From qualitative to quantitative formal methods for biochemical signalling pathways

Oana Andrei
University of Glasgow

LIFL, February 25, 2010

joint work with Hélène Kirchner (INRIA Bordeaux) and Muffy Calder (University of Glasgow)
Outline

• Motivation
• Rule-based modelling
• Abstractions for CTMCs
• Conclusion and perspectives
Formal methods for modelling biological systems

- lab experiments
- computational model
- results/analysis

Goals: to understand, to predict, to control
Cell signalling

• communication between cells
• cellular processes: cell growth, proliferation, apoptosis...
• malfunctions may lead to diseases
Challenges

- suitable formalisms
- abstraction techniques
- analysis
- scalability
Our approaches

- qualitative: rule-based, higher-order calculus, runtime-verification
- quantitative: abstraction for CTMCs - CTMCs with levels, stochastic model checking
Higher-order rule-based modelling
Port graphs

- graphs with multiple edges and loops
- edges connect to ports of nodes
- defined over a signature \((N,P)\)
A port graph
Molecular graphs as port graphs

<table>
<thead>
<tr>
<th>Molecular complex</th>
<th>Port graph</th>
</tr>
</thead>
<tbody>
<tr>
<td>protein</td>
<td>node</td>
</tr>
<tr>
<td>site</td>
<td>port</td>
</tr>
<tr>
<td>bond</td>
<td>edge</td>
</tr>
<tr>
<td>interaction</td>
<td>rewrite rule</td>
</tr>
</tbody>
</table>
Rewrite rules

- well-suited for modelling bio-molecular interactions
- a rule $L \rightarrow R$ defines a class of reactions
Port graph rewrite rules
Port graph rewrite rules

```
\begin{align*}
\text{i\hspace{-.1em}:X} & \quad \text{i.1\hspace{-.1em}:X.1} & \quad \text{i.2\hspace{-.1em}:X.2} \\
\text{j\hspace{-.1em}:Y} & \quad \text{j\hspace{-.1em}:Y} & \quad \text{j\hspace{-.1em}:Y}
\end{align*}
```
Port graph rewrite rules
Port graph rewrite rules
Port graph rewrite rules
Port graph rewrite rules

\[
\begin{align*}
\text{i:}X & \quad \text{\rightarrow} \\
\text{j:}Y & \\
\end{align*}
\]

\[
\begin{align*}
\text{i:}X & \\
\text{j:}Y & \\
\end{align*}
\]

\[
\begin{align*}
\text{i:}X & \quad \text{\rightarrow} \\
\text{i.j:}X.Y & \\
\end{align*}
\]
A port graph rewrite rule is a port graph.
A port graph rewrite rule is a port graph
A port graph rewrite rule is a port graph
Port graph rewriting relation

\[ G \Rightarrow_{L \Rightarrow R} G' \quad \text{if} \quad \exists (g, G^-, \mathcal{B}) \in Sol(L \leftarrow G) \]

such that

\[ G = G^- \upharpoonright_{\mathcal{B}} g(L) \]

and

\[ G' = G^- \upharpoonright_{\mathcal{B}} g(R) \]

\[ \downarrow_{g \mathcal{B}} \]
Example: a fragment of the EGFR signaling pathway

Initial state:

1:S

2:S

3:S

4:S

5:R

6:R

7:A
Example: a fragment of the EGFR signaling pathway
Example: a fragment of the EGFR signaling pathway

i:S  
\[
\begin{array}{c}
2 \\
1 \\
\end{array}
\]

j:S  
\[
\begin{array}{c}
1 \\
2 \\
\end{array}
\]

DimerS

i.j:DS  
\[
\begin{array}{c}
1 \\
1 \\
2 \\
2 \\
\end{array}
\]

i:DS  
\[
\begin{array}{c}
2 \\
\end{array}
\]

j:R  
\[
\begin{array}{c}
1 \\
4 \\
\end{array}
\]

DSbindsR

i:DS  
\[
\begin{array}{c}
2 \\
\end{array}
\]

j:R  
\[
\begin{array}{c}
1 \\
4 \\
\end{array}
\]
Example: a fragment of the EGFR signaling pathway
Example: a fragment of the EGFR signaling pathway

A stable state:

1.2:DS

3.4:DS

5:R

6:R

7:A

- 2 x DimerS
- 2 x DSbindsR
- 1 x DimerR
- 2 x ActivateDR
- 1 x DRbindsA
Graph-base approaches

