**Calculi for Systems Biology**

Jane Hillston.
LFCS, University of Edinburgh

18th October 2007

Joint work with Federica Ciocchetta

---

**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, Hassel)
- 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**

- Introduction to Systems Biology
  - Motivation
  - Challenges
- Stochastic Process Algebra
  - Abstract Modelling
  - Case Study
  - Bio-PEPA
- Case Studies
  - Simple genetic network
  - Goldbeter’s model
  - Extended model
- Summary
Stochastic Process Algebra

Introduction to Systems Biology

Natural System

Measurement

Explanation

Obervation

Induction

Interpretation

Modelling

Stochastic Simulation

Note that given a large enough number of molecules an accumulation of protein is a stochastic process.

Thus it is a discrete approach with respect to the number of molecules.

Systems Analysis

In biochemical signalling pathways the events of interest are:

- when reagent concentrations start to increase;
- when concentrations pass certain thresholds;
- when a peak of concentration is reached.

E.g. in a gene network the delay from the activation of one gene until the next promoter reaches an effective level to activate the next gene 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 it is a distribution rather than a deterministic time.

Models should match wet lab experimental data.

Formal Systems

There are two alternative approaches to constructing 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.

Stochastic simulation algorithms

Gillespie's Stochastic Simulation Algorithm (SSA) gives a numerical simulation of the time evolution of a well-stirred chemically reacting system by taking proper account of the randomness inherent in such a system.

It is derived from 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.

Since each molecule is represented explicitly the number of generated states can be extremely large.

Individual vs. Population behaviour

- Biochemistry is concerned with the reactions between individual molecules and so it is often more natural to model at this level.
- However experimental data is usually more readily available in terms of populations rather than individual molecules (average reaction rates rather than the forces at play on an individual molecule in a particular physical context).
- These should be regarded as alternatives, each being appropriate for some models. The challenge then becomes when to use which approach.
- Note that given a large enough number of molecules an “individuals” model will (in many circumstances) be indistinguishable from the a “population” level model.
### Noise vs. Determinism

- With perfect knowledge the behaviour of a biochemical reaction would be deterministic.
- However, in general, we do not have the requisite knowledge of thermodynamic forces, exact relative positions, temperature, velocity etc.
- Thus a reaction appears to display stochastic behaviour.
- When a large number of such reactions occur, the randomness of the individual reactions can cancel each other out and the apparent behaviour exhibits less variability.
- However, in some systems the variability in the stochastic behaviour plays a crucial role in the dynamics of the system.

### Modularity vs. Infinite Regress

As computer scientists we are firm believers in modularity and compositionality. When it comes to biochemical pathways opinion amongst biologists is divided about whether it makes sense to take a modular view of cellular pathways.

Some biologists (e.g. Leibler) argue that there is modularity, naturally occurring, where they define a module relative to a biological function.

Others such as Cornish-Bowden are much more skeptical and cite the problem of infinite regress as being insurmountable.

### The problem of Infinite Regress

![Infinite Regress Diagram]

There is a fundamental challenge when modelling cellular pathways that little is known about some aspects of cellular processes.

In some cases this is because no experimental data is available, or that the experimental data that is available is inconsistent.

In other cases the data is unknowable because experimental techniques do not yet exist to collect the data, or those that do involve modification to the system.

Even when data exists the quality is often very poor.

### Dealing with the Unknown

- In most current work mathematics is being used directly as the formal system.
- 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.

### Formal Systems Revisited

- 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

<table>
<thead>
<tr>
<th>Concurrency</th>
<th>Molecular Biology</th>
<th>Metabolism</th>
<th>Signal Transduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concurrent</td>
<td>Molecules</td>
<td>Enzymes and metabolites</td>
<td>Interacting proteins</td>
</tr>
<tr>
<td>computational</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>processes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Synchronous</td>
<td>Molecular interaction</td>
<td>Binding and catalysis</td>
<td>Binding and catalysis</td>
</tr>
<tr>
<td>communication</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transition</td>
<td>Biochemical modification</td>
<td>Metabolite synthesis</td>
<td>Protein binding, modification or sequestration</td>
</tr>
<tr>
<td>or mobility</td>
<td>or relocation</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

[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.

