Modelling Biochemical Pathways with Stochastic Process Algebra

Jane Hillston.
LFCS, University of Edinburgh

9th May 2007

The PEPA project

- The PEPA project started in Edinburgh in 1991.

The PEPA project

- The PEPA project started in Edinburgh in 1991.
- It was motivated by problems encountered when carrying out performance analysis of large computer and communication systems, based on numerical analysis of Markov processes.

The PEPA project

- The PEPA project started in Edinburgh in 1991.
- It was motivated by problems encountered when carrying out performance analysis of large computer and communication systems, based on numerical analysis of Markov processes.
- Process algebras offered a compositional description technique supported by apparatus for formal reasoning.

The PEPA project

- The PEPA project started in Edinburgh in 1991.
- It was motivated by problems encountered when carrying out performance analysis of large computer and communication systems, based on numerical analysis of Markov processes.
- Process algebras offered a compositional description technique supported by apparatus for formal reasoning.
- Performance Evaluation Process Algebra (PEPA) sought to address these problems by the introduction of a suitable process algebra.

The PEPA project

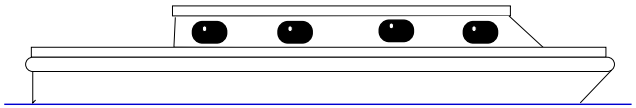
- The PEPA project started in Edinburgh in 1991.
- It was motivated by problems encountered when carrying out performance analysis of large computer and communication systems, based on numerical analysis of Markov processes.
- Process algebras offered a compositional description technique supported by apparatus for formal reasoning.
- Performance Evaluation Process Algebra (PEPA) sought to address these problems by the introduction of a suitable process algebra.
- The project has sought to investigate and exploit the interplay between the process algebra and the continuous time Markov chain (CTMC).

PEPA Case Studies (1)

- Multiprocessor access-contention protocols (Gilmore, Hillston and Ribaudo, Edinburgh and Turin)
- Protocols for fault-tolerant systems (Clark, Gilmore, Hillston and Ribaudo, Edinburgh and Turin)
- Multimedia traffic characteristics (Bowman et al, Kent)
- Database systems (The STEADY group, Heriot-Watt University)
- Software Architectures (Pooley, Bradley and Thomas, Heriot-Watt and Durham)
- Switch behaviour in active networks (Hillston, Kloul and Mokhtari, Edinburgh and Versailles)

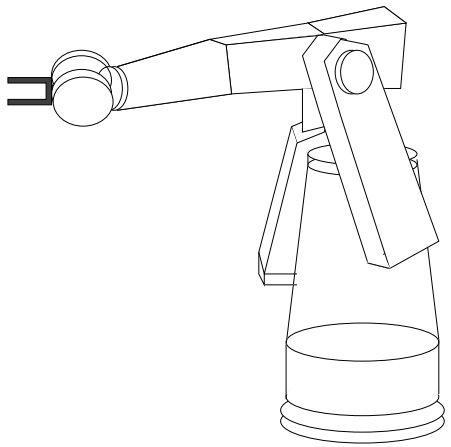
PEPA Case Studies (2)

- Locks and movable bridges in inland shipping in Belgium (Knapen, Hasselt)



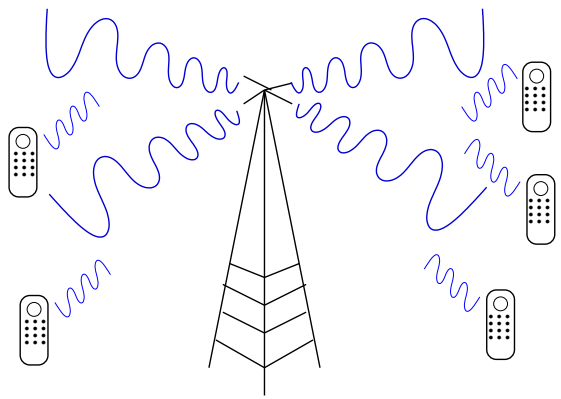
PEPA Case Studies (2)

- Locks and movable bridges in inland shipping in Belgium (Knapen, Hasselt)
- Robotic workcells (Holton, Gilmore and Hillston, Bradford and Edinburgh)



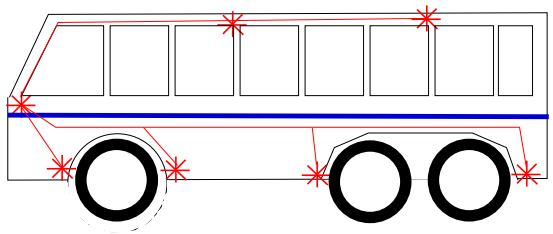
PEPA Case Studies (2)

- Locks and movable bridges in inland shipping in Belgium (Knapen, Hasselt)
- Robotic workcells (Holton, Gilmore and Hillston, Bradford and Edinburgh)
- Cellular telephone networks (Kloul, Fourneau and Valois, Versailles)



PEPA Case Studies (2)

- Locks and movable bridges in inland shipping in Belgium (Knapen, Hasselt)
- Robotic workcells (Holton, Gilmore and Hillston, Bradford and Edinburgh)
- Cellular telephone networks (Kloul, Fourneau and Valois, Versailles)
- Automotive diagnostic expert systems (Console, Picardi and Ribaudo, Turin)



Outline

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.

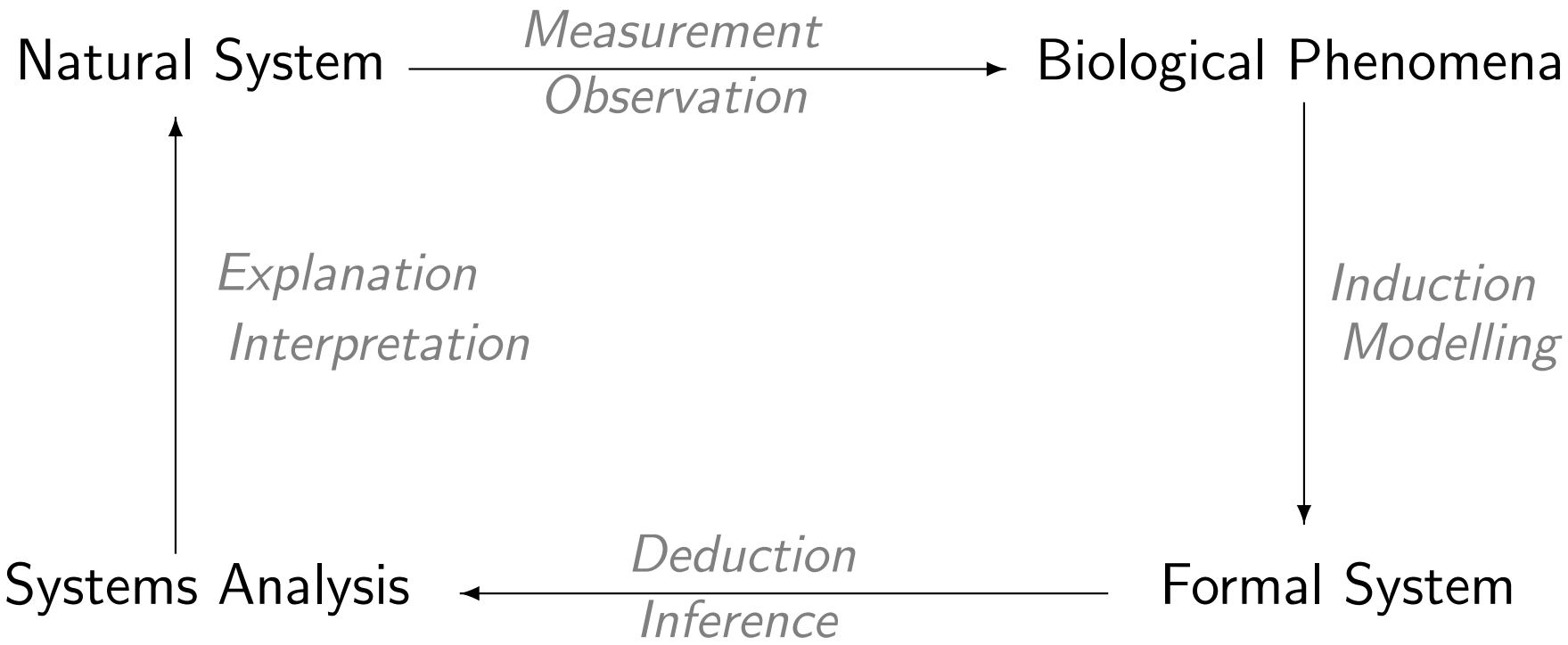
Systems Biology

- 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.
- It involves taking a systems theoretic view of biological processes — analysing inputs and outputs and the relationships between them.

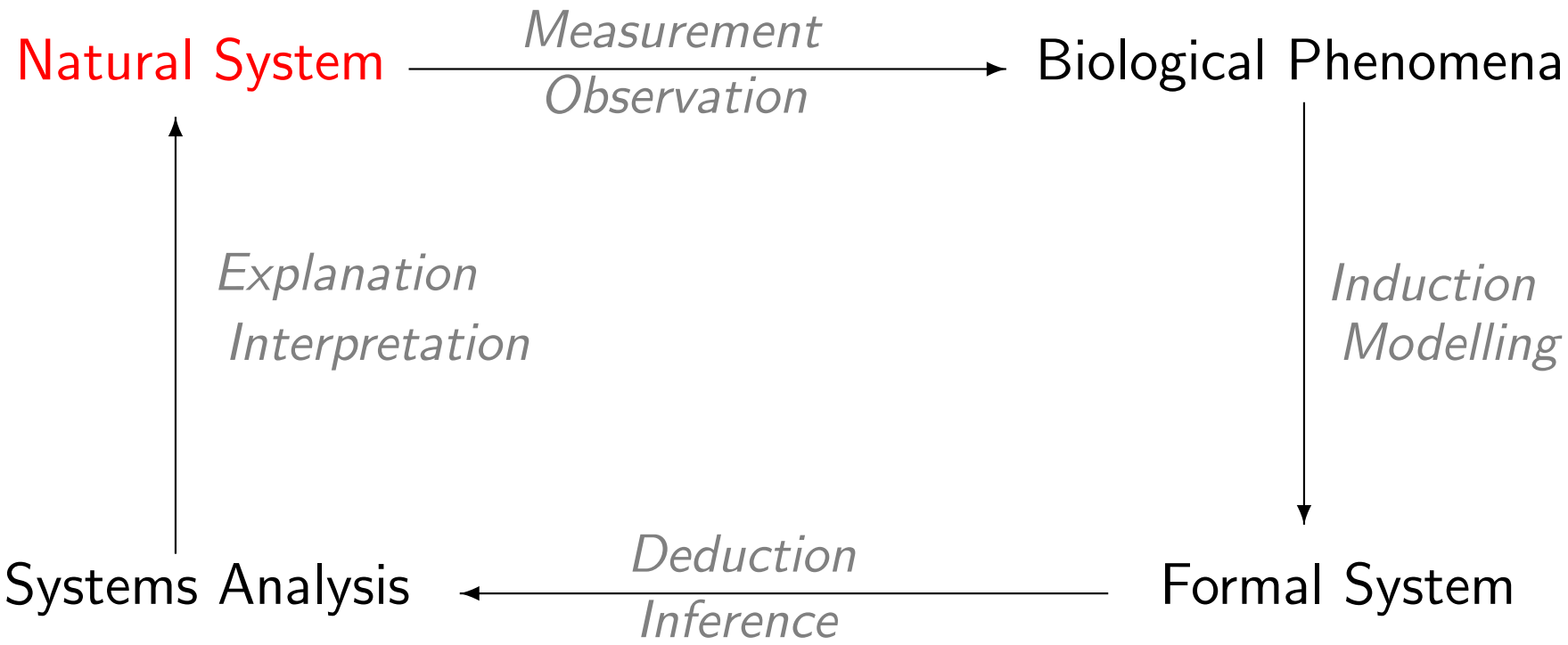
Systems Biology

- 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.
- It involves taking a systems theoretic view of biological processes — analysing inputs and outputs and the relationships between them.
- A radical shift from earlier reductionist approaches, systems biology aims to provide a conceptual basis and a methodology for reasoning about biological phenomena.