- $\kappa$-calculus, Kappa factory [Danos et al.]
- BioNetGen [Hlavacek et al.]
- Pathway Logic [Talcott et al.]
Chemical programming

• γ-calculus = λ-calculus + chemical paradigm [BanatreFR04-07]

• a chemical solution where molecules interact freely according to reaction rules

• everything is a molecule

\[ \text{prod} = \text{replace } X, Y \text{ by } X \times Y \]

\[ \langle \text{prod}, 3, 1, 4, 5, 2 \rangle \rightarrow \langle \text{prod}, 1, 4, 15, 2 \rangle \rightarrow^* \langle \text{prod}, 120 \rangle \]
Rewriting calculus

• extends first-order term rewriting and the \( \lambda \)-calculus [CirsteaK01]

• terms, rules, rule application are explicit objects of the calculus

\[(s(x)+y \xrightarrow{} s(x+y)) \ (s(5)+s(2)) \xrightarrow{\rho} s(5+s(2))\]
Biochemical calculus

• add biochemical flavour to the chemical calculus - structures (like port graphs)

• rewrite strategies for controlling the rule application (Identity, Failure, Sequence, Not, First, ...)

• verification techniques
Syntax

- objects: port graphs
- rewrite rules
- abstractions
- application

\[
\begin{align*}
\text{(Objects)} & \quad \mathcal{O} ::= \mathcal{OBJ} \mid X \mid \mathcal{O} \cdot \mathcal{O} \\
\text{(Rule)} & \quad \mathcal{R} ::= \mathcal{O} \Rightarrow \mathcal{O} \\
\text{(Molecule)} & \quad \mathcal{M} ::= \mathcal{O} \mid \mathcal{R} \mid \mathcal{M} \cdot \mathcal{M} \\
\text{(Abstraction)} & \quad \mathcal{A} ::= \mathcal{M} \Rightarrow \mathcal{M} \\
\text{(Configuration)} & \quad \mathcal{K} ::= \mathcal{M} \mid \mathcal{A} \mid \mathcal{K} \cdot \mathcal{K} \\
\text{(System)} & \quad \mathcal{S} ::= [\mathcal{K}]
\end{align*}
\]
Semantics

(Interaction) \[ K \cdot (M \Rightarrow N) \cdot M' \] \[ \rightarrow_i [K \cdot \varsigma(N)] \]
if \( \varsigma \in Sol(M \Leftarrow M') \)
More control? Use strategies

• provide control over the composition or choice of the abstraction to apply

• enforce confluence and termination

★ Identity, Failure, Sequence, Not, First, Repeat...

\[
First(S_1, S_2)(G) = S_1(G) \text{ if } S_1 \text{ does not fail, } S_2(G) \text{ otherwise}
\]

• encoded as abstractions in the calculus
Strategies-based extensions

- tackling application failure

(InteractionR) \[ [K \cdot T \cdot M] \rightarrow_{ir} [K \cdot \text{seq}(T, \text{try(stk \Rightarrow T \cdot M)))@M] \]

✨ persistent strategies $S!$
Invariant verification

• invariant:
  • rule $G \Rightarrow G$

• strategy \texttt{first}(G$\Rightarrow$G, X$\Rightarrow$”Failure”)

• remove (G$\Rightarrow$”Failure”)! or “repair” (G$\Rightarrow$H)!

• but we can do more...
Structural formulas
Structural formulas:

\[ \varphi ::= T \mid \bot \mid \gamma \mid \neg \varphi \mid \varphi_1 \land \varphi_2 \mid \varphi_1 \lor \varphi_2 \mid \varphi_1 \rightarrow \varphi_2 \mid \Diamond \varphi \]
Structural formulas:

\[ \varphi ::= T \mid \bot \mid \gamma \mid \neg \varphi \mid \varphi_1 \land \varphi_2 \mid \varphi_1 \lor \varphi_2 \mid \varphi_1 \rightarrow \varphi_2 \mid \diamond \varphi \]