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:

- Process algebra-based analyses such as comparing models (e.g., for equivalence or simulation) and model checking are only possible if 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

- ODEs population view
- Abstract SPA model
- Stochastic Simulation individual view
- Abstract PEPA model
- CTMC with M levels abstract view
- Model checking and Markovian analysis
- Stochastic Simulation individual view
SPA Languages

- integrated time
- orthogonal time

SPA Languages

exponential only
PEPA, Stochastic π-calculus

exponential + instantaneous
EMPA, Markovian TIPP

general distributions
TIPP, SPADES, GSMPA

orthogonal time

orthogonal time

Exponential only

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

\[
S := (\alpha, r).S | S + S | A
\]

\[
P := S | P \parallel P | P / L
\]

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

PEPA: Performance Evaluation Process Algebra

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.

Reagent-centric modelling [CGH04]

<table>
<thead>
<tr>
<th>Reagent role</th>
<th>Impact on reagent</th>
<th>Impact on reaction rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Producer</td>
<td>decreases concentration</td>
<td>has a positive impact, i.e. proportional to current concentration</td>
</tr>
<tr>
<td>Product</td>
<td>increases concentration</td>
<td>has no impact on the rate, except at saturation</td>
</tr>
<tr>
<td>Enzyme</td>
<td>concentration unchanged</td>
<td>has a positive impact, i.e. proportional to current concentration</td>
</tr>
<tr>
<td>Inhibitor</td>
<td>concentration unchanged</td>
<td>has a negative impact, i.e. inversely proportional to current concentration</td>
</tr>
</tbody>
</table>

PEPA reagent-centric example

\[
A_H := (ab_c, \alpha).A_L
\]

\[
A_{il} := (b_a, \beta).A_H + (c_a, \gamma).A_H
\]

\[
B_H := (ab_c, \alpha).B_L + (b_a, \beta).B_L
\]

\[
B_{il} := (c_b, \delta).B_H
\]

\[
C_H := (c_a, \gamma).C_L + (c_b, \delta).C_L
\]

\[
C_{il} := (ab_c, \alpha).C_H
\]

\[
A_H \parallel B_H \parallel C_H
\]
Example: The Ras/Raf-1/MEK/ERK pathway

PEPA components of the reagent-centric model

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

Commentary on the model

- Here we have shown the model with only high and low levels of concentration.
- In general we would discretise the concentration into more levels, say 6 or 7 levels. As we add levels we are capturing the concentration at finer levels of granularity.
- In fact to generate ODE and SSA models we only need two levels as this is sufficient to record the impact of each reaction on each reagent.

The state space

Alternative models

- When a molecular mapping is used in general a CTMC state space is too large to permit anything but stochastic simulation.
- The ODE model can be regarded as an approximation of a CTMC in which the number of molecules is large enough that the randomness averages out and the system is essentially deterministic.
- In reagent PEPA models with levels, each level of granularity gives rise to a CTMC, and the behaviour of this sequence of Markov processes converges to the behaviour of the system of ODEs.
- Some analyses which can be carried out via numerical solution of the CTMC are not readily available from ODEs or stochastic simulation.

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).
Quantified analysis – $k_{8product}$

Approximating a variation in the initial concentration of RKIP by varying the rate constant $k_1$, we can assess the impact on the production of ERK-PP.

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

**ODE analysis**

Solving a system of ODEs will show how the concentrations of reagents vary over time.

Solution is (relatively) fast and definitive...

... but no variability is captured, unlike Markovian analyses (and real systems).

**ODEs from SPA**

There are advantages to be gained by using 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 process algebra model allows us to derive properties of the model, such as freedom from deadlock, before numerical analysis is carried out.
- The algebraic formulation of the model emphasises interactions between the biochemical entities.

**ODE Analysis of the MAPK example**
Some drawbacks of PEPA

Not all the features of biological systems can be represented into PEPA.

- stoichiometry is not represented explicitly
- general kinetic laws different from Mass Action are not considered.

The latter assumption is restrictive since general kinetic laws are widely-used in the models.