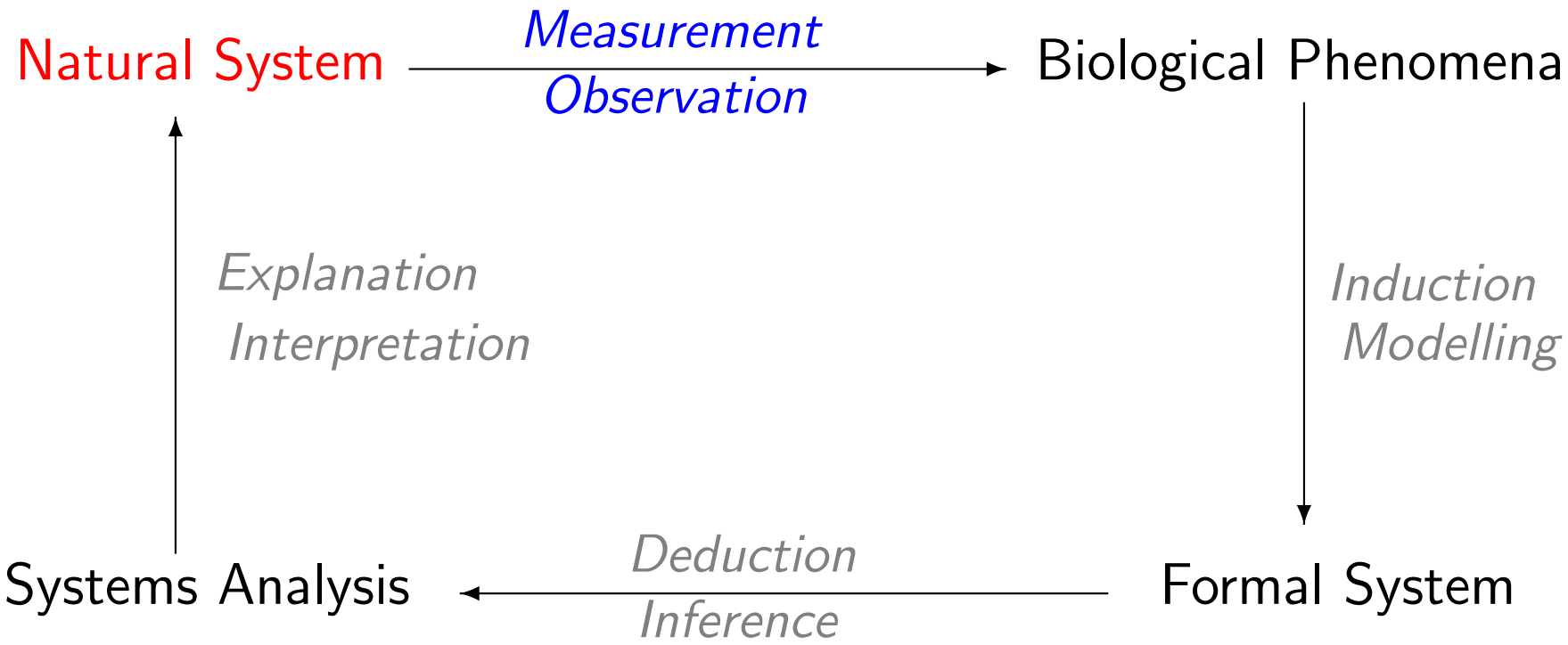
Systems Biology Methodology



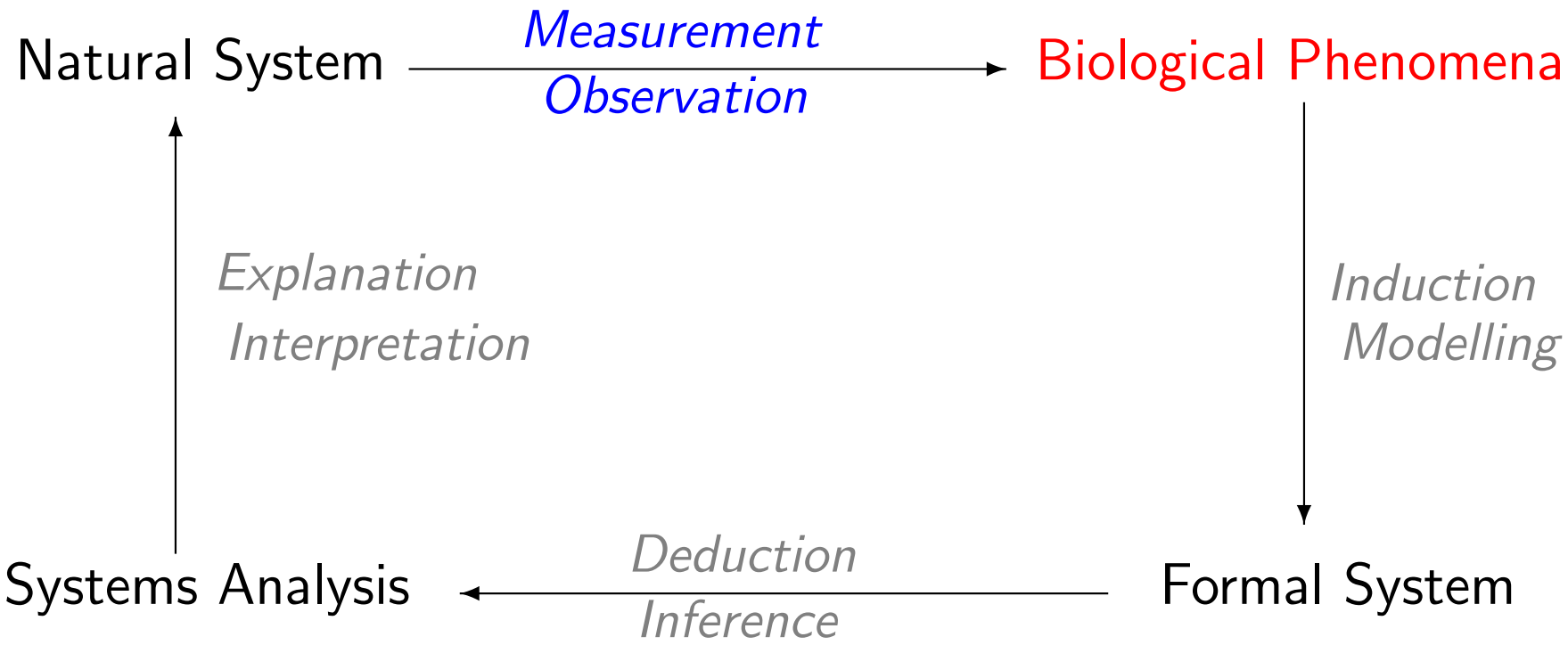
Systems Biology Methodology



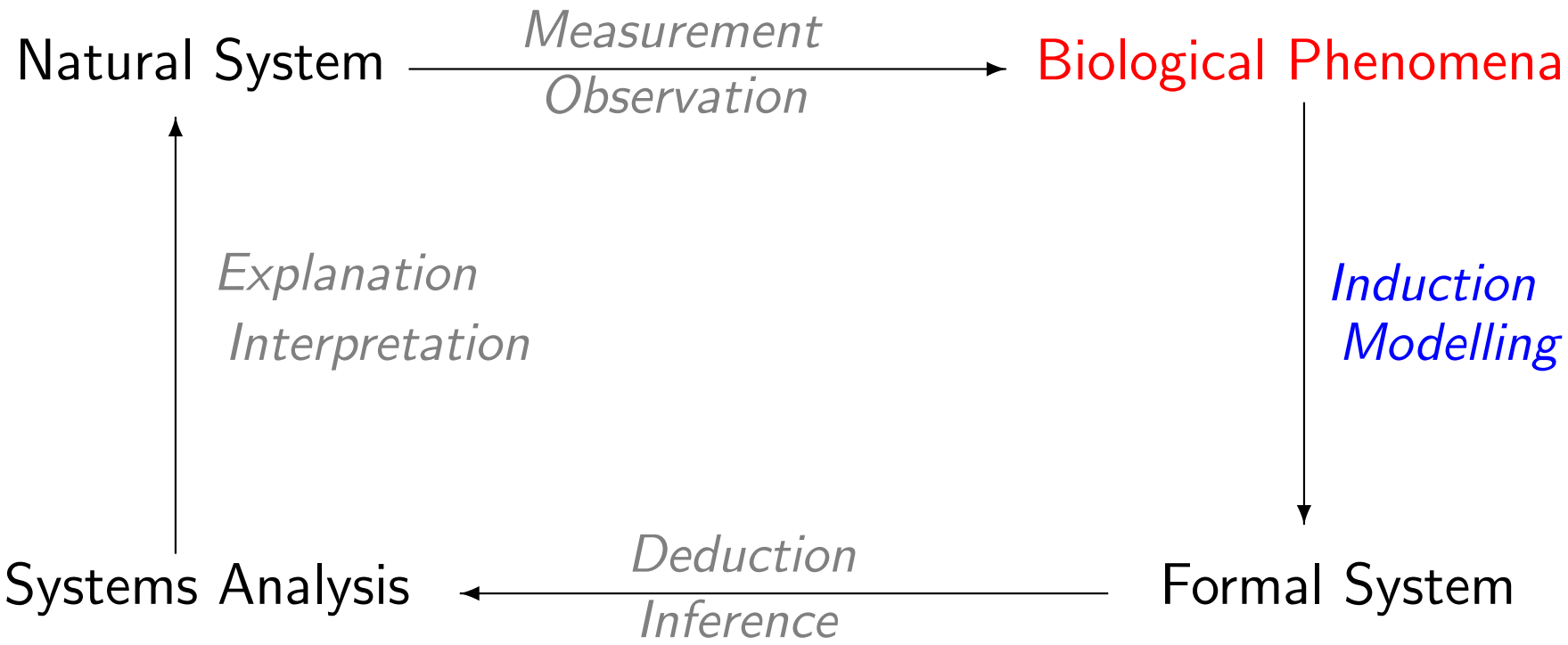
Systems Biology Methodology



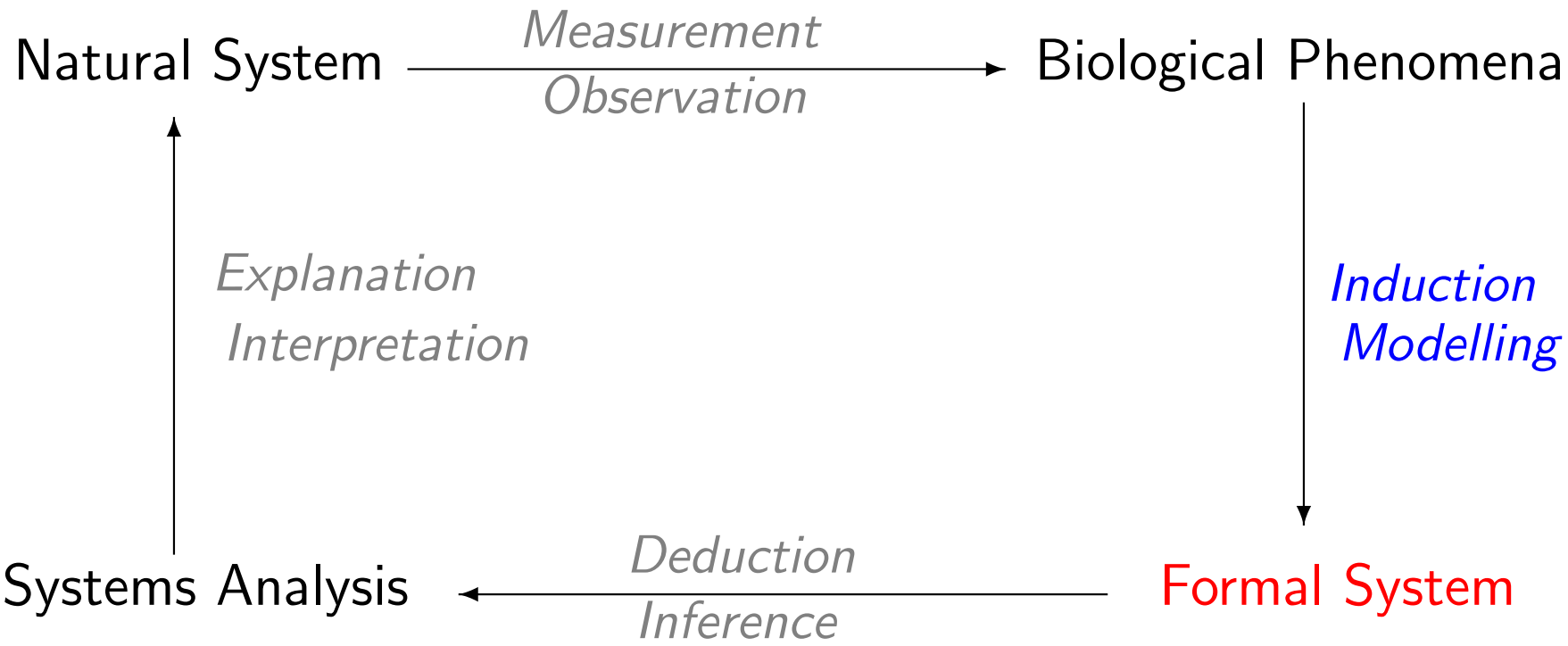
Systems Biology Methodology



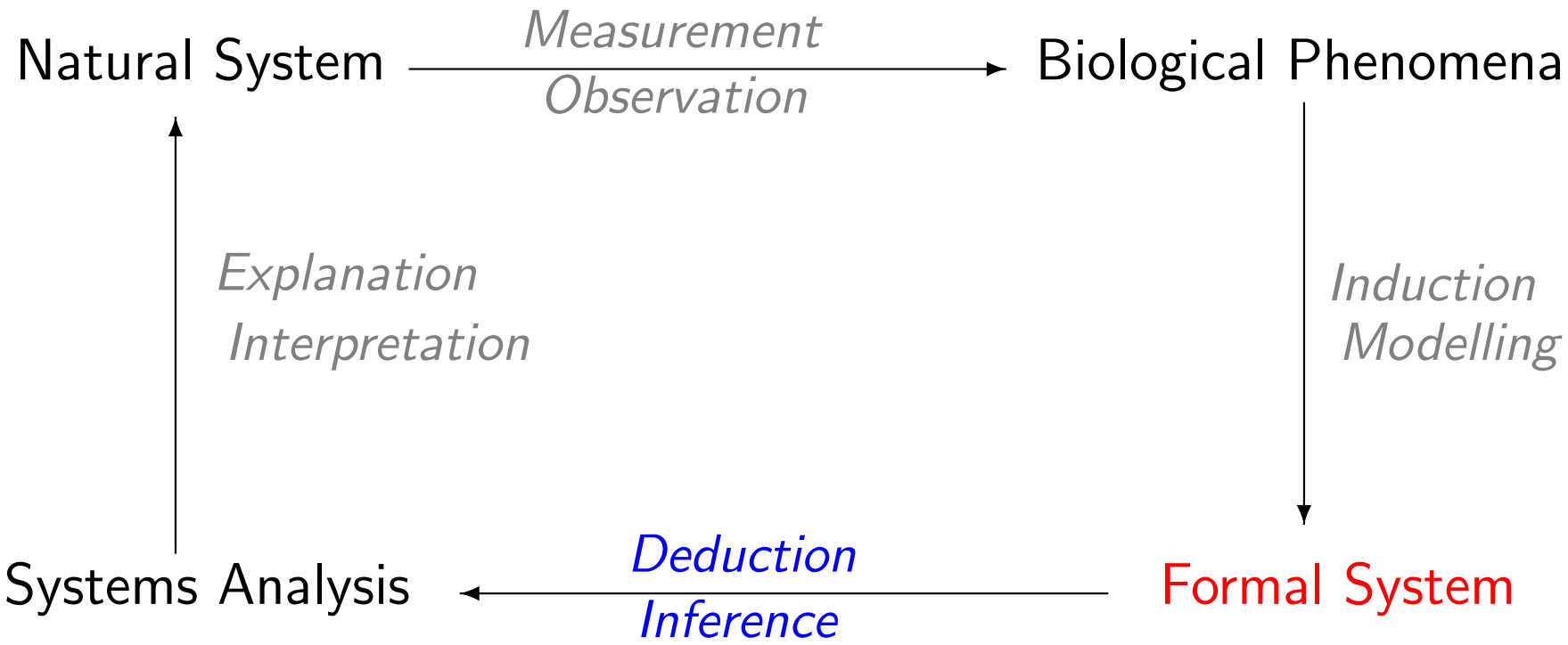
Systems Biology Methodology



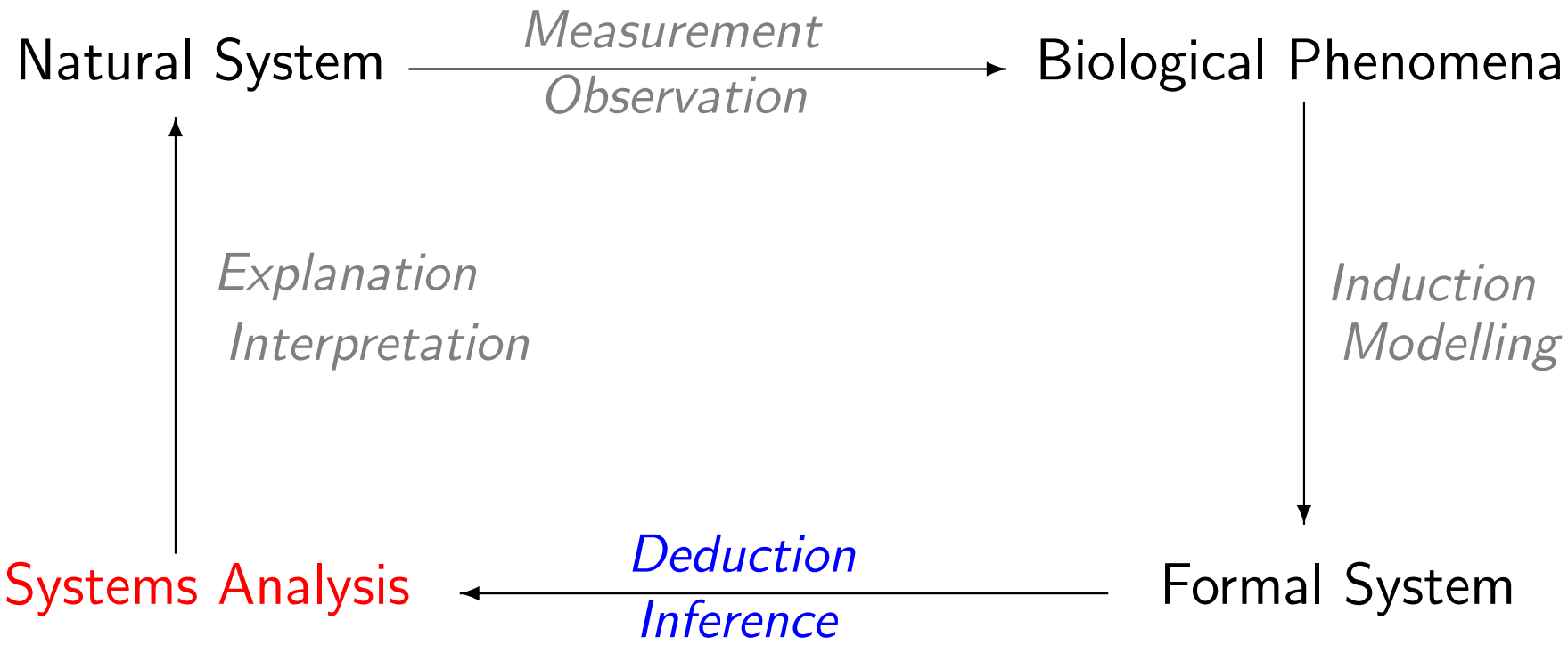
Systems Biology Methodology



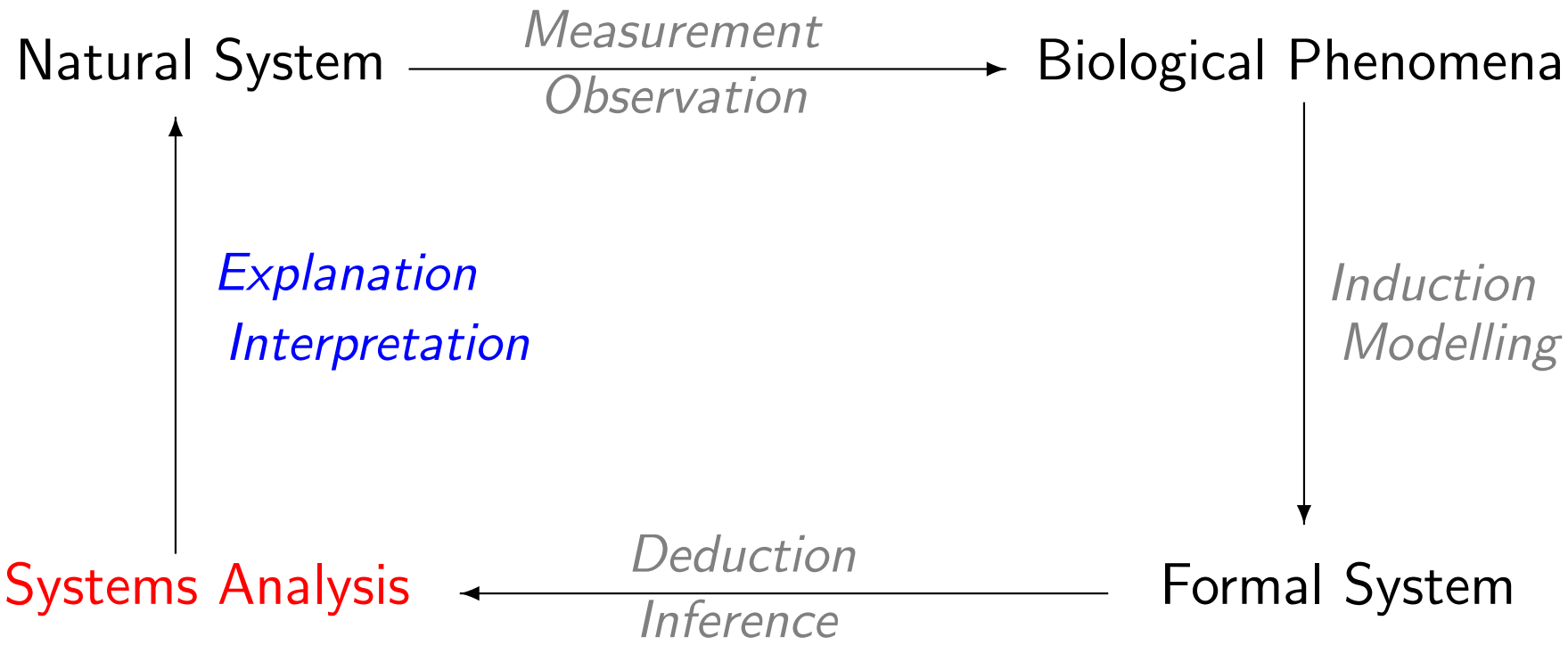
Systems Biology Methodology



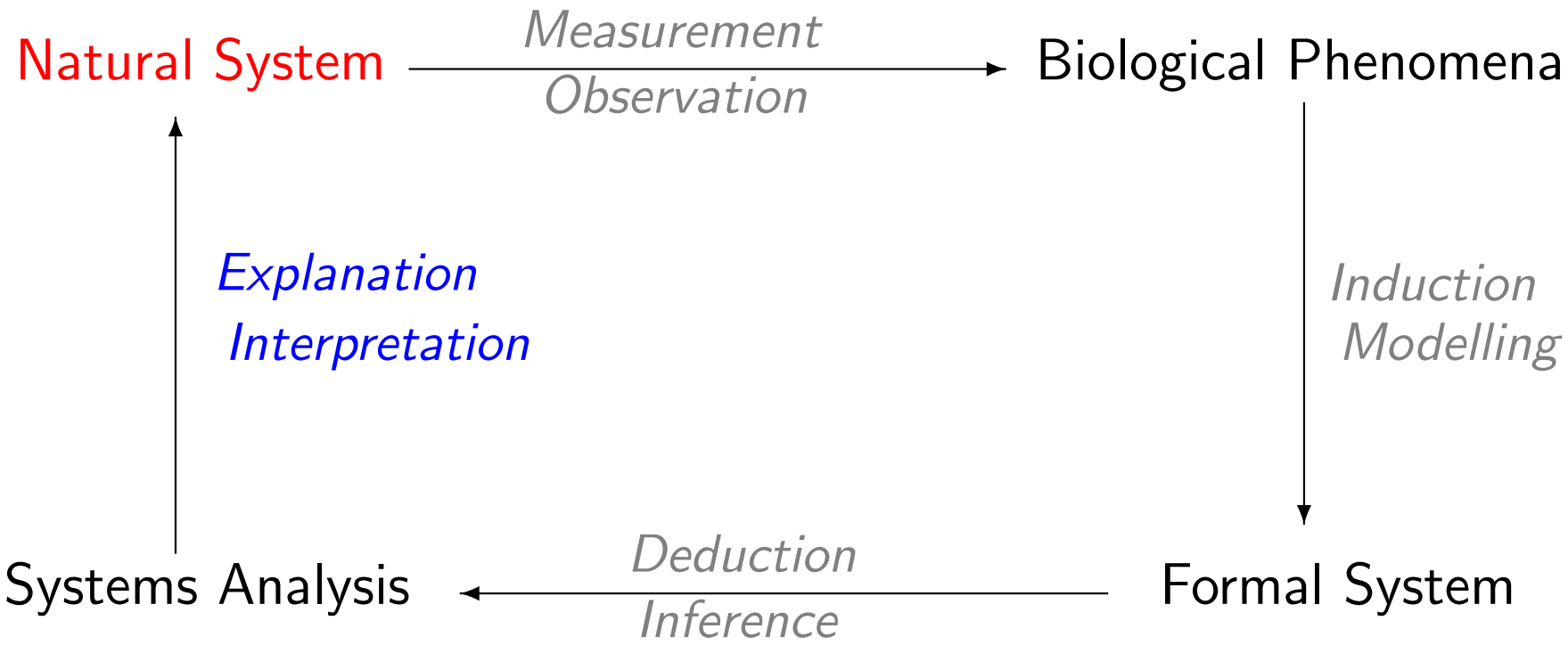
Systems Biology Methodology



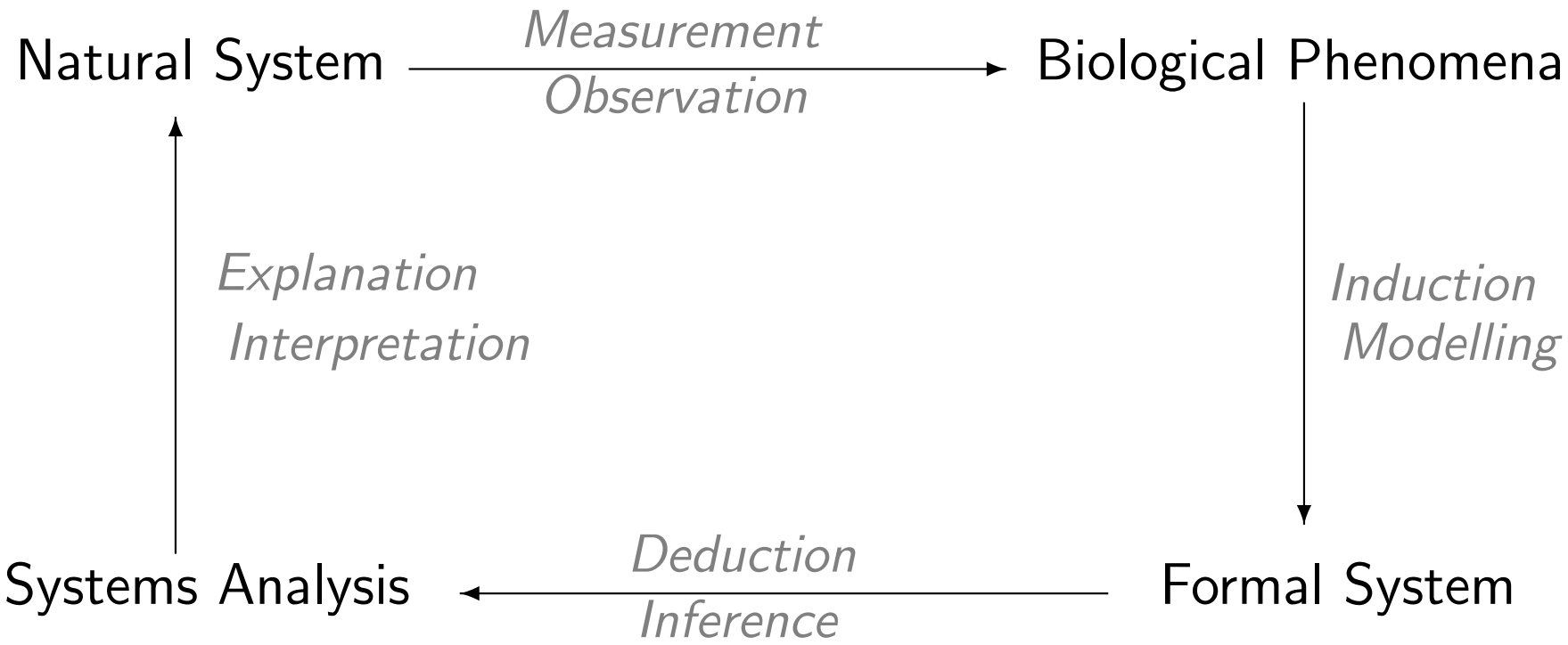
Systems Biology Methodology



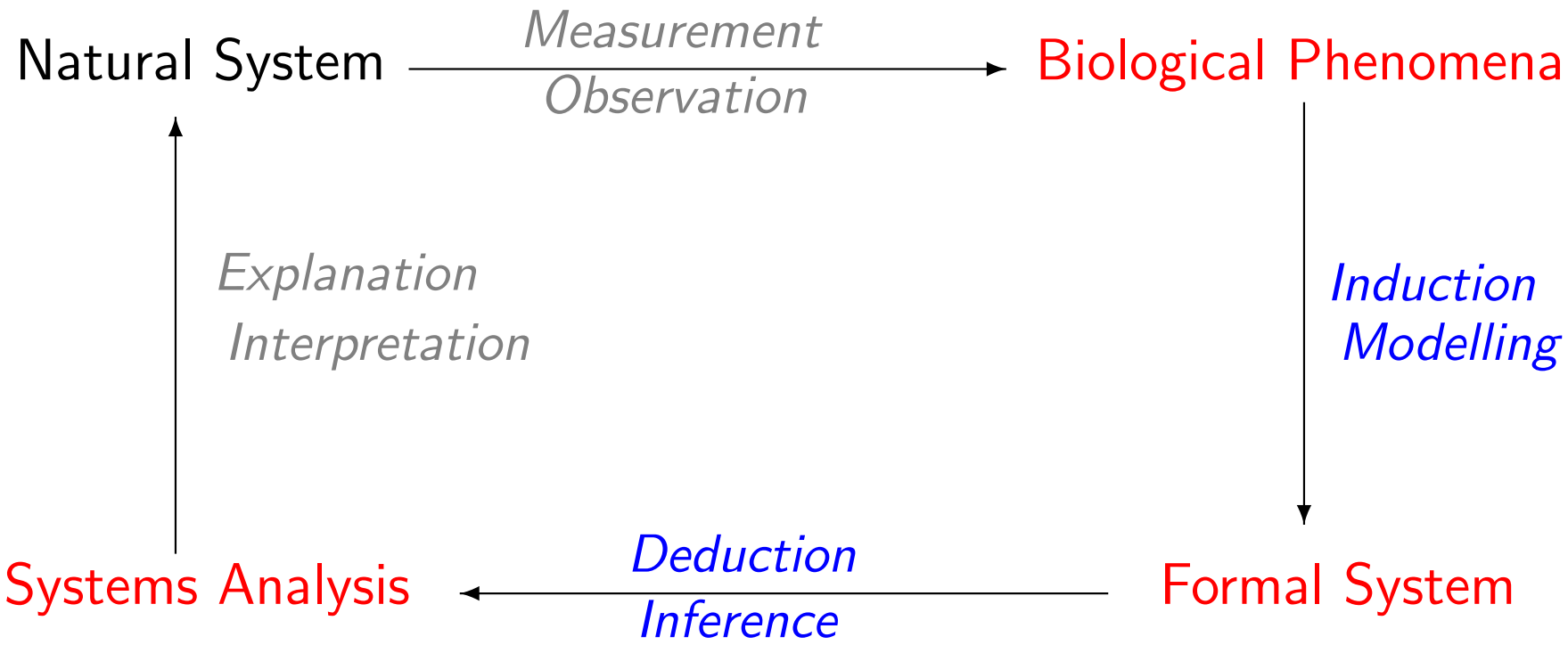
Systems Biology Methodology



Systems Biology Methodology



Systems Biology Methodology



Biochemical Pathways

At the intra-cellular level we can distinguish three distinct types of pathways or networks

Biochemical Pathways

At the intra-cellular level we can distinguish three distinct types of pathways or networks

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

Biochemical Pathways

At the intra-cellular level we can distinguish three distinct types of pathways or networks

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.

Biochemical Pathways

At the intra-cellular level we can distinguish three distinct types of pathways or networks

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.

Biochemical Pathways

At the intra-cellular level we can distinguish three distinct types of pathways or networks

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

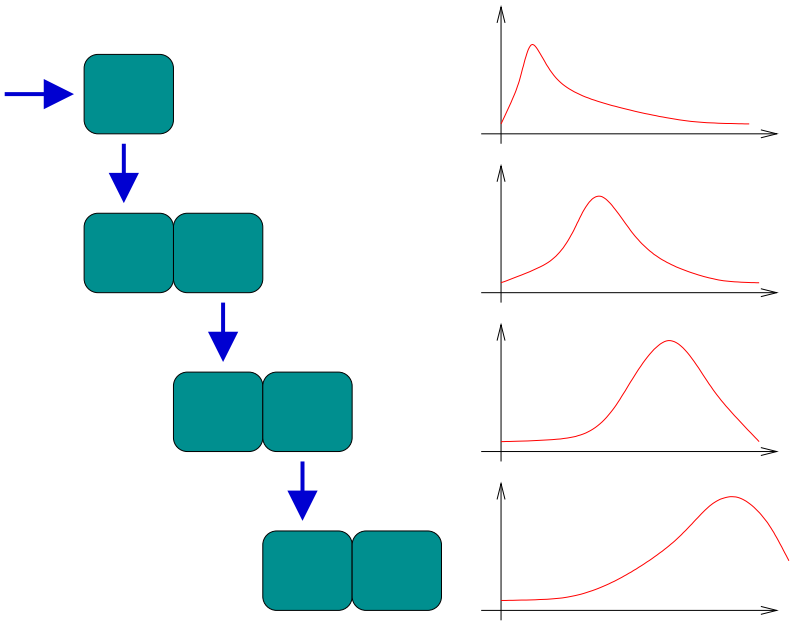
Signal transduction networks: External stimuli initiate messages that are carried through a cell via a cascade of biochemical reactions.

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

But these distinctions are to some extent arbitrary as models may include elements of more than one pathway type.

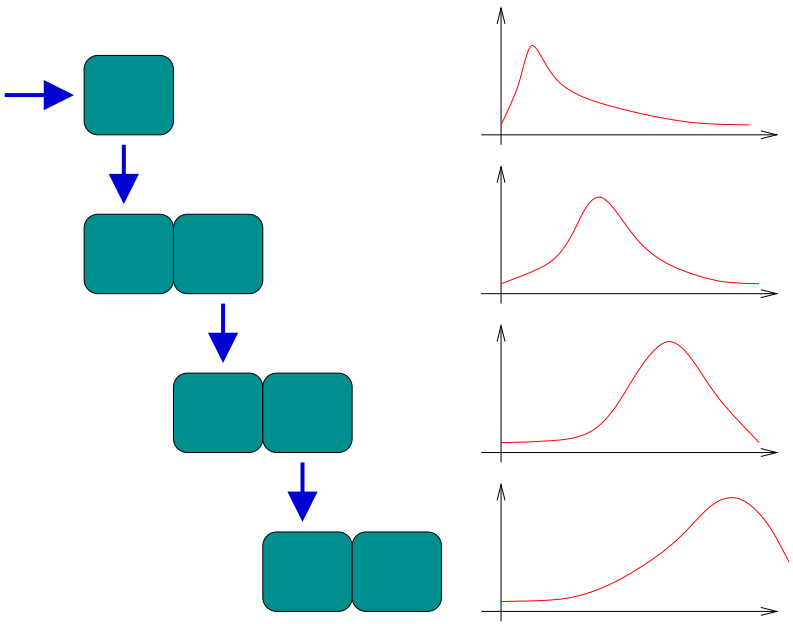
Signal transduction pathways

- All signalling is biochemical:



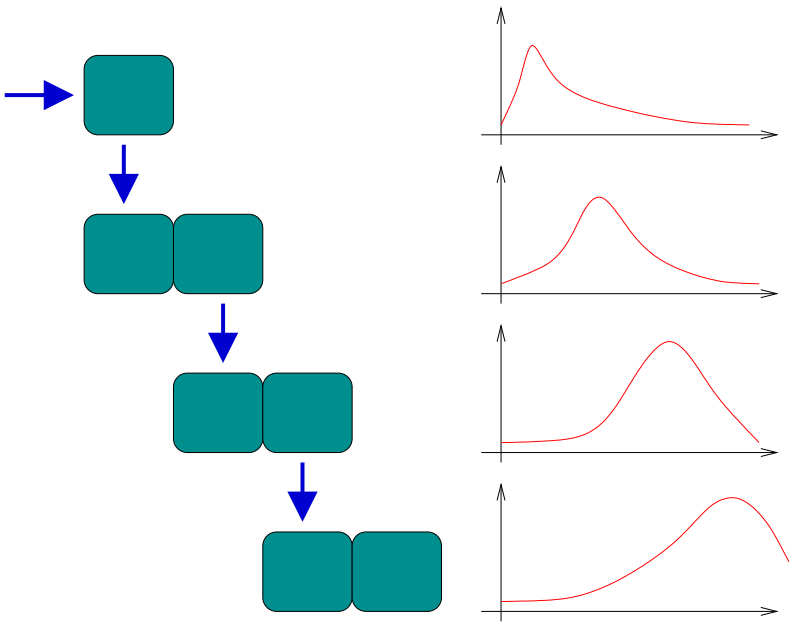
Signal transduction pathways

- All signalling is biochemical:
- Increasing protein concentration broadcasts the information about an event.



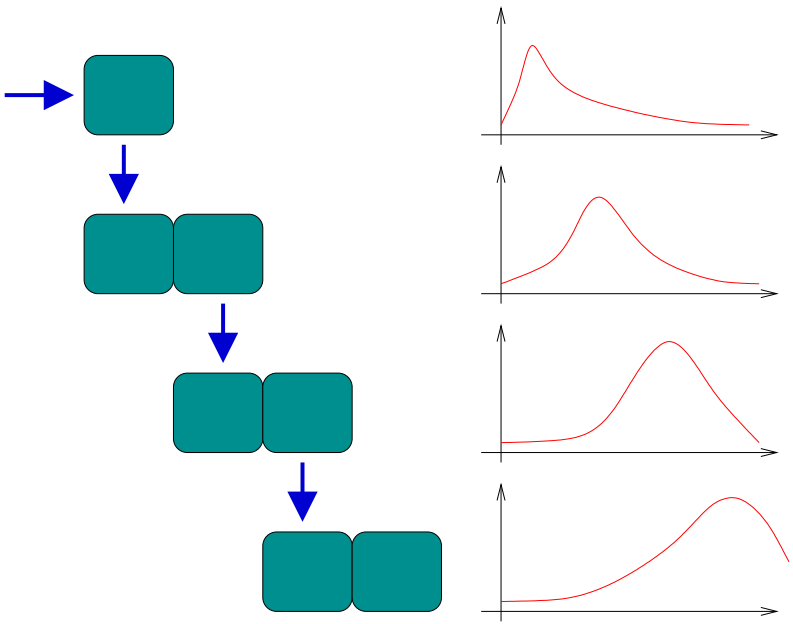
Signal transduction pathways

- All signalling is biochemical:
- Increasing protein concentration broadcasts the information about an event.
- The message is "received" by a concentration dependent response at the protein signal's site of action.



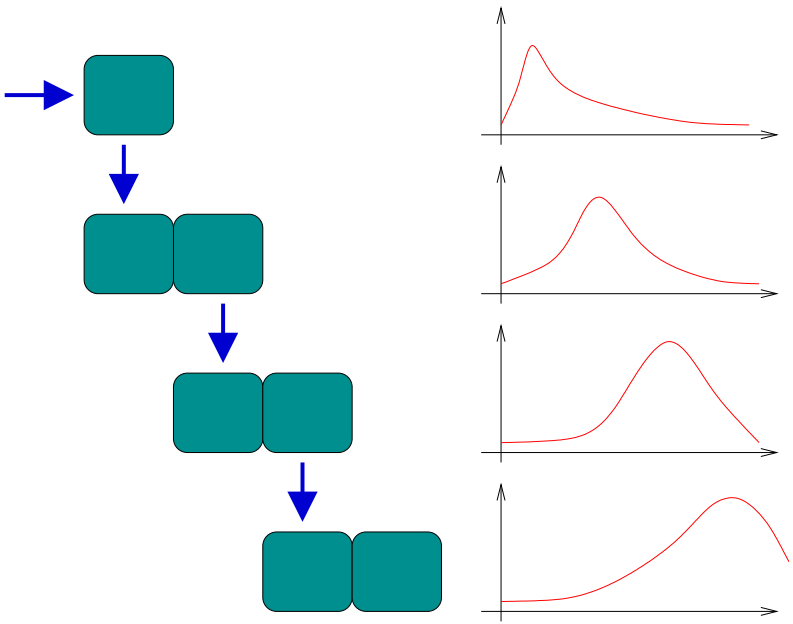
Signal transduction pathways

- All signalling is biochemical:
- Increasing protein concentration broadcasts the information about an event.
- 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.



Signal transduction pathways

- All signalling is biochemical:
- Increasing protein concentration broadcasts the information about an event.
- 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.



Formal Systems

There are two alternative approaches to contructing dynamic models of biochemical pathways commonly used by biologists:

Formal Systems

There are two alternative approaches to contructing dynamic models of biochemical pathways commonly used by biologists:

- Ordinary Differential Equations:
  - continuous time,
  - continuous behaviour (concentrations),
  - deterministic.

Formal Systems

There are two alternative approaches to contructing dynamic models of biochemical pathways commonly used by biologists:

- Ordinary Differential Equations:
  - continuous time,
  - continuous behaviour (concentrations),
  - deterministic.
- Stochastic Simulation:
  - continuous time,
  - discrete behaviour (no. of molecules),
  - stochastic.

Ordinary Differential Equations

- This deterministic approach has at its core 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.

Ordinary Differential Equations

- This deterministic approach has at its core 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.

For example, for a reaction $A + B \xrightarrow{k} C$:

$$\frac{d[A]}{dt} = \frac{d[B]}{dt} = -k[A][B]$$

$$\frac{d[C]}{dt} = k[A][B]$$

Limitations of Ordinary Differential Equations

- Given knowledge of initial molecular concentrations, the law of mass action provides a complete picture of the component concentrations at all future time points.

Limitations of Ordinary Differential Equations

- Given knowledge of initial molecular concentrations, the law of mass action provides a complete picture of the component concentrations at all future time points.
- This is based on the assumption that chemical reactions to be macroscopic under convective or diffusive stirring, continuous and deterministic.

Limitations of Ordinary Differential Equations

- Given knowledge of initial molecular concentrations, the law of mass action provides a complete picture of the component concentrations at all future time points.
- This is based on the assumption that chemical reactions to be macroscopic under convective or diffusive stirring, continuous and deterministic.
- This is a simplification, because in reality chemical reactions involve discrete, random collisions between individual molecules.

Limitations of Ordinary Differential Equations

- Given knowledge of initial molecular concentrations, the law of mass action provides a complete picture of the component concentrations at all future time points.
- This is based on the assumption that chemical reactions to be macroscopic under convective or diffusive stirring, continuous and deterministic.
- This is a simplification, because in reality 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.

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: Chemical Master Equation

Applying this leads us to an important *partial differential equation* (PDE) known as the Chemical Master Equation (CME).

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

Does not lend itself to either analytic nor numerical solutions.

Stochastic: Chemical Master Equation

Applying this leads us to an important *partial differential equation* (PDE) known as the Chemical Master Equation (CME).

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

Does not lend itself to either analytic nor numerical solutions.

Stochastic: Chemical Master Equation

