Chapter 3.4
BioSimGrid Biomolecular Simulation Database

Kaihsu Tai
University of Oxford, UK

Mark S. P. Sansom
University of Oxford, UK

ABSTRACT

BioSimGrid is a distributed biomolecular simulation database. It is a general-purpose database for trajectories from molecular dynamics simulations. Though initially designed as a distributed data grid, BioSimGrid allows for installation as a stand-alone instance. This can later be integrated into a wider, networked system. This presentation of BioSimGrid follows a scenario in biological research to demonstrate how to install the system, and how to deposit, query, and analyze trajectories in this system, with real Python code examples for each step. What then follow are explanations of the underlying concepts in the implementation of BioSimGrid: relational database, distributed computing, and the input/output (deposit and analysis) modules. Finishing the presentation is a discussion of the emerging trends and concerns in the further development of BioSimGrid and similar biological databases. This discussion touches on quality-assurance issues and the use of BioSimGrid as a back-end for other speciality databases. The experience of developing BioSimGrid compels the conclusion: In the development and maintenance of biomolecular simulation databases, it is essential that sustainability be asserted as a key principle.

DOI: 10.4018/978-1-60960-195-9.ch304

Copyright © 2011, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

1. INTRODUCTION AND BACKGROUND: A REPOSITORY OF BIOMOLECULAR SIMULATIONS

Since the first application of the molecular dynamics on proteins in 1976 (Adcock and McCammon, 2006), this simulation methodology has added value to experimental structural biology by making biomolecules 'come alive' and by compensating in the nanosecond time-scale where experimental methods are only beginning to be able to access. Adding to this, we have the method of comparison, a precursor to the process of classification, which is fundamental to biology (Brooks and McLennan, 2006). Insights into the internal motions of proteins can come from comparing the results of molecular dynamics simulations, namely the trajectories (Pang et al., 2005; Tai et al., 2007). This process can be facilitated by having a database of trajectories. We have developed in the past few years (2003 to 2006) such a database called BioSimGrid (http://www.biosimgrid.org/; Tai et al., 2004).

Comparable endeavours to BioSimGrid include the Ascona B-DNA Consortium (http://humphry.chem.wesleyan.edu:8080/MDDNA/; Dixit et al., 2005), Dynameomics (http://www.dynameomics.org/; Scott et al., 2007), GEMS (http://gipse.cse.nd.edu/GEMS/; Wozniak et al., 2005), and SimDB (http://simdb.bch.msu.edu/; Feig et al., 1999). To differentiate, BioSimGrid is a general-purpose database for trajectories from molecular dynamics simulations. It is free software licensed under the terms of GNU General Public License (Stallman 2002). It can take advantage of distributed ('grid') computing, to enhance reliability and ensure longevity of the trajectory content (Berman et al., 2003a).

By 'general-purpose', we mean the following. Firstly, BioSimGrid can admit trajectories generated by different simulation packages, such as Amber (Pearlman et al., 1995), Gromacs (Lindahl et al., 2001), Charmm (Brooks et al., 1983), NWChem (Straatsma et al., 2000), and NAMD (Kalé et al., 1999). Secondly, BioSimGrid is not restricted to a special kind of system or granularity. It can host simulations for nucleic acids, proteins, small molecules, or even non-biological polymers; simulations at the all-atom level or coarse-grain molecular dynamics (Bond and Sansom, 2006; Marrink et al., 2004; Nielsen et al., 2004). It can also store non-sequential 'trajectories', or rather ensembles, generated by other methods such as Monte Carlo, homology modelling (Šali and Blundell, 1993), and CONCOORD (de Groot et al., 1997).

Here we briefly describe the implementation of BioSimGrid, so the reader can have a conceptual understanding of the system: aspects such as installation, interfaces, architecture, data schema, and the deposit and analysis modules. We present an end-to-end case scenario where a scientist can deposit a trajectory into BioSimGrid, query the database, and analyze a trajectory. Finally, we discuss the prospects of such a database: the quality-assurance and sustainability issues, and customization of BioSimGrid as 'back-ends' to specialist databases.

2. BIOSIMGRID: INSTALLATION AND INTERFACES

As we introduced above, BioSimGrid is a distributed database for biomolecular simulations. The BioSimGrid system was initially designed to be a distributed system, now deployed over a internet-based consortium including universities of Oxford, Southampton, Bristol, Nottingham, and York. A node has been added in the Pacific Northwest National Laboratory in Richland, Washington.

In the larger grid context, the National Grid Service (NGS, United Kingdom; http://www.grid-support.ac.uk/) provides the Storage Resource Broker (SRB) as a middleware service. Since Many large-scale grids elsewhere in the world also provide such a service for SRB. Since SRB is the

15 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/biosimgrid-biomolecular-simulation-database/49409

Related Content

Comparison of Image Decompositions Through Inverse Difference and Laplacian Pyramids
Roumen Kountchev, Stuart Rubin, Mariofanna Milanovaand Roumiana Kountcheva (2015). *International Journal of Multimedia Data Engineering and Management (pp. 19-38).*
www.irma-international.org/article/comparison-of-image-decompositions-through-inverse-difference-and-laplacian-pyramids/124243

Virtual Learning Environment Blends
Robert J. McClelland (2008). *Handbook of Research on Digital Information Technologies: Innovations, Methods, and Ethical Issues (pp. 324-344).*
www.irma-international.org/chapter/virtual-learning-environment-blends/19851

Peer-to-Peer Networks: Protocols, Cooperation and Competition
Hyunggon Park, Rafit Izhak Ratzinand Mihaela van der Schaar (2011). *Streaming Media Architectures, Techniques, and Applications: Recent Advances (pp. 262-294).*
www.irma-international.org/chapter/peer-peer-networks/47522

Interactive Multimedia Technologies for Distance Education Systems
Hakikur Rahman (2005). *Encyclopedia of Multimedia Technology and Networking (pp. 454-460).*
www.irma-international.org/chapter/interactive-multimedia-technologies-distance-education/17283

Rank-Pooling-Based Features on Localized Regions for Automatic Micro-Expression Recognition
Trang Thanh Quynh Le, Thuong-Khanh Tranand Manjeet Rege (2020). *International Journal of Multimedia Data Engineering and Management (pp. 25-37).*
www.irma-international.org/article/rank-pooling-based-features-on-localized-regions-for-automatic-micro-expression-recognition/267765