Proceedings of the 2006 Winter Simulation Conference
L. F. Perrone, F. P. Wieland, J. Liu, B. G. Lawson, D. M. Nicol, and R. M. Fujimoto, eds.

BacGrid: LARGE SCALE SYSTEMS BIOLOGY SIMULATION ON THE GRID

Brian Logan
School of Computer Science
University of Nottingham
Nottingham U.K.

Michael Lees
School of Computer Science
University of Nottingham
Nottingham U.K.

John King
School of Mathematical Sciences
University of Nottingham
Nottingham U.K.

1 INTRODUCTION

Modelling and simulation form an integral part of a predictive and explanatory approach to computational systems biology. However, systems biology simulations are extremely challenging in requiring the modelling of many complex phenomena at multiple spatial and temporal scales. No single simulation can cover all the cases of interest. As a result, simulations must be developed "on demand" to suit each new systems biology problem, often drawing on the expertise of different research groups.

Such large scale collaborative model development presents significant challenges for the distributed simulation community. We believe these challenges can potentially be met by a combination of two emerging standards and their supporting middleware: the Grid and the High Level Architecture. The Grid supports e-Science through resource discovery, secure access to remote computational resources, data archiving and sharing etc., allowing virtual research teams to collaborate to solve research problems. The High Level Architecture (HLA) is an IEEE standard (IEEE 2000) for simulator interoperability, which supports the creation of distributed, composable simulations. The combination of these two complementary technologies offers great promise for the "on-demand" development of systems biology simulations by facilitating the reuse of existing simulation components. In addition, many of the services necessary to support dynamic composition of simulations, e.g., model discovery and matching, secure execution, migration and load balancing, sharing and archival of simulation results etc., which are not addressed by the HLA standard, can potentially be provided by the Grid infrastructure.

In recent work, some initial steps have been taken towards the composition and subsequent execution of simulation models on the Grid, e.g., HLA_Grid (Xie et al. 2005), XMSF (Pullen et al. 2005). However this work is still relatively immature and the appropriateness of the HLA for biological simulation has yet to be established. To evaluate the suitability of the HLA as a standard for systems biology simulations on the Grid, we are developing a prototype HLA-compliant, Grid-based simulator for systems biology which we call BacGrid. In this abstract, we outline the systems biology problem we have adopted as a test case, briefly describe the simulation model and the architecture of the simulator and summarise the current state of the implementation.

2 BIOFILM SIMULATION

In recent years major advances have been made in understanding the gene and signalling networks that control the behaviour of individual cells, and the need to understand the implications of these breakthroughs at the population level is increasingly widely recognised. Bacterial biofilms are sufficiently complex to exemplify many of the generic features of multi-cellular behaviour, without such complexity becoming overwhelming. They are in addition of enormous environmental, industrial and medical importance. Many processes in biofilms operate at the macroscopic scale and are thus susceptible to continuum modelling approaches. It is essential, however, that models incorporate in an appropriate way information about the micro-scale behaviour and their results must in turn be coupled back into the rules adopted in the cell-scale modelling, motivating the use of agent-based modelling.

3 MODEL DESIGN

BacGrid models a 3D biofilm reactor consisting of two compartments: bulk liquid and biofilm. The bulk liquid compartment contains a completely mixed solution of S different soluble substrates at constant concentrations. The biofilm compartement contains B different types of biomass which grow on a planar support inside a rectangular box with periodic x and y boundaries. The biofilm and bulk liquid compartments are in contact and exchange solutes only by diffusion. Bacteria, substrates and other material are assumed to be washed away once they reach the z boundary (detachment layer).

The computational domain is decomposed into *model regions*, each containing one or more volume elements or *voxels*. Pressure, substrate and Quorum Sensing Molecule (QSM) concentrations are assumed to be constant across a

voxel, and the upper bound on the size of a voxel is chosen so that the discrete values approximate the continuous case. Substrate and QSM diffuse between voxels and particles are transferred between voxels when there is a sufficient pressure difference. For reasons of computational efficiency, bacterial cells are aggregated into *particles*, each of which represents one or more cells (Picioreanu, Kreft, and van Loosdrecht 2004). The maximum size of a particle, i.e., its size at division, thus determines the resolution of the model. Particles are modelled as agents and implement a simple model of growth, division and displacement (Kreft, Booth, and Wimpenny 1998), and upregulation (e.g., the production of extracellular polymeric substances) in the presence of QSM (Ward et al. 2003).