Applying this leads us to an important *partial differential equation* (PDE) known as the Chemical Master Equation (CME).

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

Does not lend itself to either analytic nor numerical solutions.

Stochastic: Chemical Master Equation

Applying this leads us to an important *partial differential equation* (PDE) known as the Chemical Master Equation (CME).

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

Does not lend itself to either analytic nor numerical solutions.

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.

Systems Analysis

- In biochemical signalling pathways the events of interests are
  - when reagent concentrations start to increase;
  - when concentrations pass certain thresholds;
  - when a peak of concentration is reached.

Systems Analysis

- In biochemical signalling pathways the events of interests are
  - when reagent concentrations start to increase;
  - when concentrations pass certain thresholds;
  - when a peak of concentration is reached.
- For example, delay from the activation of a gene promoter until reaching an effective level to control the next promoter in a pathway depends on the rate of protein accumulation.

Systems Analysis

- In biochemical signalling pathways the events of interests are
  - when reagent concentrations start to increase;
  - when concentrations pass certain thresholds;
  - when a peak of concentration is reached.
- For example, delay from the activation of a gene promoter until reaching an effective level to control the next promoter in a pathway depends on the rate of protein accumulation.
- These data can be collected from wet lab experiments.

Systems Analysis

- In biochemical signalling pathways the events of interests are
  - when reagent concentrations start to increase;
  - when concentrations pass certain thresholds;
  - when a peak of concentration is reached.
- For example, delay from the activation of a gene promoter until reaching an effective level to control the next promoter in a pathway depends on the rate of protein accumulation.
- These data can be collected from wet lab experiments.
- The accumulation of protein is a stochastic process affected by several factors in the cell (temperature, pH, etc.).

Systems Analysis

- In biochemical signalling pathways the events of interests are
  - when reagent concentrations start to increase;
  - when concentrations pass certain thresholds;
  - when a peak of concentration is reached.
- For example, delay from the activation of a gene promoter until reaching an effective level to control the next promoter in a pathway depends on the rate of protein accumulation.
- These data can be collected from wet lab experiments.
- The accumulation of protein is a stochastic process affected by several factors in the cell (temperature, pH, etc.).
- Thus it is more realistic to talk about a distribution rather than a deterministic time.

Formal Systems Revisited

- Currently mathematics is being used directly as the formal system — even the work with the stochastic π-calculus only uses the π-calculus to describe a continuous time Markov chain (CTMC) for simulation.

Formal Systems Revisited

- Currently mathematics is being used directly as the formal system — even the work with the stochastic π-calculus only uses the π-calculus to describe a continuous time Markov chain (CTMC) for simulation.
- Previous experience in the performance arena has shown us that there can be benefits to interposing a formal model between the system and the underlying mathematical model.

Formal Systems Revisited

- Currently mathematics is being used directly as the formal system — even the work with the stochastic π-calculus only uses the π-calculus to describe a continuous time Markov chain (CTMC) for simulation.
- Previous experience in the performance arena has shown us that there can be benefits to interposing a formal model between the system and the underlying mathematical model.
- Moreover taking this "high-level programming" style approach offers the possibility of different "compilations" to different mathematical models.

Outline

Using Stochastic Process Algebras

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

Using Stochastic Process Algebras

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.

Using Stochastic Process Algebras

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.
- Structure can also be apparent.

Using Stochastic Process Algebras

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.
- Structure can also be apparent.
- Equivalence relations allow formal comparison of high-level descriptions.

Using Stochastic Process Algebras

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.
- 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.

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]

Mapping biological systems to process algebra

The work using the stochastic π-calculus and related calculi, maps a molecule to a process in the process algebra description.

Mapping biological systems to process algebra

The work using the stochastic π-calculus and related calculi, maps a molecule to a process in the process algebra description.

This is an inherently individuals-based view of the system and analysis will generally be via stochastic simulation.

Mapping biological systems to process algebra

The work using the stochastic π-calculus and related calculi, maps a molecule to a process in the process algebra description.

This is an inherently individuals-based view of the system and analysis will generally be via stochastic simulation.

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

The work using the stochastic π-calculus and related calculi, maps a molecule to a process in the process algebra description.

This is an inherently individuals-based view of the system and analysis will generally be via stochastic simulation.

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 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

The work using the stochastic π-calculus and related calculi, maps a molecule to a process in the process algebra description.

This is an inherently individuals-based view of the system and analysis will generally be via stochastic simulation.

