Calculi for Systems Biology

Jane Hillston.
LFCS, University of Edinburgh

18th October 2007

Joint work with Federica Ciocchetta

The PEPA project

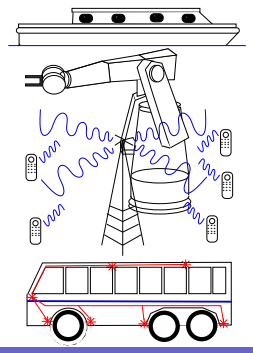
- The PEPA project started in Edinburgh in 1991.
- It was motivated by problems encountered when carrying out *performance analysis* of large computer and communication systems, based on numerical analysis of *Markov processes*.
- *Process algebras* offered a compositional description technique supported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

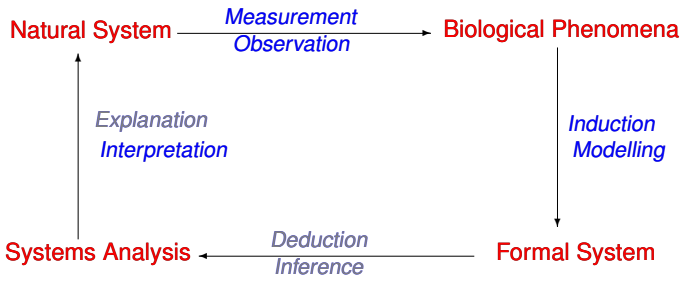


Outline

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



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.

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 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.

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.
- 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*.

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 cf. 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.

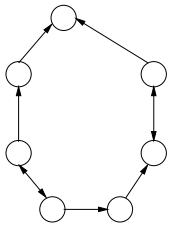
Modularity vs. Infinite Regress

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

Some biologists (e.g. Leibler) argue that there is modularity, naturally occuring, 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*



Dealing with the Unknown

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.

Formal Systems Revisited

- 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.

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]

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).

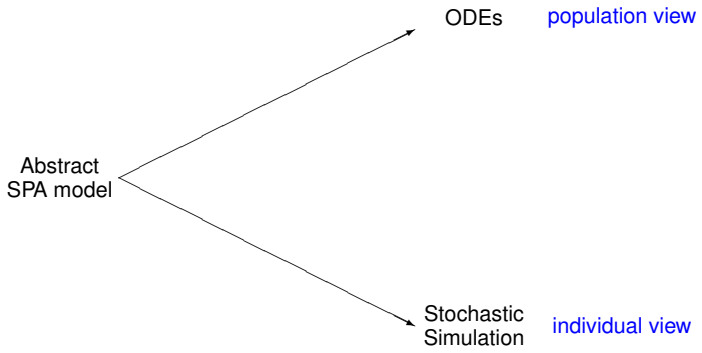
Alternative mappings from the process algebra to underlying mathematics are then readily available.

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 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.

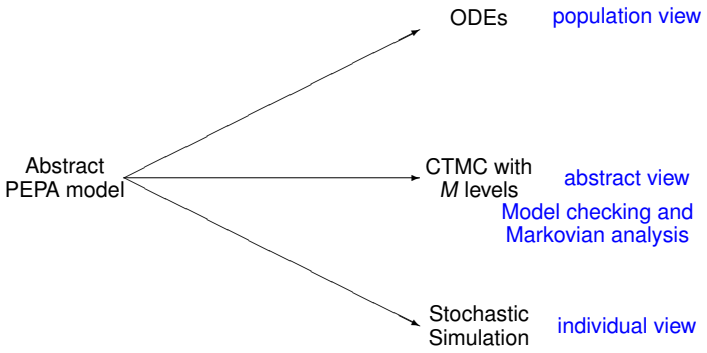
Alternative Representations



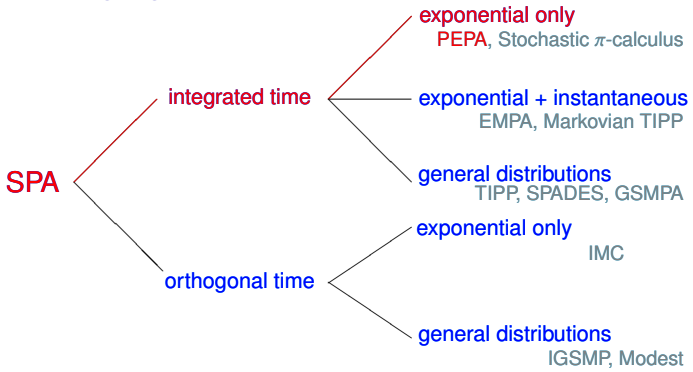
Discretising the population view

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



SPA Languages



PEPA: Performance Evaluation Process Algebra

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

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

SPA MODEL —SOS rules→ LABELLED TRANSITION SYSTEM —state transition diagram→ CTMC **Q**

Q is the infinitesimal generator matrix characterising the CTMC.