Bio-PEPA: main features

- it is based on the reagent-centric view
- it considers general kinetic laws and expresses them as functional rates
- the PEPA activities are replaced by new ones with stoichiometry and the information about the role of the species (enzyme, inhibitor, ...)
- parameters represent concentration levels
- it can be mapped for the analysis by means of ODEs, stochastic simulation, CTMC, model checking (PRISM)

The aim of the work

In order to overcome the drawbacks above, we have defined Bio-PEPA.

The main field of application is the one of biochemical networks.

Schema

Biochemical networks → Bio-PEPA system → Analysis

The syntax

Sequential component (species component)

\[ S \overset{(\alpha, \kappa)}{\text{op}} S | S + S | C \]  where \( \text{op} = \downarrow \uparrow \oplus \ominus \oslash \)

Model component

\[ P \overset{\omega}{\rightarrow} P \mid S(l) \]

Each action \( \alpha_i \) is associated with a rate \( f_i \)

The list \( \mathcal{N} \) contains the numbers of levels/maximum concentrations

Semantics: prefix rules

PrefixReac \( \overset{(\alpha, \kappa)}{\text{op}} S(l) \overset{[(\alpha, \kappa)]}{\rightarrow} S(l-1) \)  \( 0 < l \leq N \)

PrefixProd \( \overset{(\alpha, \kappa)}{\text{op}} S(l) \overset{[(\alpha, \kappa)]}{\rightarrow} S(l+1) \)  \( 0 \leq l < N \)

PrefixMod \( \overset{(\alpha, \kappa)}{\text{op}} S(l) \overset{[(\alpha, \kappa)]}{\rightarrow} S(l) \)  \( 0 \leq l \leq N \)

with \( \text{op} = \ominus \oplus \oslash \)

Semantics: constant and choice rules

Choice1 \[ \frac{S_i(l) \overset{[\omega]}{\rightarrow} S_i(l)}{(S_i + S_j)(l) \overset{[\omega]}{\rightarrow} S_j(l)} \]

Choice2 \[ \frac{S_i(l) \overset{[\omega]}{\rightarrow} S_i(l)}{(S_i + S_j)(l) \overset{[\omega]}{\rightarrow} S_j(l)} \]

Constant \[ \frac{S_i(l) \overset{[\omega]}{\rightarrow} S_i(l)}{C(l) \overset{[\omega]}{\rightarrow} S(l)} \]  with \( C \overset{\text{def}}{=} S \)
Semantics: cooperation rules

coop1: \[ P_1^{(a_1,v)} P_1' \rightarrow P_1^{(a_1,v)} P_1' \]
\[ P_1^{(a_1,v)} P_1' \rightarrow P_1^{(a_1,v)} P_1' \] \( \text{with } a \notin \mathcal{L} \)

coop2: \[ P_1^{(a_1,v)} P_1' \rightarrow P_1^{(a_1,v)} P_1' \]
\[ P_1^{(a_1,v)} P_1' \rightarrow P_1^{(a_1,v)} P_1' \] \( \text{with } a \notin \mathcal{L} \)

coopFinal: \[ P_1^{(a_1,v)} P_1' P_2^{(a_2,v)} P_2' \rightarrow P_1^{(a_1,v)} P_1' P_2^{(a_2,v)} P_2' \] \( \text{with } a \in \mathcal{L} \)

The abstraction

- each species \( i \) is described by a species component \( C_i \)
- each reaction \( j \) is associated with an action type \( a_j \) and its dynamics is described by a specific function \( f_{ij} \)
- compartments are not represented explicitly

The species components are then composed together to describe the behaviour of the system.

Example: Michaelis-Menten

The reaction \( S \underset{E}{\overset{\gamma}{\longrightarrow}} P \) represents the enzymatic reaction from the substrate \( S \) to the product \( P \) with enzyme \( E \).