Satisfaction relation:

\[ G \models \gamma \iff \exists \sigma \text{ such that } G = \sigma(\gamma) \]
\[ G \models \diamond \varphi \iff \exists G' \subseteq G \text{ such that } G' \models \varphi \]
Structural formulas as strategies

\[
\begin{align*}
\tau(\top) & = \text{id} \\
\tau(\bot) & = \text{fail} \\
\tau(\Diamond \gamma) & = \gamma \Rightarrow \gamma \\
\tau(\neg \varphi) & = \text{not}(\tau(\varphi)) \\
\tau(\varphi_1 \land \varphi_2) & = \text{seq}(\tau(\varphi_1), \tau(\varphi_2)) \\
\tau(\varphi_1 \lor \varphi_2) & = \text{first}(\tau(\varphi_1), \tau(\varphi_2)) \\
\tau(\varphi_1 \rightarrow \varphi_2) & = X \Rightarrow \text{seq}(\tau(\varphi_1), \text{first}(\text{stk} \Rightarrow X, \tau(\varphi_2)))@X
\end{align*}
\]

\[
G \models \varphi \text{ if and only if } \tau(\varphi)@G \rightarrow^* G
\]

\[
G \not\models \varphi \text{ if and only if } \tau(\varphi)@G \rightarrow^* \text{stk}
\]
Guarded systems

• define a new reduction relation

\[ [K]_\varphi \iff [K']_\varphi \text{ if } [K] \Rightarrow [K'] \text{ and } K' \models \varphi \]

• use strategies

\[ [K]_\varphi \iff \text{ifThenElse}(\tau(\varphi), X_1 \Rightarrow [K']_\varphi, X_2 \Rightarrow \text{error\_message})@K' \]

if \([K] \Rightarrow [K']\)
Conclusions (first part)

• port graphs: a biologically-inspired graphical structure

• biochemical calculus: a higher-order rule-based formalism

• verification of invariant properties

• applications to protein-protein interactions and autonomic systems
Future work

• embed runtime verification
  • diagnose faults at execution and repair faults (adaptive behaviour)
  • identify properties to monitor
  • choose temporal logic: $\text{LTL}_3 (T, \perp, ?)$

• add a stochastic semantics

• robustness analysis
Abstractions for continuous-time Markov chains
CTMCs

- state-based formalisms for describing dynamic systems: $C = (S, s_0, R, L)$
- discrete steps, continuous time-steps
- suitable for modelling signalling pathways: stochastic, computational, concurrent
CTMCs with levels

- population (species) based modelling
- discrete levels of concentrations
  - maximum molar concentration $M$
  - choose granularity $N$ for the abstraction, concentration step size $H = M/N$
  - $0, 1, ..., N$ levels of concentrations correspond to $0, (0, H], (H, 2H], ..., ((N-1)H, NH]$
Formal model

- continuous time Markov chains with levels
- properties expressed as formulas in Continuous Stochastic Logic (CSL)
- symbolic probabilistic model checker PRISM
Formal model

• mass-action kinetics

• reaction $A + B \rightarrow C$ with $k$ constant rate

• transition rate: $k \ast (L_A \ast H) \ast (L_B \ast H)/H$

[rct1] $L_A > 0 \rightarrow (L_A \ast H) : L_A' = L_A - 1$  // (in module for A)

[rct1] $L_B > 0 \rightarrow (L_B \ast H) : L_B' = L_B - 1$  // (in module for B)

[rct1] $L_C < \max C \rightarrow 1 : L_C' = L_C + 1$  // (in module for C)

[rct1] true $\rightarrow k/H : true$  // (in module for const)
Signalling and scaffold proteins
Signalling and scaffold proteins
Signalling and scaffold proteins
Signalling and scaffold proteins
Signalling and scaffold proteins
Signalling and scaffold proteins
AKAP
(A-kinase anchoring protein)
AKAP
(A-kinase anchoring protein)
AKAP
(A-kinase anchoring protein)
AKAP
(A-kinase anchoring protein)
AKAP
(A-kinase anchoring protein)
AKAP
(A-kinase anchoring protein)
AKAP
(A-kinase anchoring protein)
AKAP
(A-kinase anchoring protein)
AKAP
(A-kinase anchoring protein)
AKAP
(A-kinase anchoring protein)
Expected behaviour

Q1: \( \uparrow p\text{PDE8A1} \rightarrow \downarrow \text{cAMP} \rightarrow \downarrow \text{PKA}^+ \rightarrow \uparrow \text{Raf activity} \rightarrow \downarrow p\text{Raf}_{S259} \)
Expected behaviour

\[ Q_1: \uparrow \text{pPDE8A1} \rightarrow \downarrow \text{cAMP} \rightarrow \downarrow \text{PKA}^+ \rightarrow \uparrow \text{Raf activity} \rightarrow \downarrow \text{pRaf}_{S259} \]

\[ Q_2: \text{Pulsating behaviour} \]
PRISM model

- modules for cAMP, scaffold, free PDE8A1, PP
- mass action kinetics
- information on constant rates ratios
Continuous Stochastic Logic