PEPA: Performance Evaluation Process Algebra

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

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

SPA MODEL —syntactic analysis→ ACTIVITY MATRIX —continuous interpretation→ ODEs

PEPA: Performance Evaluation Process Algebra

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

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

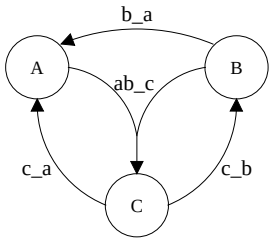
SPA MODEL —syntactic analysis→ RATE EQUATIONS —Gillespie's algorithm→ STOCHASTIC SIMULATION

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

Reagent-centric modelling [CGH04]

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

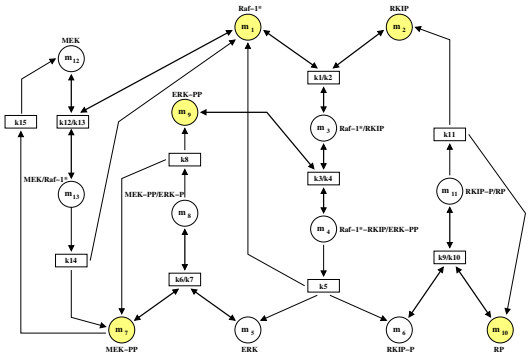
PEPA reagent-centric example



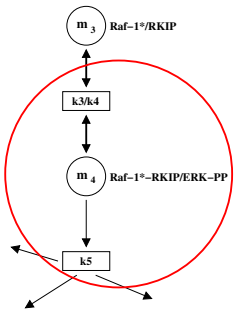
$$A_H \overset{\text{def}}{=} (ab\_c, \alpha).A_L$$
$$A_L \overset{\text{def}}{=} (b\_a, \beta).A_H + (c\_a, \gamma).A_H$$
$$B_H \overset{\text{def}}{=} (ab\_c, \alpha).B_L + (b\_a, \beta).B_L$$
$$B_L \overset{\text{def}}{=} (c\_b, \delta).B_H$$
$$C_H \overset{\text{def}}{=} (c\_a, \gamma).C_L + (c\_b, \delta).C_L$$
$$C_L \overset{\text{def}}{=} (ab\_c, \alpha).C_H$$

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

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



PEPA components of the reagent-centric model



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

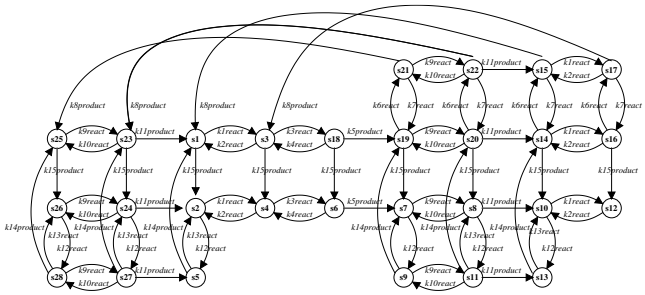
$$\text{Raf-1}^*/\text{RKIP}/\text{ERK-PP}_\text{L} \stackrel{def}{=} (k3react, k_3).\text{Raf-1}^*/\text{RKIP}/\text{ERK-PP}_\text{H}$$

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

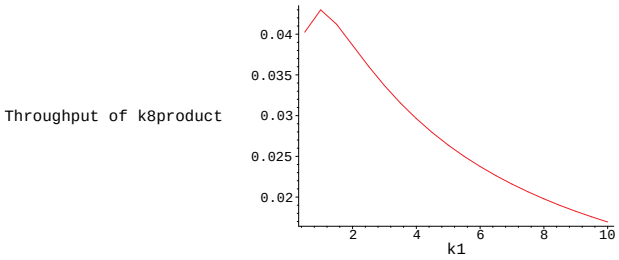
- When a molecular mapping is used in general a CTMC state space is too large to permit anything but stochastic simulation.
- The ODE model can be regarded as an approximation of a CTMC in which the number of molecules is large enough that the randomness *averages out* and the system is essentially deterministic.
- 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).