4 SYSTEM ARCHITECTURE

BacGrid is implemented using the Mason agent toolkit (Luke et al. 2005) and the DMSO RTI 1.3NGv6. The Grid interface is based on HLA_Grid (Xie et al. 2005), which allows HLA-compliant simulators to be instantiated and linked using Grid services. HLA_Grid employs a Federate-Proxy-RTI architecture, in which different participants (clients) in the same simulation run their federate codes at their local sites, and the RTIExec and FedExec are executed at a remote site. Proxies act on behalf of the client federate's code and communicate with the proxies of other clients through the RTI. Federate codes and their respective proxies communicate with each other through Grid services and a Grid-enabled HLA library, which provides the standard HLA API to the federate codes, and translates RTI calls into Grid service invocations. HLA_Grid includes additional Grid services to support the creation of the RTI, discovery of federations, etc.

The BacGrid prototype adopts a spatial decomposition in which federates implement model regions. In addition there is a diffusion federate which handles diffusion of substrate and QSMs and a visualisation federate for run-time monitoring of simulation progress (and ultimately simulation steering). While the current prototype is relatively simple, it creates a starting point for the integration of other simulators and models, e.g., at the cellular level through the inclusion of different kinds of bacteria, and at the sub-cellular level through the inclusion of models of gene expression and signalling.

5 IMPLEMENTATION AND FUTURE WORK

At the current stage of the project, a non-distributed version of the simulator has been implemented to allow validation of the biofilm model. Our initial experiments have focused on the role of QSM in the development of bacterial colonies, and in particular how QSM inhibitors can be used to prevent the up-regulation of cells within a colony (Koerber et al. 2002).

Work on distributing the simulator (which involves integrating Mason with the HLA and reimplementing HLA_Grid for GT4) is well underway. In the longer term, we plan to investigate the integration of a more complex cell model and recent extensions to HLA_Grid which support model discovery. We will also evaluate the suitability of the HLA_Grid protocols for biological simulations, and in particular the routing space approach to data distribution management utilised by the HLA.

ACKNOWLEDGEMENTS

This work is supported by BBSRC project number BB/D006619/1 and by EPSRC projects GR/S82862/01, EP/C549406/1 and EP/C549414/1.

REFERENCES

IEEE 2000. IEEE Standard for modeling and simulation (M&S) High Level Architecture (HLA) — Framework and rules. IEEE. (IEEE Standard No.: 1516-2000).

Koerber, A. J., J. R. King, J. P. Ward, P. Williams, J. M. Croft, and R. E. Sockett. 2002. A mathematical model of partial-thickness burn-wound infection by *Pseudomonas aeruginosa*: Quorum sensing and the build-up to invasion. *Bulletin of Mathematical Biology* 64:239–259.

Kreft, J.-U., G. Booth, and J. W. T. Wimpenny. 1998. BacSim, a simulator for individual-based modelling of bacterial colony growth. *Microbiology* 144:3275–3287.

Luke, S., C. Cioffi-Revilla, L. Panait, K. Sullivan, and G. Balan. 2005. MASON: A multiagent simulation environment. *Simulation* 81 (7): 517–527.

Minson, R., and G. Theodoropoulos. 2004, June. Distributing RePast agent-based simulations with HLA. In *Proceedings of the 2004 European Simulation Interoperability Workshop*, Number 04E-SIW-046. Edinburgh: Simulation Interoperability Standards Organisation and Society for Computer Simulation International. Paper No. 04E-SIW-046.

Picioreanu, C., J.-U. Kreft, and M. C. M. van Loosdrecht. 2004, May. Particle-based multidimensional multispecies biofilm model. *Applied and Environmental Microbiology* 70 (5): 3024–3040.

Pullen, J. M., R. Brunton, D. Brutzman, D. Drake, M. Hieb, K. Morse, and A. Tolk. 2005. Using Web services to integrate heterogeneous simulations in a Grid environment. *Future Generation Computer Systems* 21:97–106.

Ward, J. P., J. R. King, A. J. Koerber, J. M. Croft, R. E. Sockett, and P. Williams. 2003. Early development and quorum sensing in bacterial biofilms. *Journal of Mathematical Biology* (47): 23–55.

Xie, Y., Y. M. Teo, W. Cai, and S. Turner. 2005, June. Service provisioning for HLA-based distributed simulation on the Grid. In *Proceedings of the Nineteenth ACM/IEEE/SCS Workshop on Principles of Advanced and Distributed Simulation (PADS 2005)*, 282–291. Monterey, CA, USA.