The dynamics is described by the law \( f_{MM}(v, K, S, E) = \frac{v E S}{K + S} \).

\[
\begin{align*}
S & \overset{(a, 1)}{\Longrightarrow} (a, 1) \downarrow S \\
E & \overset{(a, 1)}{\Longrightarrow} (a, 1) \downarrow E \\
P & \overset{(a, 1)}{\Longrightarrow} (a, 1) \downarrow P \\
(S(l_{S0}) \underset{E(l_{E0})}{\overset{E(l_{I0})}{\longrightarrow}} P(l_0)) & \underset{(a, 1)}{\Longrightarrow}
\end{align*}
\]

Example: Competitive Inhibition

Binding of the inhibitor to the enzyme prevents binding of the substrate and vice versa.

\[
S + E + I \leftrightarrow SE \Rightarrow P + E
\]
\[ \Downarrow \]
\[ \Downarrow EI \]

Under OSSA (the intermediate species \( SE \) and \( EI \) are constant) we can approximate the reactions above by a unique reaction

\[
S \overset{E(l_{I})}{\longrightarrow} P \quad \text{with rate } f_{ij} = \frac{w S + E}{S + K_m(1 + l \downarrow)}
\]

where \( w \): turnover number (catalytic constant), \( K_m \): Michaelis-constant and \( K_i \): inhibition constant.
Equivalence relations

We are seeking to define a number of equivalence relations for BioPEPA — both those that are expected from the computer science perspective and those that are useful from the biological perspective.

From the computer science perspective we have defined an isomorphism and a (strong) bisimulation.

From a biological perspective we are investigating the situations in which biologists regard models or elements of models to be equivalent, particularly when this is employed for model simplification.

Analysis

A Bio-PEPA system is a formal, intermediate and compositional representation of the system. From it we can obtain

- a CTMC (with levels)
- a ODE system for simulation and other kinds of analysis
- a Gillespie model for stochastic simulation
- a PRISM model for model checking

Each of these kinds of analysis can be of help for studying different aspects of the biological model.

Simple genetic network model

The biological entities are:

- the mRNA molecule \((M)\),
- the protein in monomer form \((P)\) and
- the protein in dimeric form \((P2)\).

All the reactions are described by mass action kinetics with the exception of the first reaction, that has an inhibition kinetics.

Translation into Bio-PEPA

1-Definition of the list \(\mathcal{N}\)

\[ \{M : N_M, M_P; P : N_P, M_P; P2 : N_{P2}, M_{P2}\} \]

2-Definition of functional rates

\[
\begin{align*}
 f_{\gamma_1} &= \frac{v(M)}{K_{M} + P} \\
 f_{\gamma_2} &= \frac{mMA(k_2, [M])}{K_{M} + P} \\
 f_{\gamma_3} &= \frac{mMA(k_3, [M])}{K_{M} + P} \\
 f_{\gamma_4} &= \frac{mMA(k_4, [P])}{K_{M} + P} \\
 f_{\gamma_5} &= \frac{mMA(k_5, [P])}{K_{M} + P2}
\end{align*}
\]
Simple genetic network

Introduction to Systems Biology
Stochastic Process Algebra
Case Studies
Summary

Translation into Bio-PEPA (cont.)

3-Definition of the system components

\[
M = (x_{2,1} \uparrow M + (x_{3,1} \downarrow M + (x_{1,1} \uparrow M;
\]
\[
P = (x_{4,1} \downarrow P + (x_{5,2} \downarrow P + (x_{2,3} \uparrow P) + (x_{2,0} \downarrow P;
\]
\[
P2 = (x_{1,1} \uparrow P2 + (x_{3,5} \downarrow P2 + (x_{3,1} \uparrow P2;
\]
\[
Res = (x_{3,1} \uparrow Res + (x_{4,1} \uparrow Res;
\]
\[
CF = (x_{1,1} \uparrow CF;
\]
\[
4-Definitions of the system
\]
\[
(((CF(1) \uparrow \downarrow M(0)) \downarrow \uparrow P(0)) \downarrow \uparrow P2(0)) \downarrow \uparrow Res(0)
\]