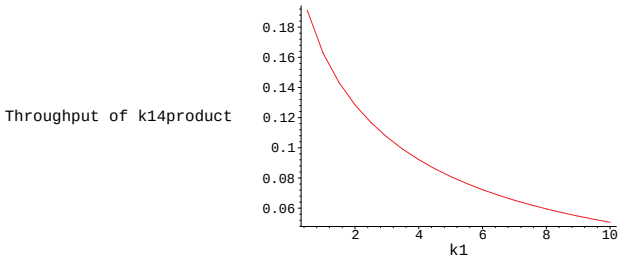
Quantified analysis – *k8product*

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



Quantified analysis – *k14product*

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



ODE analysis

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

Solution is (relatively) fast and definitive....

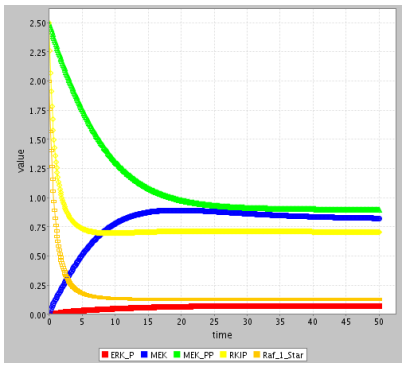
... but no variability is captured, unlike Markovian analyses (and real systems).

ODEs from SPA

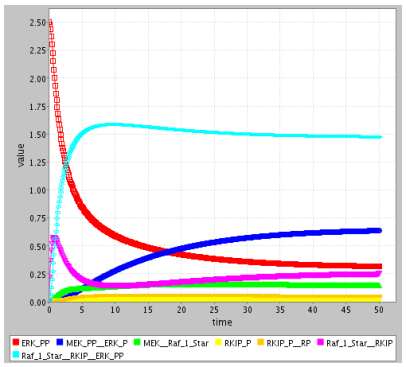
There are advantages to be gained by using a process algebra model as an intermediary to the derivation of the ODEs.

- The ODEs can be *automatically* generated from the descriptive process algebra model, thus reducing human error.
- The process algebra model allow us to *derive properties* of the model, such as freedom from deadlock, before numerical analysis is carried out.
- The algebraic formulation of the model emphasises interactions between the biochemical entities.

ODE Analysis of the MAPK example



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.

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 $\longrightarrow$ *Bio-PEPA system* $\longrightarrow$ *Analysis*

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 syntax

Sequential component (species component)

$$S \stackrel{def}{=} (\alpha, \kappa) \text{ op } S \mid S + S \mid C \qquad \text{where op} = \downarrow \mid \uparrow \mid \oplus \mid \ominus \mid \odot$$

Model component

$$P \stackrel{def}{=} P \underset{\mathcal{L}}{\bowtie} P \mid S(l)$$

Each action α_j is associated with a rate f_{α_j}

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

Semantics: prefix rules

$$\texttt{prefixReac} \quad ((\alpha,\kappa)\downarrow S)(l) \xrightarrow{(\alpha,[S:\downarrow(l,\kappa)])} S(l-1) \quad 0 < l \leq N$$

$$\texttt{prefixProd} \quad ((\alpha,\kappa)\uparrow S)(l) \xrightarrow{(\alpha,[S:\uparrow(l,\kappa)])} S(l+1) \quad 0 \leq l < N$$

$$\texttt{prefixMod} \quad ((\alpha,\kappa)\ op\ S)(l) \xrightarrow{(\alpha,[S:op(l,\kappa)])} S(l) \quad 0 \leq l \leq N$$

with $op = \odot, \oplus$, or $\ominus$

Semantics: constant and choice rules

$$\texttt{Choice1} \quad \frac{S_1(l) \xrightarrow{(\alpha,v)} S'_1(l')}{(S_1 + S_2)(l) \xrightarrow{(\alpha,v)} S'_1(l')}$$

$$\texttt{Choice2} \quad \frac{S_2(l) \xrightarrow{(\alpha,v)} S'_2(l')}{(S_1 + S_2)(l) \xrightarrow{(\alpha,v)} S'_2(l')}$$

$$\texttt{Constant} \quad \frac{S(l) \xrightarrow{(\alpha,S':[op(l,\kappa))]} S'(l')}{C(l) \xrightarrow{(\alpha,C:[op(l,\kappa))]} S'(l')} \quad \text{with } C \stackrel{def}{=} S$$

