + Q \xrightarrow{\alpha} Q'}$$



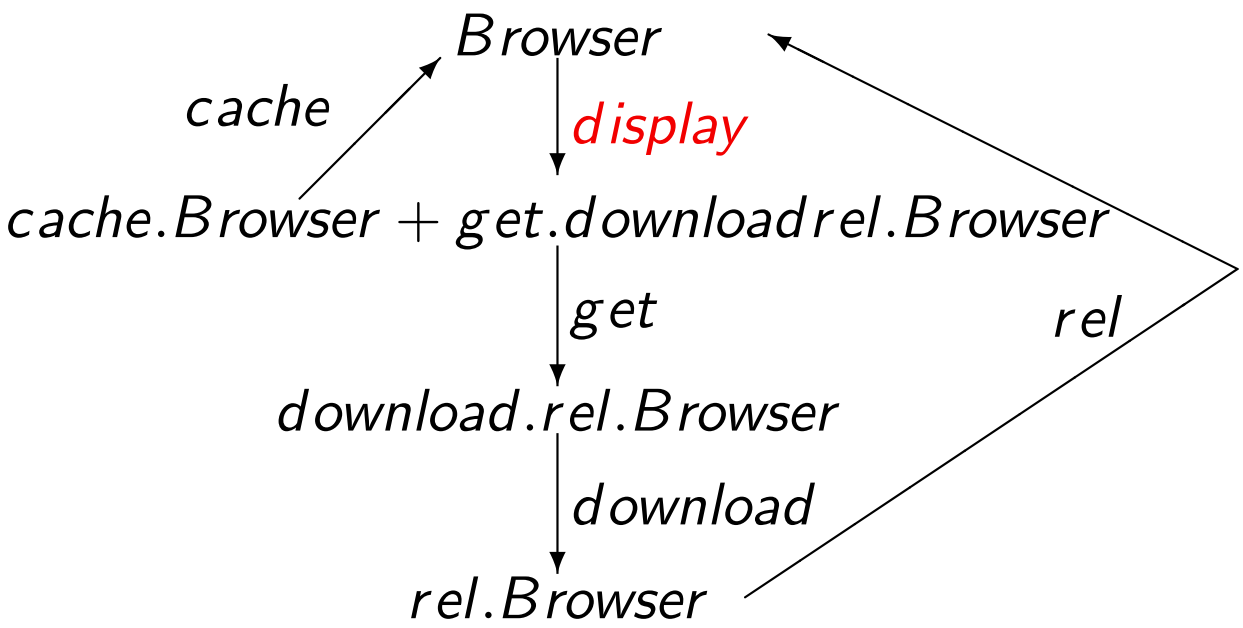
Dynamic behaviour

$$Browser \stackrel{def}{=} display.(cache.Browser + get.download.rel.Browser)$$

$$\frac{}{\alpha.P \xrightarrow{\alpha} P}$$

$$\frac{P \xrightarrow{\alpha} P'}{P + Q \xrightarrow{\alpha} P'}$$

$$\frac{Q \xrightarrow{\alpha} Q'}{P + Q \xrightarrow{\alpha} Q'}$$



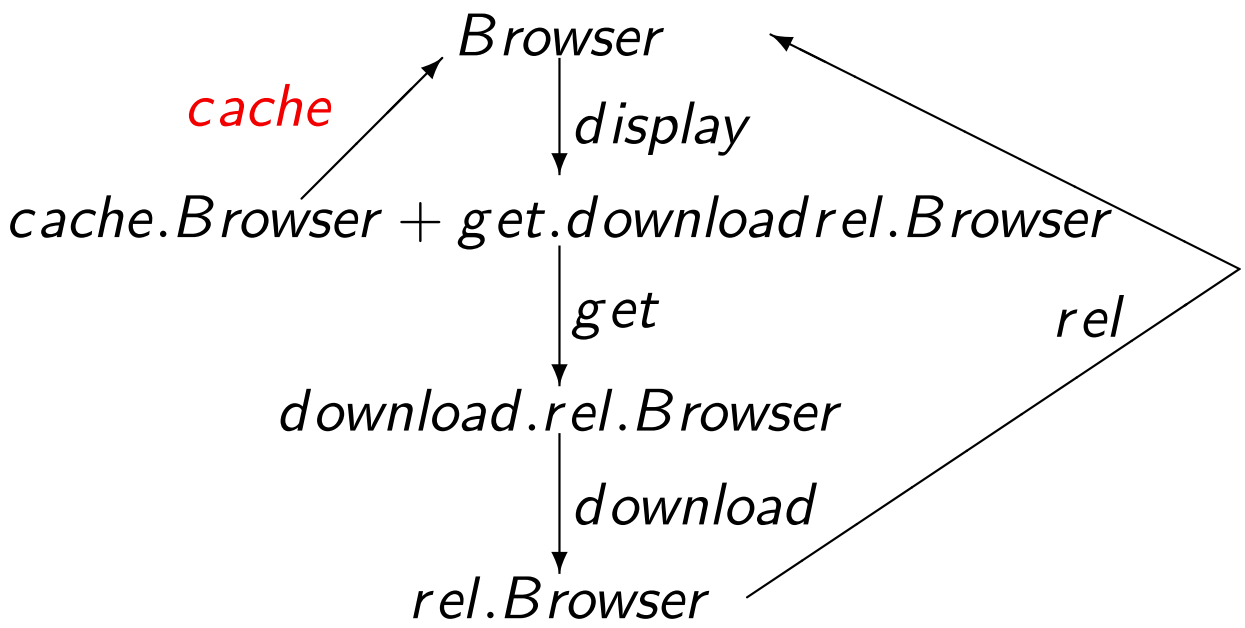
Dynamic behaviour

$$Browser \stackrel{def}{=} display.(cache.Browser + get.download.rel.Browser)$$

$$\frac{}{\alpha.P \xrightarrow{\alpha} P}$$

$$\frac{P \xrightarrow{\alpha} P'}{P + Q \xrightarrow{\alpha} P'}$$

$$\frac{Q \xrightarrow{\alpha} Q'}{P + Q \xrightarrow{\alpha} Q'}$$



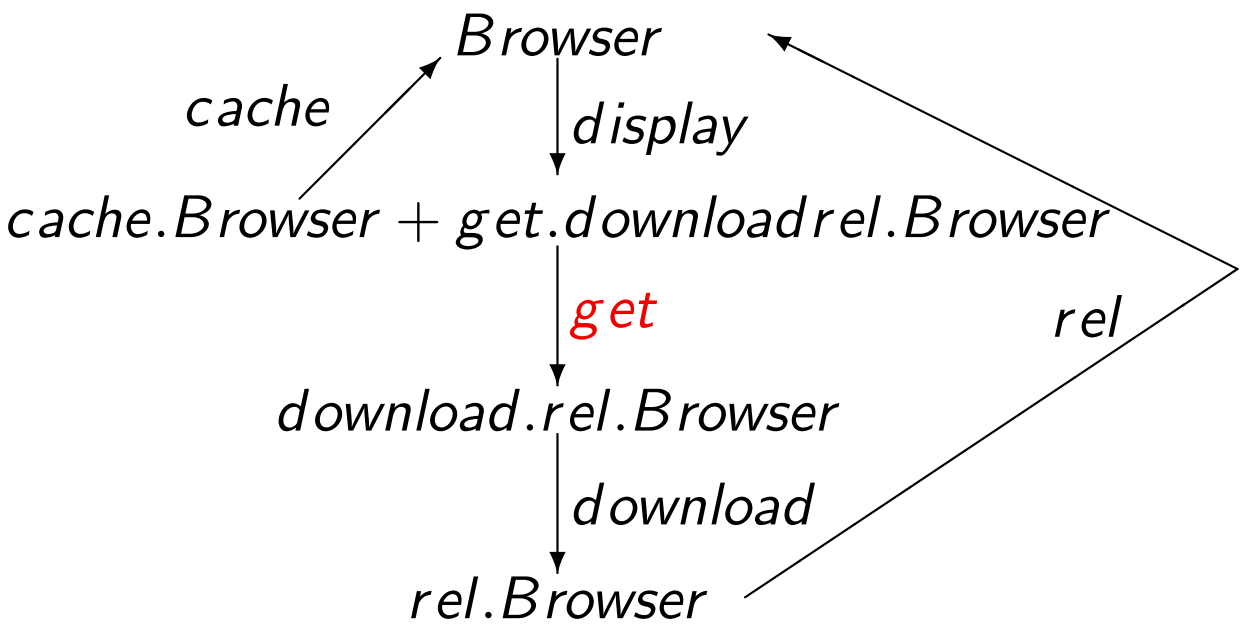
Dynamic behaviour

$$Browser \stackrel{def}{=} display.(cache.Browser + get.download.rel.Browser)$$

$$\frac{}{\alpha.P \xrightarrow{\alpha} P}$$

$$\frac{P \xrightarrow{\alpha} P'}{P + Q \xrightarrow{\alpha} P'}$$

$$\frac{Q \xrightarrow{\alpha} Q'}{P + Q \xrightarrow{\alpha} Q'}$$



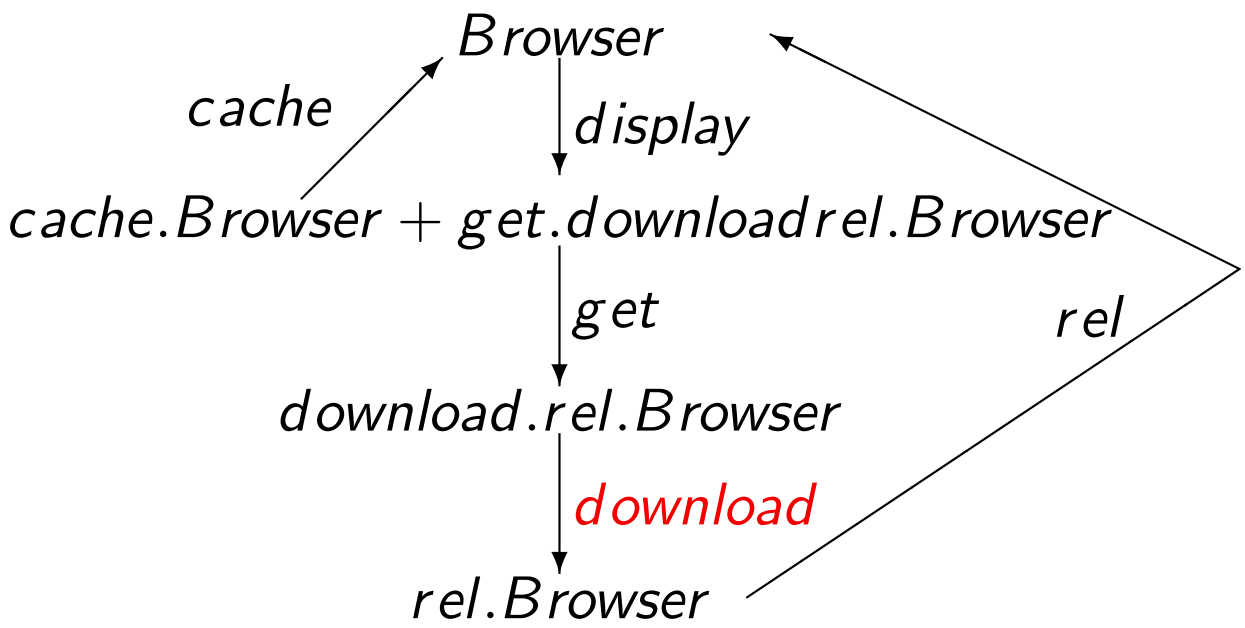
Dynamic behaviour

$$Browser \stackrel{def}{=} display.(cache.Browser + get.download.rel.Browser)$$

$$\frac{}{\alpha.P \xrightarrow{\alpha} P}$$

$$\frac{P \xrightarrow{\alpha} P'}{P + Q \xrightarrow{\alpha} P'}$$

$$\frac{Q \xrightarrow{\alpha} Q'}{P + Q \xrightarrow{\alpha} Q'}$$



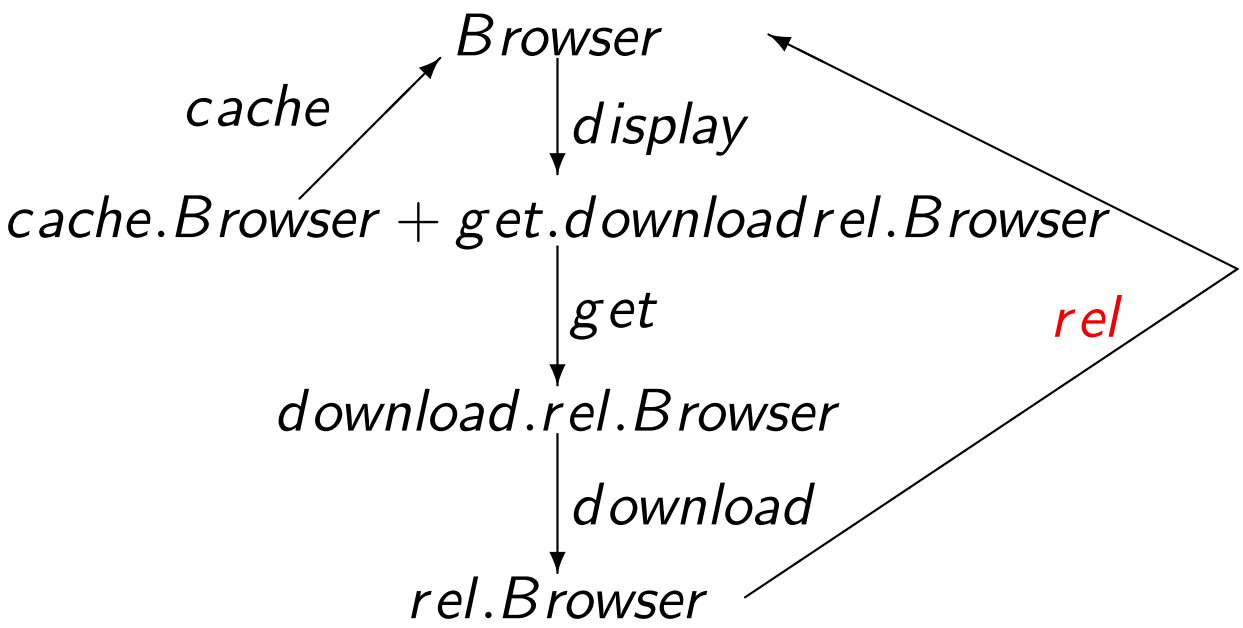
Dynamic behaviour

$$Browser \stackrel{def}{=} display.(cache.Browser + get.download.rel.Browser)$$

$$\frac{}{\alpha.P \xrightarrow{\alpha} P}$$

$$\frac{P \xrightarrow{\alpha} P'}{P + Q \xrightarrow{\alpha} P'}$$

$$\frac{Q \xrightarrow{\alpha} Q'}{P + Q \xrightarrow{\alpha} Q'}$$

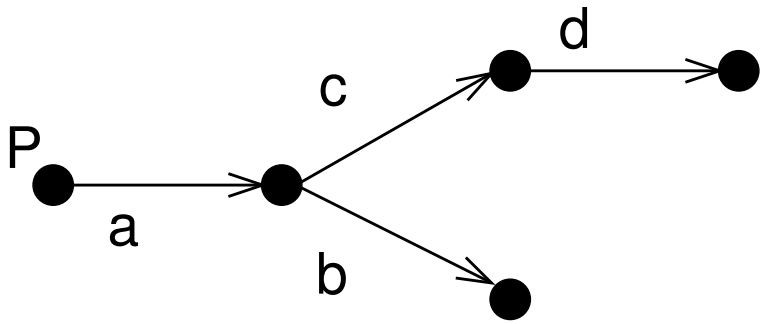


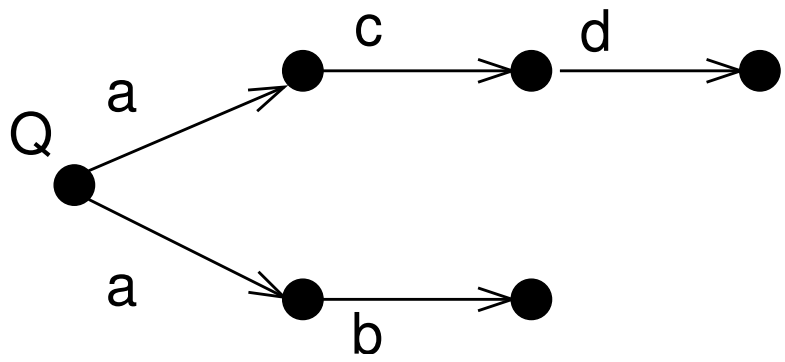
Observational Equivalence

In process algebra equivalence relations are defined based on the notion of observability:

Observational Equivalence

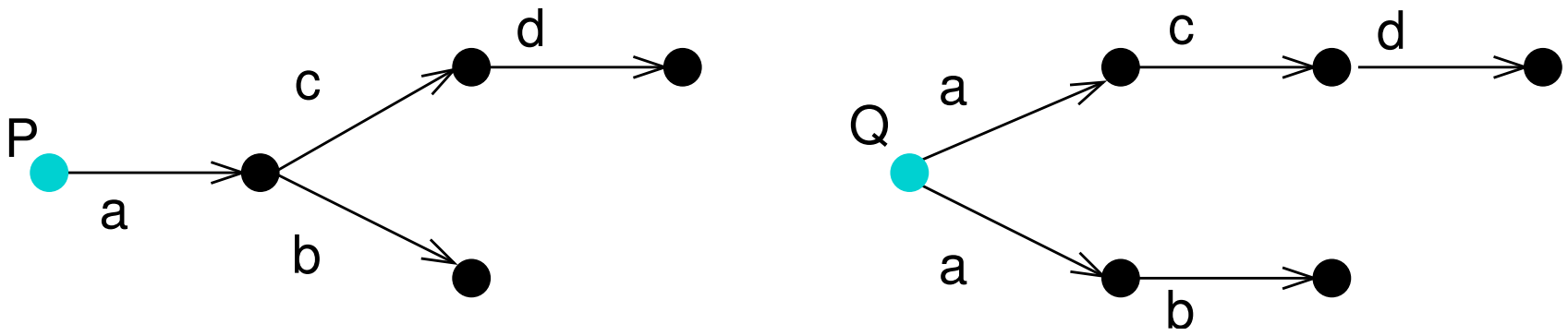
In process algebra equivalence relations are defined based on the notion of observability:





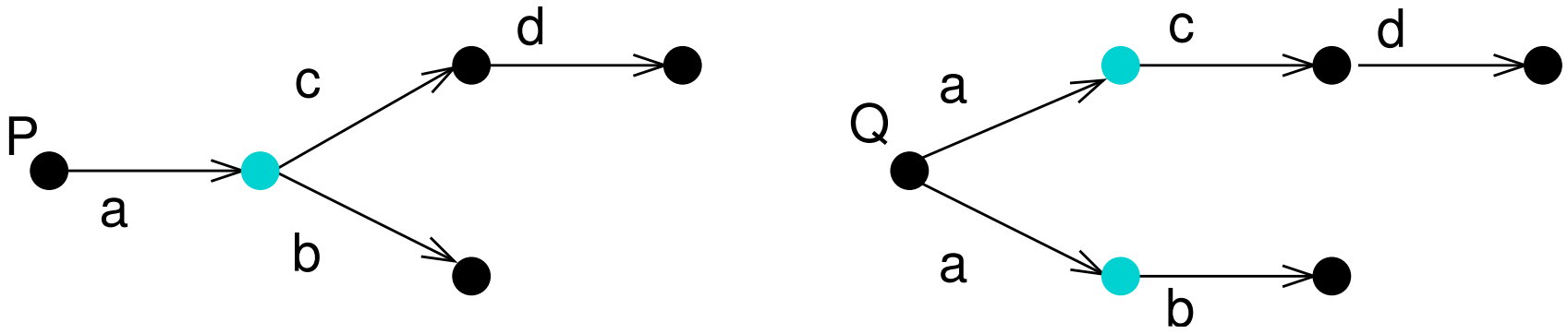
Observational Equivalence

In process algebra equivalence relations are defined based on the notion of observability:



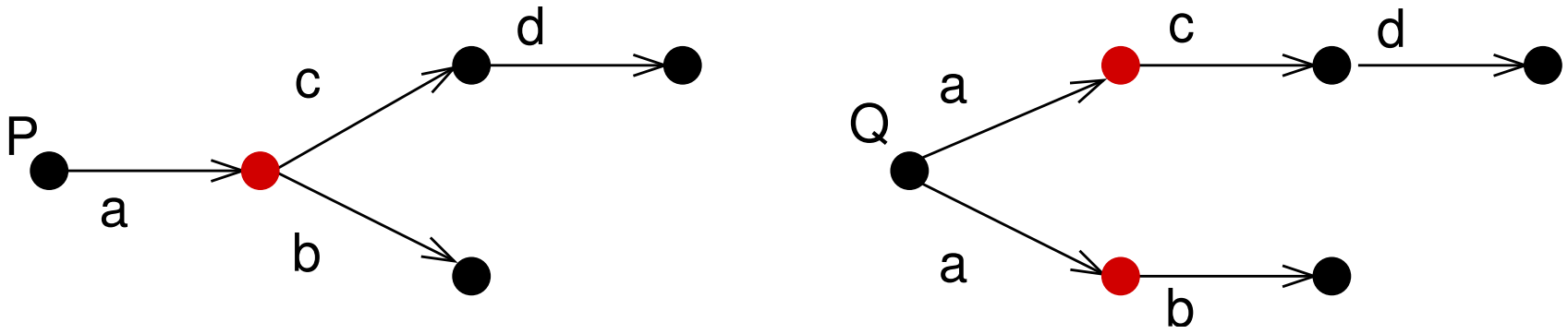
Observational Equivalence

In process algebra equivalence relations are defined based on the notion of observability:



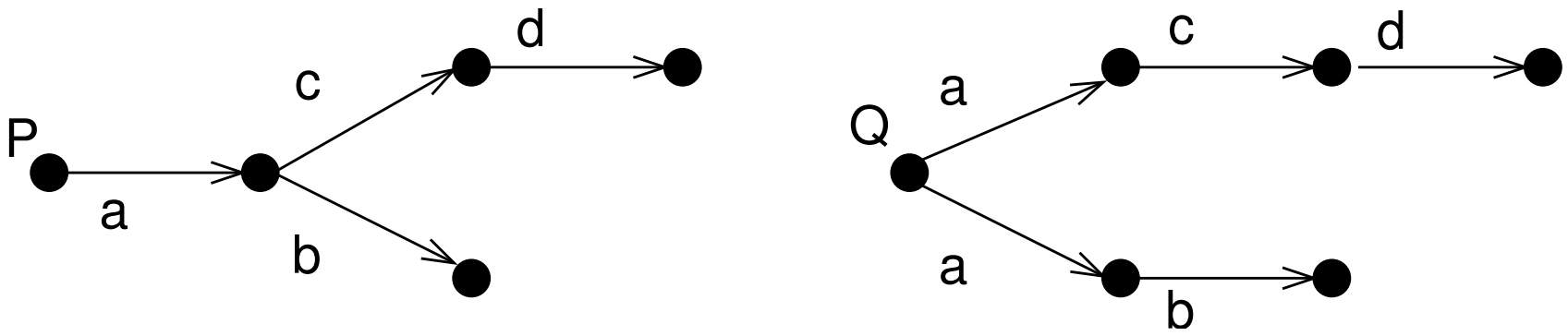
Observational Equivalence

In process algebra equivalence relations are defined based on the notion of observability:



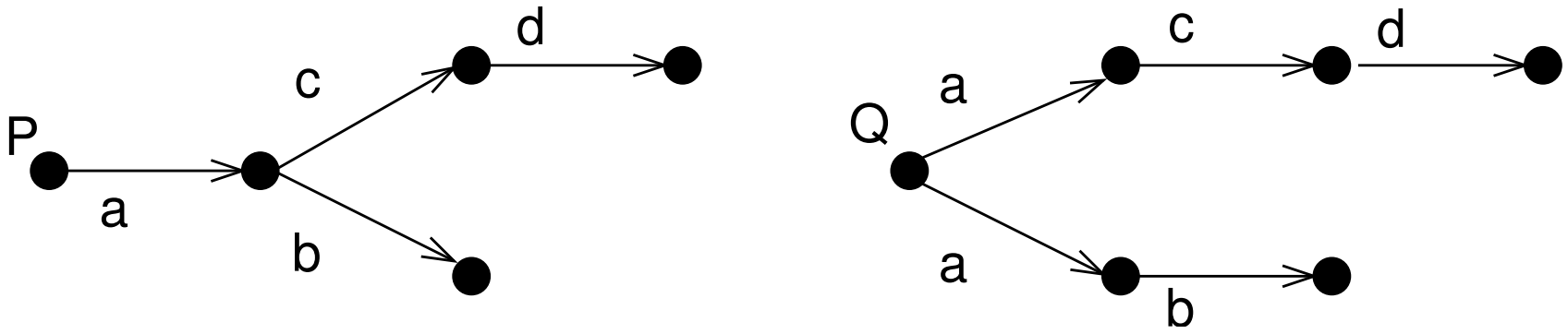
Observational Equivalence

In process algebra equivalence relations are defined based on the notion of observability:



Observational Equivalence

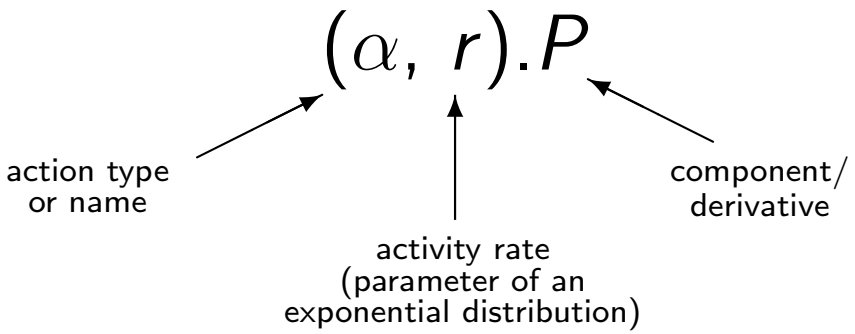
In process algebra equivalence relations are defined based on the notion of observability:



Processes are equivalent if they can match actions and arrive at states that also match actions.

Performance Evaluation Process Algebra

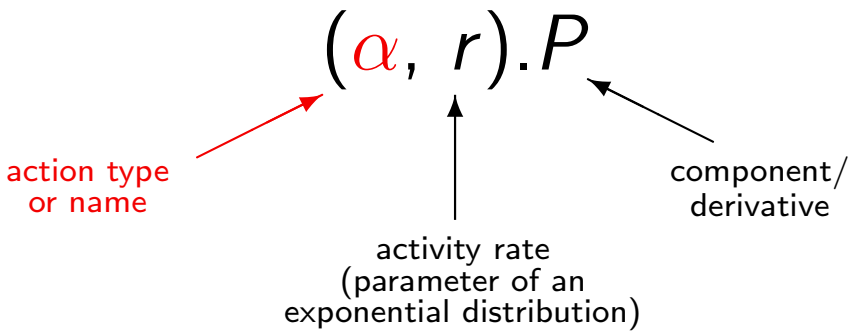
- Models are constructed from components which engage in activities.



The language may be used to generate a Markov Process (CTMC).

Performance Evaluation Process Algebra

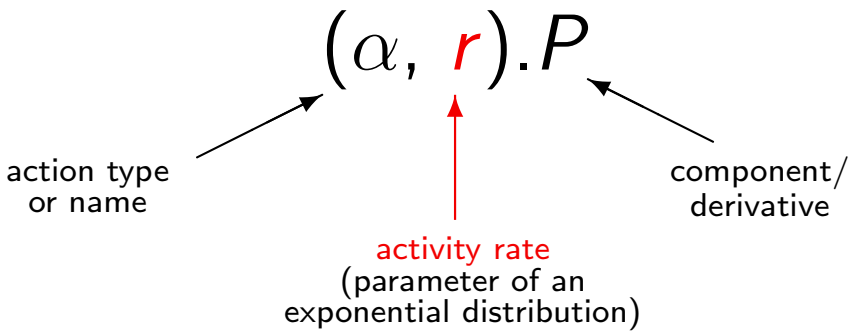
- Models are constructed from components which engage in activities.



The language may be used to generate a Markov Process (CTMC).

Performance Evaluation Process Algebra

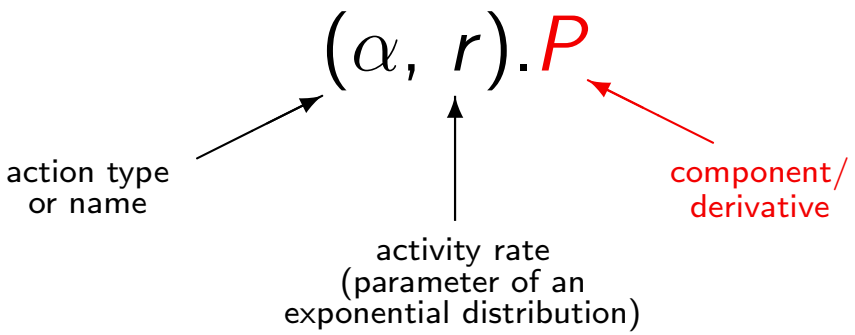
- Models are constructed from components which engage in activities.



The language may be used to generate a Markov Process (CTMC).

Performance Evaluation Process Algebra

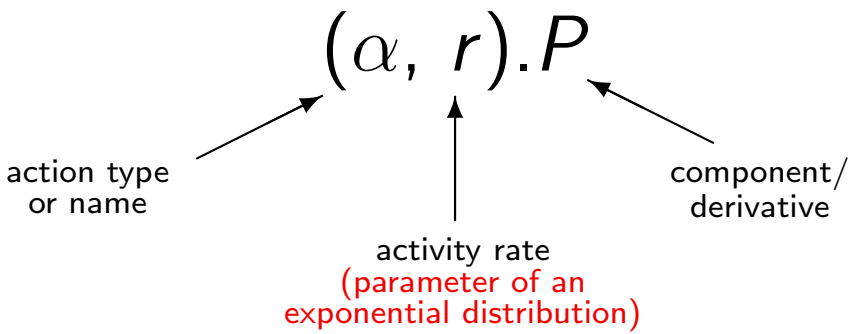
- Models are constructed from components which engage in activities.



The language may be used to generate a Markov Process (CTMC).

Performance Evaluation Process Algebra

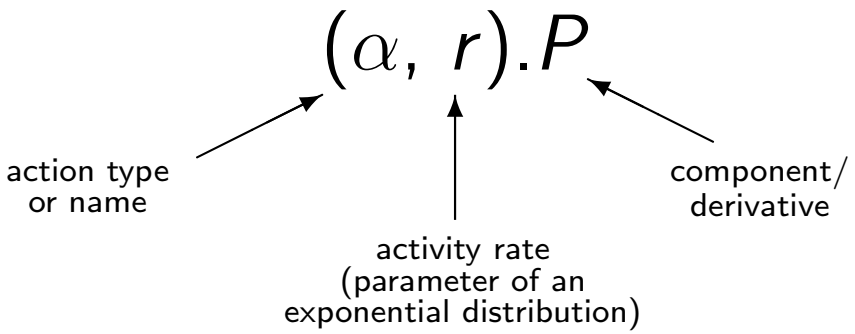
- Models are constructed from components which engage in activities.



The language may be used to generate a Markov Process (CTMC).

Performance Evaluation Process Algebra

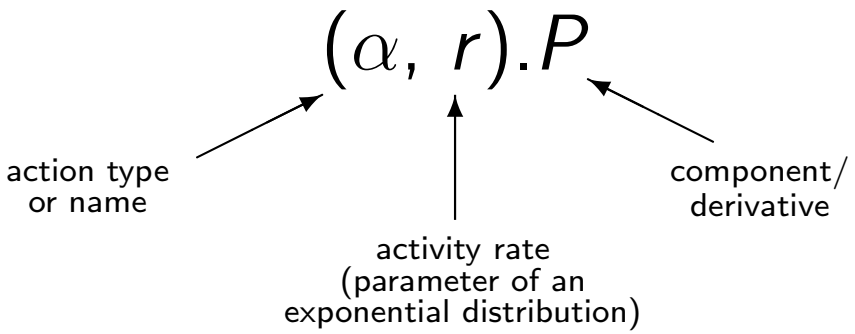
- Models are constructed from components which engage in activities.



The language may be used to generate a Markov Process (CTMC).

Performance Evaluation Process Algebra

- Models are constructed from components which engage in activities.

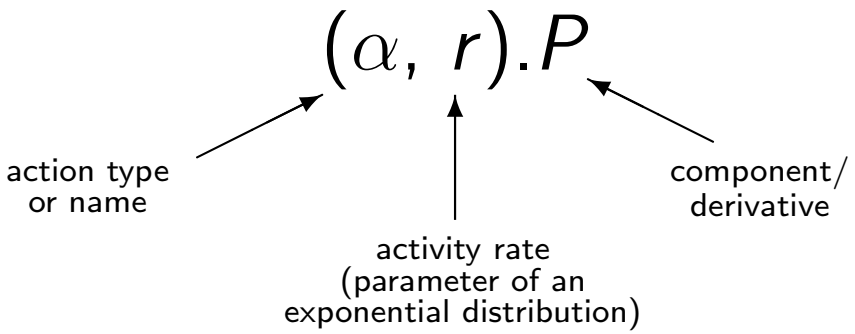


The language may be used to generate a Markov Process (CTMC).

