

Replicating the MAP Kinase Cascade in PEPA

Muffy Calder¹, Adam Duguid², Stephen Gilmore², and Jane Hillston²

¹ Department of Computer Science, The University of Glasgow
muffy@dcs.gla.ac.uk

² School of Informatics, The University of Edinburgh
{jeh,stg}@inf.ed.ac.uk , a.j.duguid@sms.ed.ac.uk

Abstract. Ordinary Differential Equations (ODEs) represent a powerful technique for analysing biological systems. ODEs can cope with large numbers of components involved in multiple interactions. ODEs also bring a certain level of complexity through the reactions and the combinations of components within these reactions. Using the Mitogen-Activated Protein (MAP) Kinase cascade, it will be shown that use of a high level language such as Performance Evaluation Process Algebra (PEPA) allows modellers to build clearer models while also bringing the advantage of multiple forms of analysis.

1 Introduction

Systems Biology, in no more than ten words, could be described as *the modelling of biological systems through their interactions*. The belief is that this will bring a better understanding of the biological entities through the behaviour of their components, rather than a static view listing components and concentrations. The popular approach to systems biology has been through the use of ODEs which, while powerful, can become complex and offers only one view of the system. The use of a higher level language can allow abstraction away from the details of ODEs, and offer not only a cleaner view of the system, but also the possibility to analyse or simulate the model with a variety of techniques.

Recently the PEPA language has defined a mapping to ODEs [1, 2], using the Extracellular signal Regulated Kinase (ERK) signalling pathway (consisting of 13 components) as an example. Larger models, such as the Schoeberl *et al.* [3] which model the MAP Kinase cascade with internalized and surface EGF receptors, contains 94 components with an updated version consisting of 103. The motivation for replicating the Schoeberl *et al.* model in PEPA is quite simple: does PEPA scale when modelling large biological systems?

2 MAP Kinase Cascade

Figure 1 shows the graphical representation³ of the MAP Kinase cascade from the Schoeberl *et al.* paper [3]. The diagram hides several complexities, two of which are highlighted in Fig. 2.