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

Alternative mappings from the process algebra to underlying mathematics are then readily available.

Motivations for Abstraction

Our motivations for seeking more abstraction in process algebra models for systems biology are:

Motivations for Abstraction

Our motivations for seeking more abstraction in process algebra models for systems biology are:

- 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 are:

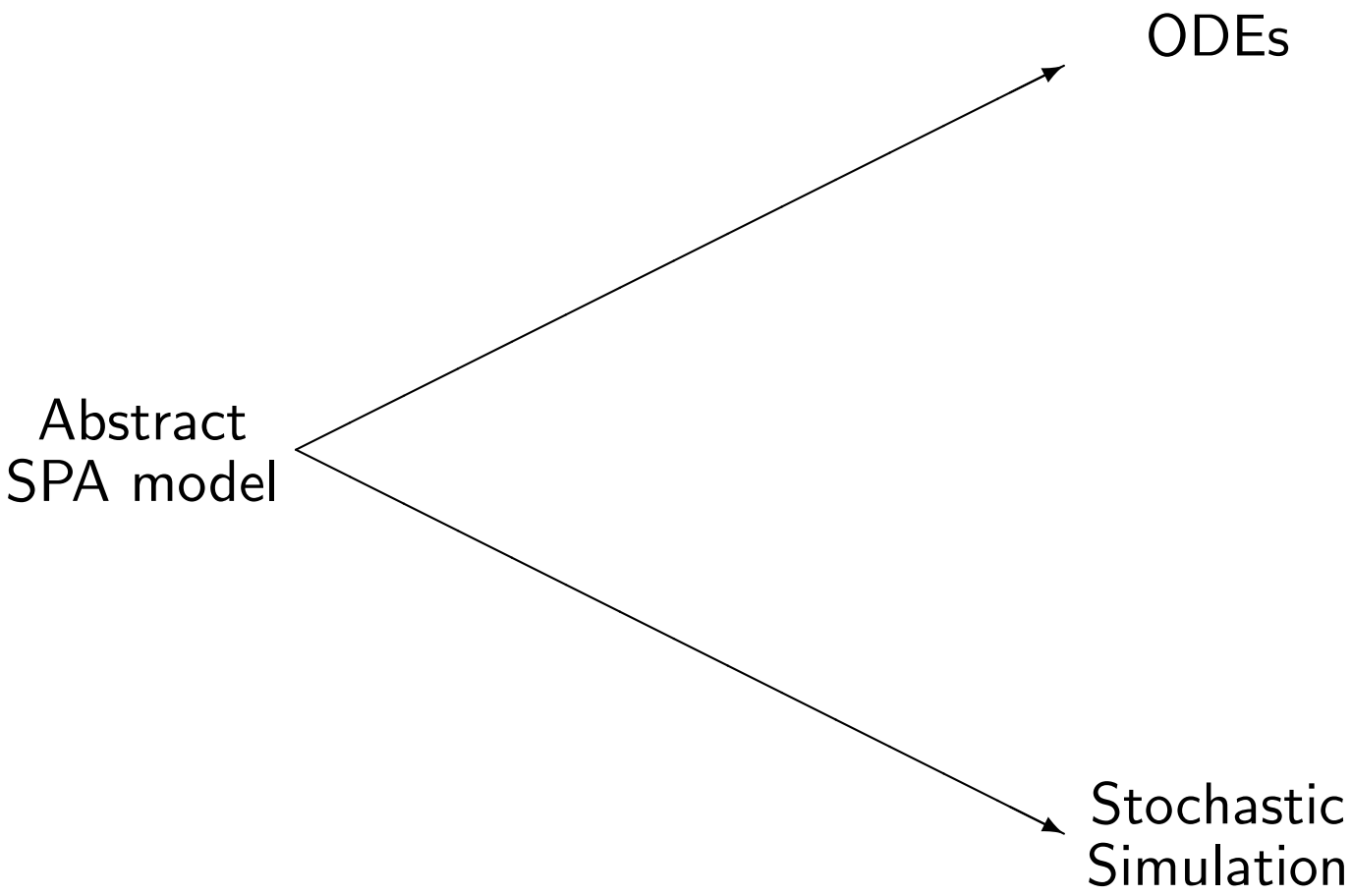
- Process algebra-based analyses such as comparing models (e.g. for equivalence or simulation) and model checking are only possible is the state space is not prohibitively large.
- The data that we have available to parameterise models is sometimes speculative rather than precise.

Motivations for Abstraction

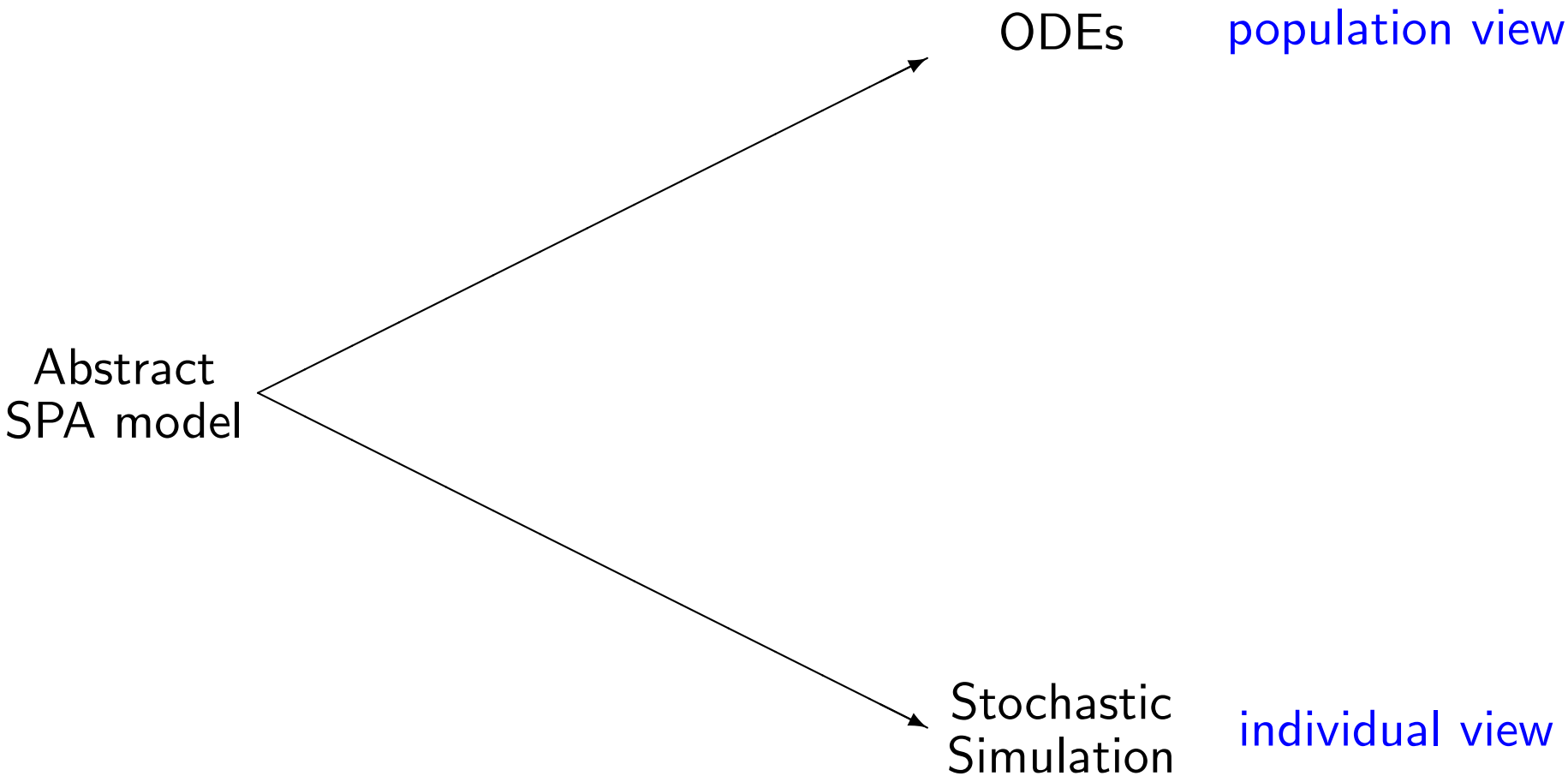
Our motivations for seeking more abstraction in process algebra models for systems biology are:

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

Alternative Representations



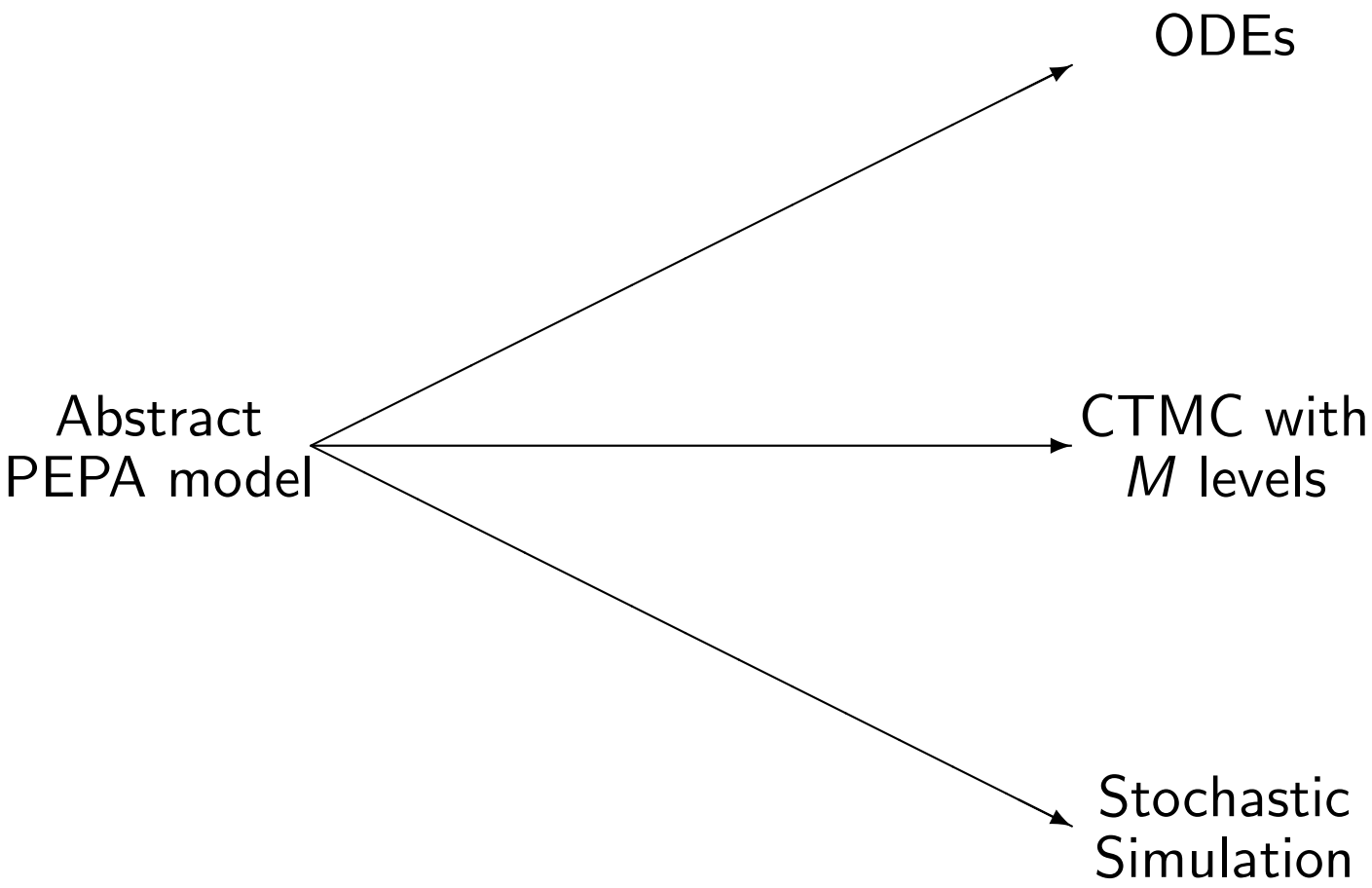
Alternative Representations



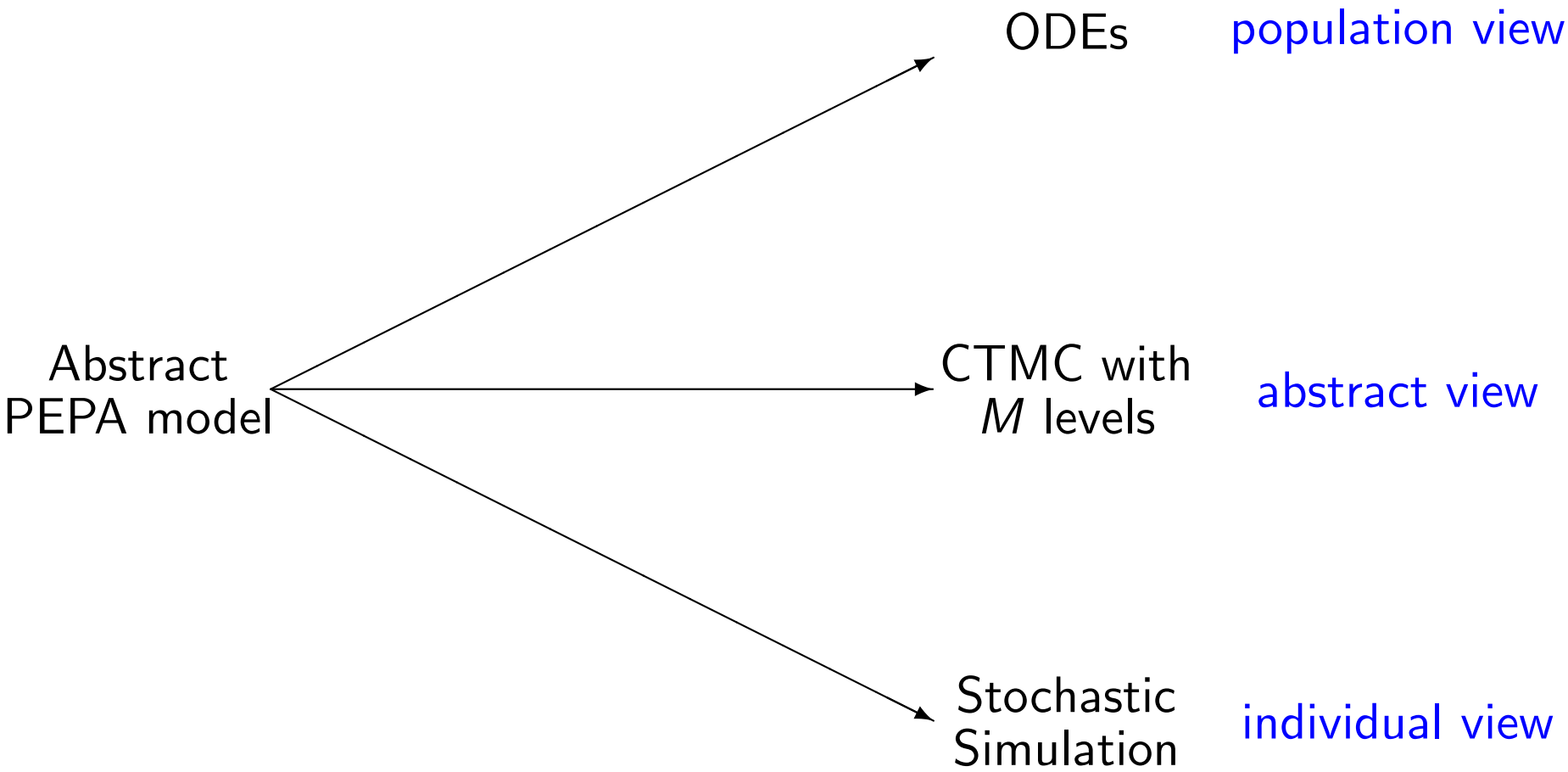
Discretising the population view

We can discretise the continuous range of possible concentration values into a number of distinct states. These form the possible states of the component representing the reagent.

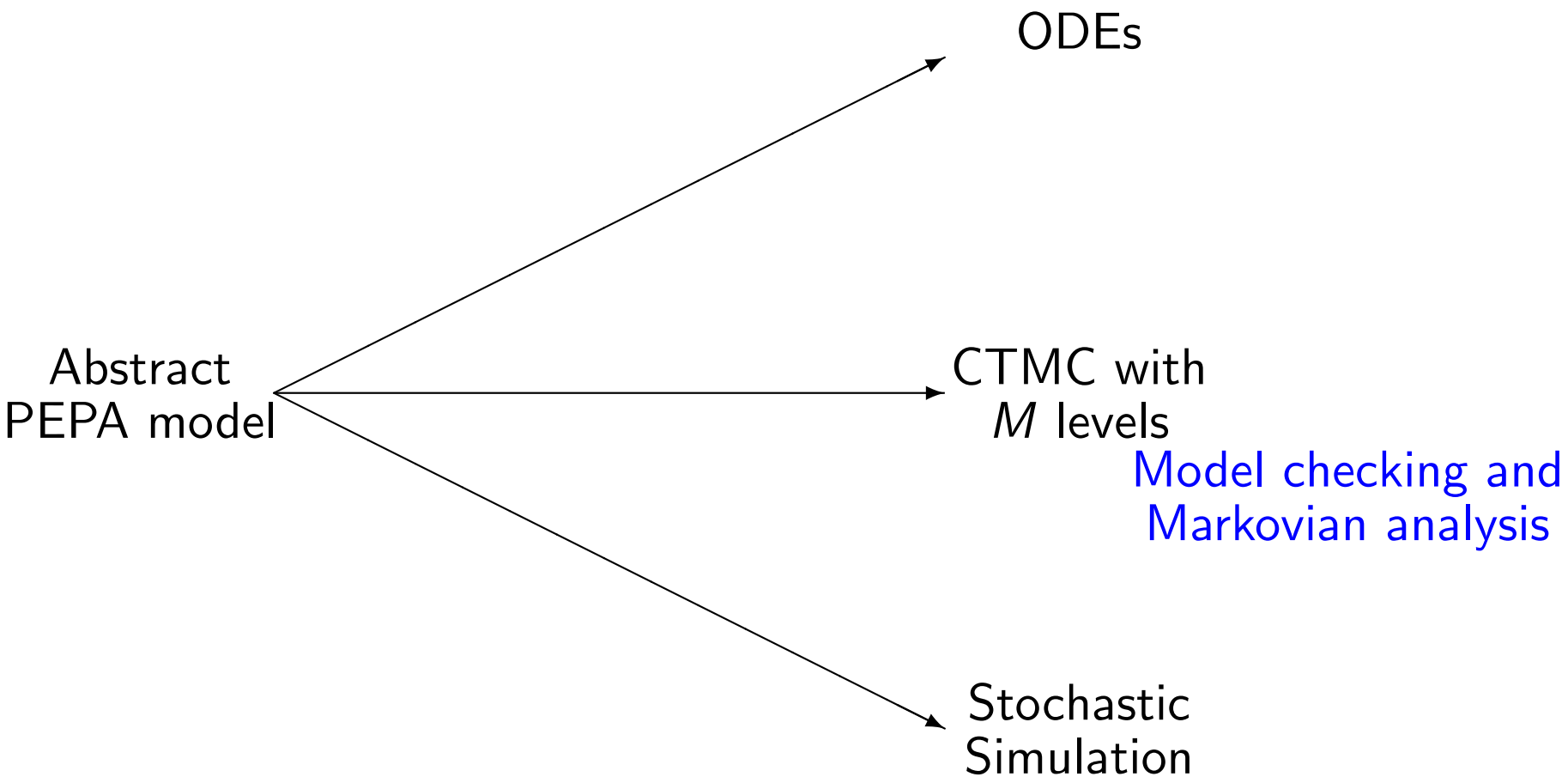
Alternative Representations



Alternative Representations



Alternative Representations



PEPA: Performance Evaluation Process Algebra

$$
\begin{aligned}
S &::= (\alpha, r).S \mid S + S \mid A \\
P &::= S \mid P \underset{L}{\bowtie} P \mid P/L
\end{aligned}
$$

PEPA: Performance Evaluation Process Algebra

$$
\begin{aligned}
S &::= (\alpha, r).S \mid S + S \mid A \\
P &::= S \mid P \underset{L}{\bowtie} P \mid P/L
\end{aligned}
$$

PEPA: Performance Evaluation Process Algebra

$$
\begin{aligned}
S &::= (\alpha, r).S \mid S + S \mid A \\
P &::= S \mid P \underset{L}{\bowtie} P \mid P/L
\end{aligned}
$$

PEPA: Performance Evaluation Process Algebra

$$
\begin{aligned}
S &::= (\alpha, r).S \mid S + S \mid A \\
P &::= S \mid P \underset{L}{\bowtie} P \mid P/L
\end{aligned}
$$

PEPA: Performance Evaluation Process Algebra

$$
\begin{aligned}
S &::= (\alpha, r).S \mid S + S \mid A \\
P &::= S \mid P \underset{L}{\bowtie} P \mid P/L
\end{aligned}
$$

PEPA: Performance Evaluation Process Algebra

$$
\begin{aligned}
S &::= (\alpha, r).S \mid S + S \mid A \\
P &::= S \mid P \underset{L}{\bowtie} P \mid P/L
\end{aligned}
$$

PEPA: Performance Evaluation Process Algebra

$$
\begin{aligned}
S &::= (\alpha, r).S \mid S + S \mid A \\
P &::= S \mid P \underset{L}{\bowtie} P \mid P/L
\end{aligned}
$$

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

PEPA: Performance Evaluation Process Algebra

$$
\begin{aligned}
S &::= (\alpha, r).S \mid S + S \mid A \\
P &::= S \mid P \underset{L}{\bowtie} P \mid P/L
\end{aligned}
$$

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

SPA
MODEL

PEPA: Performance Evaluation Process Algebra

$$\begin{aligned} S &::= (\alpha, r).S \mid S + S \mid A \\ P &::= S \mid P \underset{L}{\bowtie} P \mid P/L \end{aligned}$$

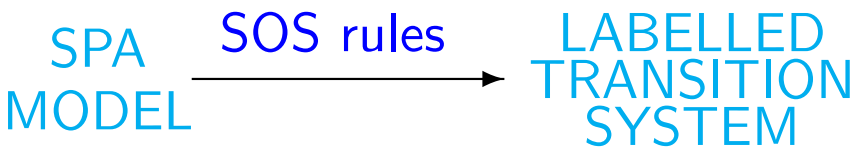
The language may be used to generate a Markov Process (CTMC).