PEPA
MODEL

Performance Evaluation Process Algebra

- Models are constructed from components which engage in activities.

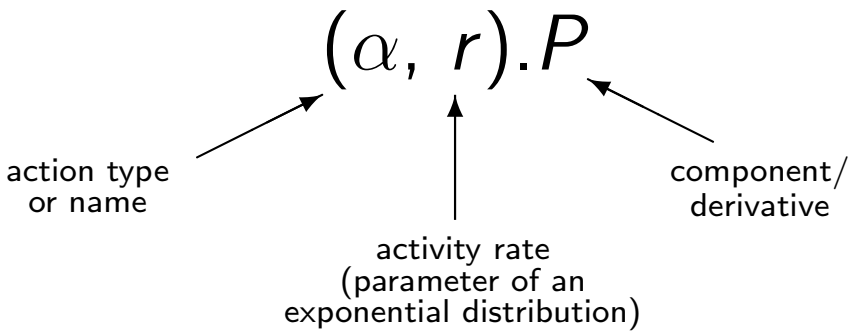


The language may be used to generate a Markov Process (CTMC).

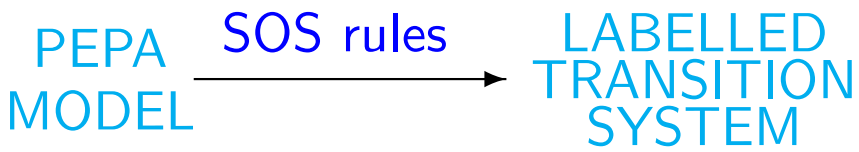
PEPA MODEL $\xrightarrow{\text{SOS rules}}$

Performance Evaluation Process Algebra

- Models are constructed from components which engage in activities.

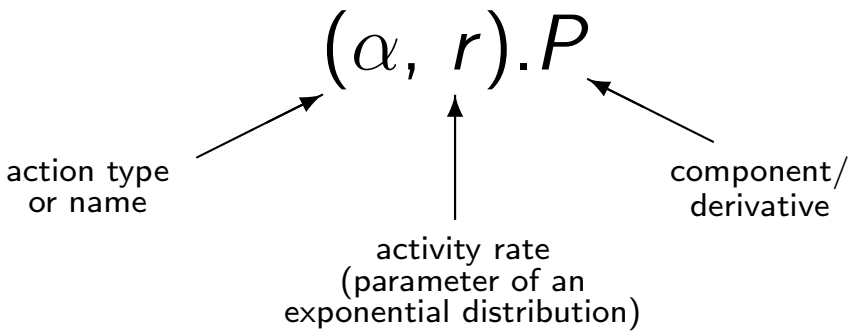


The language may be used to generate a Markov Process (CTMC).



Performance Evaluation Process Algebra

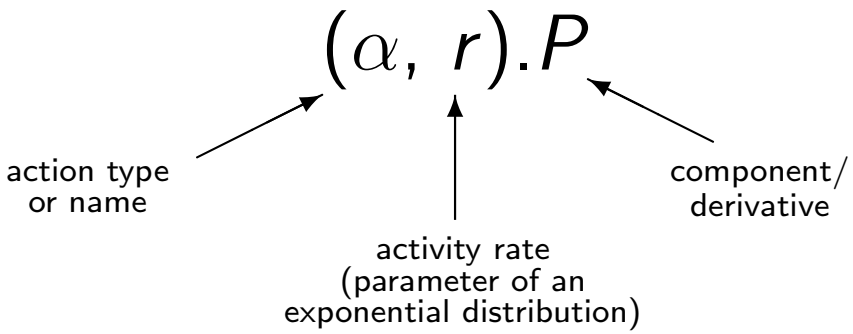
- Models are constructed from components which engage in activities.



The language may be used to generate a Markov Process (CTMC).

Performance Evaluation Process Algebra

- Models are constructed from components which engage in activities.



The language may be used to generate a Markov Process (CTMC).

Performance Evaluation Process Algebra (PEPA)

Performance Evaluation Process Algebra (PEPA)

PREFIX: $(\alpha, r).S$ designated first action

Performance Evaluation Process Algebra (PEPA)

PREFIX:	$(\alpha, r).S$	designated first action
CHOICE:	$S + S$	competing components (determined by race policy)

Performance Evaluation Process Algebra (PEPA)

PREFIX:	$(\alpha, r).S$	designated first action
CHOICE:	$S + S$	competing components (determined by race policy)
CONSTANT:	$A \stackrel{def}{=} S$	assigning names

Performance Evaluation Process Algebra (PEPA)

PREFIX:	$(\alpha, r).S$	designated first action
CHOICE:	$S + S$	competing components (determined by race policy)
CONSTANT:	$A \stackrel{def}{=} S$	assigning names
COOPERATION:	$P \underset{L}{\bowtie} P$	$\alpha \notin L$ concurrent activity (*individual actions*) $\alpha \in L$ cooperative activity (*shared actions*)

Performance Evaluation Process Algebra (PEPA)

PREFIX:	$(\alpha, r).S$	designated first action
CHOICE:	$S + S$	competing components (determined by race policy)
CONSTANT:	$A \stackrel{def}{=} S$	assigning names
COOPERATION:	$P \underset{L}{\bowtie} P$	$\alpha \notin L$ concurrent activity (*individual actions*) $\alpha \in L$ cooperative activity (*shared actions*)

Performance Evaluation Process Algebra (PEPA)

PREFIX:	$(\alpha, r).S$	designated first action
CHOICE:	$S + S$	competing components (determined by race policy)
CONSTANT:	$A \stackrel{def}{=} S$	assigning names
COOPERATION:	$P \underset{L}{\bowtie} P$	$\alpha \notin L$ concurrent activity (*individual actions*) $\alpha \in L$ cooperative activity (*shared actions*)
HIDING:	P/L	abstraction $\alpha \in L \Rightarrow \alpha \rightarrow \tau$

Molecular processes as concurrent computations

Concurrency	**Molecular Biology**	**Metabolism**	**Signal Transduction**
Concurrent computational processes	Molecules	Enzymes and metabolites	Interacting proteins
Synchronous communication	Molecular interaction	Binding and catalysis	Binding and catalysis
Transition or mobility	Biochemical modification or relocation	Metabolite synthesis	Protein binding, modification or sequestration

[Regev *et al* 2000]

Molecular processes as concurrent computations

Concurrency	**Molecular Biology**	**Metabolism**	**Signal Transduction**
Concurrent computational processes	Molecules	Enzymes and metabolites	Interacting proteins
Synchronous communication	Molecular interaction	Binding and catalysis	Binding and catalysis
Transition or mobility	Biochemical modification or relocation	Metabolite synthesis	Protein binding, modification or sequestration

[Regev *et al* 2000]

Molecular processes as concurrent computations

Concurrency	**Molecular Biology**	**Metabolism**	**Signal Transduction**
Concurrent computational processes	Molecules	Enzymes and metabolites	Interacting proteins
Synchronous communication	Molecular interaction	Binding and catalysis	Binding and catalysis
Transition or mobility	Biochemical modification or relocation	Metabolite synthesis	Protein binding, modification or sequestration

[Regev *et al* 2000]

Molecular processes as concurrent computations

Concurrency	**Molecular Biology**	**Metabolism**	**Signal Transduction**
Concurrent computational processes	Molecules	Enzymes and metabolites	Interacting proteins
Synchronous communication	Molecular interaction	Binding and catalysis	Binding and catalysis
Transition or mobility	Biochemical modification or relocation	Metabolite synthesis	Protein binding, modification or sequestration

[Regev *et al* 2000]

PEPA

- Use of PEPA has focused primarily on signal transduction.

PEPA

- Use of PEPA has focused primarily on signal transduction.
- Analysis may be conducted based on Continuous Time Markov Chains, Ordinary Differential Equations and Stochastic Simulation.

PEPA

- Use of PEPA has focused primarily on signal transduction.
- Analysis may be conducted based on Continuous Time Markov Chains, Ordinary Differential Equations and Stochastic Simulation.
- The abstraction level chosen for PEPA is slightly different from that for the stochastic π-calculus: rather than associating a component with each molecule, we associate a component with a species or a pathway.

Mapping biological systems to process algebra

There has been much work on the use of the stochastic π-calculus and related calculi, for modelling biochemical signalling within cells

Mapping biological systems to process algebra

There has been much work on the use of the stochastic π-calculus and related calculi, for modelling biochemical signalling within cells

This work treats a molecule in a pathway as corresponding to the component in the process algebra description.

Mapping biological systems to process algebra

There has been much work on the use of the stochastic π-calculus and related calculi, for modelling biochemical signalling within cells

This work treats a molecule in a pathway as corresponding to the component in the process algebra description.

In the PEPA modelling we have been doing we have experimented with more abstract mappings between process algebra constructs and elements of signalling pathways.

Mapping biological systems to process algebra

There has been much work on the use of the stochastic π-calculus and related calculi, for modelling biochemical signalling within cells

This work treats a molecule in a pathway as corresponding to the component in the process algebra description.

In the PEPA modelling we have been doing we have experimented with more abstract mappings between process algebra constructs and elements of signalling pathways.

In our first mapping we focus on species (c.f. a type rather than an instance, or a class rather than an object).

Mapping biological systems to process algebra

There has been much work on the use of the stochastic π-calculus and related calculi, for modelling biochemical signalling within cells

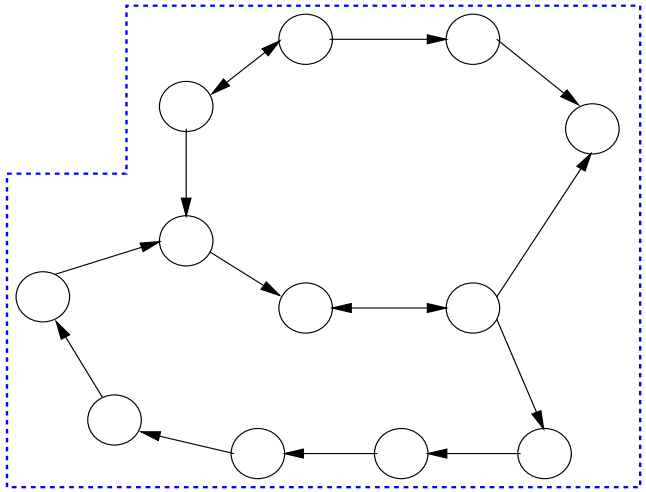
This work treats a molecule in a pathway as corresponding to the component in the process algebra description.

In the PEPA modelling we have been doing we have experimented with more abstract mappings between process algebra constructs and elements of signalling pathways.

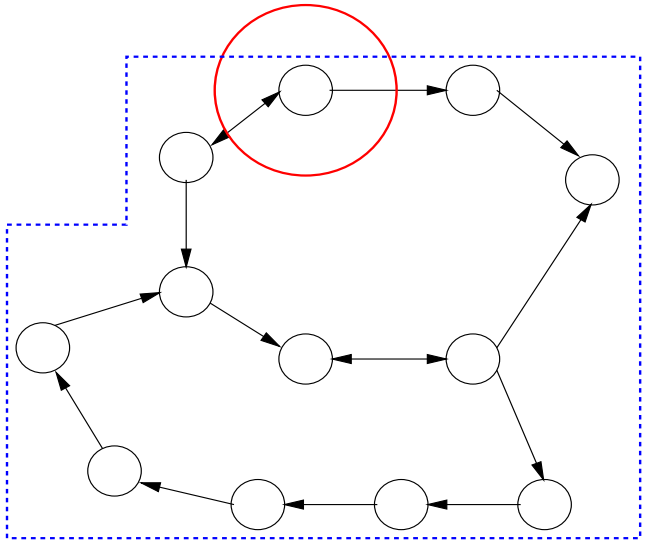
In our first mapping we focus on species (c.f. a type rather than an instance, or a class rather than an object).

In our second we focus on sub-pathways.

Alternative Mappings: illustration

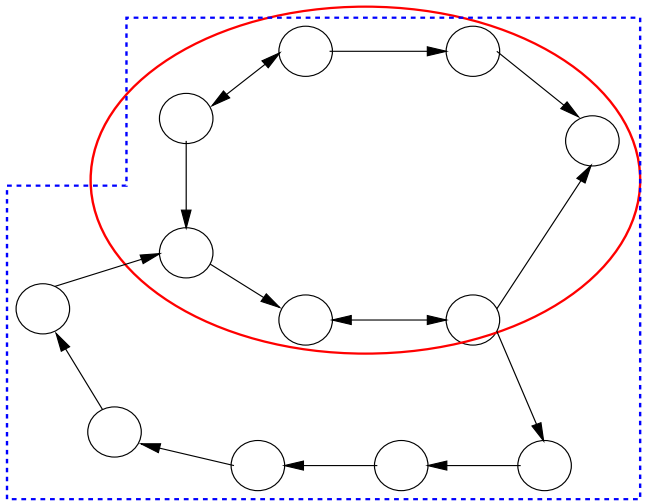


Alternative Mappings: illustration



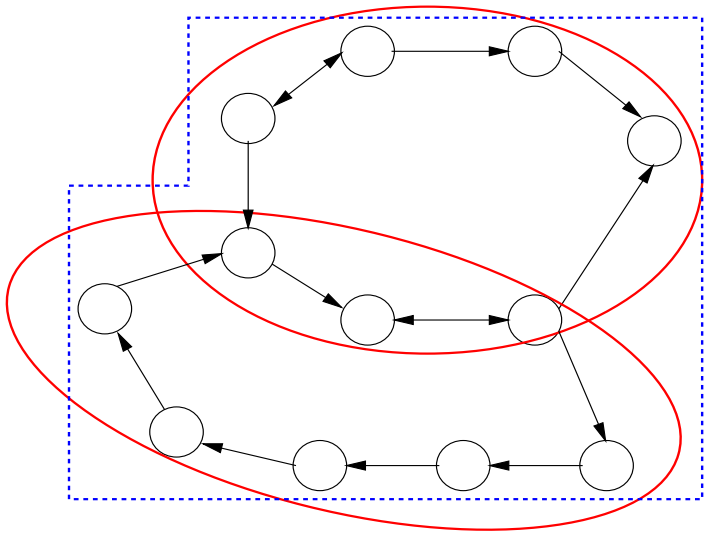
Reagent mapping: Each species is a distinct component in the model with local states to capture differing levels of concentration

Alternative Mappings: illustration



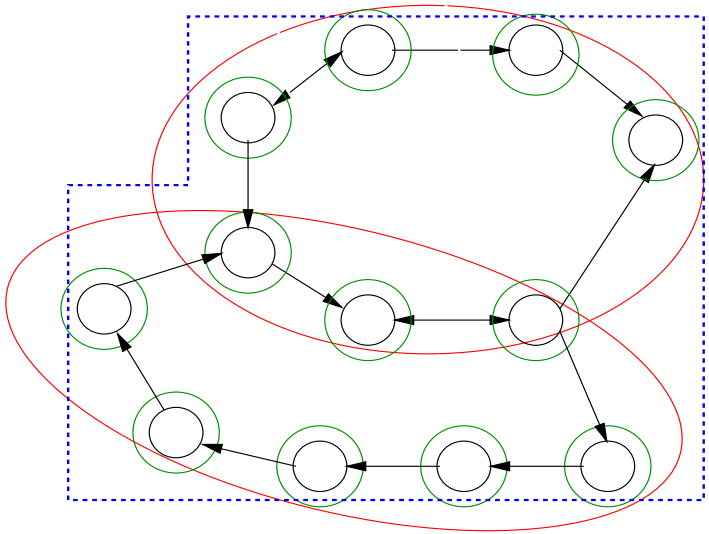
Pathway mapping: Each sub-pathway is a distinct component in the model with local states to capture progress through the pathway

Alternative Mappings: illustration



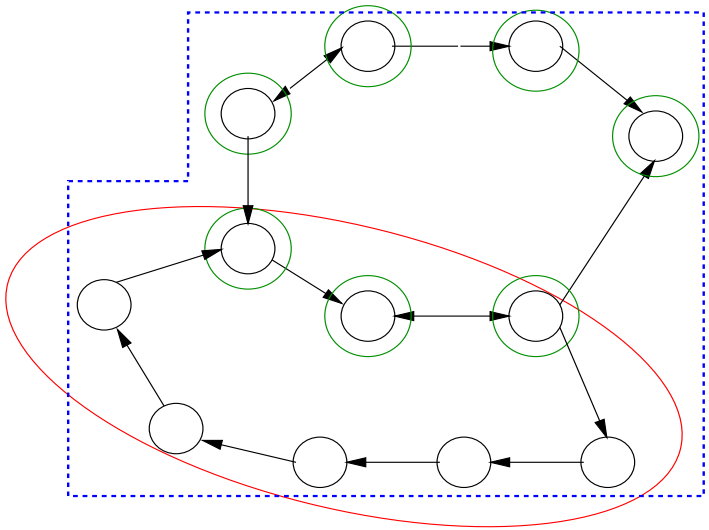
Pathway mapping: Each sub-pathway is a distinct component in the model with local states to capture progress through the pathway

Alternative Mappings: illustration



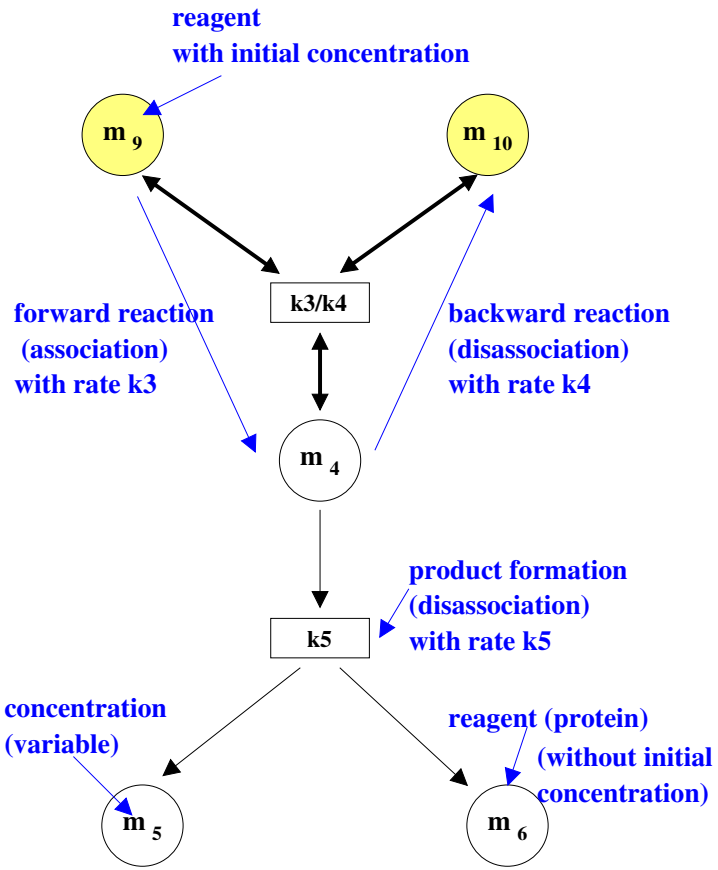
Reasoning based on bisimulation equivalence is able to prove that the two representations are equivalent.