Jane Hillston. LFCS, University of Edinburgh.
Calculi for Systems Biology

Introduction to Systems Biology
Stochastic Process Algebra
Case Studies
Summary

Derivation of ODEs and Gillespie model

The stoichiometry matrix \( D \) associated with the system is

\[
\begin{array}{ccccccc}
R1 & R2 & R3 & R4 & R5 & R6 \\
\hline
CF & 0 & 0 & 0 & 0 & 0 & x_{CF} \\
Res & 0 & 0 & 0 & 0 & 0 & x_{Res} \\
M & +1 & 0 & -1 & 0 & 0 & x_1 \\
P & 0 & +1 & 0 & -1 & -2 & +2 & x_2 \\
P2 & 0 & 0 & 0 & 0 & +1 & -1 & x_3 \\
\end{array}
\]

The kinetic-law vector is

\[
w^T = \begin{pmatrix}
\frac{v}{K + x_3}; k_2 \times x_1; k_3 \times x_1; k_4 \times x_2; k_5 \times x_2; k_5 \times x_3
\end{pmatrix}
\]

Jane Hillston. LFCS, University of Edinburgh.
Calculi for Systems Biology

Introduction to Systems Biology
Stochastic Process Algebra
Case Studies
Summary

Derivation of ODEs (2)

The system of ODEs is obtained as \( \frac{dx}{dt} = D \times w^T \):

\[
\begin{align*}
\frac{dx_1}{dt} &= \frac{v \times 1}{K + x_3} - k_3 \times x_1 \\
\frac{dx_2}{dt} &= k_2 \times x_1 - k_4 \times x_2 - 2 \times k_5 \times x_2 + 2 \times k_5 \times x_3 \\
\frac{dx_3}{dt} &= k_5 \times x_2^2 - k_5 \times x_3
\end{align*}
\]

Jane Hillston. LFCS, University of Edinburgh.
Calculi for Systems Biology

Introduction to Systems Biology
Stochastic Process Algebra
Case Studies
Summary

Simulation results

ODE results

Stochastic simulation results (10 runs)

Jane Hillston. LFCS, University of Edinburgh.
Calculi for Systems Biology
PRISM model

Each species is represented as a PRISM module. For example, the protein is represented as:

```plaintext
module p
  p : [0..Np] init 0;
  [a2]p < Np → (p' = p + 1);
  [a4]p > 0 → (p' = p - 1);
  [a5]p > 0 → (p' = p - 2);
  [a5]p < Np → (p' = p + 2);
endmodule
```

PRISM analysis

- Frequency of monomer P over total P (in terms of levels).
  We need to define a reward structure in the PRISM file as:
  ```plaintext
  rewards
  true : p / (p + Np);
  endrewards
  ```
  We can ask for the frequency of monomer P (in terms of levels) by using the query:
  ```plaintext
  R = ?[i = T]
  ```
- Probability that P is at level i at time T
  ```plaintext
  P = ?[trueU{T,T}]p = i
  ```

PRISM model (2)

An additional (dummy) module is needed to capture the kinetic rates.

```plaintext
module Functional_rates
  dummy: bool init true;
  [a1]dummy = true → \( \frac{v}{k} \) : (dummy' = dummy);
  [a2]dummy = true → r2 : (dummy' = dummy);
  [a3]dummy = true → r3 : (dummy' = dummy);
  [a4]dummy = true → r4 : (dummy' = dummy);
  [a5]dummy = true → r5 : (dummy' = dummy);
  [a5]dummy = true → r5i : (dummy' = dummy);
endmodule
```

Goldbeter’s model

Goldbeter’s model describes the activity of the cyclin in the cell cycle.
- The cyclin promotes the activation of a cdk (cdc2) which in turn activates a cyclin protease.
- This protease promotes cyclin degradation.
- This leads to a negative feedback loop.
- In the model most of the kinetic laws are of kind Michaelis-Menten and this can be reflected in the Bio-PEPA model.

Goldbeter’s model [Goldbeter 91]

- Goldbeter’s model describes the activity of the cyclin in the cell cycle.
- The cyclin promotes the activation of a cdk (cdc2) which in turn activates a cyclin protease.
- This protease promotes cyclin degradation.
- This leads to a negative feedback loop.
- In the model most of the kinetic laws are of kind Michaelis-Menten and this can be reflected in the Bio-PEPA model.