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

PEPA: Performance Evaluation Process Algebra

$$
\begin{aligned}
S &::= (\alpha, r).S \mid S + S \mid A \\
P &::= S \mid P \underset{L}{\bowtie} P \mid P/L
\end{aligned}
$$

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



PEPA: Performance Evaluation Process Algebra

$$
\begin{aligned}
S &::= (\alpha, r).S \mid S + S \mid A \\
P &::= S \mid P \underset{L}{\bowtie} P \mid P/L
\end{aligned}
$$

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

PEPA: Performance Evaluation Process Algebra

$$
\begin{aligned}
S &::= (\alpha, r).S \mid S + S \mid A \\
P &::= S \mid P \underset{L}{\bowtie} P \mid P/L
\end{aligned}
$$

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

SPA MODEL $\xrightarrow{\text{SOS rules}}$ LABELLED TRANSITION SYSTEM $\xrightarrow[\text{diagram}]{\text{state transition}}$ CTMC **Q**

Q is the infinitesimal generator matrix characterising the CTMC.

PEPA: Performance Evaluation Process Algebra

$$
\begin{aligned}
S &::= (\alpha, r).S \mid S + S \mid A \\
P &::= S \mid P \underset{L}{\bowtie} P \mid P/L
\end{aligned}
$$

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

PEPA: Performance Evaluation Process Algebra

$$\begin{aligned} S &::= (\alpha, r).S \mid S + S \mid A \\ P &::= S \mid P \underset{L}{\bowtie} P \mid P/L \end{aligned}$$

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

SPA
MODEL

PEPA: Performance Evaluation Process Algebra

$$
\begin{aligned}
S &::= (\alpha, r).S \mid S + S \mid A \\
P &::= S \mid P \underset{L}{\bowtie} P \mid P/L
\end{aligned}
$$

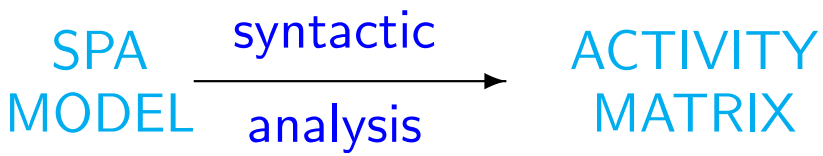
The language may be used to generate a system of ordinary differential equations (ODEs).

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

PEPA: Performance Evaluation Process Algebra

$$
\begin{aligned}
S &::= (\alpha, r).S \mid S + S \mid A \\
P &::= S \mid P \underset{L}{\bowtie} P \mid P/L
\end{aligned}
$$

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



PEPA: Performance Evaluation Process Algebra

$$
\begin{aligned}
S &::= (\alpha, r).S \mid S + S \mid A \\
P &::= S \mid P \underset{L}{\bowtie} P \mid P/L
\end{aligned}
$$

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

SPA MODEL $\xrightarrow[\text{analysis}]{\text{syntactic}}$ ACTIVITY MATRIX $\xrightarrow[\text{interpretation}]{\text{continuous}}$

PEPA: Performance Evaluation Process Algebra

$$
\begin{aligned}
S &::= (\alpha, r).S \mid S + S \mid A \\
P &::= S \mid P \underset{L}{\bowtie} P \mid P/L
\end{aligned}
$$

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

SPA MODEL $\xrightarrow[\text{analysis}]{\text{syntactic}}$ ACTIVITY MATRIX $\xrightarrow[\text{interpretation}]{\text{continuous}}$ ODEs

PEPA: Performance Evaluation Process Algebra

$$
\begin{aligned}
S &::= (\alpha, r).S \mid S + S \mid A \\
P &::= S \mid P \underset{L}{\bowtie} P \mid P/L
\end{aligned}
$$

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

PEPA: Performance Evaluation Process Algebra

$$\begin{aligned} S &::= (\alpha, r).S \mid S + S \mid A \\ P &::= S \mid P \underset{L}{\bowtie} P \mid P/L \end{aligned}$$

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

SPA
MODEL

PEPA: Performance Evaluation Process Algebra

$$\begin{aligned} S &::= (\alpha, r).S \mid S + S \mid A \\ P &::= S \mid P \underset{L}{\bowtie} P \mid P/L \end{aligned}$$

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

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

PEPA: Performance Evaluation Process Algebra

$$
\begin{aligned}
S &::= (\alpha, r).S \mid S + S \mid A \\
P &::= S \mid P \underset{L}{\bowtie} P \mid P/L
\end{aligned}
$$

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

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

PEPA: Performance Evaluation Process Algebra

$$
\begin{aligned}
S &::= (\alpha, r).S \mid S + S \mid A \\
P &::= S \mid P \underset{L}{\bowtie} P \mid P/L
\end{aligned}
$$

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

SPA MODEL $\xrightarrow[\text{analysis}]{\text{syntactic}}$ RATE EQUATIONS $\xrightarrow[\text{algorithm}]{\text{Gillespie's}}$

PEPA: Performance Evaluation Process Algebra

$$
\begin{aligned}
S &::= (\alpha, r).S \mid S + S \mid A \\
P &::= S \mid P \underset{L}{\bowtie} P \mid P/L
\end{aligned}
$$

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

PEPA: Performance Evaluation Process Algebra

$$
\begin{aligned}
S &::= (\alpha, r).S \mid S + S \mid A \\
P &::= S \mid P \underset{L}{\bowtie} P \mid P/L
\end{aligned}
$$

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.

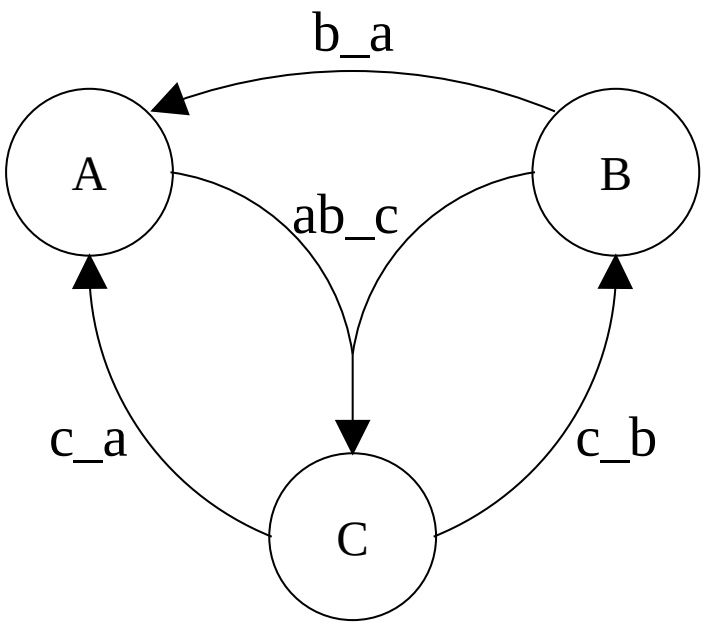
Markovian analysis

- 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.
- Stochastic model checking is available via the PRISM model checker, assessing the probable validity of properties expressed in CSL (Continuous Stochastic Logic).

Reagent-centric modelling [CGH04]

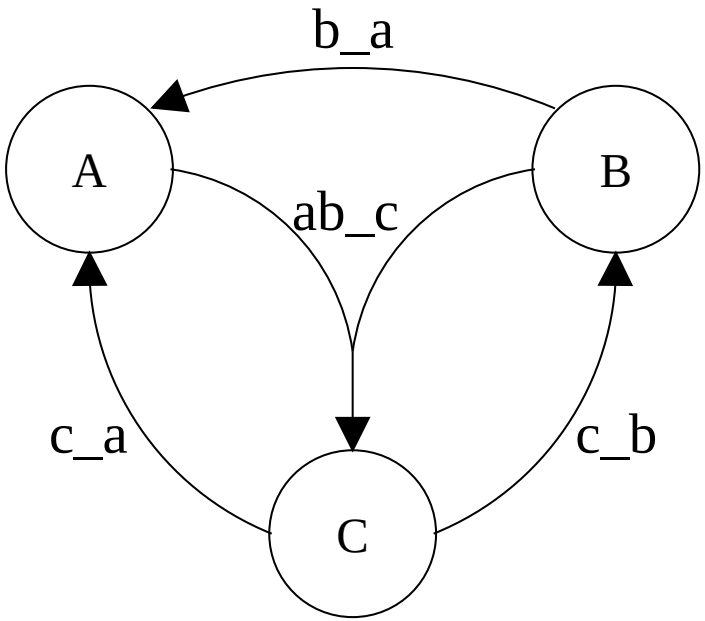
Reagent role	*Impact on reagent*	*Impact on reaction rate*
Producer	decreases concentration	has a positive impact, i.e. proportional to current concentration
Product	increases concentration	has no impact on the rate, except at saturation
Enzyme	concentration unchanged	has a positive impact, i.e. proportional to current concentration
Inhibitor	concentration unchanged	has a negative impact, i.e. inversely proportional to current concentration

PEPA reagent-centric example



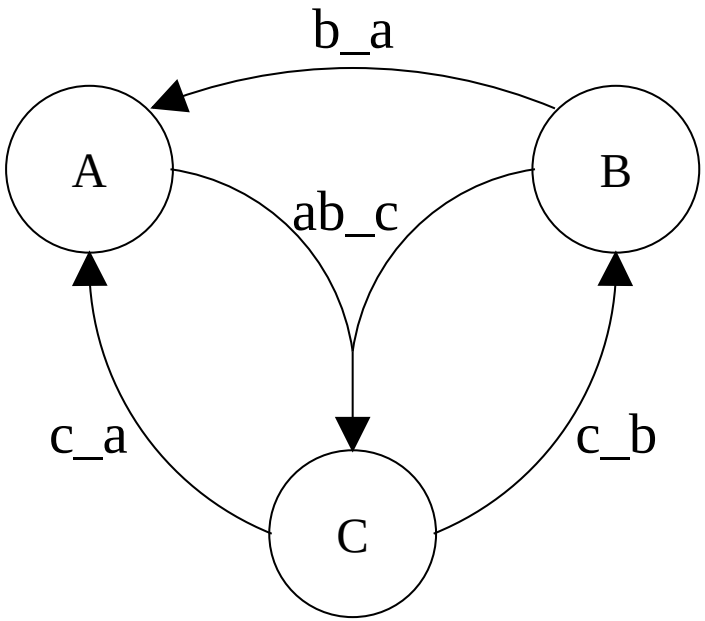
$$
\begin{aligned}
\mathrm{A_H} &\stackrel{def}{=} (ab\_c, \alpha).\mathrm{A_L} \\
\mathrm{A_L} &\stackrel{def}{=} (b\_a, \beta).\mathrm{A_H} + (c\_a, \gamma).\mathrm{A_H} \\
\mathrm{B_H} &\stackrel{def}{=} (ab\_c, \alpha).\mathrm{B_L} + (b\_a, \beta).\mathrm{B_L} \\
\mathrm{B_L} &\stackrel{def}{=} (c\_b, \delta).\mathrm{B_H} \\
\mathrm{C_H} &\stackrel{def}{=} (c\_a, \gamma).\mathrm{C_L} + (c\_b, \delta).\mathrm{C_L} \\
\mathrm{C_L} &\stackrel{def}{=} (ab\_c, \alpha).\mathrm{C_H}
\end{aligned}
$$

PEPA reagent-centric example



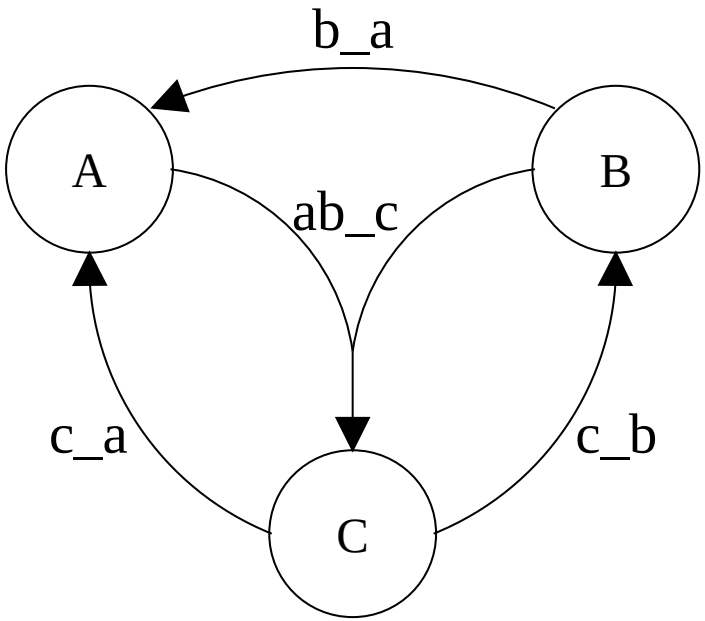
$$
\begin{aligned}
A_H &\stackrel{def}{=} (ab\_c, \alpha).A_L \\
A_L &\stackrel{def}{=} (b\_a, \beta).A_H + (c\_a, \gamma).A_H \\
B_H &\stackrel{def}{=} (ab\_c, \alpha).B_L + (b\_a, \beta).B_L \\
B_L &\stackrel{def}{=} (c\_b, \delta).B_H \\
C_H &\stackrel{def}{=} (c\_a, \gamma).C_L + (c\_b, \delta).C_L \\
C_L &\stackrel{def}{=} (ab\_c, \alpha).C_H
\end{aligned}
$$

PEPA reagent-centric example



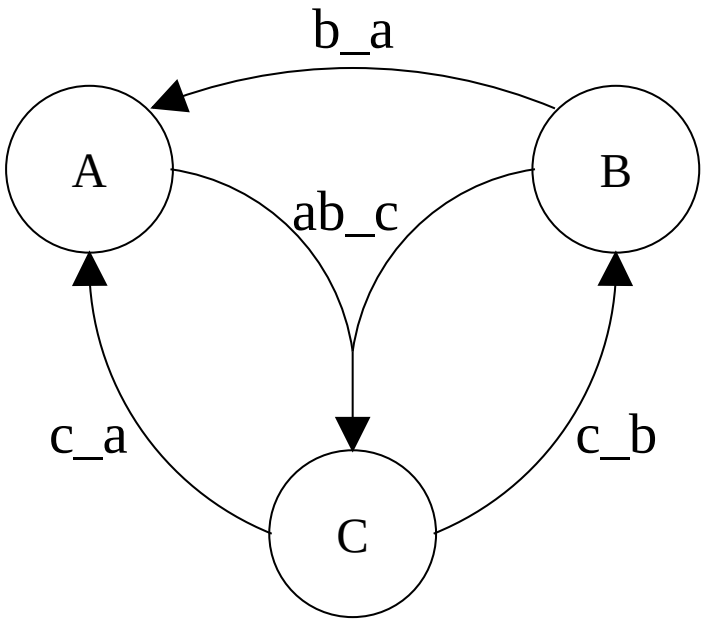
$$
\begin{aligned}
A_H &\stackrel{def}{=} (ab\_c, \alpha).A_L \\
A_L &\stackrel{def}{=} (b\_a, \beta).A_H + (c\_a, \gamma).A_H \\
B_H &\stackrel{def}{=} (ab\_c, \alpha).B_L + (b\_a, \beta).B_L \\
B_L &\stackrel{def}{=} (c\_b, \delta).B_H \\
C_H &\stackrel{def}{=} (c\_a, \gamma).C_L + (c\_b, \delta).C_L \\
C_L &\stackrel{def}{=} (ab\_c, \alpha).C_H
\end{aligned}
$$

PEPA reagent-centric example



$$
\begin{aligned}
A_H &\stackrel{def}{=} (ab\_c, \alpha).A_L \\
A_L &\stackrel{def}{=} (b\_a, \beta).A_H + (c\_a, \gamma).A_H \\
B_H &\stackrel{def}{=} (ab\_c, \alpha).B_L + (b\_a, \beta).B_L \\
B_L &\stackrel{def}{=} (c\_b, \delta).B_H \\
C_H &\stackrel{def}{=} (c\_a, \gamma).C_L + (c\_b, \delta).C_L \\
C_L &\stackrel{def}{=} (ab\_c, \alpha).C_H
\end{aligned}
$$