Alternative Mappings: illustration



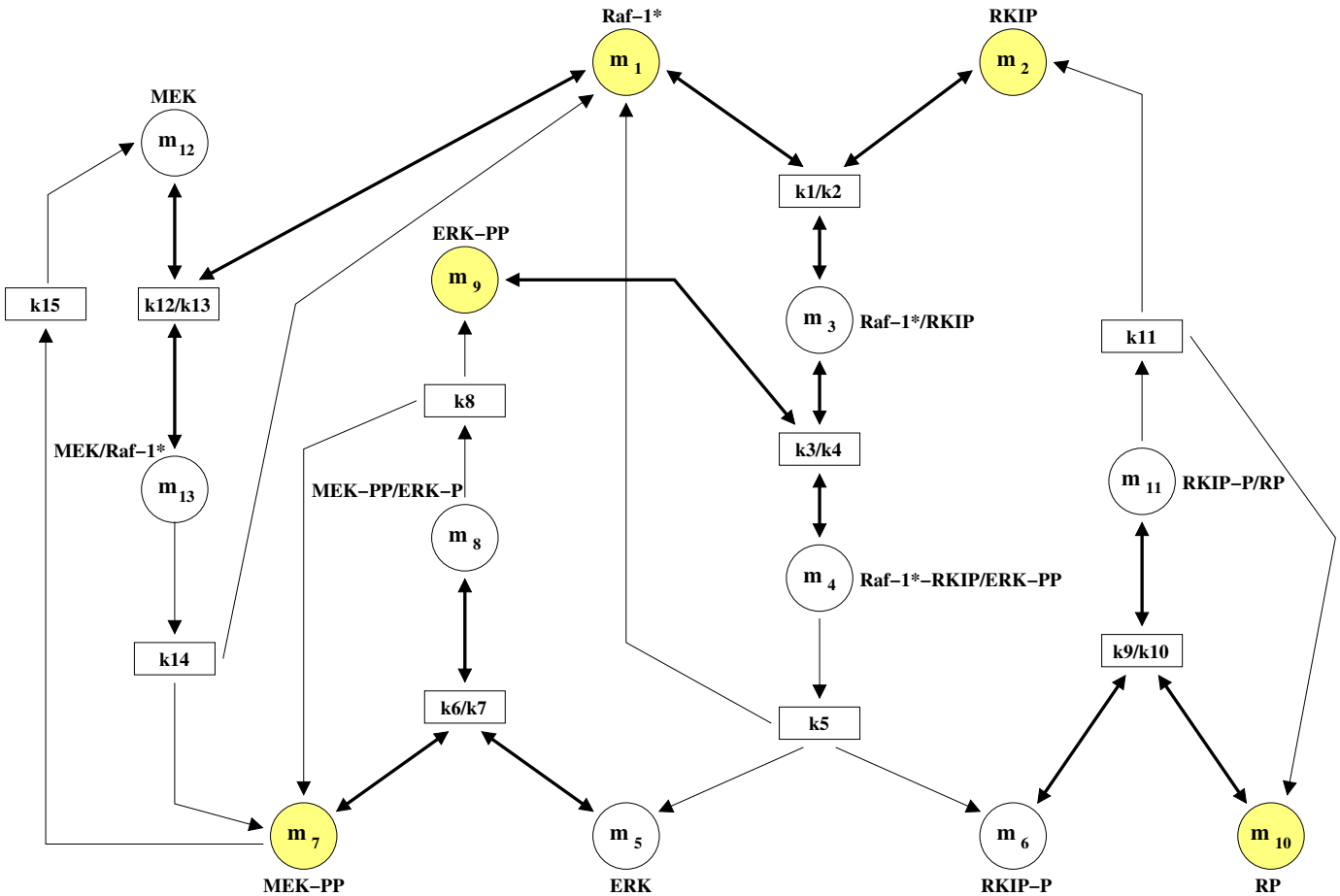
Different parts of the system may use different mappings, reflecting perhaps the level of knowledge (data) available, or the primary interests of the modeller.

Dynamics of cell signalling

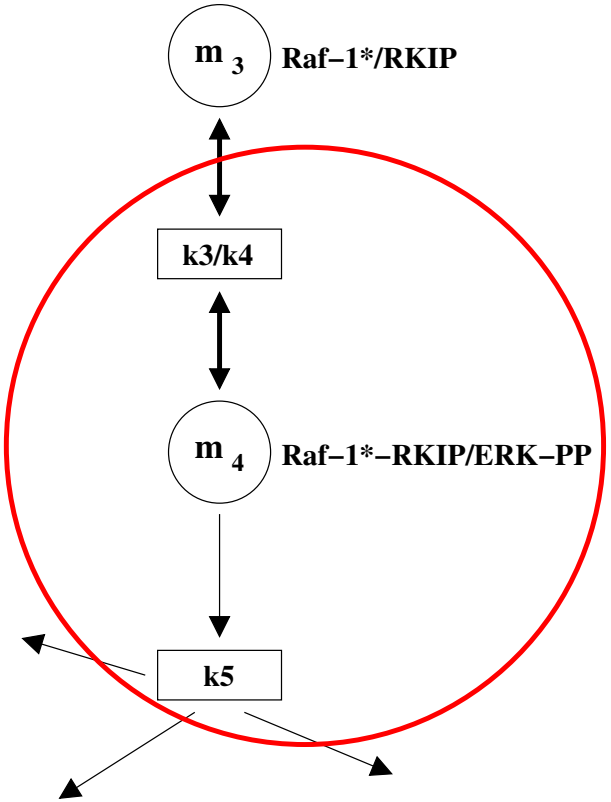


- Bi-directional arrows denote both forward and backward reactions;
- Uni-directional arrows denote reactions which are disassociations.
- Each reagent has a variable concentration, denoted m_i.
- Each reaction has a corresponding rate constant, e.g. k3, but the rate at which the reaction takes place is the product of this rate constant and the current concentrations of the *used* substrates.

Example: The Ras/Raf-1/MEK/ERK pathway



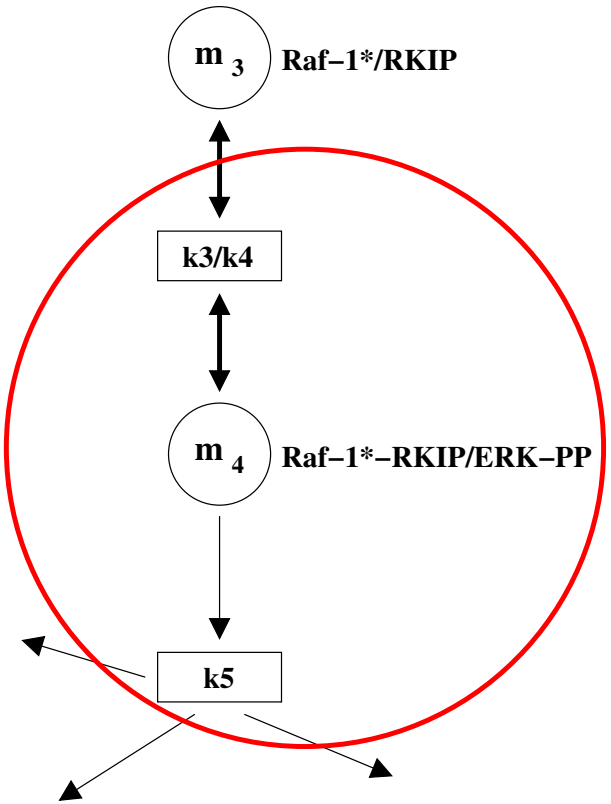
PEPA components of the reagent-centric model



$$\text{Raf-1}^*/\text{RKIP}/\text{ERK-PP}_\text{H} \stackrel{def}{=} (k5product, k_5).\text{Raf-1}^*/\text{RKIP}/\text{ERK-PP}_\text{L} + (k4react, k_4).\text{Raf-1}^*/\text{RKIP}/\text{ERK-PP}_\text{L}$$

$$\text{Raf-1}^*/\text{RKIP}/\text{ERK-PP}_\text{L} \stackrel{def}{=} (k3react, k_3).\text{Raf-1}^*/\text{RKIP}/\text{ERK-PP}_\text{H}$$

PEPA components of the reagent-centric model

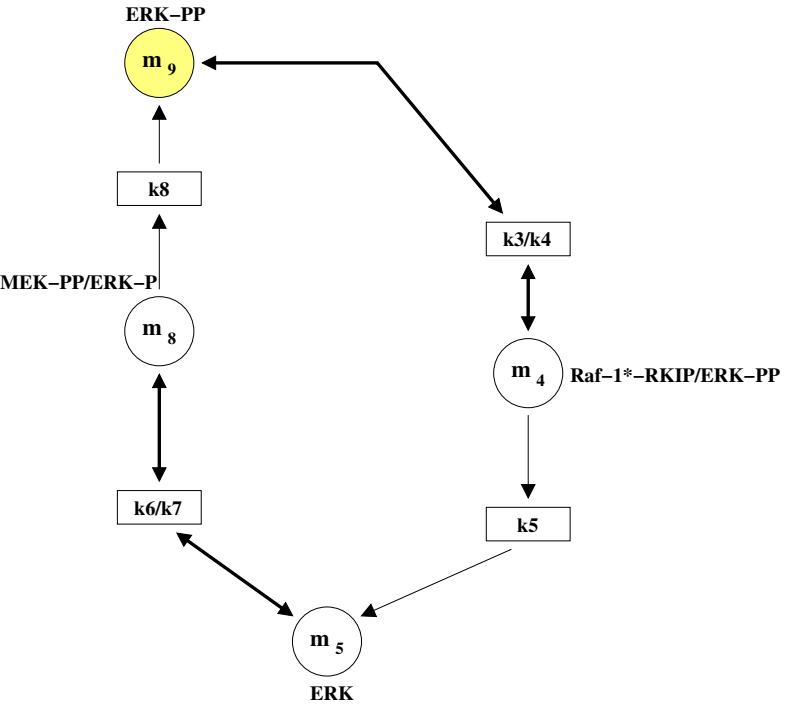


$$\text{Raf-1}^*/\text{RKIP}/\text{ERK-PP}_\text{H} \stackrel{def}{=} (k5product, k_5).\text{Raf-1}^*/\text{RKIP}/\text{ERK-PP}_\text{L} + (k4react, k_4).\text{Raf-1}^*/\text{RKIP}/\text{ERK-PP}_\text{L}$$

$$\text{Raf-1}^*/\text{RKIP}/\text{ERK-PP}_\text{L} \stackrel{def}{=} (k3react, k_3).\text{Raf-1}^*/\text{RKIP}/\text{ERK-PP}_\text{H}$$

Each reagent gives rise to a pair of PEPA definitions, one for high concentration and one for low concentration.

PEPA components of the pathway-centric model



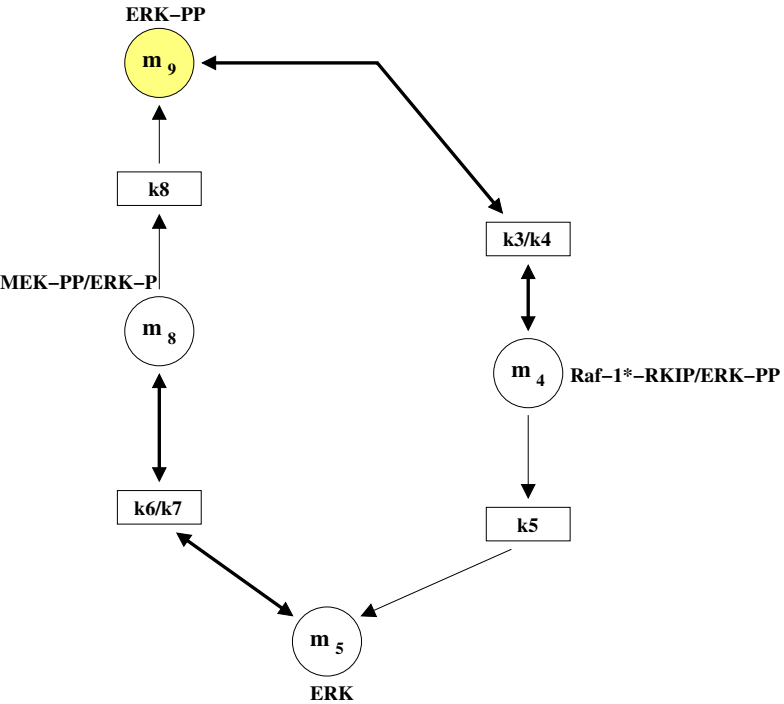
$$Pathway_{30} \stackrel{def}{=} (k3react, k_3).Pathway_{31}$$

$$Pathway_{31} \stackrel{def}{=} (k5product, k_5).Pathway_{32} + (k4react, k_4).Pathway_{30}$$

$$Pathway_{32} \stackrel{def}{=} (k6react, k_6).Pathway_{33}$$

$$Pathway_{33} \stackrel{def}{=} (k8product, k_8).Pathway_{30} + (k7react, k_7).Pathway_{32}$$

PEPA components of the pathway-centric model



$$Pathway_{30} \stackrel{def}{=} (k3react, k_3).Pathway_{31}$$

$$Pathway_{31} \stackrel{def}{=} (k5product, k_5).Pathway_{32} + (k4react, k_4).Pathway_{30}$$

$$Pathway_{32} \stackrel{def}{=} (k6react, k_6).Pathway_{33}$$

$$Pathway_{33} \stackrel{def}{=} (k8product, k_8).Pathway_{30} + (k7react, k_7).Pathway_{32}$$

For each reagent that has an initial concentration we define the sub-pathway generated by the progression of that reagent.

Commentary on the models

- Neither model currently "correctly" captures the rate of interaction – concentrations are discretized and rates are assumed to be constant within levels:

Commentary on the models

- Neither model currently "correctly" captures the rate of interaction – concentrations are discretized and rates are assumed to be constant within levels: In these examples only using high and low to modify rates in the sense of enabling or disabling activities.

Commentary on the models

- Neither model currently "correctly" captures the rate of interaction – concentrations are discretized and rates are assumed to be constant within levels: In these examples only using high and low to modify rates in the sense of enabling or disabling activities.
- The reagent-centric model can be regarded as a fine-grained view of the system.

