

Web-services for the biology community: the BASIS project

Colin S. Gillespie¹, Carole J. Proctor², Daryl P. Shanley², Darren J. Wilkinson¹,
Richard J. Boys¹, Thomas B.L. Kirkwood²

¹School of Mathematics & Statistics, University of Newcastle, Newcastle upon Tyne,
NE1 7RU, UK

²Henry Wellcome Laboratory for Biogerontology Research, School of Clinical and Medical
Sciences-Gerontology, University of Newcastle, Newcastle upon Tyne, NE4 6BE, UK

Abstract

BASIS is a UK e-Science pilot project which delivers a GRID enabled system that serves the biology of ageing research community by helping to integrate data and hypotheses from diverse biological sources. Our system is primarily designed to allow researchers to extend their working models from a qualitative to a quantitative level to enable in silico experimentation. Web-service technology offers the ideal platform to offer our tool in a distributed and flexible way.

1. Introduction

BASIS (Biology of Ageing e-Science Integration and Simulation system) is a UK e-Science pilot project which delivers a GRID enabled system that serves the biology of ageing research community by helping to integrate data and hypotheses from diverse biological sources[1]. The system is primarily designed to allow researchers to extend their working models from a qualitative to a quantitative level to enable in-silico experimentation. The project couples web-service technology with the latest methods in describing models in XML to offer a complete modelling environment in a distributed and flexible way.

2. Modelling Biological Processes

2.1 Modelling

In general, our models can be envisaged as networks of individual biochemical mechanisms, represented by a system of chemical equations, quantified by substrate and product concentrations and the associated reaction rates (see Figure 1). These models are described using the Systems Biology Markup Language SBML[2], which is essentially an eXtensible Markup Language (XML) encoding [3] of the reaction, species and compartment lists, together with the additional information required for quantitative modelling and simulation. Although there is still a great deal to add to the SBML specification, (for example it

is not easy to represent tissues composed of detailed cell models), the current version of SBML allows us to encode and distribute a large class of biochemical network models easily.

2.2 Virtual Ageing Cell

Current attention focuses on building a library of fully characterised SBML modules of key intracellular mechanisms relevant to ageing, such as telomere degradation, the role of chaperones and reactive oxygen species. The BASIS system allows users to ‘pick and choose’ components and construct a virtual ageing cell. Furthermore, it allows the user to share and simulate their model through a web-services interface. As the model database grows, it will provide a valuable resource for the ageing research community that will complement other pathway databases such as KEGG and BRENDA. Although the present focus is on intracellular interactions[4, 5], once the framework is in place we will extend the models to tissues and eventually a whole organism.

2.3 Sharing Models

One of the aims of the BASIS project is to allow users to build models “in private” and once complete, share their models with other users. BASIS assigns each model an *URN* of the form

urn:basis.ncl:model:id#

Once a model has been made public, further alteration is not possible.

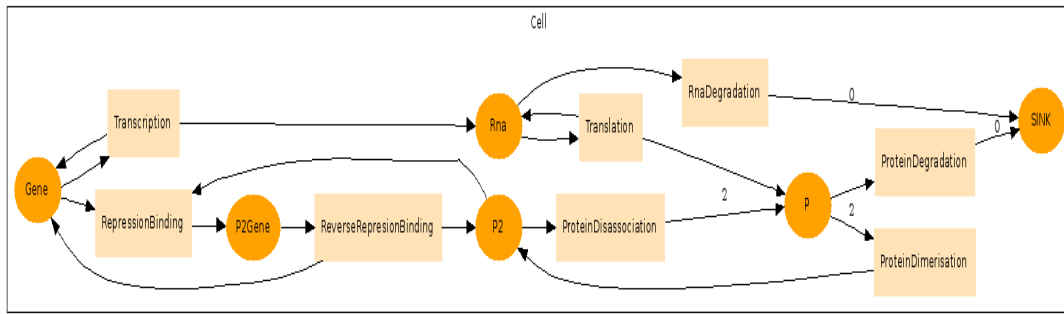


Figure 1: Diagram of a simple biochemical network.

3. The BASIS System

3.1 General Outline

The BASIS system of model definition, simulation and visualisation is exposed through several web-services that are served through Apache (see Figure 2).

To provide an initial entry point to the BASIS system we have constructed a user-friendly web portal interface for simple model adjustment and to demonstrate the range of services available [6].

The web-services interact with a postgresSQL[7] database and the job scheduler Condor[8]. Again, all details of the underlying technology are hidden from the user.

3.2 Web-services Using Python

All web-services provided by BASIS are dispatched by Python [9] using the ZSI library [10]. Provided that an apache web-server is up and running then a basic web-service can be set up with less than four short lines of codes, e.g.

```

from ZSI import dispatch
def helloWorld():
    return 'hello world'
dispatch.AsCGI()

```

Furthermore, Python is ideally suited for interacting with databases so progressing to more complex web-services is straightforward. However, this functionality comes at cost. First, due to Python being dynamically typed, there is no tool available that will automatically construct a WSDL file from Python code. Although hand-writing a WSDL file for simple web-services is straightforward, for more complex web-services this is not desirable. Second, Python ZSI does not yet support WS-Security, so we are currently investigating 'Java-

wrapped' web-services and the GT4 python bindings so that we can incorporate advanced web-service functionality. Currently our services are protected by a basic level of security based on secure sockets, username, passwords and sessions ids.

