

# Integrated Analysis from Abstract Stochastic Process Algebra Models

Jane Hillston, Federica Ciocchetta, Adam Duguid, and Stephen Gilmore

Laboratory for Foundations of Computer Science,  
The University of Edinburgh, Scotland

## Extended Abstract

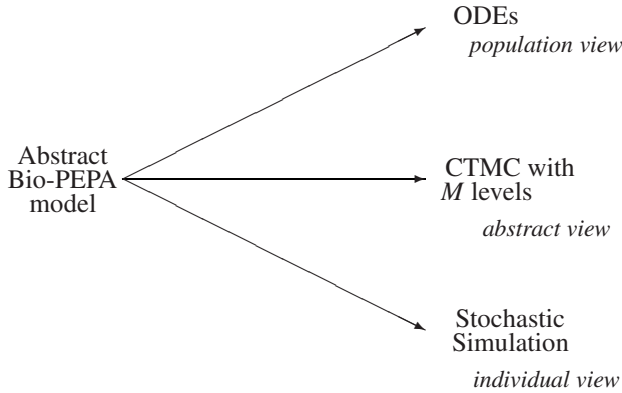
Bio-PEPA is a novel stochastic process algebra which has been recently developed for modelling biochemical pathways [5,6]. In Bio-PEPA a reagent-centric style of modelling is adopted, and a variety of analysis techniques can be applied to a single model expression. Such an approach facilitates easy validation of analysis results when the analyses address the same issues [3] and enhanced insight when the analyses are complementary [4]. Currently supported analysis techniques include stochastic simulation at the molecular level, ordinary differential equations, probabilistic model checking and numerical analysis of a continuous time Markov chain.

Process algebras are a well-established modelling approach for representing concurrent systems facilitating both qualitative and quantitative analysis. Within the last decade they have also been proposed as the basis for several modelling techniques applied to biological problems, particularly intracellular signalling pathways, e.g. [13,12,10,7,2,1].

A process algebra model captures the behaviour of a system as the actions and interactions between a number of entities, usually termed *processes* or *components*. In stochastic process algebras, such as PEPA [9] or the stochastic  $\pi$ -calculus [11], a random variable representing average duration is associated with each action. In the stochastic  $\pi$ -calculus, interactions are strictly binary whereas in PEPA and Bio-PEPA the more general, multiway synchronisation is supported.

The original motivation for the use of process algebras for modelling intracellular pathways was the recognition of the clear mapping that can be made between *molecules*, within a biochemical pathway, and *processes*, within concurrent systems [14]. The mapping is then elaborated with reactions between molecules represented by communication between processes, etc.

This mapping has been extremely influential with much subsequent work on process algebras for systems biology following its lead. It takes an inherently *individuals*-based view of a pathway or cell, and suffers the problem of individuals-based modelling, namely *state-space explosion*. When each individual within a system is represented explicitly and all transitions within or between individuals are captured as discrete events, the number of states becomes prohibitively high. This problem prohibits the use of techniques which rely on access to the state space in its entirety, such as model checking or numerical solution of a Markov chain. Essentially analysis is restricted to stochastic simulation where the state space is explored iteratively.



**Fig. 1.** Alternative modelling approaches: a single Bio-PEPA description of a system may be used to derive alternative mathematical representations offering different analysis possibilities

In contrast, biologists often take a *population*-based view of cellular systems, representing them as systems of ordinary differential equations (ODEs). These mathematical models are continuous or *fluid* approximations of the discrete state, individuals-based models of the system. In many circumstances the approximation is extremely good. In the biological context, where the exact number of molecules is often difficult to obtain but is known to be extremely large, this more abstract view is both intellectually and computationally appealing. The continuous state space models, in the form of systems of ODEs, are much more efficiently solved than their discrete state space counterparts.

In Bio-PEPA we wanted to be able to use a single model description to access both an individuals-based and a population-based view of a system. Thus we adopt an abstract style of modelling which we term, *reagent-centric*. We use the term *reagent* to mean an entity which engages in a reaction. In the basic case this will be a biochemical species such as a protein, receptor molecule or an enzyme. However it may be more abstract, capturing the behaviour of a group of entities or a whole subsidiary pathway. In this style of modelling the focus of the process algebra model is no longer the individual molecules, but rather the species or similar entities. This subtle change gives us much more flexibility in how a model may be interpreted, facilitating mappings into a number of different mathematical representations, as illustrated in Figure 1. Viewing the reagents as species, it is straightforward to use the BioPEPA description to derive the stoichiometry matrix, and the corresponding ODE model. Conversely, knowing the forms of interactions which can be engaged in by the reagent, allows an individuals-based or molecular model to be derived, suitable for solution by Gillespie’s stochastic simulation algorithm [8].