Commentary on the models

- Neither model currently "correctly" captures the rate of interaction – concentrations are discretized and rates are assumed to be constant within levels: In these examples only using high and low to modify rates in the sense of enabling or disabling activities.
- The reagent-centric model can be regarded as a fine-grained view of the system.
- The pathway-centric model can be regarded as a more structural, coarse-grained view of the system.

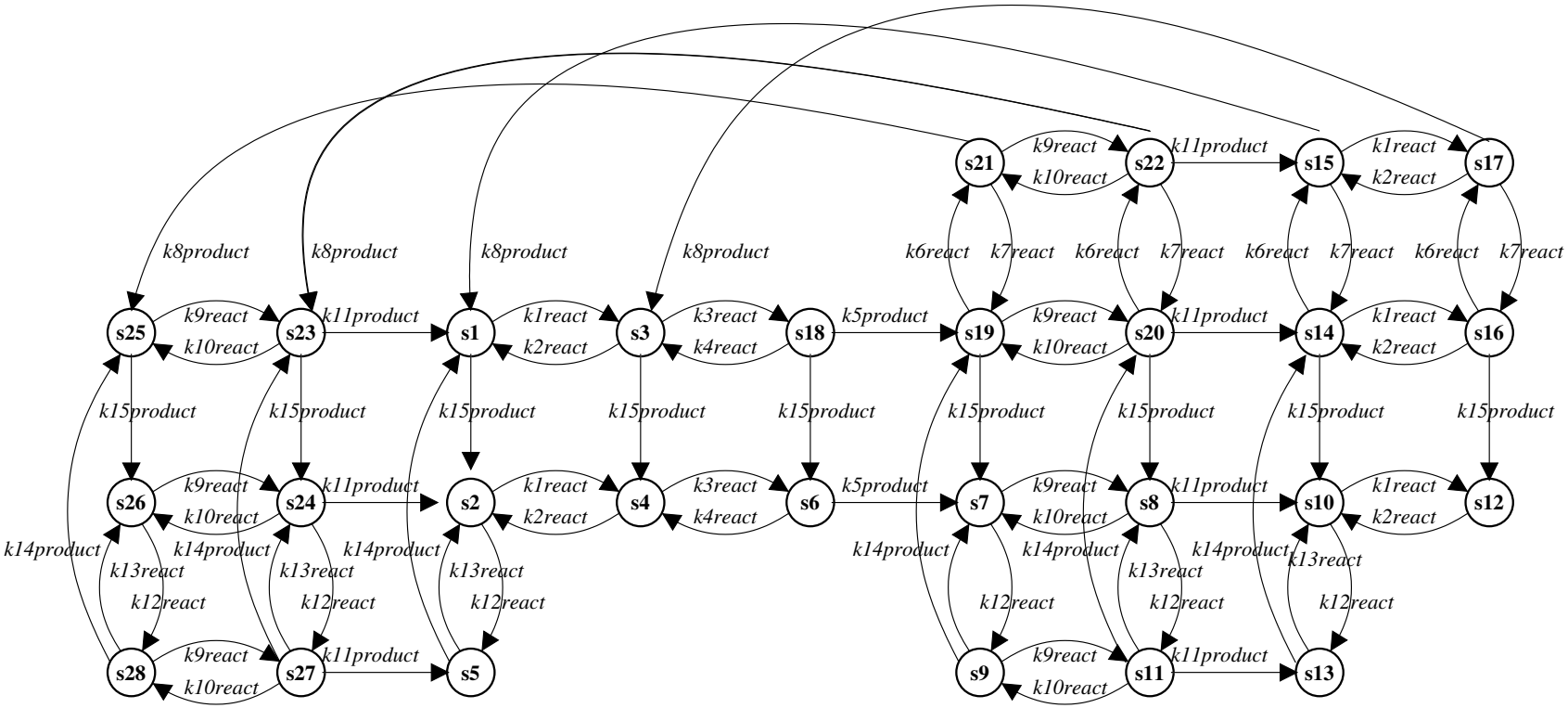
Commentary on the models

- Neither model currently "correctly" captures the rate of interaction – concentrations are discretized and rates are assumed to be constant within levels: In these examples only using high and low to modify rates in the sense of enabling or disabling activities.
- The reagent-centric model can be regarded as a fine-grained view of the system.
- The pathway-centric model can be regarded as a more structural, coarse-grained view of the system.
- Applying the structured operational semantics reveals that they are strongly bisimilar

Commentary on the models

- Neither model currently "correctly" captures the rate of interaction – concentrations are discretized and rates are assumed to be constant within levels: In these examples only using high and low to modify rates in the sense of enabling or disabling activities.
- The reagent-centric model can be regarded as a fine-grained view of the system.
- The pathway-centric model can be regarded as a more structural, coarse-grained view of the system.
- Applying the structured operational semantics reveals that they are strongly bisimilar (in fact, in this case, isomorphic).

The state space



The bisimulation

s_1	$(\text{Raf-1}^*_{\text{H}}, \text{RKIP}_{\text{H}}, \text{Raf-1}^*/\text{RKIP}_{\text{L}}, \text{Raf-1}^*/\text{RKIP/ERK-PP}_{\text{L}}, \text{ERK}_{\text{L}}, \text{RKIP-P}_{\text{L}}, \text{RKIP-P/RP}_{\text{L}}, \text{RP}_{\text{H}}, \text{MEK}_{\text{L}}, \text{MEK/Raf-1}^*_{\text{L}}, \text{MEK-PP}_{\text{H}}, \text{MEK-PP/ERK}_{\text{L}}, \text{ERK-PP}_{\text{H}})$	$(Pathway_{50}, Pathway_{40}, Pathway_{30}, Pathway_{20}, Pathway_{10})$

The bisimulation

s_1	$(\text{Raf-1}^*_{\text{H}}, \text{RKIP}_{\text{H}}, \text{Raf-1}^*/\text{RKIP}_{\text{L}}, \text{Raf-1}^*/\text{RKIP/ERK-PP}_{\text{L}}, \text{ERK}_{\text{L}}, \text{RKIP-P}_{\text{L}}, \text{RKIP-P}/\text{RP}_{\text{L}}, \text{RP}_{\text{H}}, \text{MEK}_{\text{L}}, \text{MEK}/\text{Raf-1}^*_{\text{L}}, \text{MEK-PP}_{\text{H}}, \text{MEK-PP}/\text{ERK}_{\text{L}}, \text{ERK-PP}_{\text{H}})$	$(Pathway_{50}, Pathway_{40}, Pathway_{30}, Pathway_{20}, Pathway_{10})$
s_2	$(\text{Raf-1}^*_{\text{H}}, \text{RKIP}_{\text{H}}, \text{Raf-1}^*/\text{RKIP}_{\text{L}}, \text{Raf-1}^*/\text{RKIP/ERK-PP}_{\text{L}}, \text{ERK}_{\text{L}}, \text{RKIP-P}_{\text{L}}\text{RKIP-P}/\text{RP}_{\text{L}}, \text{RP}_{\text{H}}, \text{MEK}_{\text{H}}, \text{MEK}/\text{Raf-1}^*_{\text{L}}, \text{MEK-PP}_{\text{L}}, \text{MEK-PP}/\text{ERK}_{\text{L}}, \text{ERK-PP}_{\text{H}})$	$(Pathway_{51}, Pathway_{40}, Pathway_{30}, Pathway_{20}, Pathway_{10})$

The bisimulation

s_1	$(\text{Raf-1}^*_{\text{H}}, \text{RKIP}_{\text{H}}, \text{Raf-1}^*/\text{RKIP}_{\text{L}}, \text{Raf-1}^*/\text{RKIP/ERK-PP}_{\text{L}}, \text{ERK}_{\text{L}}, \text{RKIP-P}_{\text{L}}, \text{RKIP-P}/\text{RP}_{\text{L}}, \text{RP}_{\text{H}}, \text{MEK}_{\text{L}}, \text{MEK}/\text{Raf-1}^*_{\text{L}}, \text{MEK-PP}_{\text{H}}, \text{MEK-PP}/\text{ERK}_{\text{L}}, \text{ERK-PP}_{\text{H}})$	$(Pathway_{50}, Pathway_{40}, Pathway_{30}, Pathway_{20}, Pathway_{10})$
s_2	$(\text{Raf-1}^*_{\text{H}}, \text{RKIP}_{\text{H}}, \text{Raf-1}^*/\text{RKIP}_{\text{L}}, \text{Raf-1}^*/\text{RKIP/ERK-PP}_{\text{L}}, \text{ERK}_{\text{L}}, \text{RKIP-P}_{\text{L}}\text{RKIP-P}/\text{RP}_{\text{L}}, \text{RP}_{\text{H}}, \text{MEK}_{\text{H}}, \text{MEK}/\text{Raf-1}^*_{\text{L}}, \text{MEK-PP}_{\text{L}}, \text{MEK-PP}/\text{ERK}_{\text{L}}, \text{ERK-PP}_{\text{H}})$	$(Pathway_{51}, Pathway_{40}, Pathway_{30}, Pathway_{20}, Pathway_{10})$
s_3	$(\text{Raf-1}^*_{\text{L}}, \text{RKIP}_{\text{L}}, \text{Raf-1}^*/\text{RKIP}_{\text{H}}, \text{Raf-1}^*/\text{RKIP/ERK-PP}_{\text{L}}, \text{ERK}_{\text{L}}, \text{RKIP-P}_{\text{L}}, \text{RKIP-P}/\text{RP}_{\text{L}}, \text{RP}_{\text{H}}, \text{MEK}_{\text{L}}, \text{MEK}/\text{Raf-1}^*_{\text{L}}, \text{MEK-PP}_{\text{H}}, \text{MEK-PP}/\text{ERK}_{\text{L}}, \text{ERK-PP}_{\text{H}})$	$(Pathway_{50}, Pathway_{41}, Pathway_{30}, Pathway_{21}, Pathway_{10})$

The bisimulation

s_1	$(\text{Raf-1}^*_{\text{H}}, \text{RKIP}_{\text{H}}, \text{Raf-1}^*/\text{RKIP}_{\text{L}}, \text{Raf-1}^*/\text{RKIP/ERK-PP}_{\text{L}}, \text{ERK}_{\text{L}}, \text{RKIP-P}_{\text{L}}, \text{RKIP-P}/\text{RP}_{\text{L}}, \text{RP}_{\text{H}}, \text{MEK}_{\text{L}}, \text{MEK}/\text{Raf-1}^*_{\text{L}}, \text{MEK-PP}_{\text{H}}, \text{MEK-PP}/\text{ERK}_{\text{L}}, \text{ERK-PP}_{\text{H}})$	$(Pathway_{50}, Pathway_{40}, Pathway_{30}, Pathway_{20}, Pathway_{10})$
s_2	$(\text{Raf-1}^*_{\text{H}}, \text{RKIP}_{\text{H}}, \text{Raf-1}^*/\text{RKIP}_{\text{L}}, \text{Raf-1}^*/\text{RKIP/ERK-PP}_{\text{L}}, \text{ERK}_{\text{L}}, \text{RKIP-P}_{\text{L}}\text{RKIP-P}/\text{RP}_{\text{L}}, \text{RP}_{\text{H}}, \text{MEK}_{\text{H}}, \text{MEK}/\text{Raf-1}^*_{\text{L}}, \text{MEK-PP}_{\text{L}}, \text{MEK-PP}/\text{ERK}_{\text{L}}, \text{ERK-PP}_{\text{H}})$	$(Pathway_{51}, Pathway_{40}, Pathway_{30}, Pathway_{20}, Pathway_{10})$
s_3	$(\text{Raf-1}^*_{\text{L}}, \text{RKIP}_{\text{L}}, \text{Raf-1}^*/\text{RKIP}_{\text{H}}, \text{Raf-1}^*/\text{RKIP/ERK-PP}_{\text{L}}, \text{ERK}_{\text{L}}, \text{RKIP-P}_{\text{L}}, \text{RKIP-P}/\text{RP}_{\text{L}}, \text{RP}_{\text{H}}, \text{MEK}_{\text{L}}, \text{MEK}/\text{Raf-1}^*_{\text{L}}, \text{MEK-PP}_{\text{H}}, \text{MEK-PP}/\text{ERK}_{\text{L}}, \text{ERK-PP}_{\text{H}})$	$(Pathway_{50}, Pathway_{41}, Pathway_{30}, Pathway_{21}, Pathway_{10})$
$\vdots$	$\vdots$	$\vdots$

The bisimulation

s_1	$(\text{Raf-1}^*_{\text{H}}, \text{RKIP}_{\text{H}}, \text{Raf-1}^*/\text{RKIP}_{\text{L}}, \text{Raf-1}^*/\text{RKIP/ERK-PP}_{\text{L}}, \text{ERK}_{\text{L}}, \text{RKIP-P}_{\text{L}}, \text{RKIP-P/RP}_{\text{L}}, \text{RP}_{\text{H}}, \text{MEK}_{\text{L}}, \text{MEK/Raf-1}^*_{\text{L}}, \text{MEK-PP}_{\text{H}}, \text{MEK-PP/ERK}_{\text{L}}, \text{ERK-PP}_{\text{H}})$	$(Pathway_{50}, Pathway_{40}, Pathway_{30}, Pathway_{20}, Pathway_{10})$
s_2	$(\text{Raf-1}^*_{\text{H}}, \text{RKIP}_{\text{H}}, \text{Raf-1}^*/\text{RKIP}_{\text{L}}, \text{Raf-1}^*/\text{RKIP/ERK-PP}_{\text{L}}, \text{ERK}_{\text{L}}, \text{RKIP-P}_{\text{L}}\text{RKIP-P/RP}_{\text{L}}, \text{RP}_{\text{H}}, \text{MEK}_{\text{H}}, \text{MEK/Raf-1}^*_{\text{L}}, \text{MEK-PP}_{\text{L}}, \text{MEK-PP/ERK}_{\text{L}}, \text{ERK-PP}_{\text{H}})$	$(Pathway_{51}, Pathway_{40}, Pathway_{30}, Pathway_{20}, Pathway_{10})$
s_3	$(\text{Raf-1}^*_{\text{L}}, \text{RKIP}_{\text{L}}, \text{Raf-1}^*/\text{RKIP}_{\text{H}}, \text{Raf-1}^*/\text{RKIP/ERK-PP}_{\text{L}}, \text{ERK}_{\text{L}}, \text{RKIP-P}_{\text{L}}, \text{RKIP-P/RP}_{\text{L}}, \text{RP}_{\text{H}}, \text{MEK}_{\text{L}}, \text{MEK/Raf-1}^*_{\text{L}}, \text{MEK-PP}_{\text{H}}, \text{MEK-PP/ERK}_{\text{L}}, \text{ERK-PP}_{\text{H}})$	$(Pathway_{50}, Pathway_{41}, Pathway_{30}, Pathway_{21}, Pathway_{10})$
$\vdots$	$\vdots$	$\vdots$
s_{28}	$(\text{Raf-1}^*_{\text{L}}, \text{RKIP}_{\text{L}}, \text{Raf-1}^*/\text{RKIP}_{\text{L}}, \text{Raf-1}^*/\text{RKIP/ERK-PP}_{\text{L}}, \text{ERK}_{\text{L}}, \text{RKIP-P}_{\text{H}}, \text{RKIP-P/RP}_{\text{L}}, \text{RP}_{\text{H}}, \text{MEK}_{\text{L}}, \text{MEK/Raf-1}^*_{\text{H}}, \text{MEK-PP}_{\text{L}}, \text{MEK-PP/ERK}_{\text{L}}, \text{ERK-PP}_{\text{H}})$	$(Pathway_{50}, Pathway_{40}, Pathway_{30}, Pathway_{20}, Pathway_{10})$

Deriving quantitative data

PEPA models can be analysed for quantified dynamic behaviour in a number of different ways.

Deriving quantitative data

PEPA models can be analysed for quantified dynamic behaviour in a number of different ways.

The language may be used to generate a Markov Process (CTMC).

Deriving quantitative data

PEPA models can be analysed for quantified dynamic behaviour in a number of different ways.

The language may be used to generate a Markov Process (CTMC).

SPA
MODEL

Deriving quantitative data

PEPA models can be analysed for quantified dynamic behaviour in a number of different ways.

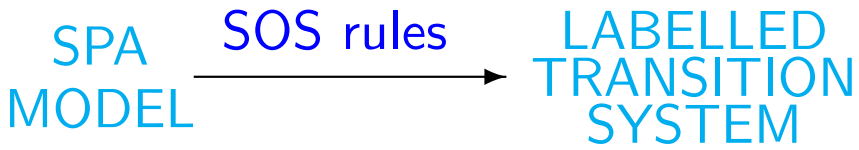
The language may be used to generate a Markov Process (CTMC).

SPA MODEL $\xrightarrow{\text{SOS rules}}$

Deriving quantitative data

PEPA models can be analysed for quantified dynamic behaviour in a number of different ways.

The language may be used to generate a Markov Process (CTMC).