PEPA reagent-centric example



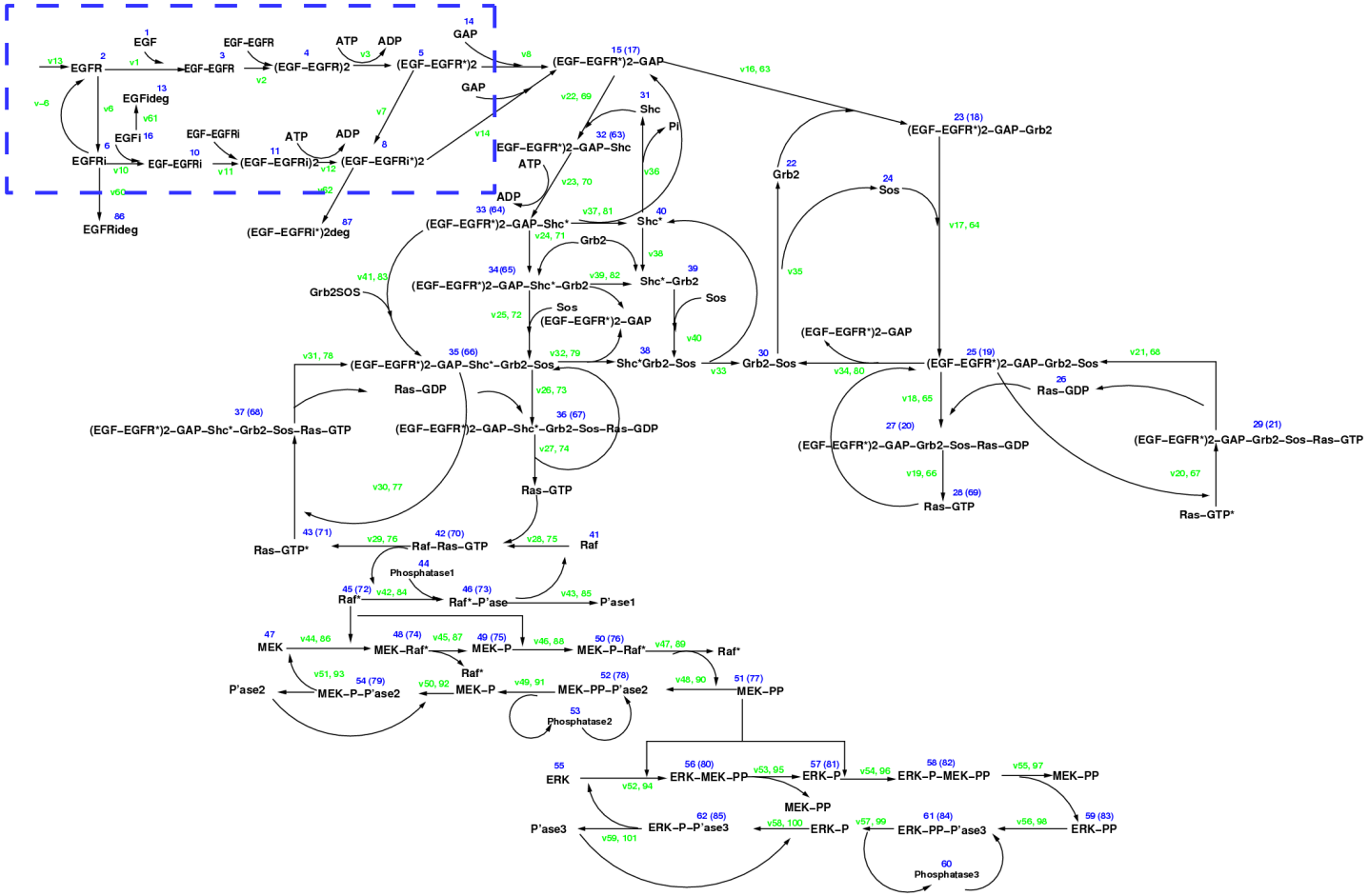
$$
\begin{aligned}
A_H &\stackrel{def}{=} (ab\_c, \alpha).A_L \\
A_L &\stackrel{def}{=} (b\_a, \beta).A_H + (c\_a, \gamma).A_H \\
B_H &\stackrel{def}{=} (ab\_c, \alpha).B_L + (b\_a, \beta).B_L \\
B_L &\stackrel{def}{=} (c\_b, \delta).B_H \\
C_H &\stackrel{def}{=} (c\_a, \gamma).C_L + (c\_b, \delta).C_L \\
C_L &\stackrel{def}{=} (ab\_c, \alpha).C_H
\end{aligned}
$$

$$(A_H \underset{\{ab\_c, b\_a\}}{\bowtie} B_H) \underset{\{ab\_c, c\_a, c\_b\}}{\bowtie} C_L$$

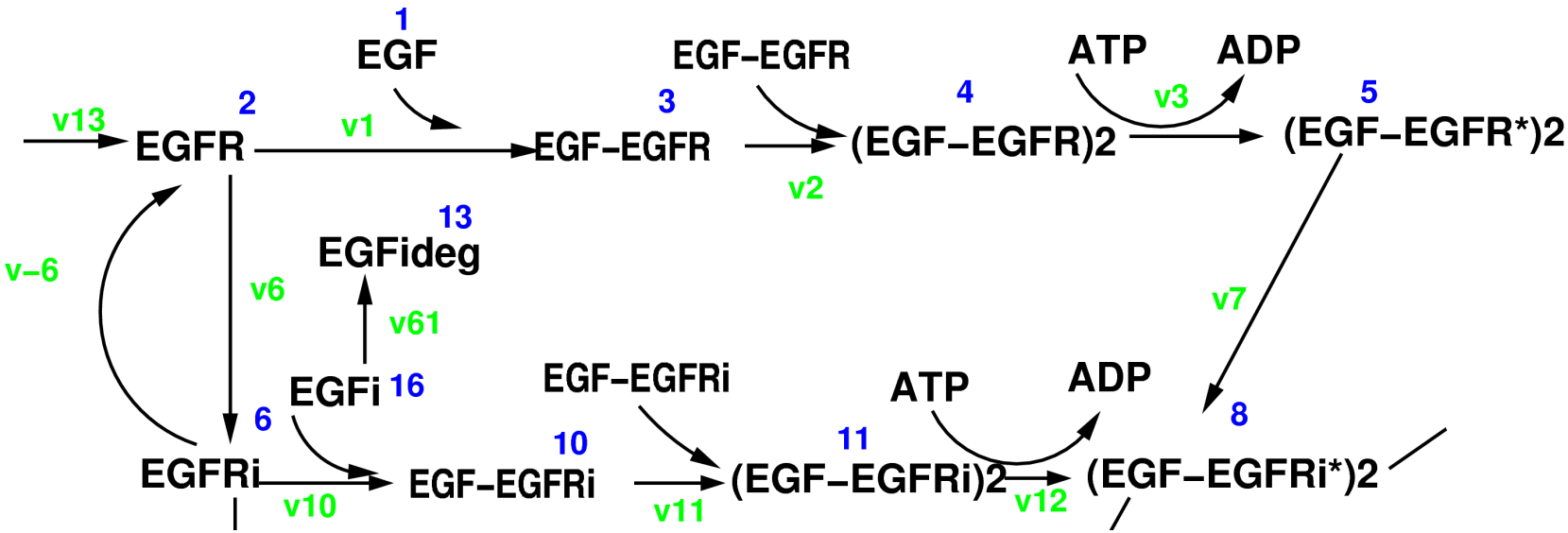
Case Study: Schoeberl *et al.*'s model of the MAPK Cascade [CDGH06]

- Published in *Nature Biotechnology* 20:370-375 in 2002.
- Influential, cited by more than 150 subsequent published papers.
- Consists of 94 reagent species involved in 125 reactions.
- Substantial ODE model consisting of 94 state variables and 95 parameters.
- Original model constructed "by hand", with help of a graphical representation.
- Original analysis based on numerical integration platform of the Matlab numerical computing platform.

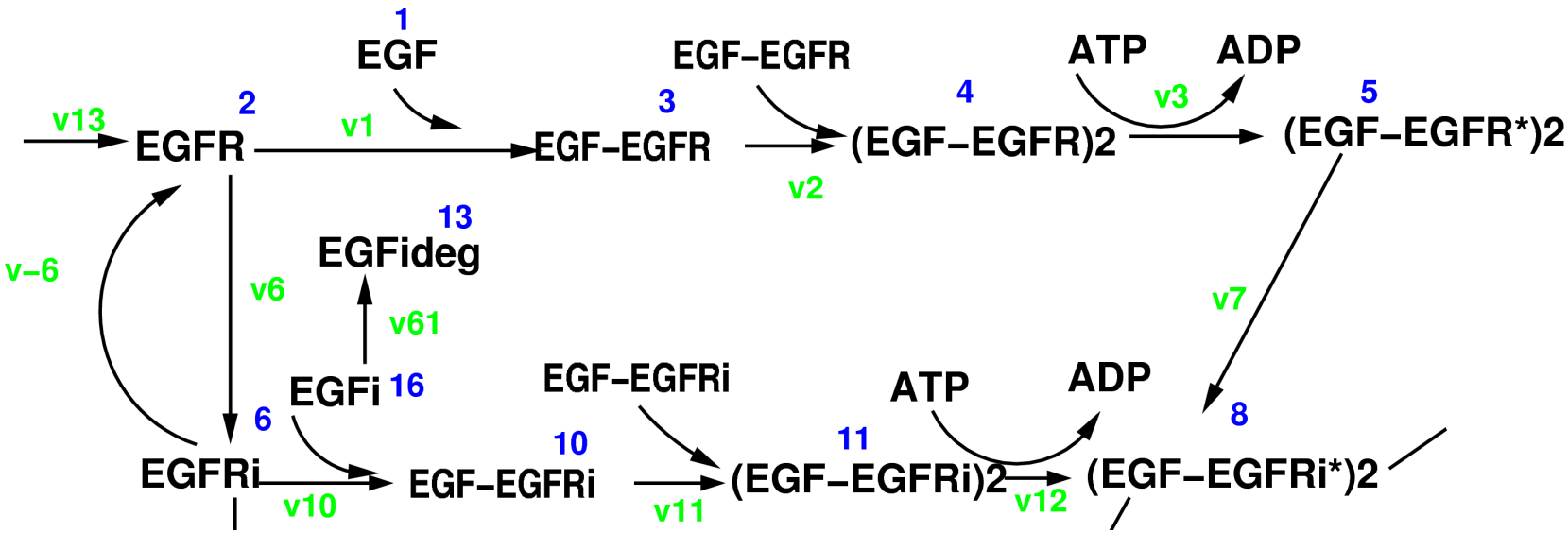
The MAP Kinase Cascade



The MAP Kinase Cascade

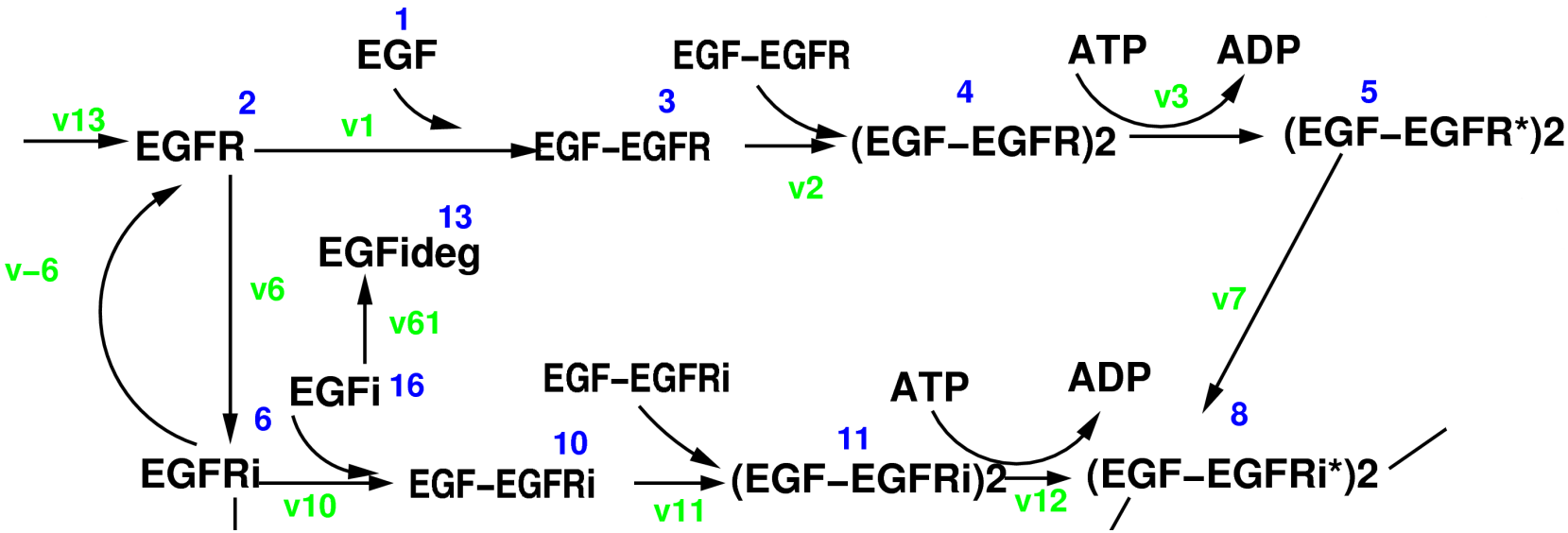


The MAP Kinase Cascade



There are many ambiguities in the graphical representation, e.g.

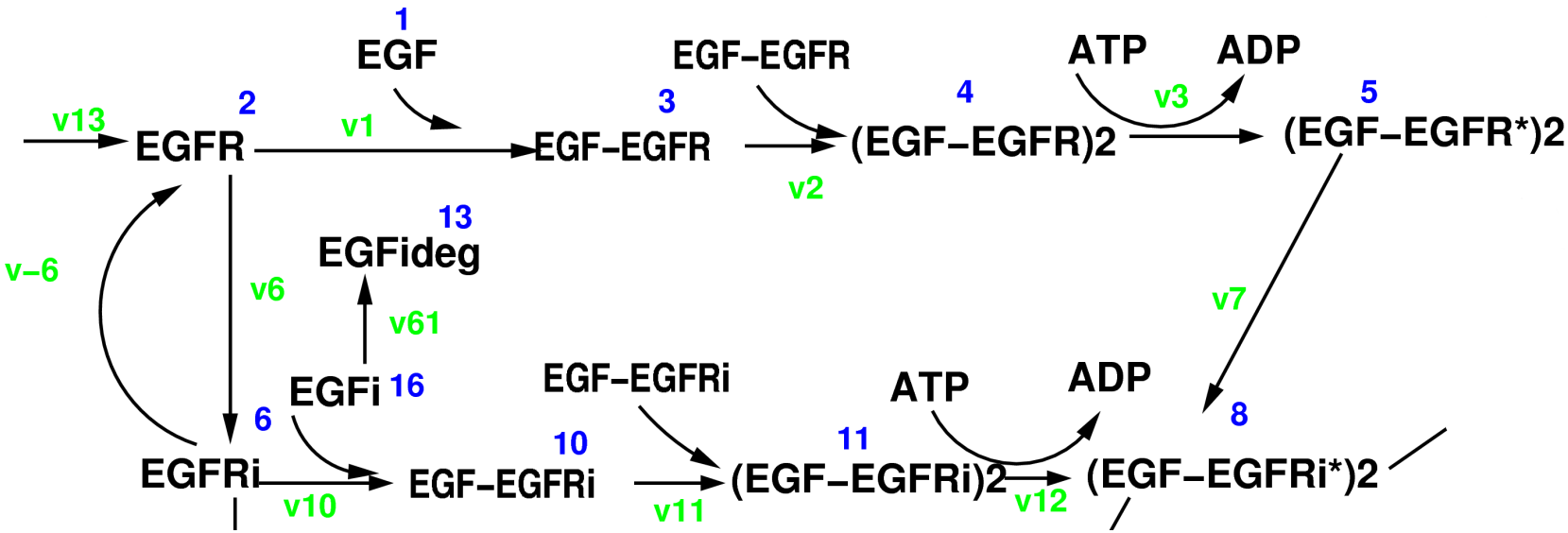
The MAP Kinase Cascade



There are many ambiguities in the graphical representation, e.g.

- An infinite supply of EGF is assumed;

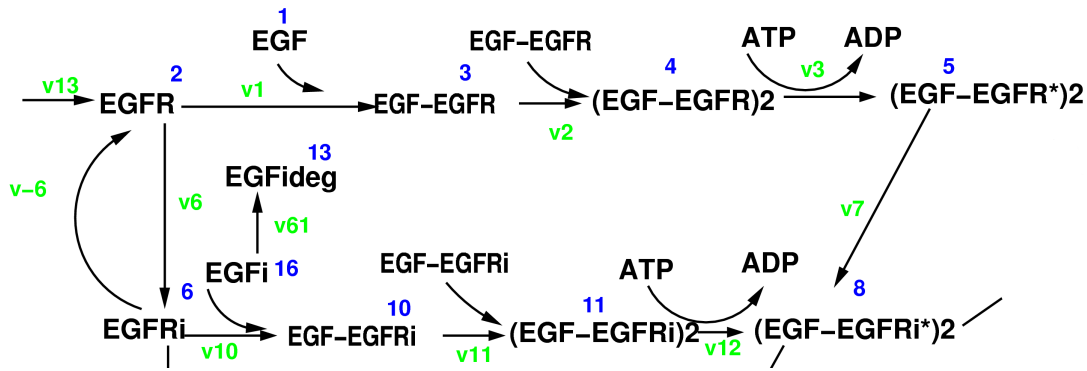
The MAP Kinase Cascade



There are many ambiguities in the graphical representation, e.g.

- An infinite supply of EGF is assumed;
- Reaction *v7* is uni-directional whereas all others are reversible.

Extracts from the model of the MAP Kinase Cascade



$$\mathrm{EGF_H} \stackrel{def}{=} (v_1, k_1).\mathrm{EGF_H}$$

$$\mathrm{EGFR_H} \stackrel{def}{=} (v_1, k_1).\mathrm{EGFR_L} + (v_6, k_6).\mathrm{EGFR_L}$$

$$\mathrm{EGFR_L} \stackrel{def}{=} (v_{\text{-}1},\ k_{\text{-}1}).\mathrm{EGFR_H} + (v_{\text{-}6},\ k_{\text{-}6}).\mathrm{EGFR_H} + (v_{13}, k_{13}).\mathrm{EGFR_H}$$

$$\text{EGF-EGFR}_\mathrm{H} \stackrel{def}{=} (v_2, k_2).\text{EGF-EGFR}_\mathrm{L} + (v_{\text{-}1}, k_{\text{-}1}).\text{EGF-EGFR}_\mathrm{L}$$

$$\text{EGF-EGFR}_\mathrm{L} \stackrel{def}{=} (v_1, k_1).\text{EGF-EGFR}_\mathrm{H} + (v_{\text{-}2}, k_{\text{-}2}).\text{EGF-EGFR}_\mathrm{H}$$

The PEPA model

Similar PEPA definitions were constructed for each of the 94 species in the pathway.

The PEPA model

Similar PEPA definitions were constructed for each of the 94 species in the pathway.

This was tedious, but not difficult, although care was needed to handle the points of ambiguity in the graphical representation.

The PEPA model

Similar PEPA definitions were constructed for each of the 94 species in the pathway.

This was tedious, but not difficult, although care was needed to handle the points of ambiguity in the graphical representation.

In order to complete the model we also needed to capture the interactions (i.e. cooperations) between the reagents. In this case we assumed that whenever reagents participated in reactions with the same name they did so in cooperation.

Validation of the PEPA model

- Once the PEPA model was constructed, we wanted to ensure that it was generating the same mathematical representation of the system.

Validation of the PEPA model

- Once the PEPA model was constructed, we wanted to ensure that it was generating the same mathematical representation of the system.
- In the first instance we derived a set of ODEs in a format suitable for Matlab.

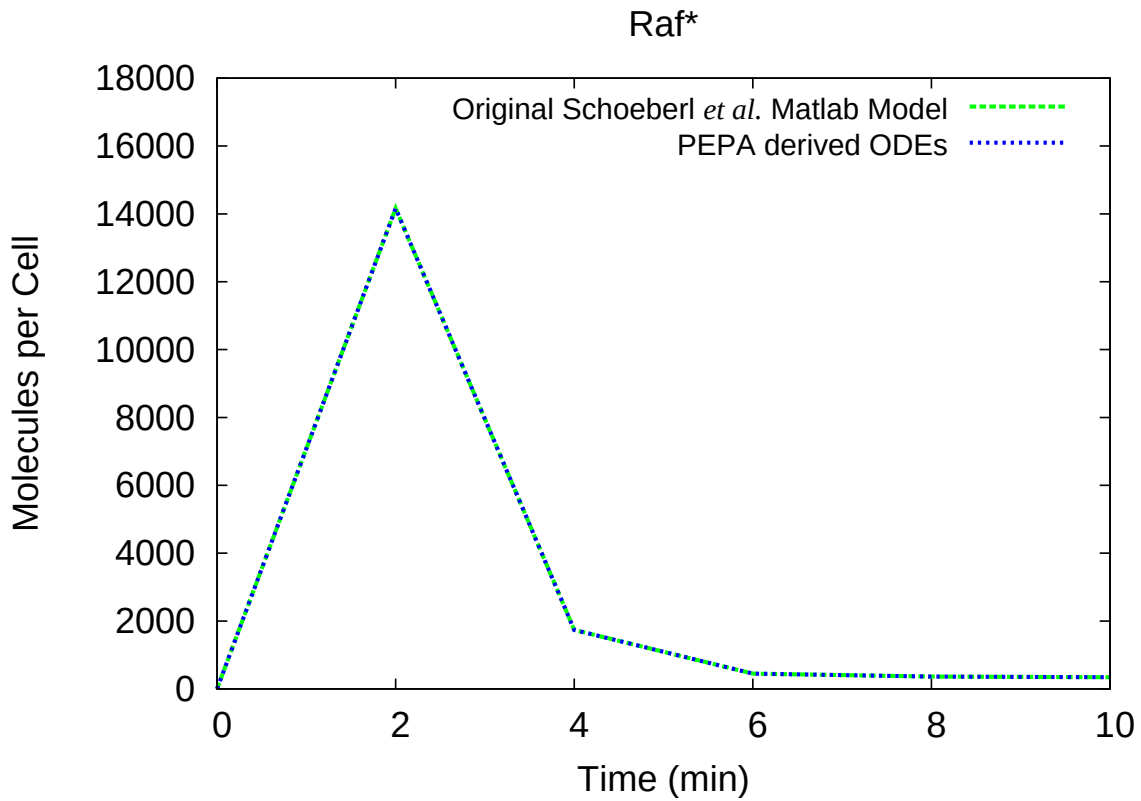
Validation of the PEPA model

- Once the PEPA model was constructed, we wanted to ensure that it was generating the same mathematical representation of the system.
- In the first instance we derived a set of ODEs in a format suitable for Matlab.
- These could not be compared directly with Schoeberl *et al*'s ODEs due to different representations being used, but we compared them empirically in terms of the results.