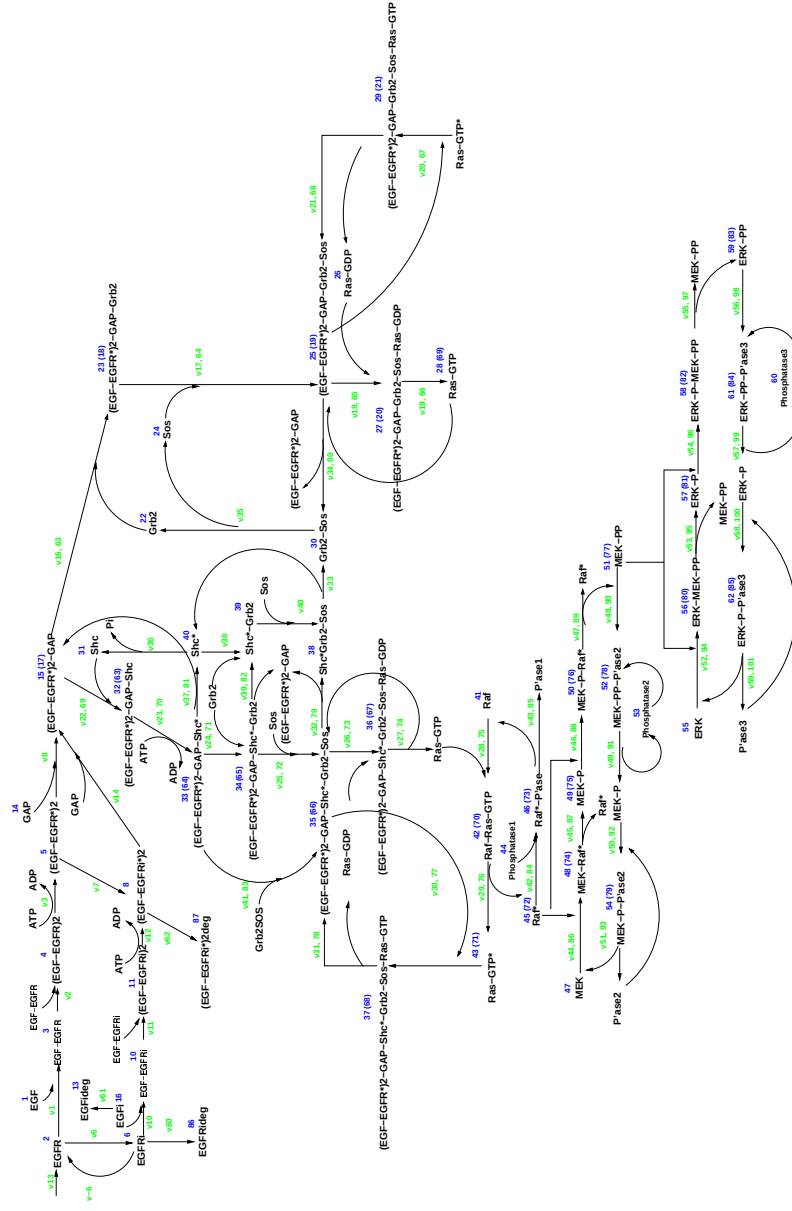


Fig. 1. The MAP Kinase Cascade

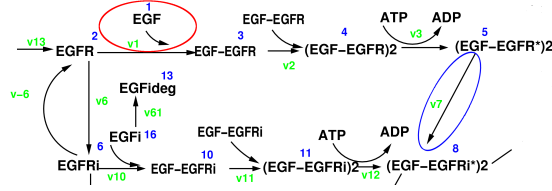


Fig. 2. Extract from Fig. 1

First, the Epidermal Growth Factor (EGF), circled in the upper left of Fig. 2 binds to the EGF receptors. The reactions presented in [3] suggest that EGF is consumed in this binding i.e. EGF is a standard reactant. This assumption is enforced when viewing the defined reactions within the Matlab model. However, this definition is later overwritten with the single line $\text{SM}(1, :) = 0$; changing the observable behaviour by zeroing the rate of change for the EGF component. In this instance EGF has been likened to a reservoir; the quantity of EGF at this given concentration is large enough that it will not change appreciably after reaction v_1 has occurred. This behaviour would seem to contradict the initial beliefs of the presented model.

Second, the reaction v_7 , is highlighted in Fig. 2 to show the confusion of reversibility within reactions. Reactions such as v_3 (autophosphorylation) are reversible, and are often represented by two definitions, each with their own rate. Reaction v_7 (internalization) is a single directed reaction which cannot be reversed, a property not clear from the diagram.

The use of a high level language aids in presentation, and allows several types of analysis to be applied to the same model. For this model the reagent-centric approach [1] will be used which allows conversion to both ODEs and Stochastic Simulation Algorithms (SSAs). An example of the syntax can be seen in (1)

$$\begin{aligned} \text{EGF-EGFR}2_H &\stackrel{\text{def}}{=} (v_3, k_3) \cdot \text{EGF-EGFR}2_L + (v_{-2}, k_{-2}) \cdot \text{EGF-EGFR}2_L \\ \text{EGF-EGFR}2_L &\stackrel{\text{def}}{=} (v_{-3}, k_{-3}) \cdot \text{EGF-EGFR}2_H + (v_2, k_2) \cdot \text{EGF-EGFR}2_H \end{aligned} \quad (1)$$

This can be read as follows. Reactions v_3 and v_{-2} *consume* the component (EGF-EGFR)2, taking it from a high level of concentration to a low level as signified by the subscript identifiers H and L . Reactions v_{-3} and v_2 replenish said component, returning the concentration from a low state to a high state. When combined with other component definitions it allows the form seen in (2) to be constructed.