Deriving quantitative data

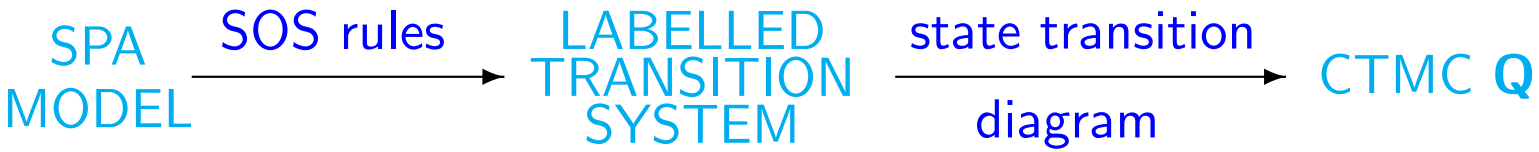
PEPA models can be analysed for quantified dynamic behaviour in a number of different ways.

The language may be used to generate a Markov Process (CTMC).

Deriving quantitative data

PEPA models can be analysed for quantified dynamic behaviour in a number of different ways.

The language may be used to generate a Markov Process (CTMC).



Deriving quantitative data

PEPA models can be analysed for quantified dynamic behaviour in a number of different ways.

The language may be used to generate a system of ordinary differential equations (ODEs).

Deriving quantitative data

PEPA models can be analysed for quantified dynamic behaviour in a number of different ways.

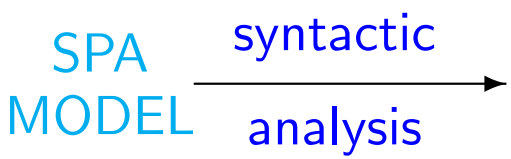
The language may be used to generate a system of ordinary differential equations (ODEs).

SPA
MODEL

Deriving quantitative data

PEPA models can be analysed for quantified dynamic behaviour in a number of different ways.

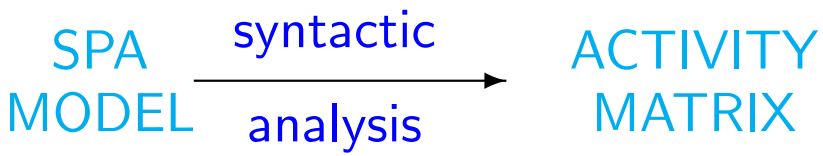
The language may be used to generate a system of ordinary differential equations (ODEs).



Deriving quantitative data

PEPA models can be analysed for quantified dynamic behaviour in a number of different ways.

The language may be used to generate a system of ordinary differential equations (ODEs).



Deriving quantitative data

PEPA models can be analysed for quantified dynamic behaviour in a number of different ways.

The language may be used to generate a system of ordinary differential equations (ODEs).

Deriving quantitative data

PEPA models can be analysed for quantified dynamic behaviour in a number of different ways.

The language may be used to generate a system of ordinary differential equations (ODEs).

Deriving quantitative data

PEPA models can be analysed for quantified dynamic behaviour in a number of different ways.

The language also may be used to generate a stochastic simulation.

Deriving quantitative data

PEPA models can be analysed for quantified dynamic behaviour in a number of different ways.

The language also may be used to generate a stochastic simulation.

SPA
MODEL

Deriving quantitative data

PEPA models can be analysed for quantified dynamic behaviour in a number of different ways.

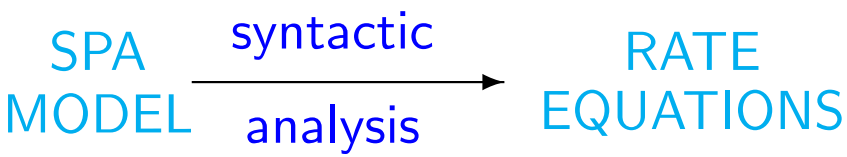
The language also may be used to generate a stochastic simulation.

SPA MODEL $\xrightarrow[\text{analysis}]{\text{syntactic}}$

Deriving quantitative data

PEPA models can be analysed for quantified dynamic behaviour in a number of different ways.

The language also may be used to generate a stochastic simulation.



Deriving quantitative data

PEPA models can be analysed for quantified dynamic behaviour in a number of different ways.

The language also may be used to generate a stochastic simulation.

Deriving quantitative data

PEPA models can be analysed for quantified dynamic behaviour in a number of different ways.

The language also may be used to generate a stochastic simulation.

Deriving quantitative data

PEPA models can be analysed for quantified dynamic behaviour in a number of different ways.

The language also may be used to generate a stochastic simulation.

Each of these has tool support so that the underlying model is derived automatically according to the predefined rules.

Markovian analysis

- Analysis of the Markov process can yield quite detailed information about the dynamic behaviour of the model.

Markovian analysis

- Analysis of the Markov process can yield quite detailed information about the dynamic behaviour of the model.
- A steady state analysis provides statistics for average behaviour over a long run of the system, when the bias introduced by the initial state has been lost.

Markovian analysis

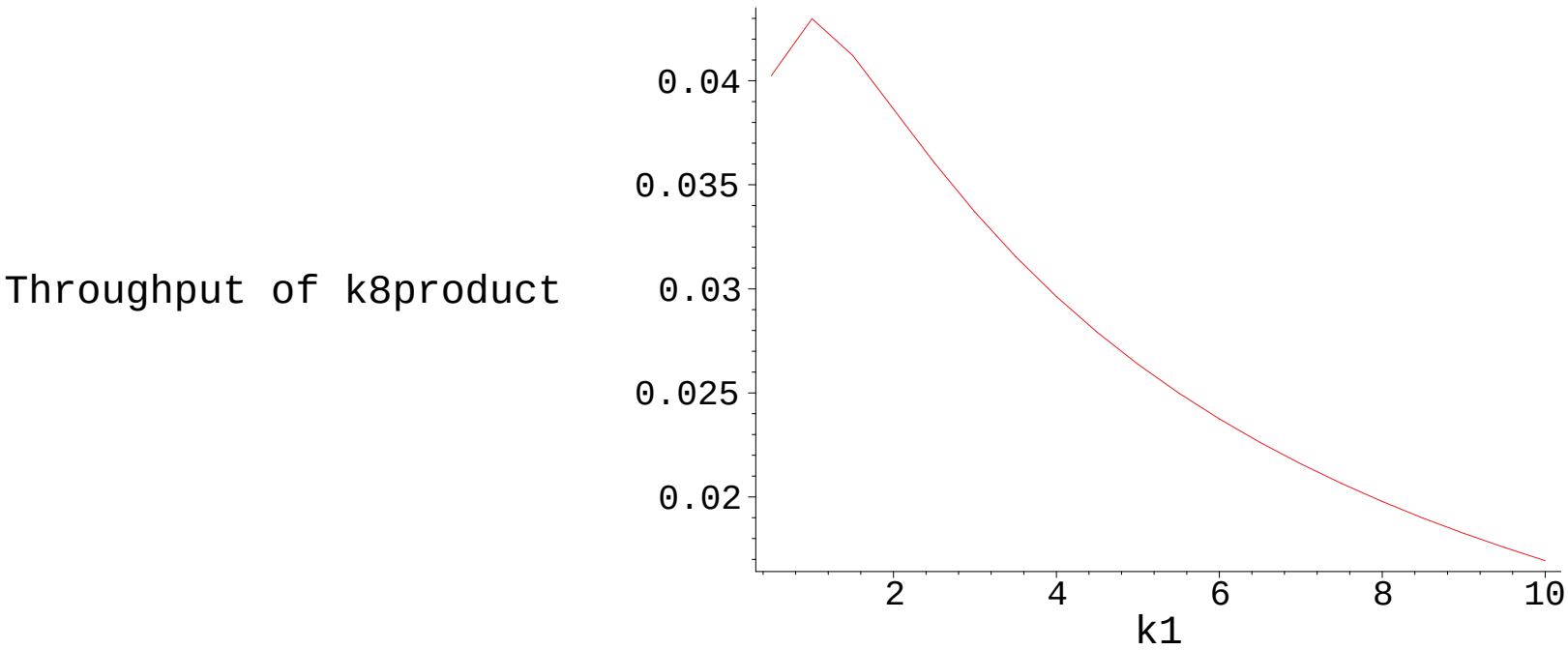
- Analysis of the Markov process can yield quite detailed information about the dynamic behaviour of the model.
- A steady state analysis provides statistics for average behaviour over a long run of the system, when the bias introduced by the initial state has been lost.
- A transient analysis provides statistics relating to the evolution of the model over a fixed period. This will be dependent on the starting state.

Markovian analysis

- Analysis of the Markov process can yield quite detailed information about the dynamic behaviour of the model.
- A steady state analysis provides statistics for average behaviour over a long run of the system, when the bias introduced by the initial state has been lost.
- A transient analysis provides statistics relating to the evolution of the model over a fixed period. This will be dependent on the starting state.
- Note, however, that a transient Markovian analysis is exact because it takes account of all possible evolutions, unlike a stochastic simulation which considers only one possible evolution in each run.

Quantified analysis – *k8product*

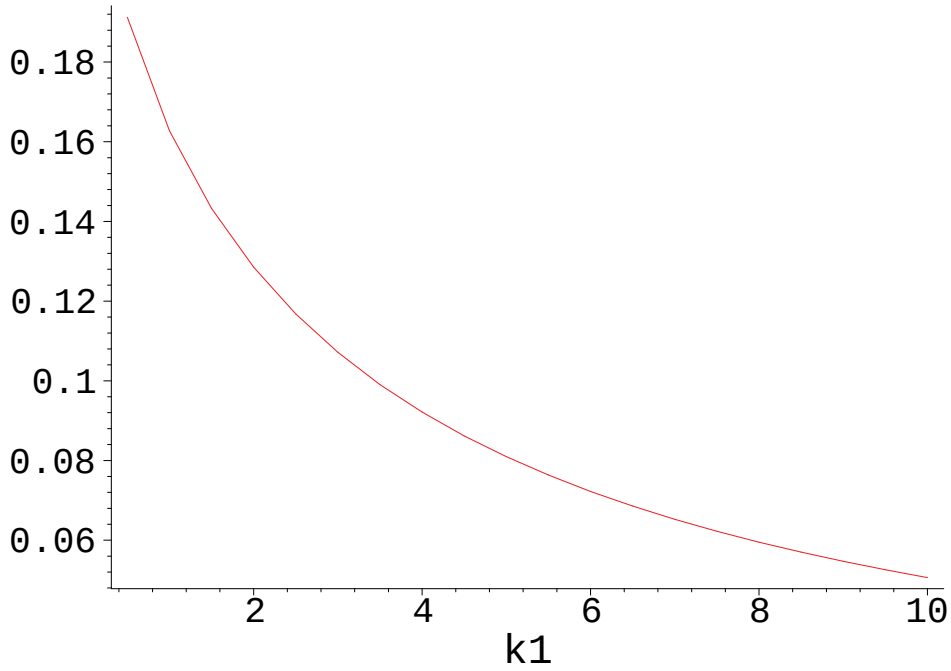
Approximating a variation in the initial concentration of RKIP by varying the rate constant k1, we can assess the impact on the production of ERK-PP.



Quantified analysis – *k14product*

Similarly we can assess the impact on the production of MEK-PP.

Throughput of k14product



Deriving differential equation: overview

- The ODEs are the familiar mathematical model for the biochemists.

Deriving differential equation: overview

- The ODEs are the familiar mathematical model for the biochemists.
- There should be one equation for each reagent/concentration, indicating how the concentration varies over time.

Deriving differential equation: overview

- The ODEs are the familiar mathematical model for the biochemists.
- There should be one equation for each reagent/concentration, indicating how the concentration varies over time.
- Standard solution tools are available for solution of this equations, which are known and trusted by the biologists.

Deriving differential equation: overview

- The ODEs are the familiar mathematical model for the biochemists.
- There should be one equation for each reagent/concentration, indicating how the concentration varies over time.
- Standard solution tools are available for solution of this equations, which are known and trusted by the biologists.
- From the reagent-centric PEPA model we can see the influence of the reactions on each concentration – this is recorded in a matrix termed the activity matrix.

Activity Matrix

For a pathway with R reactions and S reagents, the activity matrix M_a is an $S \times R$ matrix, and the entries are defined as follows.

$$(s_i, r_j) = \begin{cases} +1 & \text{if } \xrightarrow{r_j} s_i \in \mathcal{L} \\ -1 & \text{if } s_i \xrightarrow{r_j} \in \mathcal{L} \\ 0 & \text{if } s_i \xrightarrow{r_j} \notin \mathcal{L} \\ & \cup \xrightarrow{r_j} s_i \notin \mathcal{L} \end{cases}$$

Activity matrix for the MAPK pathway

	k1	k2	k3	k4	k5	k6	k7	k8	k9	k10	k11	k12	k13	k14	k15
Raf-1*	-1	+1	0	0	+1	0	0	0	0	0	0	-1	+1	+1	0
RKIP	-1	+1	0	0	0	0	0	0	0	0	+1	0	0	0	0
Raf-1*/RKIP	+1	-1	-1	+1	0	0	0	0	0	0	0	0	0	0	0
Raf-1*/RKIP/ERK-PP	0	0	+1	-1	-1	0	0	0	0	0	0	0	0	0	0
ERK	0	0	0	0	+1	-1	+1	0	0	0	0	0	0	0	0
RKIP-P	0	0	0	0	+1	0	0	0	-1	+1	0	0	0	0	0
MEK-PP	0	0	0	0	0	-1	+1	+1	0	0	0	0	0	+1	-1
MEK-PP/ERK	0	0	0	0	0	+1	-1	-1	0	0	0	0	0	0	0
ERK-PP	0	0	-1	+1	0	0	0	+1	0	0	0	0	0	0	0
RP	0	0	0	0	0	0	0	0	-1	+1	+1	0	0	0	0
RKIP-P/RP	0	0	0	0	0	0	0	0	+1	-1	-1	0	0	0	0
MEK	0	0	0	0	0	0	0	0	0	0	0	-1	+1	0	+1
MEK/Raf-1*	0	0	0	0	0	0	0	0	0	0	0	+1	-1	-1	0

Deriving differential equations: activity matrix

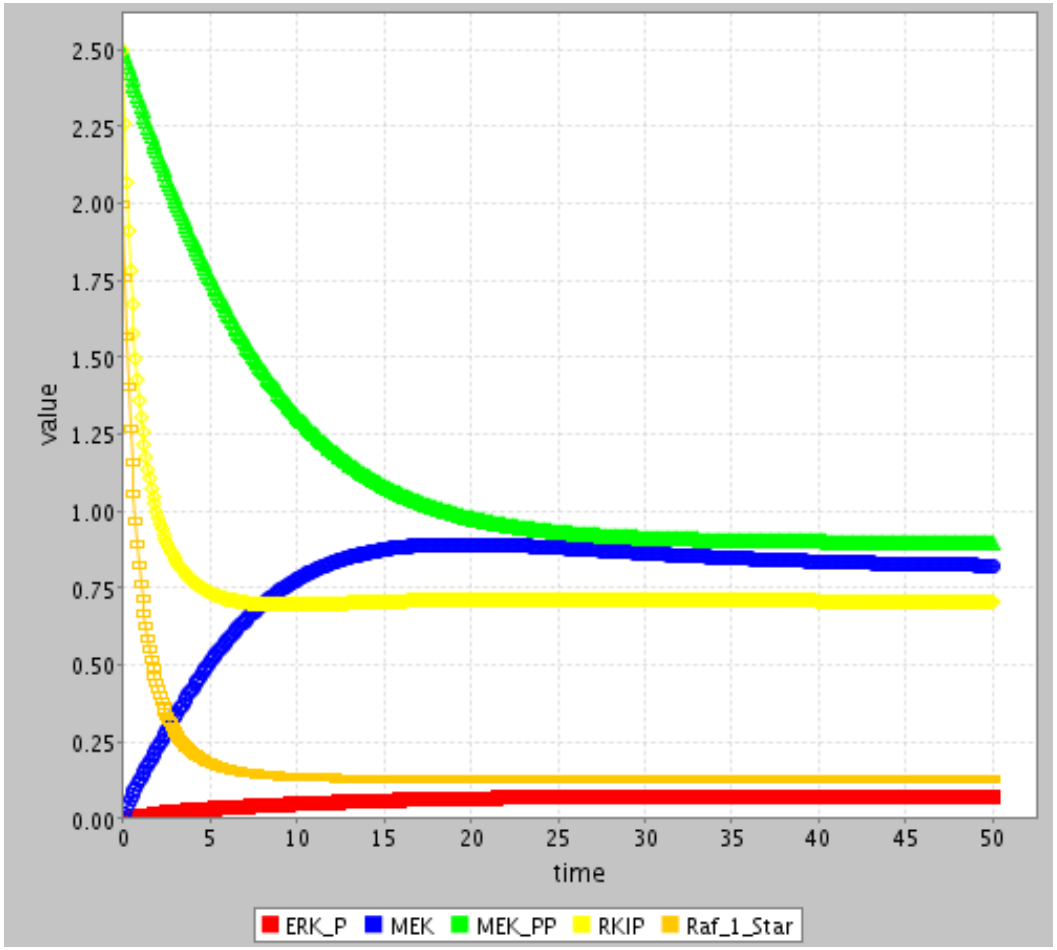
	$k1$	$k2$	$k3$	$k4$	$k5$	$k6$	...	$conc.$
Raf-1*	-1	$+1$	0	0	$+1$	0	...	m_1
RKIP	-1	$+1$	0	0	0	0	...	m_2
Raf-1*/RKIP	$+\mathbf{1}$	$-\mathbf{1}$	$-\mathbf{1}$	$+\mathbf{1}$	0	0	...	$(\mathbf{m_3})$
Raf-1*/RKIP/ERK-PP	0	0	$+1$	-1	-1	0	...	m_4
ERK	0	0	0	0	$+1$	-1	...	m_5
RKIP-P	0	0	0	0	$+1$	0	...	m_6
MEK-PP	0	0	0	0	0	-1	...	m_7
MEK-PP/ERK	0	0	0	0	0	$+1$	...	m_8
ERK-PP	0	0	-1	$+1$	0	0	...	m_9
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

$$\frac{d\,m_3(t)}{dt} = k_1\, m_1(t) m_2(t) - k_2\, m_3(t) - k_3\, m_3(t) m_9(t) + k_4\, m_4(t)$$