Semantics: cooperation rules

$$\text{coop1} \quad \frac{P_1 \xrightarrow{(\alpha,v)} P_1'}{P_1 \underset{\mathcal{L}}{\bowtie} P_2 \xrightarrow{(\alpha,v)} P_1' \underset{\mathcal{L}}{\bowtie} P_2} \quad \text{with } \alpha \notin \mathcal{L}$$

$$\text{coop2} \quad \frac{P_2 \xrightarrow{(\alpha,v)} P_2'}{P_1 \underset{\mathcal{L}}{\bowtie} P_2 \xrightarrow{(\alpha,v)} P_1 \underset{\mathcal{L}}{\bowtie} P_2'} \quad \text{with } \alpha \notin \mathcal{L}$$

$$\text{coopFinal} \quad \frac{P_1 \xrightarrow{(\alpha,v_1)} P_1' \quad P_2 \xrightarrow{(\alpha,v_2)} P_2'}{P_1 \underset{\mathcal{L}}{\bowtie} P_2 \xrightarrow{(\alpha,v_1 @ v_2)} P_1' \underset{\mathcal{L}}{\bowtie} P_2'} \quad \text{with } \alpha \in \mathcal{L}$$

Semantics: rates and transition system

In order to associate the rates we consider a new relation $\mapsto \subseteq C \times \Gamma \times C$, with $\gamma \in \Gamma := (\alpha, r)$ and $r \in \mathbb{R}^+$.

The relation is defined in terms of the previous one:

$$\text{Final} \quad \frac{P \xrightarrow{(\alpha_j,v)} P'}{P \xmapsto{(\alpha_j, f_{\alpha_j}(v,\mathcal{N}))} P'}$$

$f_{\alpha_j}(v, \mathcal{N})$ represents the parameter of an *exponential distribution* and the dynamic behaviour is determined by a *race condition*.

The transition system and the CTMC are defined as in PEPA.

The abstraction

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

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

Example: Michaelis-Menten

The reaction $S \xrightarrow{E} P$ represents the *enzymatic reaction* from the substrate S to the product P with enzyme E.

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

$$\begin{aligned} S &\overset{def}{=} (\alpha, 1)\downarrow S \\ E &\overset{def}{=} (\alpha, 1) \oplus E \\ P &\overset{def}{=} (\alpha, 1)\uparrow P \end{aligned}$$

$$(S(l_{S0}) \underset{\{\alpha\}}{\bowtie} E(l_{E0})) \underset{\{\alpha\}}{\bowtie} P(l_{P0})$$

Example: Competitive Inhibition

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

$$\begin{array}{ccccc} S + E + I & \Longleftrightarrow & SE & \Longrightarrow & P + E \\ \Updownarrow & & & & \\ EI & & & & \end{array}$$

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

$$S \xrightarrow{E,I:f_I} P \qquad \text{with rate } f_I = \frac{w * S * E}{S + K_M(1 + \frac{I}{K_I})}$$

where w: turnover number (catalytic constant), K_M: Michaelis-constant and K_I: inhibition constant.

Example: Competitive Inhibition (2)

The specification in Bio-PEPA is:

$$S = (\alpha, 1)\downarrow S \qquad P = (\alpha, 1)\uparrow P \qquad E = (\alpha, 1) \oplus E \qquad I = (\alpha, 1) \ominus I$$

The system is described by

$$(S(l_{S0}) \underset{\{\alpha\}}{\bowtie} E(l_{E0})) \underset{\{\alpha\}}{\bowtie} I(l_{I0}) \underset{\{\alpha\}}{\bowtie} P(l_{P0})$$

with functional rate

$$f_\alpha = f_{CI}((w, K_M, K_I), S, E, I) = \frac{w * S * E}{S + K_M(1 + \frac{I}{K_I})}$$

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.

Bisimulation

Definition

A binary relation $\mathcal{R} \subseteq C \times C$ is a strong bisimulation with respect to $\mapsto$, if $(P, Q) \in \mathcal{R}$ implies for all $\alpha \in \mathcal{A}$:

- if $P \stackrel{\gamma_1}{\longmapsto} P'$ then, for some Q' and γ_2, $Q \stackrel{\gamma_2}{\longmapsto} Q'$ with $(P', Q') \in \mathcal{R}$ and
  1. $action(\gamma_1) = action(\gamma_2) = \alpha$
  2. $rate(\gamma_1) = rate(\gamma_2)$
- symmetric definition for $Q \stackrel{\gamma_2}{\longmapsto} Q'$

Analysis

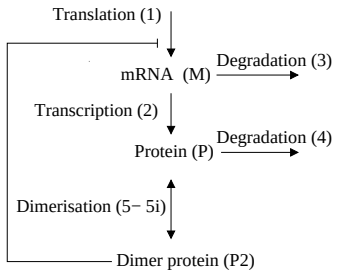
A Bio-PEPA system is a formal, intermediate and compositional representation of the system.
From it we can obtain

- 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

The biological model

Consider a **genetic network** with negative feedback through dimers.