The definition in (2) states that the reaction labelled v_3 transforms (EGF-EGFR)2 into (EGF-EGFR*)2 at a rate linked to k_3 . The actual rate will be

³ <http://www.nature.com/nbt/journal/v20/n4/extref/nbt0402-370-S1.pdf>

dependent on the existing concentration of all of the reactants, in this case (EGF-EGFR)2. The concentration of (EGF-EGFR)2 at time t will be reduced by this amount and the concentration of (EGF-EGFR*)2 increased by the same.

Through the use of PEPA, ambiguity over the reversibility of any given reaction is removed as every reaction must be explicitly defined using the PEPA Prefix. Similarly, if EGF is to act as a catalyst it must also be explicitly defined within the PEPA model. Either it is defined as simply another rate (making the concentration a constant over the entire experiment) and as such acts as a modifier to a reaction, or defined as $\text{EGF}_H \stackrel{\text{def}}{=} (v_1, k_1).\text{EGF}_H$. This states that the component EGF is involved in the reaction v_1 but its level is not affected by the reaction physically happening. This allows EGF to act as a normal component in other reactions while still affecting the rate at which v_1 occurs without being consumed in the process.

3 Results

To confirm that the defined PEPA model was indeed correct it was converted to Matlab form and solved with the same parameters as the Matlab model⁴ from [3]. The results returned were identical as can be seen in Fig. 3, but more importantly, comparison of the two different representations of ODEs shows the reactions in both are identical.

PEPA in this form also allows conversion to a representation allowing stochastic simulation. Under certain conditions simulating the PEPA model using one of Gillespie's SSAs [4, 5] should also return the same result, but it was not known if this was the case with the MAP Kinase cascade. As can be seen in Fig. 3, the results from the SSA (τ -leap simulation) differ from the results returned from the original model; in the case of Ras-GTP the peak concentration is double that shown by the ODEs. Possible answers included the increased accuracy of SSAs when dealing with very small population sizes, or even an error in the model used to perform the stochastic simulation. Investigation of this discrepancy led to the surprising finding that neither explanation was correct, instead the time step used in solving the ODEs was actually masking the true peaks in the results. Altering the time step for the ODE solver then returns results identical to those returned by the τ -leap algorithm.

4 Conclusion

Two conclusions can be drawn from this experience. The goal behind replicating the MAP Kinase cascade was to investigate the suitability of PEPA as a high level language for modelling biological systems. The decision to replicate the MAP Kinase cascade was based on the size and complexity of original model, as well as the belief shown by the biological community in the accuracy of this model. With 94 components within the cascade this results in approximately 188