The biological model
The biological model

There are three different species involved:
- **cyclin**, the protein protagonist of the cycle;
- **cdc2 kinase**, in both active (i.e. dephosphorylated) and inactive form (i.e. phosphorylated). The variables used to represent them are \( M \) and \( M' \), respectively;
- **cyclin protease**, in both active (i.e. phosphorylated) and inactive form (i.e. phosphorylated). The variable are \( X \) and \( X' \).

The Bio-PEPA model

**Definition of the list \( N' \):**

\[
N = [\text{Res} : 0,1; \ CF : 1,1; \ C : M_{C}, N_{C}; \ M : M_{M}, N_{M}; \ M' : M_{M'}, N_{M'}; \ X : M_{X}, N_{X}; \ X' : M_{X'}, N_{X'}] \quad (1)
\]

\( \text{Res} \) and \( \text{CF} \) represent degradation and synthesis respectively.

**Definition of functional rates (\( f' - f_{4} \)):**

\[
\begin{align*}
    f_{11} &= f_{\text{MA}}(v_{1}) \\
    f_{12} &= f_{\text{MA}}(v_{2}) \\
    f_{21} &= f_{\text{MM}}((v_{1}, K_{C}, K_{1}), M') \\
    f_{22} &= f_{\text{MM}}((v_{1}, K_{C}, K_{1}), M''') \\
    f_{31} &= f_{\text{MM}}(V_{3}, K_{C}) \\
    f_{32} &= f_{\text{MM}}(V_{3}, K_{D}) \\
    f_{41} &= f_{\text{MM}}(V_{4}, K_{C}) \\
    f_{42} &= f_{\text{MM}}(V_{4}, K_{D}) \\
    f_{51} &= f_{\text{MM}}(V_{5}, K_{C}) \\
    f_{52} &= f_{\text{MM}}(V_{5}, K_{D}) \\
    f_{61} &= f_{\text{MM}}(V_{6}, K_{C}) \\
    f_{62} &= f_{\text{MM}}(V_{6}, K_{D}) \\
    f_{71} &= f_{\text{MM}}(V_{7}, K_{C}) \\
    f_{72} &= f_{\text{MM}}(V_{7}, K_{D}) \\
\end{align*}
\]

**Definition of the component model \( P \):**

\[
\begin{align*}
    C(l_{0C}) &= f_{\text{MA}}(v_{1}) \\
    M(l_{0M}) &= f_{\text{MA}}(v_{2}) \\
    M'(l_{0M'}) &= f_{\text{MA}}(v_{3}) \\
    X(l_{0X}) &= f_{\text{MA}}(v_{4}) \\
    X'(l_{0X'}) &= f_{\text{MA}}(v_{5}) \\
\end{align*}
\]

Assume two levels for each species and initially \( C, M \) and \( X \) present (level 1) and the other elements not present (level 0).

The initial state is \((l_{2}(1), l_{4}(0), l_{6}(1), l_{8}(0), l_{10}(1))\).

R1 and R2 have Mass-Action kinetics, whereas all others are Michaelis-Menten.

ODEs

The stoichiometry matrix \( D \):

\[
\begin{align*}
    &R1 & R2 & R3 & R4 & R5 & R6 & R7 \\
    C & +1 & 0 & 0 & 0 & 0 & 0 & -1
    \end{align*}
\]

\[
\begin{align*}
    M' & 0 & 0 & -1 & +1 & 0 & 0 & 0
    \end{align*}
\]

\[
\begin{align*}
    M & 0 & 0 & +1 & -1 & 0 & 0 & 0
    \end{align*}
\]

\[
\begin{align*}
    X' & 0 & 0 & 0 & 0 & -1 & +1 & 0
    \end{align*}
\]

\[
\begin{align*}
    X & 0 & 0 & 0 & 0 & +1 & -1 & 0
    \end{align*}
\]

The vector that contains the kinetic laws is:

\[
\begin{align*}
    w &= (v_{1}, v_{2}, x_{C}, x_{M}, x_{M'}, x_{X}, x_{X'}, x_{C}, x_{M} + x_{C} + x_{X}, x_{M'} + x_{C} + x_{X}, x_{X'} + x_{C} + x_{X})
    \end{align*}
\]
Gardner et al. [Gardner 98] proposed an extension of the Goldbeter’s model in order to represent a control mechanism for the cell division cycle. They introduce a protein that binds to and inhibits one of the proteins involved in the cell division cycle. This influences the start and the stop of the cell division and modulates the frequency of oscillations.