3.3 Simulation Through Web-services

At present users can submit jobs through a web-service which will run on our cluster. The job queue is managed with a combination of Condor and PostgreSQL. The database manages the amount of jobs a user can submit at one time whereas Condor manages the order that the jobs submitted are executed. A special feature of BASIS is that it provides a sophisticated stochastic simulation service, as stochastic effects are particularly import in the context of ageing research. Since our cluster has currently only sixteen nodes the maximum number of concurrent users is limited. One future

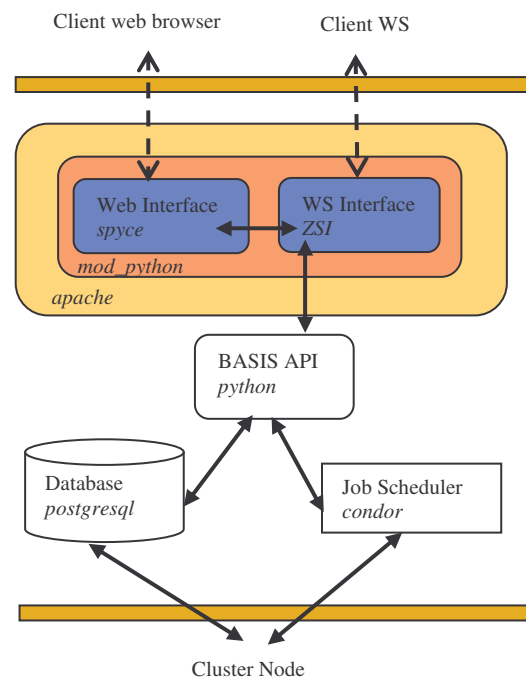


Figure 2: A diagram of the BASIS architecture.

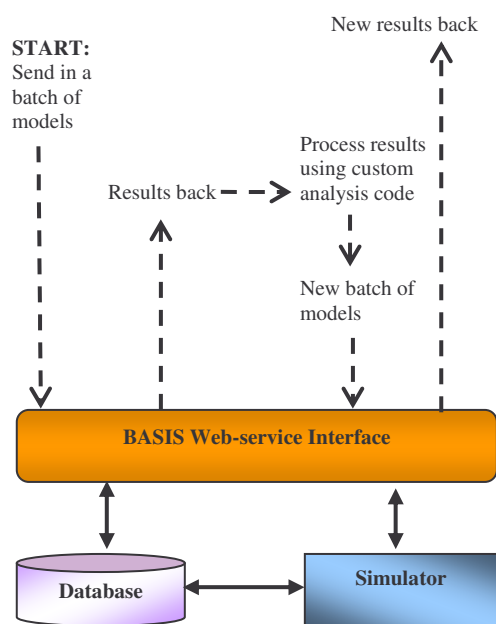


Figure 3: Example calls to the BASIS system.

possibility is for us to join the ‘Newcastle University Grid.’ This Grid will essentially link all the unused computing resources in Newcastle and allow jobs to be scheduled when free resource time is available.

4. Using the BASIS System

The benefit of building BASIS using web-services is that it provides a number of services that can be easily combined by a user to provide a flexible platform for biological modelling.

For example BASIS can be used to perform a simple parameter scan. A general algorithm for this is:

1. Submit a model to BASIS;
2. Use the alterSBML web-service to generate models with different parameters;
3. Call the Simulate web-service;
4. Get the Results.

A key point here is that it is the user who decides exactly how the analysis should be carried out.

This simple example can easily be extended to include a degree of automation. For instance a parameter scan could be carried out on the model and the results used to determine if further scans (or other analysis) are necessary (see Figure 3). This is the general idea behind the recently initiated CaliBayes[11] project. Currently, and for the foreseeable future, biological data is stored in many different data repositories. However, web-service access for these databases is now on the increase. So by pulling the data from these sources and

combining with a Bayesian Calibration engine and a number of BASIS web-services, parameters can be estimated from the data in an efficient fashion.

Overall, BASIS does not provide a single solution to all modelling problems. Rather, it provides a set of flexible tools and that allows each user to determine how best to tackle their individual and often unique problem.

Acknowledgements

We thank the BBSRC, MRC, DTI and Unilever for financial support.

References

1. Kirkwood, T.B.L., et al., *Towards an e-biology of ageing: integrating theory and data*. Nature Reviews Molecular Cell Biology, 2003. **4**(3): p. 243-249.
2. Hucka, M., et al., *The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models*. Bioinformatics, 2003. **19**(4): p. 524-531.
3. DuCharme, R., *XML: The Annotated Specification*. 1999: Prentice Hall PTR, Upper Saddle River, NJ.
4. Proctor, C.J., et al., *Modelling the actions of chaperones and their role in ageing*. Mechanisms of Ageing and Development, 2005. **126**(1): p. 119-131.
5. Gillespie, C.S., et al., *A mathematical model of ageing in yeast*. Journal of Theoretical Biology, 2004. **229**(2): p. 189-196.
6. BASIS, <http://www.basis.ncl.ac.uk/>.
7. PostgreSQL, <http://www.postgresql.org>.
8. Thain, D., T. Tannenbaum, and M. Livny, *Condor and the Grid*, in *Grid Computing: Making the Global Infrastructure a Reality*, F. Berman, G. Fox, and T. Hey, Editors. 2002, John Wiley & Sons Inc.
9. Python, <http://www.python.org>.
10. ZSI, <http://pywebsvcs.sourceforge.net/>.
11. CaliBayes, *Integration of GRID-based post-genomic data resources through Bayesian calibration of biological simulators* (<http://www.calibayes.ncl.ac.uk/>).