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 $\mathcal{N}$

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

2-Definition of functional rates

$$\begin{array}{lcl} f_{\alpha_1} & = & fI((v, K_M), [P2, CF]) = \frac{v * CF}{K_M + P2}; \\ f_{\alpha_2} & = & fMA(k_2, [M]); \quad f_{\alpha_3} = fMA(k_3, [M]); \quad f_{\alpha_4} = fMA(k_4, [P]); \\ f_{\alpha_5} & = & fMA(k_5, [P]); \quad f_{\alpha_{5i}} = fMA(k_{5i}, [P2]); \end{array}$$

Translation into Bio-PEPA (cont.)

3-Definition of the system components

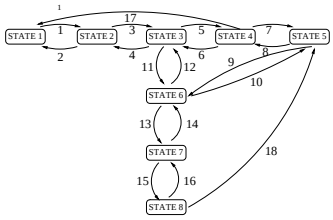
$$
\begin{array}{lcl}
\mathrm{M} & = & (\alpha_2,1) \oplus \mathrm{M} + (\alpha_3,1) \downarrow \mathrm{M} + (\alpha_1,1) \uparrow \mathrm{M}; \\
\mathrm{P} & = & (\alpha_4,1) \downarrow \mathrm{P} + (\alpha_5,2) \downarrow \mathrm{P} + (\alpha_{5i},2) \uparrow \mathrm{P}) + (\alpha_2,0) \uparrow \mathrm{P}; \\
\mathrm{P2} & = & (\alpha_1,1) \ominus \mathrm{P2} + (\alpha_{5i},1) \downarrow \mathrm{P2} + (\alpha_5,1) \uparrow \mathrm{P2}; \\
\mathrm{Res} & = & (\alpha_3,1) \odot \mathrm{Res} + (\alpha_4,1) \odot \mathrm{Res}; \\
\mathrm{CF} & = & (\alpha_1,1) \odot \mathrm{CF};
\end{array}
$$

4-Definitions of the system

$$((((CF(1) \underset{\{\alpha_1\}}{\bowtie} M(0)) \underset{\{\alpha_2\}}{\bowtie} P(0)) \underset{\{\alpha_5,\alpha_{5i}\}}{\bowtie} P2(0)) \underset{\{\alpha_3,\alpha_4\}}{\bowtie} Res(0)$$

The CTMC with levels

For **2 levels**, the CTMC consists of **8 states** and **18 transitions**.



The states are $(CF(l_1), M(l_2), P(l_3), P2(l_4), RES(l_5))$, where l_i represents the level of each species component.

Derivation of ODEs and Gillespie model

The stoichiometry matrix D associated with the system is

	R1	R2	R3	R4	R5	R6	
CF	0	0	0	0	0	0	x_{CF}
Res	0	0	0	0	0	0	x_{Res}
M	+1	0	-1	0	0	0	x_1
P	0	+1	0	-1	-2	+2	x_2
P2	0	0	0	0	+1	-1	x_3

The kinetic-law vector is

$$w^T = (\frac{v \times x_{CF}}{K + x_3};\ k_2 \times x_1; k3 \times x_1;\ k4 \times x_2;\ k5 \times {x_2}^2;\ ki5 \times x_3)$$

Derivation of ODEs (2)

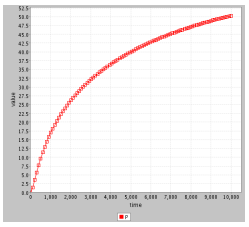
The system of ODEs is obtained as $\frac{d\bar{x}}{dt} = D \times w$:

$$
\begin{array}{lcl}
\frac{dx_1}{dt} & = & \frac{v \times 1}{K + x_3} - k3 \times x_1 \\
\frac{dx_2}{dt} & = & k2 \times x_1 - k4 \times x_2 - 2 \times k5 \times {x_2}^2 + 2 \times ki5 \times x_3 \\
\frac{dx_2}{dt} & = & k5 \times {x_2}^2 - ki5 \times x_3
\end{array}
$$

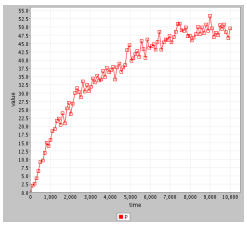
Derivation of Gillespie model

The derivation of the Gillespie model is made by creating molecules corresponding to each species and defining the possible reactions with appropriate adjustment of kinetic rates.