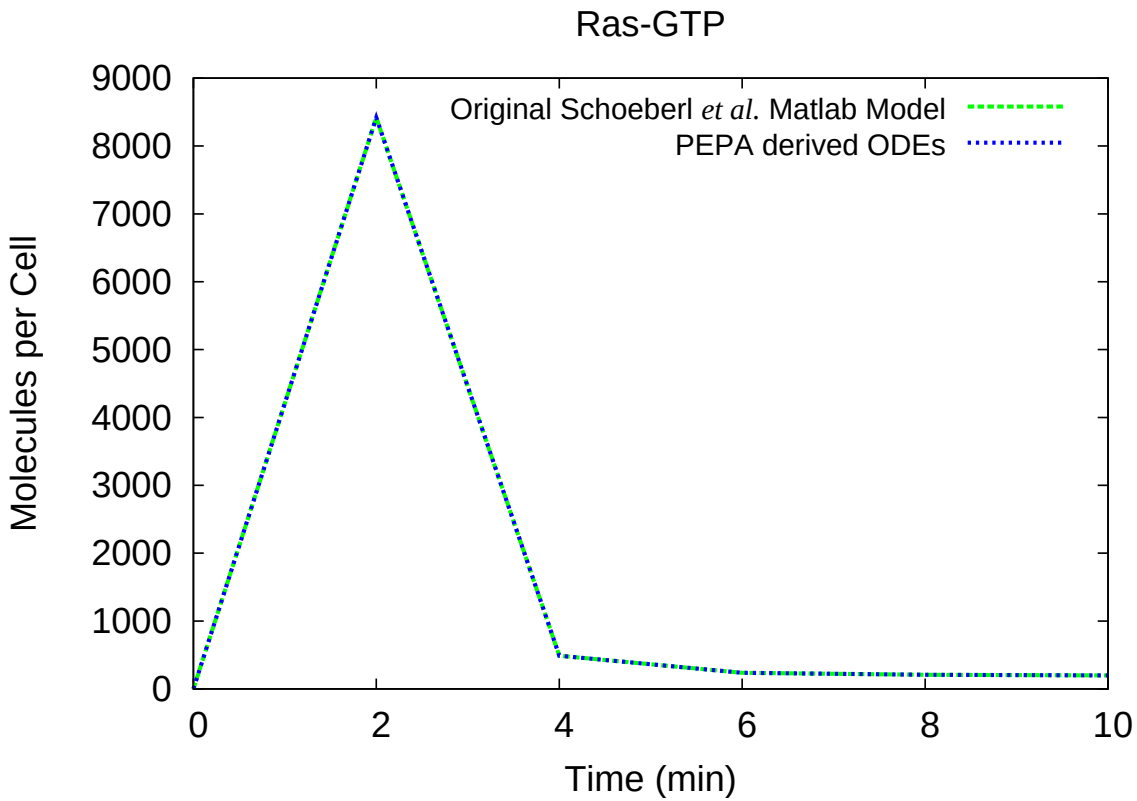
Validation of the PEPA model

- Once the PEPA model was constructed, we wanted to ensure that it was generating the same mathematical representation of the system.
- In the first instance we derived a set of ODEs in a format suitable for Matlab.
- These could not be compared directly with Schoeberl *et al*'s ODEs due to different representations being used, but we compared them empirically in terms of the results.
- Then we used an alternative mapping from the PEPA to generate a stochastic simulation of the system.

Comparing Original Results and PEPA Derived ODEs

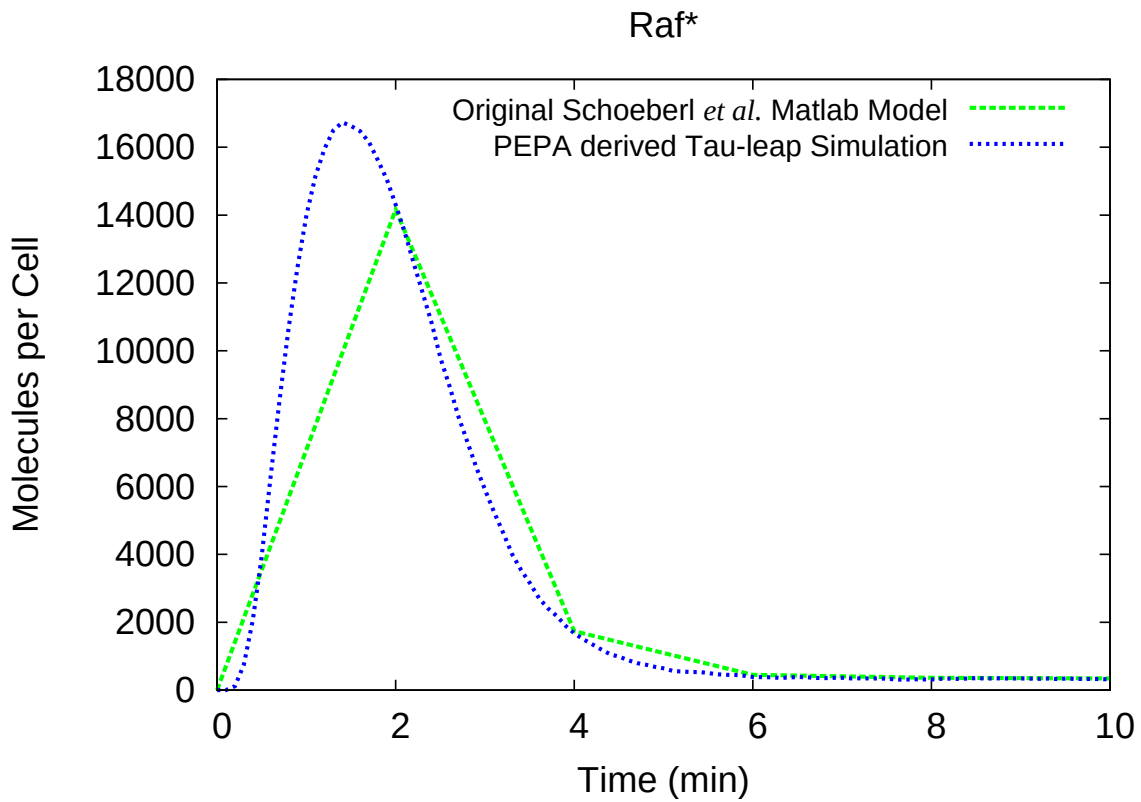


Comparing Original Results and PEPA Derived ODEs

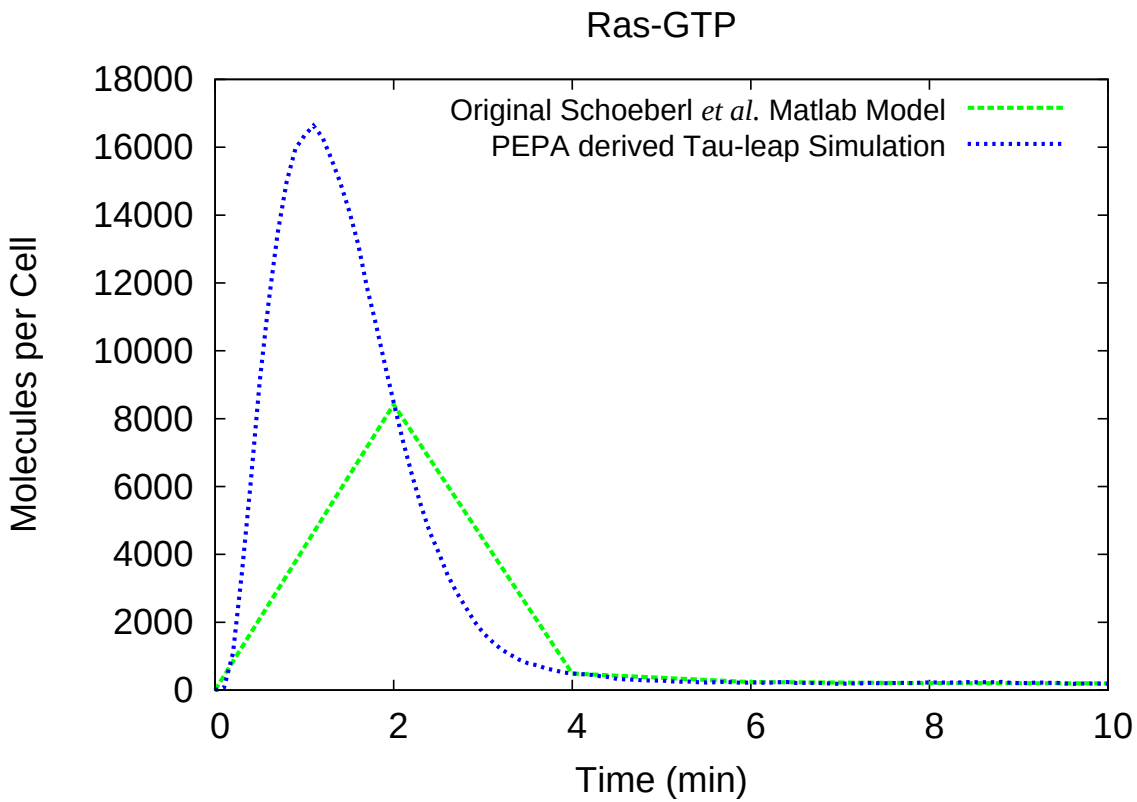


The PEPA derived ODEs return the same results as the Schoeberl *et al.* Matlab model.

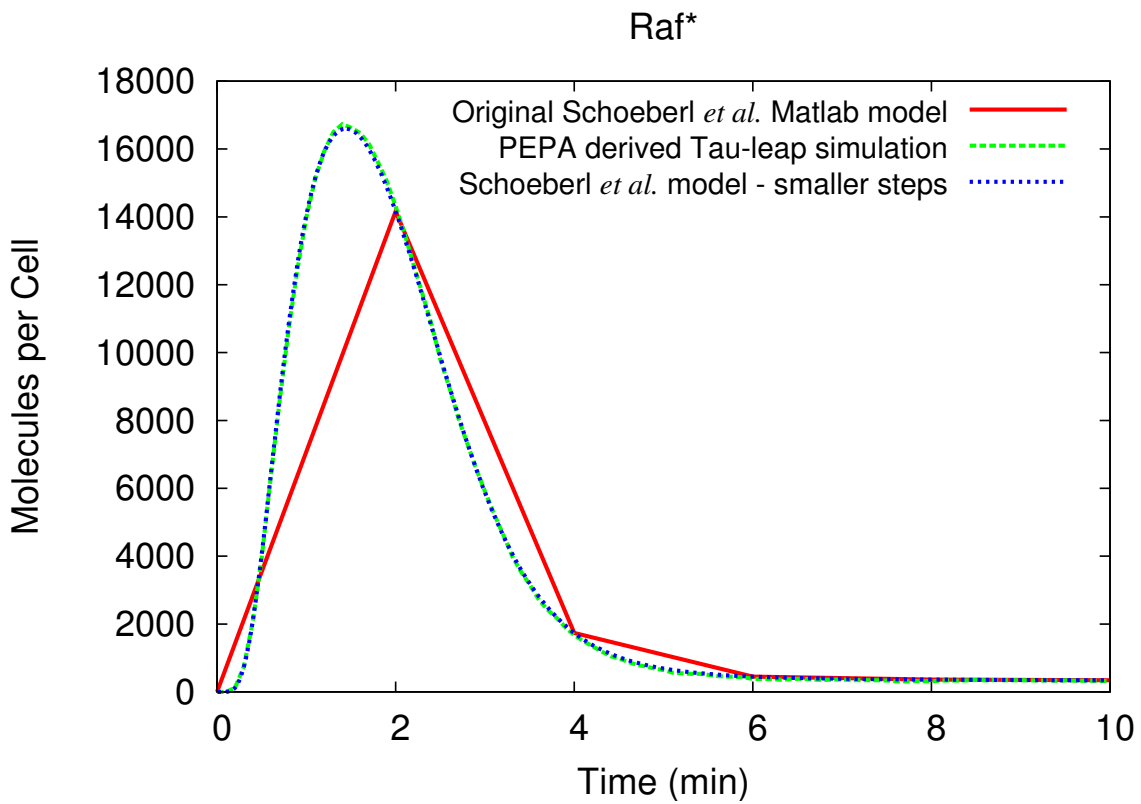
Comparing Original Results and PEPA Derived Stochastic Simulation



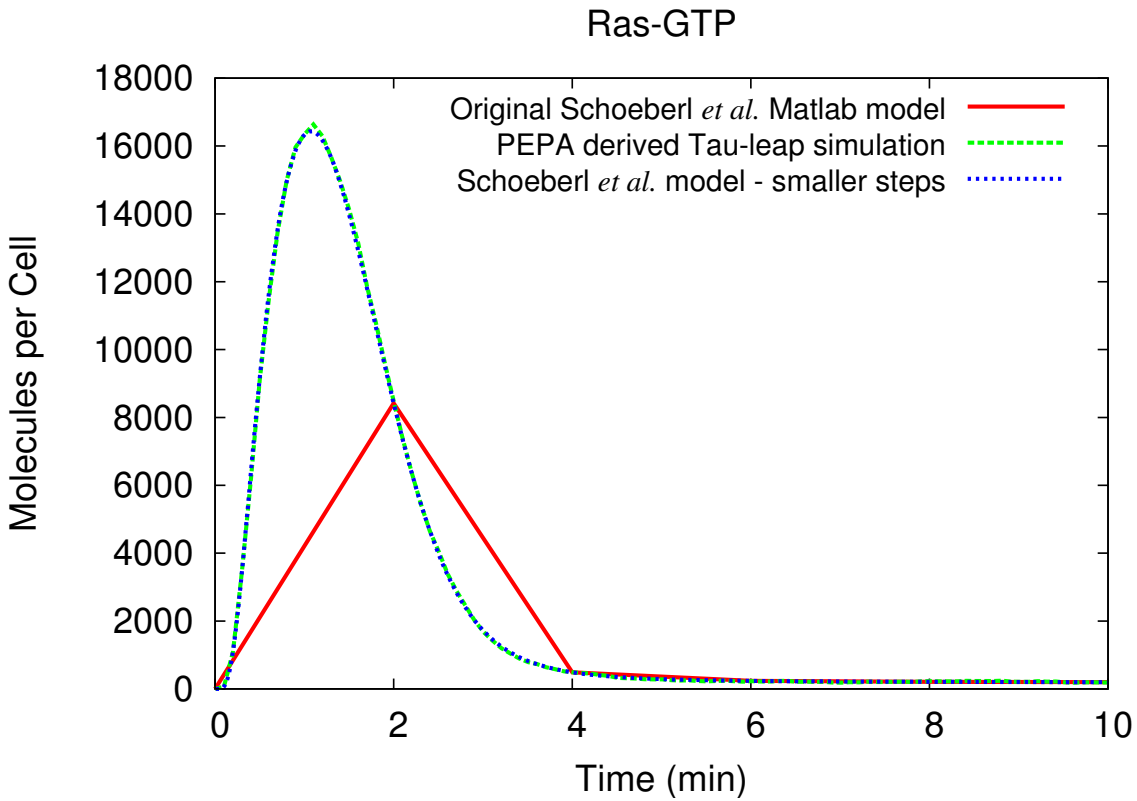
Comparing Original Results and PEPA Derived Stochastic Simulation



Corrected Time Step in Matlab Model



Corrected Time Step in Matlab Model



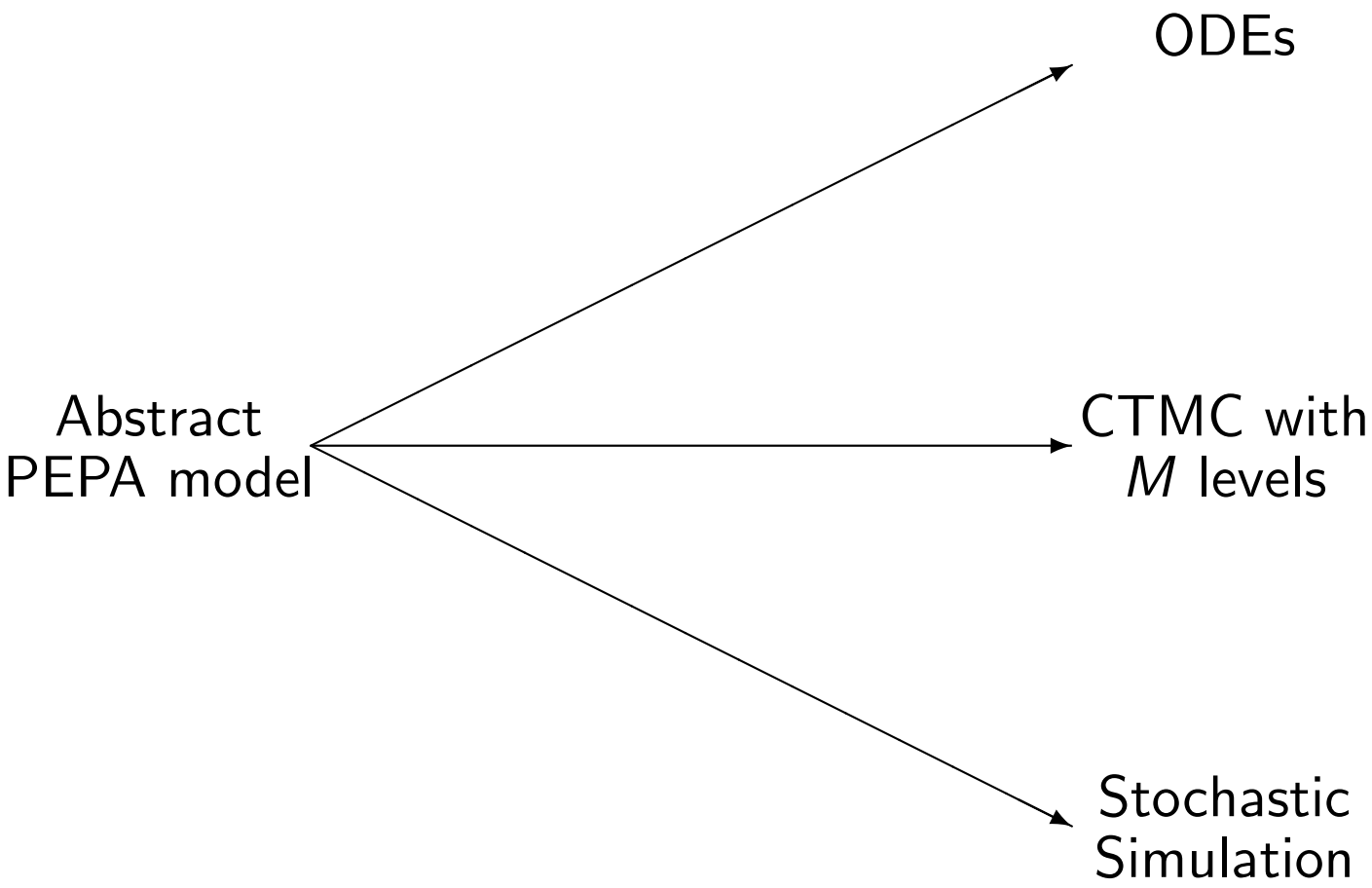
The original parameters for the Matlab model stepped over the true peak.