- extension of non-probabilistic CTL
- probability operator $P$
- steady-state operator $S$

\[
\begin{align*}
\text{State formulae} & : \Phi ::= \top \mid a \mid \neg\Phi \mid \Phi \land \Phi \mid P_{\triangleright p}[\Phi] \mid S_{\triangleright p}[\Phi] \\
\text{Path formulae} & : \phi ::= X \Phi \mid \Phi U^I \Phi
\end{align*}
\]
Reward-based properties

• use of rewards (or costs) in CSL
  - real values assigned to states or transitions
  - to track variable values in states
  - to compute the expected value of a variable at a given time
Reward-based properties

- state rewards for computing the expected levels for cAMP, pPDE8A1, PKA^+, pS259
Trend variables

• keep track of decreasing or increasing variable values

• define new variables in the PRISM modules:
  \[ \text{cAMP'} = \text{cAMP} - 1 \quad \& \quad \text{trend}_\text{cAMP'} = -1 \]

• \( \downarrow \text{x} \quad (\uparrow \text{x}) \) ascending (descending) trend for variable \( \text{x} \)
Necessarily preceded

[Monteiro et al. 08]

For $\varphi = \downarrow \text{cAMP} \land \downarrow \text{PKA}^+$ and $\psi = \uparrow \text{pPDE8A1}$

CTL: $(\text{EF } \varphi) \land \text{AG}((\neg \psi) \Rightarrow \text{AG}(\neg \varphi))$

CSL: $P_{>0}[F \varphi] \land P_{\leq 0}[F(\neg ((\neg \psi) \Rightarrow P_{\geq 1}[F(\neg \varphi)]))]$
Pulsations

Show that the levels of pPDE8A1 fluctuate:

- $\varphi = \uparrow_{pPDE8A1}$ and $\psi = \downarrow_{pPDE8A1}$
- pulsation in CTL [Fages05, Ballarini et al. 09]:
  \[
  \text{AG}(\varphi \Rightarrow \text{EF} \psi) \land (\psi \Rightarrow \text{EF} \varphi)
  \]
- pulsation in CSL:
  \[
  P_{\leq 0}[F (\neg(\varphi \Rightarrow P_{>0}[F\psi]) \lor \neg(\psi \Rightarrow P_{>0}[F\varphi]))
  \]
Pulsations

- for cAMP: $\varphi = \uparrow cAMP$ and $\psi = \downarrow cAMP$
- for $\text{PKA}^+$: $\varphi = \uparrow \text{PKA}^+$ and $\psi = \downarrow \text{PKA}^+$
- coordinated pulsations:
  
  $\varphi = \uparrow p\text{PDE8A1} \land \downarrow cAMP \land \downarrow \text{PKA}^+$ and
  
  $\psi = \downarrow p\text{PDE8A1} \land \uparrow cAMP \land \uparrow \text{PKA}^+$
Overview of AKAP modelling
Overview of AKAP modelling

☑️ formal model of a biological process
Overview of AKAP modelling

- formal model of a biological process
- the biologists validated our results
Overview of AKAP modelling

- formal model of a biological process
- the biologists validated our results
- refine the model with more experimental data
Overview of AKAP modelling

☑ formal model of a biological process
☑ the biologists validated our results
☐ refine the model with more experimental data
☐ trend variables, amplitude of oscillations
Overview of AKAP modelling

- formal model of a biological process
- the biologists validated our results
- refine the model with more experimental data
- trend variables, amplitude of oscillations
- formulate new properties and express them using a temporal logic
Abstractions for CTMCs with levels

- relation between two CTMCs with levels for the same system:
- aim: preserve temporal properties and do model checking on the more abstract model
- if $C^N \models \varphi$, then $C^{kN} \models f(\varphi)$ - who is $f$?
- (weak) simulation relation [Baier et al.] does not work...
Temporal properties

• classification of temporal properties for signalling pathways

• BIOCHAM [Fages et al.]

• patterns [Monteiro et al.08]

• stochastic models, not only qualitative or probabilistic
Temporal properties

- is CSL expressive enough?
- what about LTL(R) ? [Fages et al.]
- linear versus branching time for biologists?
- satisfaction probabilities for biologists?
Acknowledgement

• Hélène Kirchner and Horatiu Cirstea
• Muffy Calder and SIGNAL project
• Walter Kolch, George Baillie and Kim Brown from the Faculty of Biomedical & Life Science, University of Glasgow
Thank you!
Questions?
Bibliography