Simulation results



ODE results



Stochastic simulation results (10 runs)

PRISM model

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

module p
$p : [0..Np]$ **init** 0;
$[a2]p < Np \rightarrow (p' = p + 1)$;
$[a4]p > 0 \rightarrow (p' = p - 1)$;
$[a5]p > 0 \rightarrow (p' = p - 2)$;
$[a5i]p < Np \rightarrow (p' = p + 2)$;
endmodule

PRISM model (2)

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

module Functional_rates
dummy: bool **init** true;
$[a1]dummy = true \rightarrow \frac{v}{(1+(pd/k))} : (dummy' = dummy)$;
$[a2]dummy = true \rightarrow r2 : (dummy' = dummy)$;
$[a3]dummy = true \rightarrow r3 : (dummy' = dummy)$;
$[a4]dummy = true \rightarrow r4 : (dummy' = dummy)$;
$[a5]dummy = true \rightarrow r5 : (dummy' = dummy)$;
$[a5i]dummy = true \rightarrow r5i : (dummy' = dummy)$;
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:

  rewards
  $true : \frac{p}{(p+pd)}$;
  endrewards

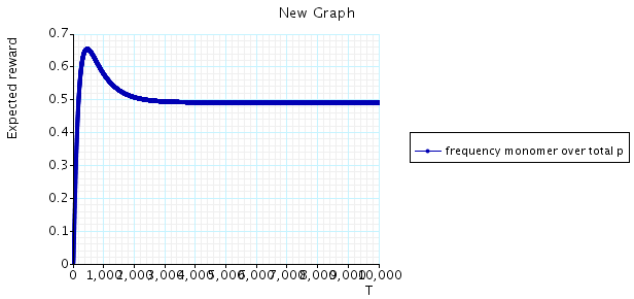
  We can ask for the frequency of monomer P (in terms of levels) by using the query:

  $$R =?[I = T]$$

- **Probability that P is at level i at time T**

  $$P =?[trueU[T, T]p = i]$$

PRISM results



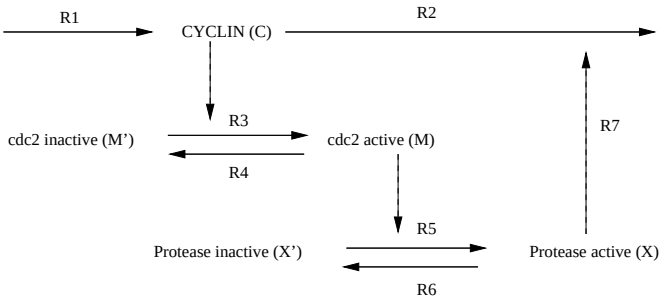
monomer frequency

New Graph

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 (2)

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'.

Reactions

id	name	react.	prod.	mod.	kinetic laws
R1	creation of cyclin	-	C	-	vi
R2	degradation of cyclin	C	-	-	$kd \times C$
R3	activation of cdc2 kinase	M'	M	-	$\frac{C*V_{M1}}{(K_c+C)} \frac{M'}{(K_1+M')}$
R4	deactivation of cdc2 kinase	M	M'	-	$\frac{M \times V_2}{(K_2+M)}$
R5	activation of cyclin protease	X'	X	M	$\frac{X' \times M \times V_{M3}}{(K_3+X')}$
R6	deactivation of cyclin protease	X	X'	-	$\frac{X \times V_4}{K_4+X}$
R7	X triggered degradation of cyclin	C	-	X	$\frac{C \times v_d \times X}{C+K_d}$

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

The Bio-PEPA model

Definition of the list $\mathcal{N}$:.

$$\mathcal{N} = [Res : 0,1;\ CF : 1,1;\ C : M_C, N_c;\ M : M_M, N_M; \\ M' : M_{M'}, N_{M'};\ X : M_X, N_X;\ X' : M_{X'}, N_{X'}] \quad (1)$$

Res and *CF* represent degradation and synthesis respectively.