On-going work

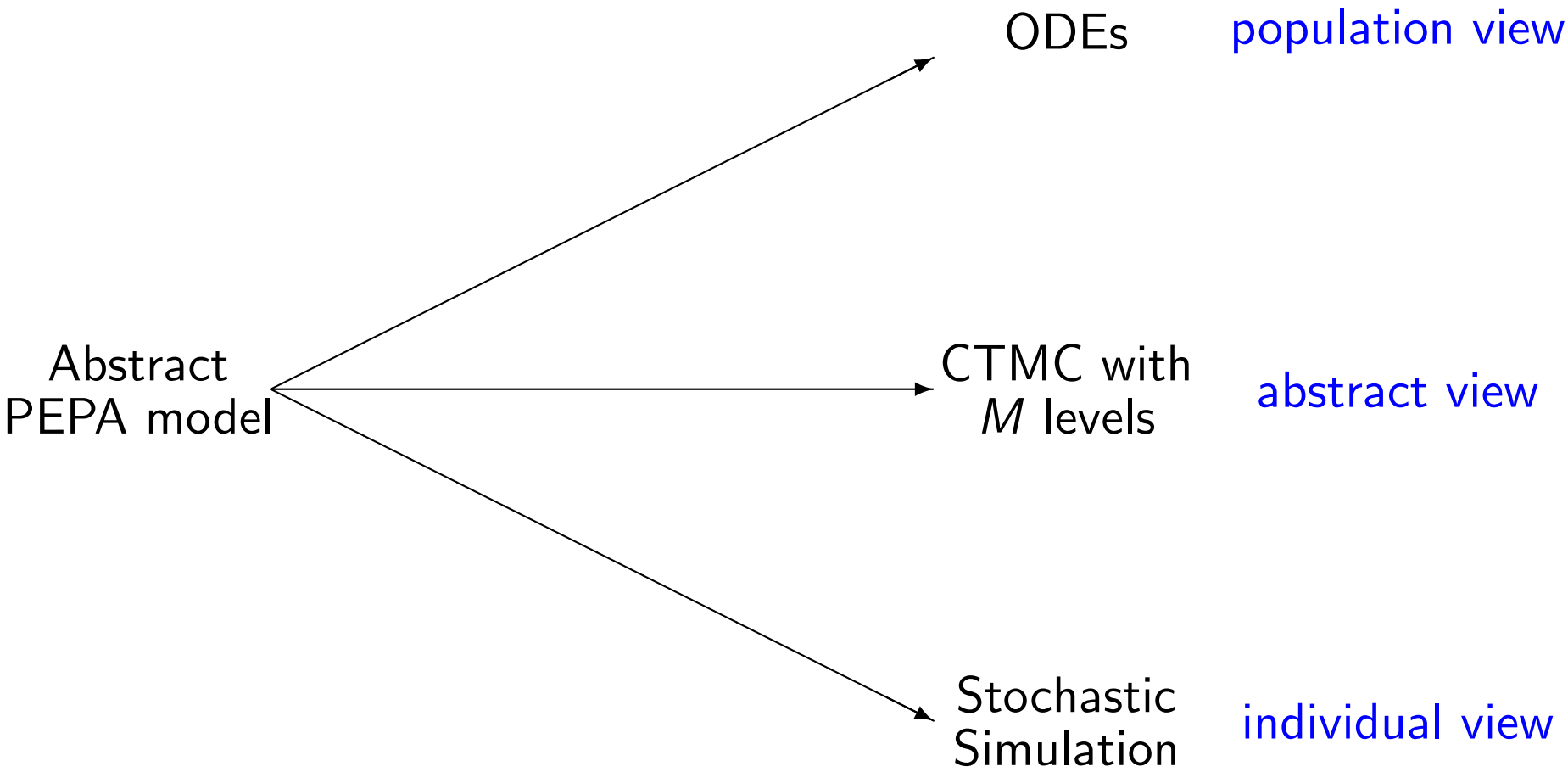
On-going work on this case study is working on a Markovian analysis of the system.

This involves developing the model to have multiple levels rather than the simple distinction between high and low which is all that is needed in order to generate the ODE and stochastic simulation models.

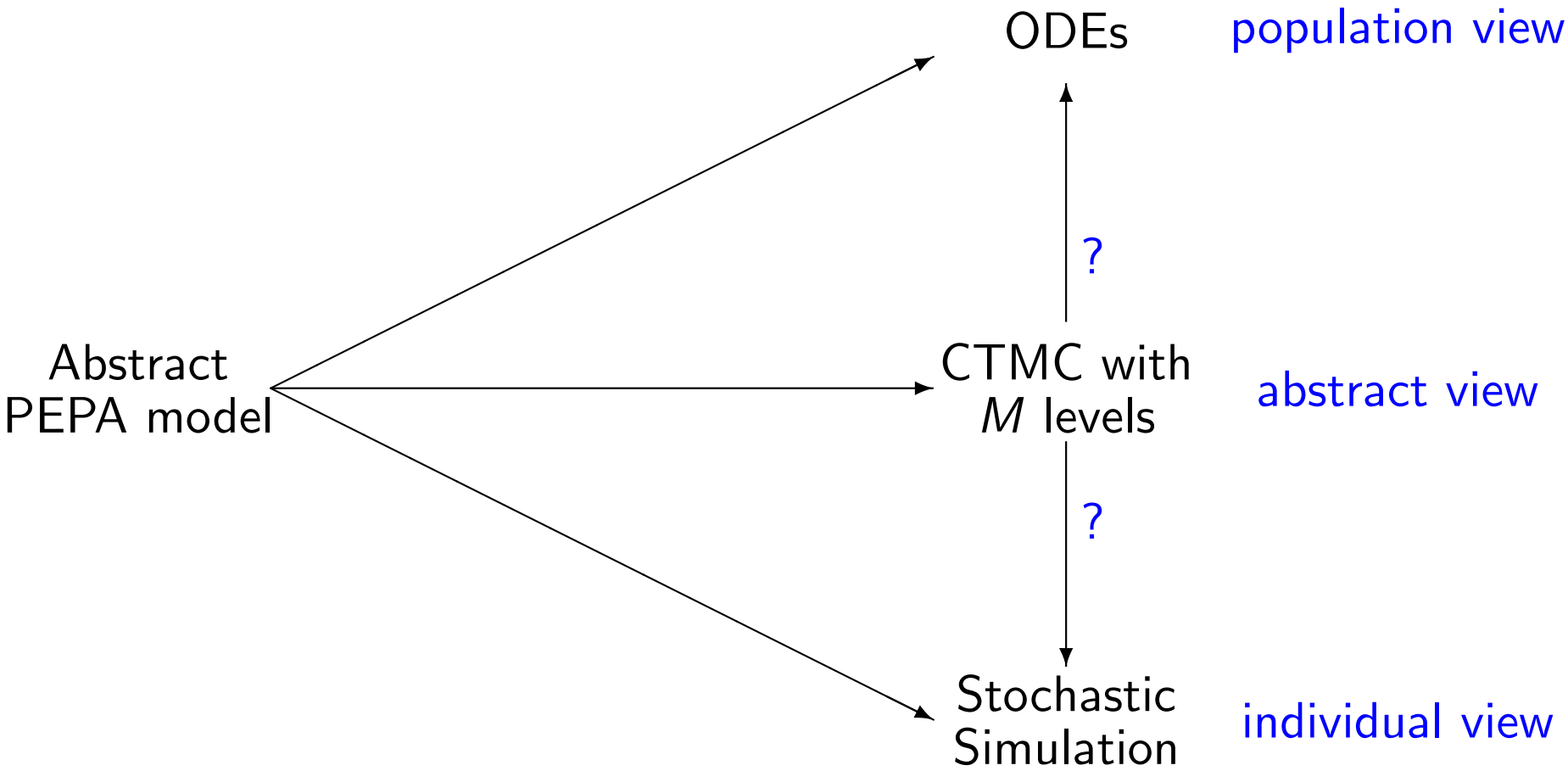
Equivalent Representations?



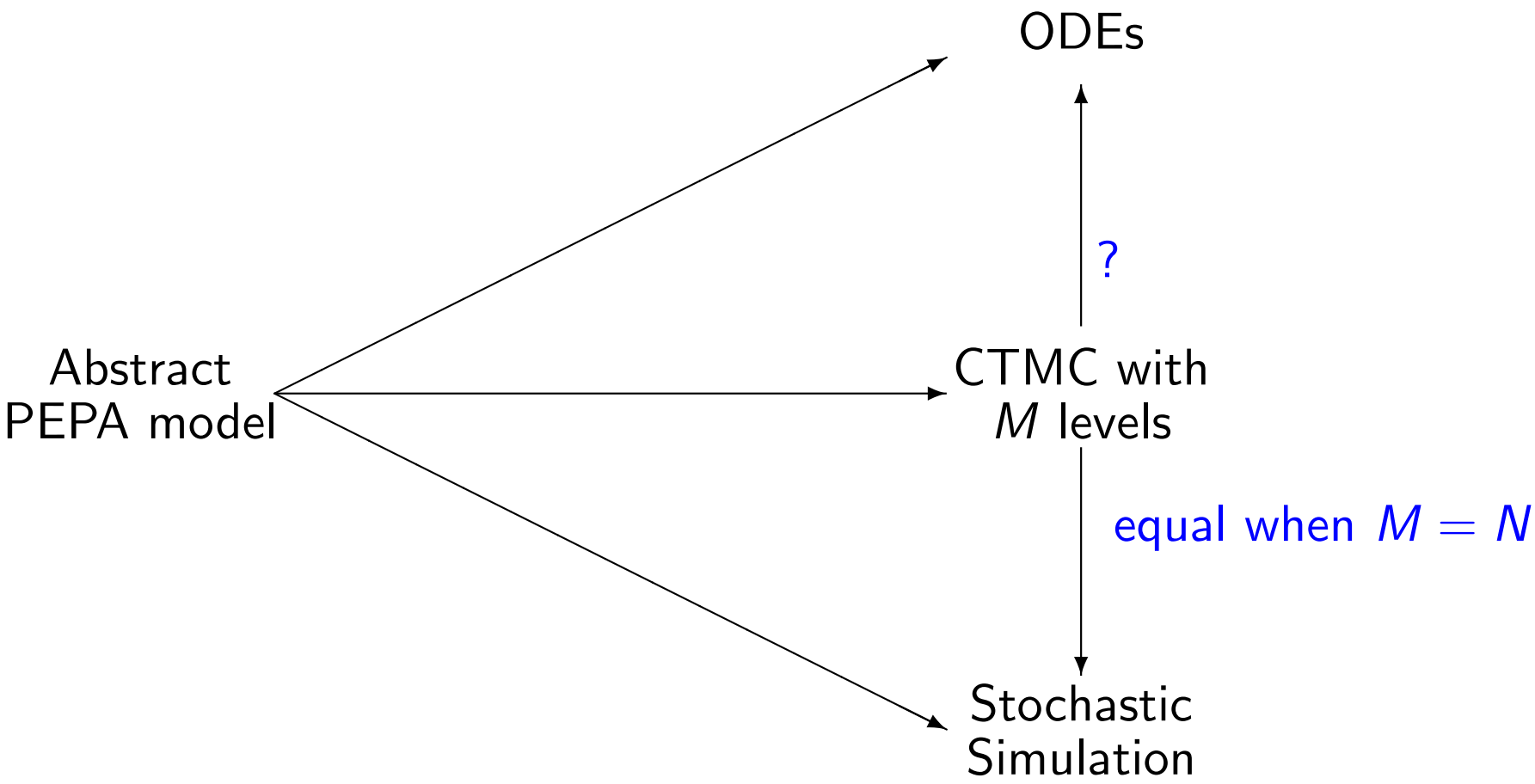
Equivalent Representations?



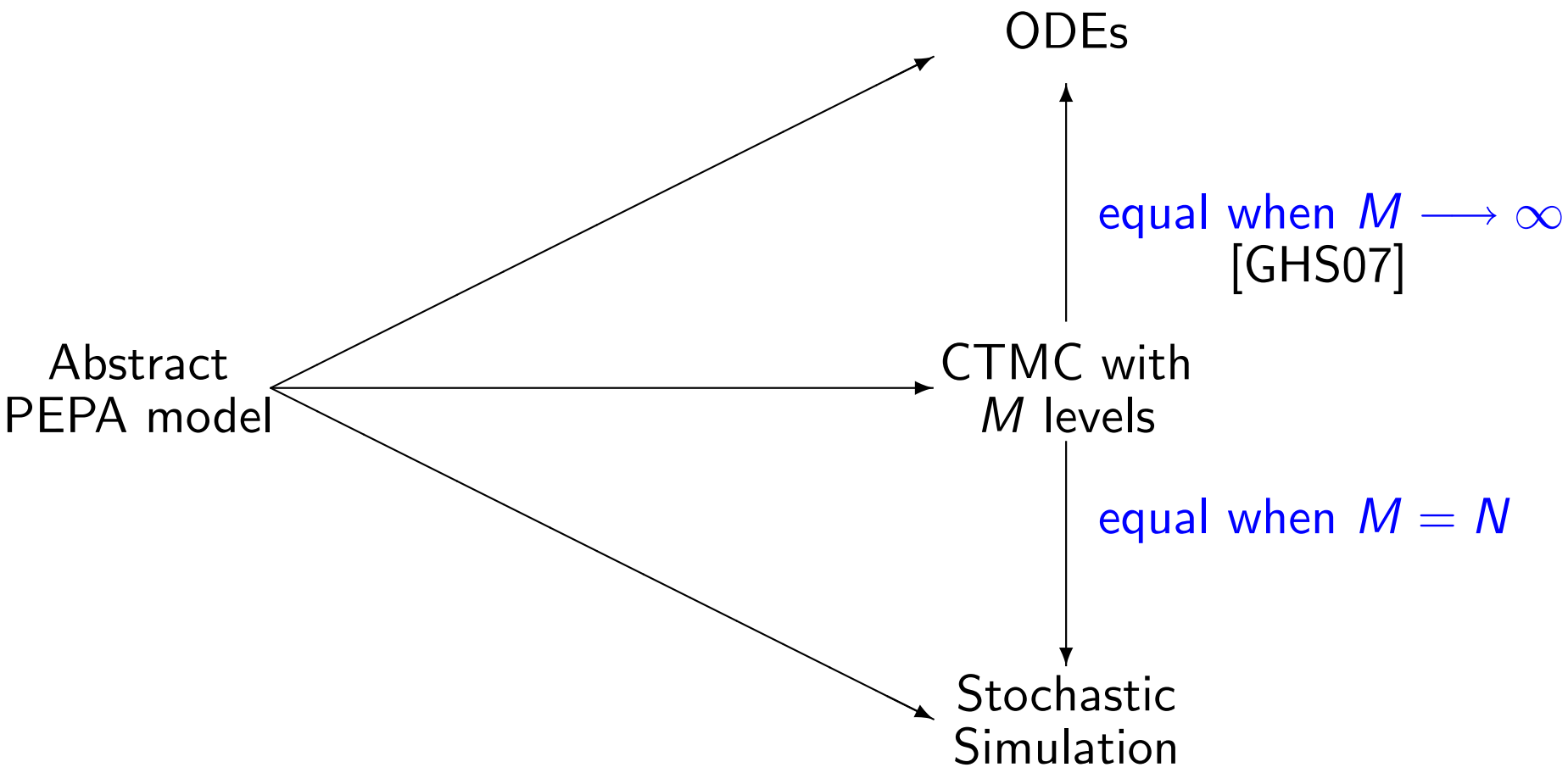
Equivalent Representations?



Equivalent Representations?



Equivalent Representations?



Relating CTMC and ODE models

- We consider an extension of PEPA, PEPA+. in which both bounded capacity and mass action kinetics are defined.

Relating CTMC and ODE models

- We consider an extension of PEPA, PEPA+. in which both bounded capacity and mass action kinetics are defined.
- We obtain a sequence of CTMCs as we consider models with finer and finer granularity — successively more levels in the SPA models.

Relating CTMC and ODE models

- We consider an extension of PEPA, PEPA+. in which both bounded capacity and mass action kinetics are defined.
- We obtain a sequence of CTMCs as we consider models with finer and finer granularity — successively more levels in the SPA models.
- Kurtz's theorem states that a sequence of pure jump Markov processes converge to a limit which coincides with a set of ODEs [Kurtz 70].

Relating CTMC and ODE models

- We consider an extension of PEPA, PEPA+. in which both bounded capacity and mass action kinetics are defined.
- We obtain a sequence of CTMCs as we consider models with finer and finer granularity — successively more levels in the SPA models.
- Kurtz's theorem states that a sequence of pure jump Markov processes converge to a limit which coincides with a set of ODEs [Kurtz 70]. In particular this holds for a class of CTMCs which are density dependent.

Relating CTMC and ODE models

- We consider an extension of PEPA, PEPA+. in which both bounded capacity and mass action kinetics are defined.
- We obtain a sequence of CTMCs as we consider models with finer and finer granularity — successively more levels in the SPA models.
- Kurtz's theorem states that a sequence of pure jump Markov processes converge to a limit which coincides with a set of ODEs [Kurtz 70]. In particular this holds for a class of CTMCs which are density dependent.
- We show that the CTMCs we construct from the PEPA+ models are density dependent and so satisfy Kurtz's theorem.

Density Dependent CTMC

A family of CTMCs is called density dependent if and only if there exists a continuous function $f(x, l), x \in \mathbb{R}^h, l \in \mathbb{Z}^h$, such that the infinitesimal generators of X_N are given by:

$$q_{k,k+l} = N\, f\left(\frac{k}{N}, l\right), \quad l \neq 0$$

where

Density Dependent CTMC

A family of CTMCs is called density dependent if and only if there exists a continuous function $f(x, l), x \in \mathbb{R}^h, l \in \mathbb{Z}^h$, such that the infinitesimal generators of X_N are given by:

$$q_{k,k+l} = N\, f\left(\frac{k}{N}, l\right), \quad l \neq 0$$

where

- $q_{k,k+1}$ denotes an entry in the infinitesimal generator matrix;

Density Dependent CTMC

A family of CTMCs is called density dependent if and only if there exists a continuous function $f(x, l), x \in \mathbb{R}^h, l \in \mathbb{Z}^h$, such that the infinitesimal generators of X_N are given by:

$$q_{k,k+l} = N\, f\left(\frac{k}{N}, l\right), \quad l \neq 0$$

where

- $q_{k,k+1}$ denotes an entry in the infinitesimal generator matrix;
- k is a numerical state vector and

Density Dependent CTMC

A family of CTMCs is called density dependent if and only if there exists a continuous function $f(x, l), x \in \mathbb{R}^h, l \in \mathbb{Z}^h$, such that the infinitesimal generators of X_N are given by:

$$q_{k,k+l} = N\, f\left(\frac{k}{N}, l\right), \quad l \neq 0$$

where

- $q_{k,k+1}$ denotes an entry in the infinitesimal generator matrix;
- k is a numerical state vector and
- l is a transition vector i.e. it records the adjustment to the number of copies of each state of each entity (species) after the transition is taken.

Outline

Summary

- Abstract modelling offers a compromise between the individual-based and population-based views of systems which biologists commonly take.

Summary

- Abstract modelling offers a compromise between the individual-based and population-based views of systems which biologists commonly take.
- Moveover we can undertake additional analysis based on the discretised population view.

Summary

- Abstract modelling offers a compromise between the individual-based and population-based views of systems which biologists commonly take.
- Moveover we can undertake additional analysis based on the discretised population view.
- Further work is needed to establish a better relationship between this view and the population view — empirical evidence has shown that 6 or 7 levels are often sufficient to capture exactly the same behaviour as the ODE model.

Summary

- Abstract modelling offers a compromise between the individual-based and population-based views of systems which biologists commonly take.
- Moveover we can undertake additional analysis based on the discretised population view.
- Further work is needed to establish a better relationship between this view and the population view — empirical evidence has shown that 6 or 7 levels are often sufficient to capture exactly the same behaviour as the ODE model.
- In the future we hope to investigate the extent to which the process algebra compositional structure can be exploited during model analysis.

Challenges

- The issue of unknown and uncertain data remains to be addressed.

Challenges

- The issue of unknown and uncertain data remains to be addressed.
- The abstract Markovian models allow quantities of interest such as "response times" to be expressed as probability distributions rather than single estimates. This may allow better reflection of wet lab data which showns variability.

Challenges

- The issue of unknown and uncertain data remains to be addressed.
- The abstract Markovian models allow quantities of interest such as "response times" to be expressed as probability distributions rather than single estimates. This may allow better reflection of wet lab data which showns variability.
- Promising recent work by Girolami *et al.* on assessing candidate models which attempt to cover both unknown structure and unknown kinetic rates with respect to experimental data, using Bayesian reasoning.

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.

Thank You!

Thank You!

Collaborators: Muffy Calder, Federica Ciocchetta, Adam Duguid, Nil Geisweiller, Stephen Gilmore and Marco Stenico.

Thank You!

Collaborators: Muffy Calder, Federica Ciocchetta, Adam Duguid, Nil Geisweiller, Stephen Gilmore and Marco Stenico.

Acknowledgements: Engineering and Physical Sciences Research Council (EPSRC) and Biotechnology and Biological Sciences Research Council (BBSRC)