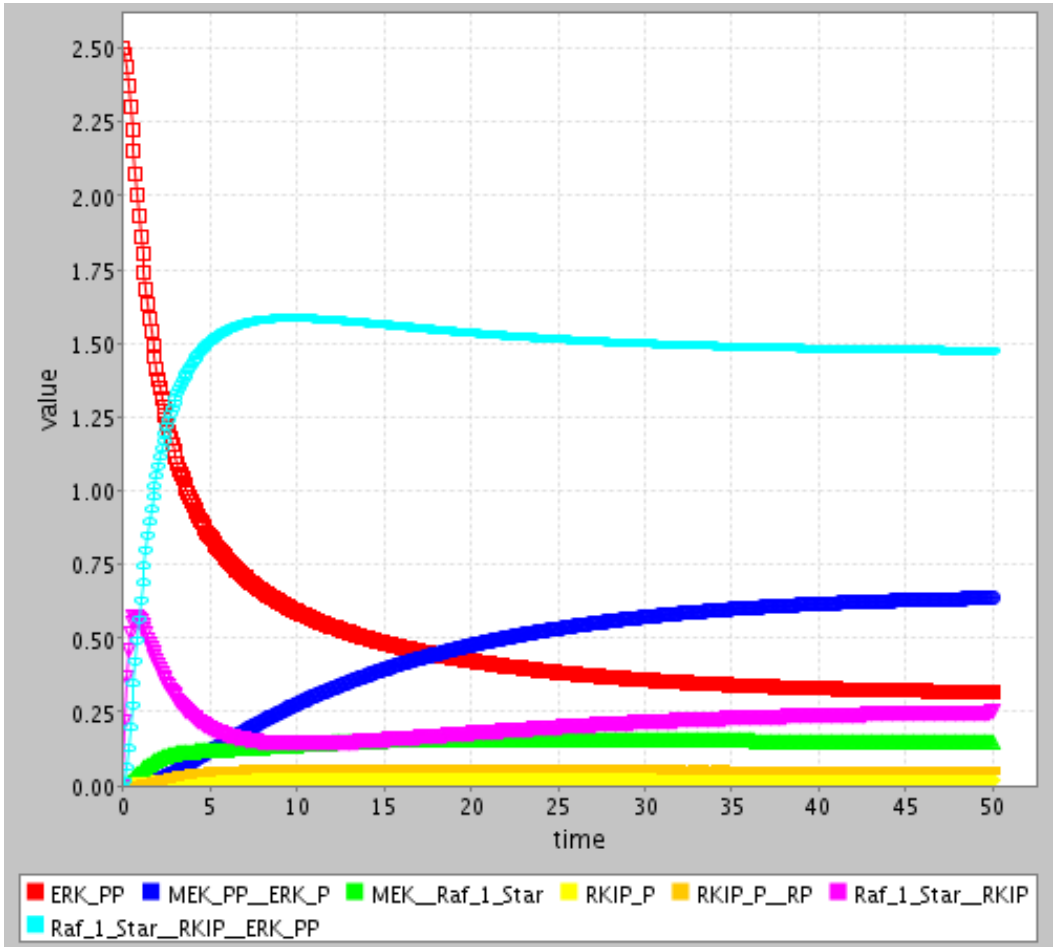
Differential equations

$$
\begin{aligned}
\frac{dm_1(t)}{dt} &= -k_1 m_1(t) m_2(t) + k_2 m_3(t) + k_5 m_4(t) - k_{12} m_1(t) m_{12}(t) \\
&\quad + k_{13} m_{13}(t) + k_{14} m_{13}(t) \\
\frac{dm_2(t)}{dt} &= -k_1 m_1(t) m_2(t) + k_2 m_3(t) + k_{11} m_{11}(t) \\
\frac{dm_3(t)}{dt} &= k_1 m_1(t) m_2(t) - k_2 m_3(t) - k_3 m_3(t) m_9(t) + k_4 m_4(t) \\
&\vdots \qquad\qquad \vdots \\
\frac{dm_{13}(t)}{dt} &= k_{12} m_1(t) m_1 2(t) - k_{13} m_{13}(t) - k_{14} m_{13}(t)
\end{aligned}
$$

ODE Analysis



ODE Analysis



ODEs via PEPA

There are several advantages to be gained by introducing a process algebra model as an intermediary to the derivation of the ODEs.

ODEs via PEPA

There are several advantages to be gained by introducing a process algebra model as an intermediary to the derivation of the ODEs.

- The ODEs can be automatically generated from the descriptive process algebra model, thus reducing human error.

ODEs via PEPA

There are several advantages to be gained by introducing a process algebra model as an intermediary to the derivation of the ODEs.

- The ODEs can be automatically generated from the descriptive process algebra model, thus reducing human error.
- The formality of the process algebra model and its underlying semantics allow us to derive properties of the model, such as freedom from deadlock, before numerical analysis is carried out.

ODEs via PEPA

There are several advantages to be gained by introducing a process algebra model as an intermediary to the derivation of the ODEs.

- The ODEs can be automatically generated from the descriptive process algebra model, thus reducing human error.
- The formality of the process algebra model and its underlying semantics allow us to derive properties of the model, such as freedom from deadlock, before numerical analysis is carried out.
- The algebraic formulation of the model makes clear the interactions between the biochemical entities, or substrates. The style of modelling is descriptive, close to informal graphical representations that biochemists already use.

References

M. Calder, S. Gilmore and J. Hillston.

Modelling the influence of RKIP on the ERK signalling pathway using the stochastic process algebra PEPA.

to appear in *Transactions on Computational Systems Biology*, 2006.

M. Calder, S. Gilmore and J. Hillston.

Deriving ODEs from PEPA models.

in *Proceedings of 3rd International Workshop on Computational Methods in Systems Biology 2005* Edinburgh, UK, April 2005.

M. Calder, A. Duguid, S. Gilmore and J. Hillston.

Stronger computational modelling of signalling pathways using both continuous and discrete-state methods.

to appear in *Proceedings of 4th International Workshop on Computational Methods in Systems Biology*. Trento, Italy. 18-19th October 2006.

Biological Reasoning

J. Heath, M. Kwiatkowska, G. Norman, D. Parker and O. Tymchyshyn.

Probabilistic model checking of complex biological pathways.

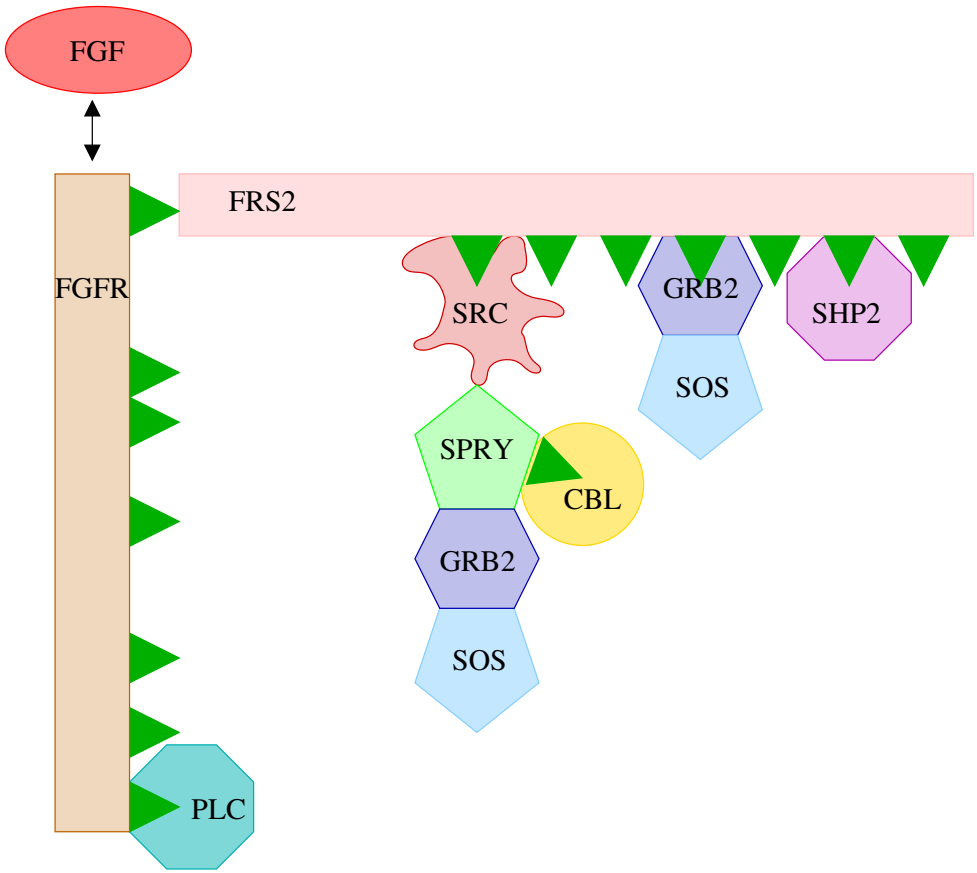
To appear in the Proceedings of 4th International Workshop on *Computational Methods in Systems Biology 2006*.

Trento, Italy, 18-19th October 2006.

The FGF Pathway

- Fibroblast Growth Factors (FGF) are a family of proteins which play a key role in the process of cell signalling in a variety of contexts, e.g. wound healing.
- The mechanisms for FGF signalling are complex and not yet fully understood.
- The model incorporates protein-protein interactions, phosphorylation, dephosphorylation, protein complex relocation and protein complex degradation.

The FGF Pathway



The binding of the signalling protein FGF to its receptor FGFR triggers a series of biochemical reactions.

Some notes about the model

The model is not intended to be a fully accurate representation.

Nevertheless it contains sufficient information to allow biological hypotheses to be evaluated — it facilitates in silico experimentation.

The abstraction has been guided by biological interests: the reactions selected are those which are currently being actively studied by the biologists.

PRISM modelling of the FGF Pathway

- The PRISM model is written in the PRISM input language of reactive modules. A companion model written in π-calculus was developed at the same time and studied via simulation.

PRISM modelling of the FGF Pathway

- The PRISM model is written in the PRISM input language of reactive modules. A companion model written in π-calculus was developed at the same time and studied via simulation.
- In the model some elements are modelled in detail with a separate state of the model for each possible state in the biochemical process (c.f. previous work with the π-calculus).

PRISM modelling of the FGF Pathway

- The PRISM model is written in the PRISM input language of reactive modules. A companion model written in π-calculus was developed at the same time and studied via simulation.
- In the model some elements are modelled in detail with a separate state of the model for each possible state in the biochemical process (c.f. previous work with the π-calculus).
- Other elements are represented more abstractly with different biochemical states represented by distinct components which have two possible states indicating only presence or absence (c.f. PEPA high-low models).

PRISM modelling of the FGF Pathway

- The PRISM model is written in the PRISM input language of reactive modules. A companion model written in π-calculus was developed at the same time and studied via simulation.
- In the model some elements are modelled in detail with a separate state of the model for each possible state in the biochemical process (c.f. previous work with the π-calculus).
- Other elements are represented more abstractly with different biochemical states represented by distinct components which have two possible states indicating only presence or absence (c.f. PEPA high-low models).
- The focus of this paper is on the role that model checking can have in testing biological hypotheses.

Model checking the FGF Pathway

- PRISM is used to check properties of the CTMC underlying the FGF model.
- Properties are expressed in the stochastic temporal logic CSL.
- For example:

Model checking the FGF Pathway

- PRISM is used to check properties of the CTMC underlying the FGF model.
- Properties are expressed in the stochastic temporal logic CSL.
- For example:
  - *What is the probability that the protein A is bound to the protein B at time instant T?* $\mathcal{P}_{=?}[\mathbf{true}\mathcal{U}^{[T,T]}ab = 1]$

Model checking the FGF Pathway

- PRISM is used to check properties of the CTMC underlying the FGF model.
- Properties are expressed in the stochastic temporal logic CSL.
- For example:
  - *What is the probability that the protein A is bound to the protein B at time instant T?* $\mathcal{P}_{=?}[\mathbf{true}\mathcal{U}^{[T,T]}ab = 1]$
  - *What is the expected number of times that the proteins A and B bind before A degrades?* $\mathcal{R}_{+?}[\mathcal{F}(a = 0 \wedge ab = 0)]$ assuming a reward of 1 is associated with any transition labelled by *bind*.

Model checking the FGF Pathway

- In the model a single instance of the pathway is represented.
- This results in a model with 80,616 states and over 560,000 transitions.
- The same signal dynamics were observed in a larger model with 100 instances of each molecule which was studied using simulation.
- Atomic propositions were defined over the model to capture events of interest, for example when bindings have been formed or degradation has been initiated.

Analysis

In silico experiments were undertaken with the model for a number of scenarios:

Analysis

In silico experiments were undertaken with the model for a number of scenarios:

1. The pathway as illustrated consisting of FGF, FGFR (unbound and unphosphorylated), FRS2 (unphosphorylated), SRC, GRB2, CBL, PLC and SOS (all unbound) and SPRY arriving after a delay.

Analysis

In silico experiments were undertaken with the model for a number of scenarios:

1 The pathway as illustrated consisting of FGF, FGFR (unbound and unphosphorylated), FRS2 (unphosphorylated), SRC, GRB2, CBL, PLC and SOS (all unbound) and SPRY arriving after a delay.

2-5 The pathway with one of SHP2, SRC, SPRY or PLC removed.

Time series plots are generated by finding the probability of a situation (e.g. GRB2 bound to FRS2) for varying values of T.

Analysis

In silico experiments were undertaken with the model for a number of scenarios:

1 The pathway as illustrated consisting of FGF, FGFR (unbound and unphosphorylated), FRS2 (unphosphorylated), SRC, GRB2, CBL, PLC and SOS (all unbound) and SPRY arriving after a delay.

2-5 The pathway with one of SHP2, SRC, SPRY or PLC removed.

Time series plots are generated by finding the probability of a situation (e.g. GRB2 bound to FRS2) for varying values of T.

In this example GRB2 spends a smaller proportion of time bound to FRS2 as time passes.

Results: GRB2-FRS2 binding

- The binding of GRB2 to FRS2 is regarded as the signal of interest.

Results: GRB2-FRS2 binding

- The binding of GRB2 to FRS2 is regarded as the signal of interest.
- In the full model this is seen to very rapidly increase but then decay.

Results: GRB2-FRS2 binding

- The binding of GRB2 to FRS2 is regarded as the signal of interest.
- In the full model this is seen to very rapidly increase but then decay.
- When SRC is missing there is a constant signal.

Results: GRB2-FRS2 binding

- The binding of GRB2 to FRS2 is regarded as the signal of interest.
- In the full model this is seen to very rapidly increase but then decay.
- When SRC is missing there is a constant signal.
- When there is no SHP2 the signal reaches a higher peak at the same speed but then decays more rapidly.

Results: GRB2-FRS2 binding

- The binding of GRB2 to FRS2 is regarded as the signal of interest.
- In the full model this is seen to very rapidly increase but then decay.
- When SRC is missing there is a constant signal.
- When there is no SHP2 the signal reaches a higher peak at the same speed but then decays more rapidly.
- When there is no SPRY the peak is reached as in the full model but the decay is more gradual.

Conclusions

The paper demonstrates that probabilistic model checking can be useful for exploring the behaviour of biochemical pathways and conducting *in silico* experiments.

Conclusions

The paper demonstrates that probabilistic model checking can be useful for exploring the behaviour of biochemical pathways and conducting *in silico* experiments.