Definition of functional rates ($\mathcal{F}$:)

$$f_{\alpha_1} = fMA(v_i); \qquad f_{\alpha_2} = fMA(k_d);$$
$$f_{\alpha_3} = fMM'((V_1, K_c, K_1), M', C) = \frac{v_1 * C}{K_c + C} \frac{M'}{K1 + M'};$$
$$f_{\alpha_4} = fMM(V_2, K_2); \qquad f_{\alpha_5} = fMM(V_3, K_3)$$
$$f_{\alpha_6} = fMM(V_4, K_4); \qquad f_{\alpha_7} = fMM(V_d, K_d)$$

The Bio-PEPA model (2)

Definition of species components (*Comp*):

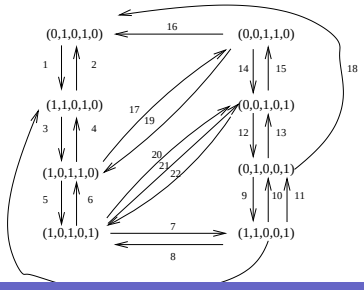
$$C = (\alpha_1,1)\uparrow C + (\alpha_2,1)\downarrow C + (\alpha_7,1)\downarrow C + (\alpha_3,1) \oplus C;$$
$$M' = (\alpha_4,1)\uparrow M' + (\alpha_3,1)\downarrow M';$$
$$M = (\alpha_3,1)\uparrow M + (\alpha_4,1)\downarrow M + (\alpha_5,1) \oplus M;$$
$$X' = (\alpha_6,1)\uparrow X' + (\alpha_5,1)\downarrow X';$$
$$X = (\alpha_5,1)\uparrow X + (\alpha_6,1)\downarrow X + (\alpha_7,1) \oplus X;$$
$$Res = (\alpha_2,1) \odot Res; \quad CF = (\alpha_1,1) \odot CF;$$

Definition of the model component (*P*):

$$C(l_{0C}) \underset{\{\alpha_3\}}{\bowtie} M(l_{0M}) \underset{\{\alpha_3,\alpha_4\}}{\bowtie} M'(l_{0M'}) \underset{\{\alpha_5,\alpha_7\}}{\bowtie} X(l_{0X}) \underset{\{\alpha_5,\alpha_6\}}{\bowtie} X'(l_{0X'}) \\ \underset{\{\alpha_2\}}{\bowtie} Deg(0) \underset{\{\alpha_1\}}{\bowtie} CF(1)$$

Analysis

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 $(l_C(1), l_{M'}(0), l_M(1), l_{X'}(0), l_X(1))$.



ODEs

The stoichiometry matrix D:

	R1	R2	R3	R4	R5	R6	R7	
C	+1	0	0	0	0	0	-1	x_C
M'	0	0	-1	+1	0	0	0	$x_{M'}$
M	0	0	+1	-1	0	0	0	x_M
X'	0	0	0	0	-1	+1	0	$x_{X'}$
X	0	0	0	0	+1	-1	0	x_X

The vector that contains the kinetic laws is:

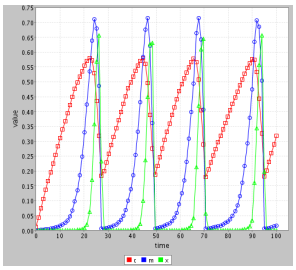
$$w = \Big(v_i * 1, k_d * x_C, \frac{V_{M1} * x_C}{K_c + x_C} \frac{x_{M'}}{(K_1 + x_{M'})}, \frac{V_2 * x_M}{(K_2 + x_M)}, \\ \frac{V_{M3} * x_M * x_{X'}}{(K_3 + x_{X'})}, \frac{V_4 * x_X}{(K_4 + x_X)}, \frac{v_d * x_C * x_X}{(K_d + x_C)}\Big)$$

ODEs (2)

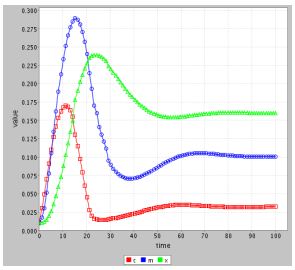
The system of ODEs is obtained as $\frac{d\bar{x}}{dt} = D \times w$, where $\bar{x}^T =: (x_C, x_{M'}, x_M, x_{X'}, x_X)$ is the vector of the species variables:

$$\begin{aligned}
\frac{dx_C}{dt} &= v_i * 1 - k_d * x_C - \frac{v_d * x_C * x_X}{(K_d + x_C)} \\
\frac{dx_{M'}}{dt} &= -\frac{V_{M1} * x_C}{K_c + x_C}\frac{x_{M'}}{(K_1 + x_{M'})} + \frac{V_2 * x_M}{(K_2 + x_M)} \\
\frac{dx_M}{dt} &= +\frac{V_{M1} * x_C}{K_c + x_C}\frac{x_{M'}}{(K_1 + x_{M'})} - \frac{V_2 * x_M}{(K_2 + x_M)} \\
\frac{dx_{X'}}{dt} &= -\frac{V_{M3} * x_M * x_{X'}}{(K_3 + x_{X'})} + \frac{V_4 * x_X}{(K_4 + x_X)} \\
\frac{dx_X}{dt} &= \frac{V_{M3} * x_M * x_{X'}}{(K_3 + x_{X'})} - \frac{V_4 * x_X}{(K_4 + x_X)}
\end{aligned}$$

ODE results



$K_1 = K_2 = K_3 = K_4 = 0.02\mu M$



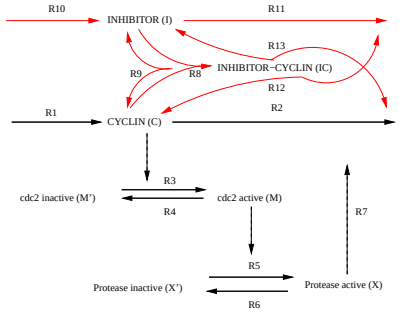
$K_1 = K_2 = K_3 = K_4 = 40\mu M$

Extended model

- 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 extension were presented; we consider one of them.

Extension of Goldbeter's model



Extended Bio-PEPA model

$$\begin{aligned}
C &= \cdots + (\alpha_8, 1)\downarrow C + (\alpha_9, 1)\uparrow C + (\alpha_{12}, 1)\uparrow C; \\
&\vdots \quad\quad \vdots \\
Res &= \cdots + (\alpha_{11}, 1) \odot Res; \quad CF = \cdots + (\alpha_{10}, 1) \odot CF; \\
I &= (\alpha_8, 1)\downarrow I + (\alpha_9, 1)\uparrow I + (\alpha_{10}, 1)\uparrow I + (\alpha_{11}, 1)\downarrow I + (\alpha_{13}, 1)\uparrow I; \\
IC &= (\alpha_8, 1)\uparrow IC + (\alpha_9, 1)\downarrow IC + (\alpha_{12}, 1)\downarrow IC + (\alpha_{13}, 1)\downarrow IC;
\end{aligned}$$

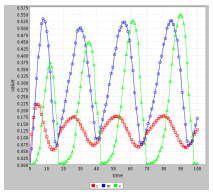
New functional rates

$$\begin{aligned}
f_{\alpha_8} &= v_s; \\
f_{\alpha_9} &= fMA(d_1); \\
f_{\alpha_{10}} &= fMA(a_1); \\
f_{\alpha_{11}} &= fMA(a_2); \\
f_{\alpha_{12}} &= fMA(\theta * d_1); \\
f_{\alpha_{13}} &= fMA(\theta * k_d)
\end{aligned}$$

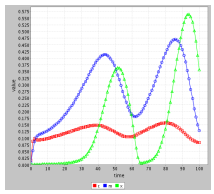
Complete Bio-PEPA model

$$C(l_{0C}) \underset{\{\alpha_3\}}{\bowtie} M(l_{0M}) \underset{\{\alpha_3,\alpha_4\}}{\bowtie} M'(l_{0M'}) \underset{\{\alpha_5,\alpha_7\}}{\bowtie} X(l_{0X}) \underset{\{\alpha_5,\alpha_6\}}{\bowtie} X'(l_{0X'}) \underset{\{\alpha_2\}}{\bowtie}$$
$$Deg(0) \underset{\{\alpha_1\}}{\bowtie} CF(1)$$
$$\underset{\{\alpha_8,\alpha_9,\alpha_{10},\alpha_{11}\}}{\bowtie} I(l_{0I}) \underset{\{\alpha_8,\alpha_9,\alpha_{12},\alpha_{13}\}}{\bowtie} IC(l_{0IC})$$

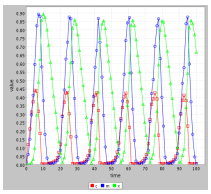
New ODE results



$a_1 = a_2 = 0.3$
$v_s = 0.6$



$a_1 = a_2 = 0.7$
$v_s = 1.4$



$a_1 = a_2 = 0.05$
$v_s = 0.1$

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.
- 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 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 showns 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)