Moreover using the *reagents-as-processes* abstraction, together with other features of the BioPEPA language, make it straightforward to capture several characteristics of biochemical reactions which can be problematic for other process algebras. These include reactions with stoichiometry greater than one, with more than two reactants, and with general kinetic laws.

We have also been keen to investigate the extent to which “classical” process algebra analyses can be used to provide insight into system biology models. In the context of stochastic process algebras such analyses include numerical analysis of the underlying continuous time Markov chain (CTMC) and probabilistic model checking. Whilst the stochastic simulation described above is based on a CTMC, the size of the state space in most examples will prohibit any state-based analysis. Thus we have also developed a third mapping from BioPEPA models to an smaller CTMC, which we term the *CTMC with levels*. In such models the concentration of each reagent is split into discrete steps, leading to a more compact state space, more readily amenable to state-based analyses.

## References

1. Bortolussi, L., Policriti, A.: Modeling Biological Systems in Stochastic Concurrent Constraint Programming. *Constraints* 13(1-2), 55–90 (2006)
2. Calder, M., Gilmore, S., Hillston, J.: Modelling the influence of RKIP on the ERK signalling pathway using the stochastic process algebra PEPA. In: Priami, C., Ingólfssdóttir, A., Mishra, B., Riis Nielson, H. (eds.) *Transactions on Computational Systems Biology VII*. LNCS (LNBI), vol. 4230, pp. 1–23. Springer, Heidelberg (2006)
3. Calder, M., Duguid, A., Gilmore, S., Hillston, J.: Stronger computational modelling of signalling pathways using both continuous and discrete-space methods. In: Priami, C. (ed.) *CMSB 2006*. LNCS (LNBI), vol. 4210, pp. 63–77. Springer, Heidelberg (2006)
4. Ciocchetta, F., Gilmore, S., Guerriero, M.-L., Hillston, J.: Stochastic Simulation and Probabilistic Model-Checking for the Analysis of Biochemical Systems (2008) (submitted for publication)
5. Ciocchetta, F., Hillston, J.: Bio-PEPA: an extension of the process algebra PEPA for biochemical networks. In: *Proc. of FBTC 2007*. *Electronic Notes in Theoretical Computer Science*, vol. 194(3), pp. 103–117 (2008)
6. Ciocchetta, F., Hillston, J.: Process algebras for Systems Biology. In: Bernardo, M., Degano, P., Zavattaro, G. (eds.) *SFM 2008*. LNCS, vol. 5016, pp. 265–312. Springer, Heidelberg (2008)
7. Danos, V., Feret, J., Fontana, W., Harmer, R., Krivine, J.: Ruled-based modelling of cellular signalling. In: Caires, L., Vasconcelos, V.T. (eds.) *CONCUR 2007*. LNCS, vol. 4703, pp. 17–41. Springer, Heidelberg (2007)
8. Gillespie, D.T.: Exact stochastic simulation of coupled chemical reactions. *Journal of Physical Chemistry* 81, 2340–2361 (1977)
9. Hillston, J.: *A Compositional Approach to Performance Modelling*. Cambridge University Press, Cambridge (1996)
10. Priami, C., Quaglia, P.: Beta-binders for biological interactions. In: Danos, V., Schachter, V. (eds.) *CMSB 2004*. LNCS (LNBI), vol. 3082, pp. 20–33. Springer, Heidelberg (2005)
11. Priami, C.: Stochastic  $\pi$ -calculus. *The Computer Journal* 38(6), 578–589 (1995)
12. Priami, C., Regev, A., Silverman, W., Shapiro, E.: Application of a stochastic name-passing calculus to representation and simulation of molecular processes. *Information Processing Letters* 80, 25–31 (2001)
13. Regev, A.: Representation and simulation of molecular pathways in the stochastic  $\pi$ -calculus. In: *Proceedings of the 2nd workshop on Computation of Biochemical Pathways and Genetic Networks* (2001)
14. Regev, A., Shapiro, E.: Cells as computation. *Nature* 419, 343 (2002)