Modelling molecular evolution with process algebras

Marek Kwiatkowski
ETH Zürich & Eawag
marek.kwiatkowski@eawag.ch

8 June 2011
WCSB 2011, Zürich

(PhD work at the University of Edinburgh, supervised by Ian Stark)
Overview

1. Introduction and motivation
   - Some existing work
   - Towards a unifying framework

2. Modelling evolution of a signalling cascade
   - Process algebras for biology
   - The MAPK cascade and its model
   - Evolutionary setup
   - Fitness distributions and model backtracking

3. Conclusions
Some recent studies


Common theme
models ≡ genotypes, execution ≡ development, results ≡ phenotypes.
Towards a unifying framework

Just like systems biology has benefited from SBML, evolutionary systems biology could benefit from a standard specification and modelling format. Ideally, it should:

1. Be agent-centric, not reaction-centric,
2. Support dynamic complex formation,
3. Have deterministic primary dynamics, but
4. Admit a variety of execution modes.

In what follows we introduce and evaluate such a prototype framework.

Process algebra and biology

Process algebras are, loosely speaking, idealised programming languages with a focus on parallel computing. They have been used to model biochemical networks since ca. 1999.

Define:

\[
A \equiv a.(A_1|A_2)
\]
\[
B \equiv b.B
\]

Compute:

\[
A | B = a.(A_1|A_2) | b.B \rightarrow A_1 | A_2 | B
\]

Benefits: formality, parsimony, compositionality, abstraction.

Case study: the MAPK cascade (1)

- Functionally conserved in most animals
- Crucial component of many signal transduction pathways
- Relays and amplifies the signal efficiently
- Benchmark for new modelling techniques
Case study: the MAPK cascade (2)

\[
\begin{align*}
Ras & \triangleq (\nu x \rightarrow \overline{x})ras(x; y).((\overline{x}.Ras + y.Ras) \\
Raf & \triangleq (\nu x \rightarrow \overline{x})raf(x; y).((\overline{x}.Raf + y.Raf^*)) \\
Raf^* & \triangleq (\nu x \rightarrow \overline{x})(\nu z \rightarrow \overline{z})(raf^*(x; y).((\overline{x}.Raf^* + y.Raf^*) + raf_b^*(z; y).((\overline{z}.Raf^* + y.Raf))) \\
PP2A1 & \triangleq (\nu x \rightarrow \overline{x})pp2a1(x; y).((\overline{x}.PP2A1 + y.PP2A1) \\
MEK & \triangleq (\nu x \rightarrow \overline{x})mek(x; y).((\overline{x}.MEK + y.MEK^*) \\
MEK^* & \triangleq (\nu x \rightarrow \overline{x})(\nu z \rightarrow \overline{z})(mek^*(x; y).((\overline{x}.MEK^* + y.MEK^{**}) + mek_b^*(z; y).((\overline{z}.MEK^{**} + y.MEK^*))) \\
MEK^{**} & \triangleq (\nu x \rightarrow \overline{x})(\nu z \rightarrow \overline{z})(mek^{**}(x; y).((\overline{x}.MEK^{**} + y.MEK^*)) + mek_b^*(z; y).((\overline{z}.MEK^{**} + y.MEK^*))) \\
PP2A2 & \triangleq (\nu x \rightarrow \overline{x})pp2a2(x; y).((\overline{x}.PP2A2 + y.PP2A2) \\
ERK & \triangleq (\nu x \rightarrow \overline{x})erk(x; y).((\overline{x}.ERK + y.ERK^*) \\
ERK^* & \triangleq (\nu x \rightarrow \overline{x})(\nu z \rightarrow \overline{z})(erk^*(x; y).((\overline{x}.ERK^* + y.ERK^{**}) + erk_b^*(z; y).((\overline{z}.ERK^{**} + y.ERK^*))) \\
ERK^{**} & \triangleq (\nu x \rightarrow \overline{x})erk_b^{**}(x; y).((\overline{x}.ERK^{**} + y.ERK^*)) \\
MKP3 & \triangleq (\nu x \rightarrow \overline{x})mkp3(x; y).((\overline{x}.MKP3 + y.MKP3) \\
\Pi & \triangleq c_1 \cdot Raf \parallel c_2 \cdot Ras \parallel c_3 \cdot MEK \parallel c_4 \cdot ERK \parallel c_5 \cdot PP2A1 \parallel c_6 \cdot PP2A2 \parallel c_7 \cdot MKP3
\end{align*}
\]
Case study: the MAPK cascade (3)

Twenty-three differential equations extracted from the $c\pi$ model and solved with Octave. Emergent Michaelis-Menten kinetics for every reaction.
Evolutionary analysis of the MAPK cascade: the plan

- Reconfigure every site in every way possible (ca. 1M variants)
- Find evolutionarily fragile and robust sites
- Compute the fitness of every variant using signal integration
- Find the distribution of mutation effects on fitness
Evolutionary analysis of the MAPK cascade: fitness function

- Rewards fast and strong response (green area)
- Punishes incomplete switching-off (red area)

Evolutionary analysis of the MAPK cascade: fitness distributions

(a) ras
similar: $raf^*, raf_b^*, mek_b^*$, $mek^*, mek_b^*, erk_b^*$, and $erk^*$$$

(b) pp2a1
similar: $raf, mek^*, pp2a2$ and $erk^*$

(c) mck
similar: $erk$ and $mkp3$

(d) all
Evolutionary analysis of the MAPK cascade: two strange peaks (left)
Evolutionary analysis of the MAPK cascade: two strange peaks (right)
Evolutionary analysis of the MAPK cascade: advantageous mutations
Conclusions

1 Introduction and motivation
   • Some existing work
   • Towards a unifying framework

2 Modelling evolution of a signalling cascade
   • Process algebras for biology
   • The MAPK cascade and its model
   • Evolutionary setup
   • Fitness distributions and model backtracking

3 Conclusions