Several possible extensions were presented; we consider one of them.

Extended model

- ODEs (2)

The system of ODEs is obtained as \( \frac{d\mathbf{x}}{dt} = \mathbf{A} \times \mathbf{w} \), where \( \mathbf{x} \) is the vector of the species variables:

\[
\begin{align*}
\frac{dx_C}{dt} &= v_1 + 1 - K_d \times x_C - \frac{v_d \times x_C \times x_Y}{(K_y + x_Y)} \\
\frac{dx_Y}{dt} &= \frac{v_M \times x_C \times x_M}{(K_c + x_C)} - \frac{v_2 \times x_M}{(K_2 + x_M)} \\
\frac{dx_M}{dt} &= \frac{K_c + x_C \times (K_1 + x_M)}{(K_1 + x_Y)} - \frac{v_2 \times x_M}{(K_2 + x_M)} \\
\frac{dx_X}{dt} &= \frac{v_M \times x_M \times x_V}{(K_3 + x_K)} + \frac{v_4 \times x_X}{(K_4 + x_X)} \\
\frac{dx_C}{dt} &= \frac{v_M \times x_M \times x_C}{(K_3 + x_K)} - \frac{v_4 \times x_X}{(K_4 + x_X)}
\end{align*}
\]

Jane Hillston. LFCS, University of Edinburgh.

Calculi for Systems Biology

Extension of Goldbeter's model

Extended Bio-PEPA model

\[
\begin{align*}
C &= \cdots + (a_8, 1)\Join C + (a_9, 1)\Join C + (a_{12}, 1)\Join C; \\
\vdots & \vdots \\
Res &= \cdots + (a_{11}, 1) \Join Res; \quad CF = \cdots + (a_{10}, 1) \Join CF; \\
I &= (a_{8}, 1)\Join I + (a_{9}, 1)\Join I + (a_{10}, 1)\Join I + (a_{11}, 1)\Join I + (a_{12}, 1)\Join I; \\
IC &= (a_{8}, 1)\Join IC + (a_{9}, 1)\Join IC + (a_{12}, 1)\Join IC + (a_{13}, 1)\Join IC;
\end{align*}
\]

Jane Hillston. LFCS, University of Edinburgh.

Calculi for Systems Biology

New functional rates

- \( f_{a_{8}} = v_8; \)
- \( f_{a_{9}} = IMA(d_1); \)
- \( f_{a_{10}} = IMA(d_2); \)
- \( f_{a_{11}} = IMA(d_3); \)
- \( f_{a_{12}} = IMA(\theta + d_1); \)
- \( f_{a_{13}} = IMA(\theta + K_2) \)
Introduction to Systems Biology

Stochastic Process Algebra

Case Studies

Summary

Extended model

Case Studies

Complete Bio-PEPA model

\[
C(l_{0C}) \uplus M(l_{0M}) \uplus M'(l_{0M}') \uplus X(l_{0X}) \uplus X'(l_{0X}') \uplus \text{Deg}(0) \uplus C(1) \uplus I(0C)
\]

\[
\uplus \text{Deg}(0) \uplus C(1) \uplus I(0C)
\]

New ODE results

\[
\begin{align*}
a_1 &= a_2 = 0.3 \\
v_2 &= 0.6 \\
a_1 &= a_2 = 0.7 \\
v_2 &= 1.4 \\
a_1 &= a_2 = 0.05 \\
v_2 &= 0.1
\end{align*}
\]

Conclusions

Bio-PEPA is a modification of the process algebra PEPA for the modelling and the analysis of biochemical networks.

Bio-PEPA allows us to represent explicitly some features of biological networks, such as stoichiometry and general kinetic laws.

Some future investigations concern:
- the definition of bisimulations and equivalences;
- the study of properties of CTMC with levels;
- the application of model checking techniques.

Challenges

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

- 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 shows variability.
- Promising recent work by Girolami et al. on assessing candidates 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.
- 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!

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)