The size of the system must be kept modest in order for the tool to be able to work since an explicit representation of the underlying continuous time Markov chain is needed.

Conclusions

The paper demonstrates that probabilistic model checking can be useful for exploring the behaviour of biochemical pathways and conducting *in silico* experiments.

The size of the system must be kept modest in order for the tool to be able to work since an explicit representation of the underlying continuous time Markov chain is needed.

The relationship between the single instance version of the model, as used here, and the more realistic model with multiple copies of each species, is yet to be formally established.

Further References

J. Heath, M. Kwiatkowska, G. Norman, D. Parker and O. Tymchynshyn

Computer assisted biological reasoning: The simulation and analysis of FGF signalling pathway dynamics.

Submitted for publication 2006.

M. Calder, V. Vyshemirsky, D. Gilbert and R. Orton

Analysis of signalling pathways using continuous time Markov chains.

to appear in *Transactions on Computational Systems Biology* 2006.

Outline

Stochastic π-calculus

Stochastic π-calculus [*Priami, 1995*] extends the π-calculus with exponentially-distributed rates.

Stochastic π-calculus

Stochastic π-calculus [*Priami, 1995*] extends the π-calculus with exponentially-distributed rates.

$$0 \qquad \text{Nil}$$

Stochastic π-calculus

Stochastic π-calculus [*Priami, 1995*] extends the π-calculus with exponentially-distributed rates.

0	Nil
$(\pi, r).P$	Prefix

Stochastic π-calculus

Stochastic π-calculus [*Priami, 1995*] extends the π-calculus with exponentially-distributed rates.

0	Nil
$(\pi, r).P$	Prefix
$(\nu\, n)\, P$	New

Stochastic π-calculus

Stochastic π-calculus [*Priami, 1995*] extends the π-calculus with exponentially-distributed rates.

0	Nil
$(\pi, r).P$	Prefix
$(\nu\, n)\, P$	New
$[x = y]P$	Matching

Stochastic π-calculus

Stochastic π-calculus [*Priami, 1995*] extends the π-calculus with exponentially-distributed rates.

0	Nil
$(\pi, r).P$	Prefix
$(\nu\, n)\, P$	New
$[x = y]P$	Matching
$P_1 \mid P_2$	Parallel

Stochastic π-calculus

Stochastic π-calculus [*Priami, 1995*] extends the π-calculus with exponentially-distributed rates.

0	Nil
$(\pi, r).P$	Prefix
$(\nu\, n)\, P$	New
$[x = y]P$	Matching
$P_1 \mid P_2$	Parallel
$P_1 + P_2$	Choice

Stochastic π-calculus

Stochastic π-calculus [*Priami, 1995*] extends the π-calculus with exponentially-distributed rates.

0	Nil
$(\pi, r).P$	Prefix
$(\nu\, n)\, P$	New
$[x = y]P$	Matching
$P_1 \mid P_2$	Parallel
$P_1 + P_2$	Choice
$P(y_1, \ldots, y_n)$	Definition

Stochastic π-calculus

Stochastic π-calculus [*Priami, 1995*] extends the π-calculus with exponentially-distributed rates.

0	Nil
$(\pi, r).P$	Prefix
$(\nu\, n)\, P$	New
$[x = y]P$	Matching
$P_1 \mid P_2$	Parallel
$P_1 + P_2$	Choice
$P(y_1, \ldots, y_n)$	Definition

where π is either $x(y)$ (input), $\overline{x}y$ (output) or τ (silent).

Stochastic π-calculus

- The stochastic π-calculus has been used to model and analyse a wide variety of biological systems.

Stochastic π-calculus

- The stochastic π-calculus has been used to model and analyse a wide variety of biological systems.
- Examples include metabolic pathways, gene transcription and signal transduction.

Stochastic π-calculus

- The stochastic π-calculus has been used to model and analyse a wide variety of biological systems.
- Examples include metabolic pathways, gene transcription and signal transduction.
- Currently all analysis is based on stochastic simulation (Gillespie's algorithm).

Stochastic π-calculus

- The stochastic π-calculus has been used to model and analyse a wide variety of biological systems.
- Examples include metabolic pathways, gene transcription and signal transduction.
- Currently all analysis is based on stochastic simulation (Gillespie's algorithm).
- Two tools: BioSPI and SPIM which implement slightly different versions of the language.

Stochastic π-calculus

- The stochastic π-calculus has been used to model and analyse a wide variety of biological systems.
- Examples include metabolic pathways, gene transcription and signal transduction.
- Currently all analysis is based on stochastic simulation (Gillespie's algorithm).
- Two tools: BioSPI and SPIM which implement slightly different versions of the language.
- There has also been some work on a graphical notation associated with the SPIM tool.

Example: The VICE project

- The aim was to construct a *minimal* cell *in silico* in order to track the dynamics of a complete metabolome.

Example: The VICE project

- The aim was to construct a *minimal* cell *in silico* in order to track the dynamics of a complete metabolome.
- Thus a VIrtual CEll was defined as a stochastic π-calculus model, which seems to behave as a simplified prokaryote.

Example: The VICE project

- The aim was to construct a *minimal* cell *in silico* in order to track the dynamics of a complete metabolome.
- Thus a VIrtual CEll was defined as a stochastic π-calculus model, which seems to behave as a simplified prokaryote.
- Started from a published minimal gene set which eliminated duplicated genes and other redundancies from the smallest known bacterial genomes. This was further reduced to 180 different genes.

Example: The VICE project

- The aim was to construct a *minimal* cell *in silico* in order to track the dynamics of a complete metabolome.
- Thus a VIrtual CEll was defined as a stochastic π-calculus model, which seems to behave as a simplified prokaryote.
- Started from a published minimal gene set which eliminated duplicated genes and other redundancies from the smallest known bacterial genomes. This was further reduced to 180 different genes.
- Experimental results were in accordance with those available from *in vivo* experiments.

Example: The VICE project

- The aim was to construct a *minimal* cell *in silico* in order to track the dynamics of a complete metabolome.
- Thus a VIrtual CEll was defined as a stochastic π-calculus model, which seems to behave as a simplified prokaryote.
- Started from a published minimal gene set which eliminated duplicated genes and other redundancies from the smallest known bacterial genomes. This was further reduced to 180 different genes.
- Experimental results were in accordance with those available from *in vivo* experiments.
- Some extensions to the stochastic π-calculus were needed.

References

A. Regev, W. Silverman and E. Shapiro.

Representation and simulation of biochemical processes using the pi-calculus process algebra.

in *Pacific Symposium on Biocomputing*, Volume 6, pages 459–470. World Scientific Press, 2001.

C. Priami, A. Regev, W. Silverman and E. Shapiro.

Application of a stochastic name passing process calculus to representation and simulation of molecular processes.

in *Information Processing Letters*, 80:25–31, 2001.

D. Chiarugi, M. Curti, P. Degano and R. Marangoni

VICE: A VIrtual CEll

in *Proceedings of the 2nd International Workshop on Computational Methods in Systems Biology* Paris, France, April 2004.

Biology-specific process calculi

The current work on defining biology-specific process calculi has focused on spatial aspects.

Biology-specific process calculi

The current work on defining biology-specific process calculi has focused on spatial aspects.

Thus each of the new calculi places emphasis on the location of components and how this impacts on their potential interactions.

Biology-specific process calculi

The current work on defining biology-specific process calculi has focused on spatial aspects.

Thus each of the new calculi places emphasis on the location of components and how this impacts on their potential interactions.

Examples include:

- The Brane Calculus
- The Bioambient Calculus
- Beta Binders

The Brane Calculus

- Originating from the group of Luca Cardelli at Microsoft Research, Cambridge.

The Brane Calculus

- Originating from the group of Luca Cardelli at Microsoft Research, Cambridge.
- Membranes are taken as primitives of the calculus:

The Brane Calculus

- Originating from the group of Luca Cardelli at Microsoft Research, Cambridge.
- Membranes are taken as primitives of the calculus:
  - Systems consist of nested membranes;

The Brane Calculus

- Originating from the group of Luca Cardelli at Microsoft Research, Cambridge.
- Membranes are taken as primitives of the calculus:
  - Systems consist of nested membranes;
  - Membranes consist of collections of actions;

The Brane Calculus

- Originating from the group of Luca Cardelli at Microsoft Research, Cambridge.
- Membranes are taken as primitives of the calculus:
  - Systems consist of nested membranes;
  - Membranes consist of collections of actions;
  - Reactions happen only at the level of systems and are caused only by actions on membranes.

The Brane Calculus

- Originating from the group of Luca Cardelli at Microsoft Research, Cambridge.
- Membranes are taken as primitives of the calculus:
  - Systems consist of nested membranes;
  - Membranes consist of collections of actions;
  - Reactions happen only at the level of systems and are caused only by actions on membranes.
  - Actions may be bitonal actions of the membrane, binding or release, or molecular interactions.

The Bioambient Calculus

- Originating from Regev *et al.*

The Bioambient Calculus

- Originating from Regev *et al.*
- An extension of the stochastic π-calculus, inspired by the Ambient Calculus which captures process location and movement through computational domains [Cardelli and Gordon].

The Bioambient Calculus

- Originating from Regev *et al.*
- An extension of the stochastic π-calculus, inspired by the Ambient Calculus which captures process location and movement through computational domains [Cardelli and Gordon].
- An ambient is a bounded place where computation can happen — within each ambient there may be component processes and sub-ambients.

The Bioambient Calculus

- Originating from Regev *et al.*
- An extension of the stochastic π-calculus, inspired by the Ambient Calculus which captures process location and movement through computational domains [Cardelli and Gordon].
- An ambient is a bounded place where computation can happen — within each ambient there may be component processes and sub-ambients.
- Entities may enter or exit an ambient and ambients may merge.

Beta Binders

- Originating from the group of Corrado Priami and Paola Quaglia at the University of Trento.

Beta Binders

- Originating from the group of Corrado Priami and Paola Quaglia at the University of Trento.
- Can be viewed as an extension of the π-calculus.

Beta Binders

- Originating from the group of Corrado Priami and Paola Quaglia at the University of Trento.
- Can be viewed as an extension of the π-calculus.
- Processes are encapsulated into boxes, that mimic biological membranes, and have interaction sites, motifs, depicting where molecules can bind together.

Beta Binders

- Originating from the group of Corrado Priami and Paola Quaglia at the University of Trento.
- Can be viewed as an extension of the π-calculus.
- Processes are encapsulated into boxes, that mimic biological membranes, and have interaction sites, motifs, depicting where molecules can bind together.
- The semantics give rules on joining and splitting boxes, as well as the affinity between interaction sites.

References

L. Cardelli

Brane Calculus

in *Proceedings of the 2nd International Workshop on Computational Methods in Systems Biology* Paris, France, April 2004.

A. Regev, E. Panina, W. Silverman, L. Cardelli and E. Shapiro

BioAmbients: an abstraction for biological compartments.

in *Theoretical Computer Science* 325(1):141–167, 2004.

C. Priami and P. Quaglia

Beta-binders for Biological Interactions

in *Proceedings of the 2nd International Workshop on Computational Methods in Systems Biology* Paris, France, April 2004.

On-going work

- System description techniques previously used for performance analysis have been demonstrated to be useful abstractions of a variety of biochemical systems.

On-going work

- System description techniques previously used for performance analysis have been demonstrated to be useful abstractions of a variety of biochemical systems.
- Particular emphasis is being paid to the use of abstraction and reasoning about models.

On-going work

- System description techniques previously used for performance analysis have been demonstrated to be useful abstractions of a variety of biochemical systems.
- Particular emphasis is being paid to the use of abstraction and reasoning about models.
- We are studying the relationship between population level models, and more individual-focused models.

On-going work

- System description techniques previously used for performance analysis have been demonstrated to be useful abstractions of a variety of biochemical systems.
- Particular emphasis is being paid to the use of abstraction and reasoning about models.
- We are studying the relationship between population level models, and more individual-focused models.
- In the future we plan to investigate the extent to which the process algebra compositional structure can be exploited during model analysis.

Models as Tools

When designing a modelling formalism it is important to consider two key aspects:

Models as Tools

When designing a modelling formalism it is important to consider two key aspects:

- the ability of the formalism to capture the behaviour of interest and the availability of data to instantiate the model;

Models as Tools

When designing a modelling formalism it is important to consider two key aspects:

- the ability of the formalism to capture the behaviour of interest and the availability of data to instantiate the model;
- the amenability of the formalism to appropriate analysis.

Models as Tools

When designing a modelling formalism it is important to consider two key aspects:

- the ability of the formalism to capture the behaviour of interest and the availability of data to instantiate the model;
- the amenability of the formalism to appropriate analysis.

What is the behaviour of interest and appropriate analysis depends on the question or problem you are seeking to address.

Models as Tools

When designing a modelling formalism it is important to consider two key aspects:

- the ability of the formalism to capture the behaviour of interest and the availability of data to instantiate the model;
- the amenability of the formalism to appropriate analysis.

What is the behaviour of interest and appropriate analysis depends on the question or problem you are seeking to address.

It is not the case the models have to be completely faithful to their subject in order to be useful.

Motivations for Abstraction

Our motivations for seeking more abstraction in process algebra models for systems biology comes from both key aspects of modelling:

Motivations for Abstraction

Our motivations for seeking more abstraction in process algebra models for systems biology comes from both key aspects of modelling:

- The data that we have available to parameterise models is sometimes speculative rather than precise.

Motivations for Abstraction

Our motivations for seeking more abstraction in process algebra models for systems biology comes from both key aspects of modelling:

- The data that we have available to parameterise models is sometimes speculative rather than precise. This suggests that we should use semiquantitative models rather than quantitative ones.

Motivations for Abstraction

Our motivations for seeking more abstraction in process algebra models for systems biology comes from both key aspects of modelling:

- The data that we have available to parameterise models is sometimes speculative rather than precise. This suggests that we should use semiquantitative models rather than quantitative ones.
- Process algebra based analyses such as comparing models (e.g. for equivalence or simulation) and model checking are only possible is the state space is not prohibitively large.

Motivations for Abstraction

Our motivations for seeking more abstraction in process algebra models for systems biology comes from both key aspects of modelling:

- The data that we have available to parameterise models is sometimes speculative rather than precise. This suggests that we should use semiquantitative models rather than quantitative ones.
- Process algebra based analyses such as comparing models (e.g. for equivalence or simulation) and model checking are only possible is the state space is not prohibitively large.

Challenges of Systems Biology

- Biologists.

Challenges of Systems Biology

- Biologists.
  - Varied approaches to capturing and representing data.

Challenges of Systems Biology

- Biologists.
  - Varied approaches to capturing and representing data.
  - Unfamiliarity with notions of *abstraction* and *quantification*.

Challenges of Systems Biology

- Biologists.
  - Varied approaches to capturing and representing data.
  - Unfamiliarity with notions of *abstraction* and *quantification*.
  - Expectations.

Challenges of Systems Biology

- Biologists.
  - Varied approaches to capturing and representing data.
  - Unfamiliarity with notions of *abstraction* and *quantification*.
  - Expectations.
- Scalability and tractability.

Conclusions

- Ultimately we want to understand the functioning of cells as useful levels of abstraction, and to predict unknown behaviour.

Conclusions

- Ultimately we want to understand the functioning of cells as useful levels of abstraction, and to predict unknown behaviour.
- It remains an open and challenging problem to define a set of basic and general primitives for modelling biological systems, inspired by biological processes.

Conclusions

- Ultimately we want to understand the functioning of cells as useful levels of abstraction, and to predict unknown behaviour.
- It remains an open and challenging problem to define a set of basic and general primitives for modelling biological systems, inspired by biological processes.
- Achieving this goal is anticipated to have two broad benefits:

Conclusions

- Ultimately we want to understand the functioning of cells as useful levels of abstraction, and to predict unknown behaviour.
- It remains an open and challenging problem to define a set of basic and general primitives for modelling biological systems, inspired by biological processes.
- Achieving this goal is anticipated to have two broad benefits:
  - Better models and simulations of living phenomena

Conclusions

- Ultimately we want to understand the functioning of cells as useful levels of abstraction, and to predict unknown behaviour.
- It remains an open and challenging problem to define a set of basic and general primitives for modelling biological systems, inspired by biological processes.
- Achieving this goal is anticipated to have two broad benefits:
  - Better models and simulations of living phenomena
  - New models of computations that are biologically inspired.

Conclusions

- Ultimately we want to understand the functioning of cells as useful levels of abstraction, and to predict unknown behaviour.
- It remains an open and challenging problem to define a set of basic and general primitives for modelling biological systems, inspired by biological processes.
- Achieving this goal is anticipated to have two broad benefits:
  - Better models and simulations of living phenomena
  - New models of computations that are biologically inspired.
- Inclusion of quantitative/stochastic elements is essential.

Thank You!