⁴ http://web.mit.edu/dllaz/egf_pap

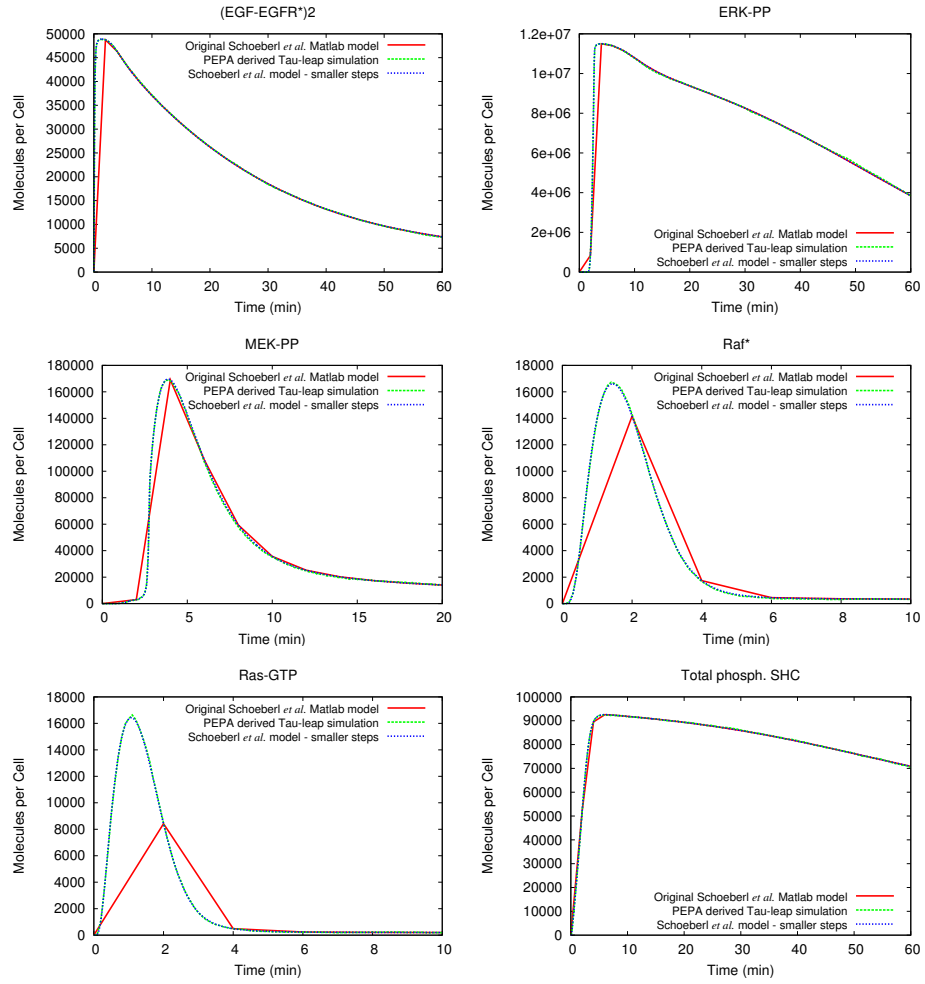


Fig. 3. Graphs showing 6 groups components from the MAP Kinase cascade. Each graph compares the original ODE analysis from [3], the τ -leap stochastic simulation and the ODE with modified timestep.

PEPA definitions, two for each component. The model also contains components that exhibit a catalytic style behaviour not seen in the signalling pathways previously modelled. The reagent-centric approach allows (in some respects) easier modification to the model. A component can be removed or added to a pre-existing reaction from one point rather than requiring editing of several reaction definitions. It could also be said that errors are easier to locate as each component consists of simple clean definitions.

The second conclusion that can be gained is the benefit from using a high level language that allows multiple forms of analysis. The original purpose of running the τ -leap algorithm was to compare the results from a SSA to those obtained from the ODE analysis. It was this use of an alternative solver that allowed the discovery of the true peak for Ras-GTP and led to the discovery of the cause of the difference (the use of an overly large time step within the original model). The claim is not that the original model is somehow flawed, nor that there are inherent issues with ODEs; rather that the use of a high level language can allow verification through different forms of analysis which may or may not discover irregularities with any one form of analysis.

References

1. M. Calder, S. Gilmore, and J. Hillston, "Modelling the influence of RKIP on the ERK signaling pathway using the stochastic process algebra PEPA," in *Proceedings of BioConcur'04*, (London, England), Aug. 2004. Extended version to appear in Transactions on Computational Systems Biology.
2. M. Calder, S. Gilmore, and J. Hillston, "Automatically deriving ODEs from process algebra models of signalling pathways," in *Proceedings of Computational Methods in Systems Biology '05*, (Edinburgh, Scotland), Apr. 2005.
3. B. Schoeberl, C. Eichler-Jonsson, E. Gilles, and G. Muller, "Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors," *Nature Biotechnology*, vol. 20, pp. 370–375, 2002.
4. D. Gillespie and L. Petzold, "Improved leap-size selection for accelerated stochastic simulation," *J. Comp. Phys.*, vol. 119, pp. 8229–8234, 2003.
5. M. Rathinam, L. Petzold, Y. Cao, and D. Gillespie, "Stiffness in stochastic chemically reacting systems: The implicit tau-leaping method," *J. Comp. Phys.*, vol. 119, no. 24, pp. 12784–12794, 2003.