The BioMAZE System: an environment for representation and analysis of biological processes.

Christian LEMER[1], Hassan ANERHOUR[1], Fabian COUCHE[1], Simon DE KEYZER[1], Yves DEVILLE,[2]
Frederic FAYS[1], Olivier HUBAUT[1], Jacques VAN HELDEN[1], Shoshana WODAK[3], Jean RICHELLE.[1]

1. Service de Conformation des Macromolecules Biologiques et de Bioinformatique - ULB
2. Department of Computing Science and Engineering - UCL
3. Centre for Computational Biology - The Hospital for Sick Children - Toronto

Abstract

The BioMAZE System is a workbench for the representation and analysis of networks of molecular interactions and cellular processes, genetic expression and regulation, enzymatic transformations and regulation, metabolic pathways, signal transduction. The BioMAZE system integrates the aMAZE database and various tools for analyzing and visualizing the stored information.

Introduction

The BioMAZE system in an integrated solution to the problem of representing and analyzing rich information on complex heterogeneous processes. The BioMAZE system integrates the aMAZE database, as well as various tools for visualizing and analyzing the stored information.

BioMAZE workbench

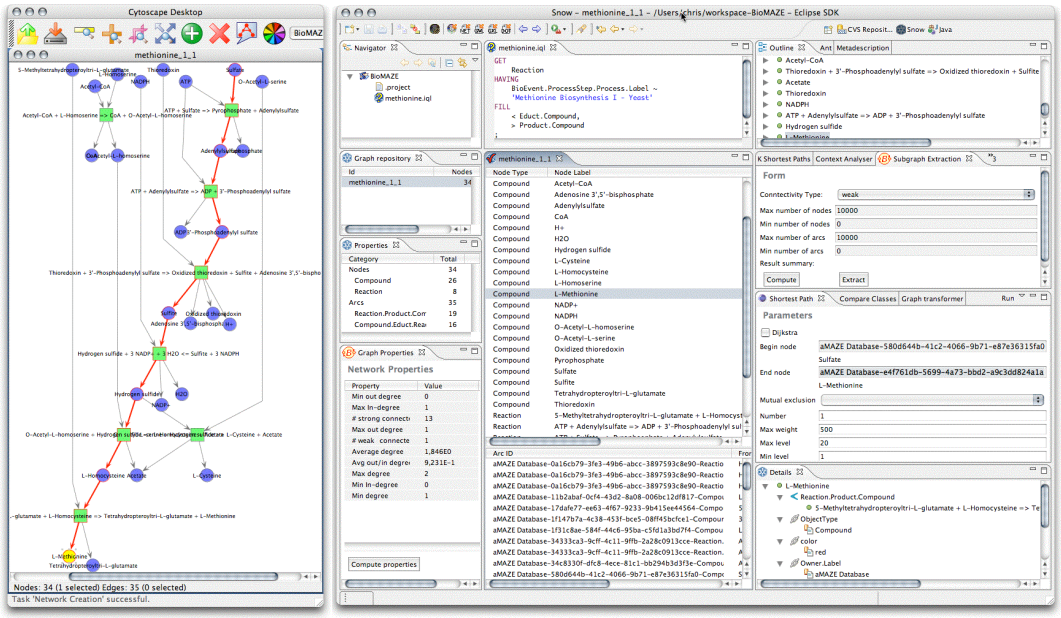
The architecture of the BioMAZE workbench is highly modular. The flexibility of the system stems from the integration of these tools as plugins into the Eclipse environment. It's Application Programming Interface is public, enabling programmers to implement their own plugins to extend the system and provide functionalities in a seamless way.

The main sub-components are: the Snow system, a workbench for graph management; the Igloo database management system dedicated to network data; the VisualBioMAZE component, a biochemical networks visualization tool and the BioEdge components, a collection of tools dedicated to the analysis of biological networks.

The Snow system is the kernel of the workbench, interconnecting the different tools and providing the basic features. Snow is in charge of data import and export, basic graph edition functionalities, data browsing. It also provides the user interface for the Igloo DBMS.

Igloo database managment system

Igloo, is the database management layer in charge of querying and editing (creating and modifying) the entities stored in the database. It also handles the process of loading and annotation of the data. In addition, Igloo provides an Application Programming Interface (API) and the Igloo Query Language (IQL) for database interrogation. Moreover, Igloo hides the underlying data storage organization, which is a PostgreSQL relational database. The data model is not hard-coded inside the application but is retrieved from an external repository. Igloo can therefore be used in many other application domains, by simply retrieving alternative data models. It is furthermore an independent database management system available as a standalone Java library.

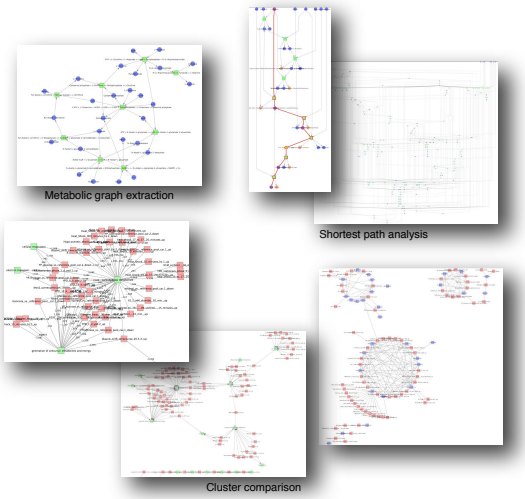


aMAZE Database

Until recently, there were no readily accessible universal sources for pathway information. For a long time, the data content of the aMAZE database was assembled by semi-automatic data loading from diverse external sources and through manual curation. Recently this situation has evolved with the adoption of the exchange format developed by the BioPAX[1] consortium. This consortium is a collaborative effort to develop a file format suitable for exchanging data on biological pathways. The BioPAX format now covers metabolic pathways and protein-protein interaction and will soon cover signal-transduction data. It is expected to evolve into the accepted standard for data exchange format for biochemical networks. Several pathway resources (KEGG, MetaCyc, Reactome) already output data in BioPAX format and most other database providers (TranPath, IHNO and others) are currently working on such export software. The BioMAZE database is able to import data provided in BioPAX format and correctly merge them using the provided unification information.

BioEdge analysis tools

The BioEdge components implement the graph analysis functionalities. BioEdge is organized as an extendable set of Eclipse plugins, providing Eclipse views within the BioMAZE workbench. Each view is thus an independent tool for analyzing biochemical network. The tools can be combined, enabling the output of one analysis to be processed further by other tools. The tools currently available in BioEdge are path and graph properties, context explorer, constrained k-shortest path extractor, sub-graph extraction, constrained path and sub-graph finding, path and graph matching, approximate path and graph constrained matching, motif extraction and analysis.



Demo

The BioMAZE System queries and analysis will be demonstrated on biological use cases at the demos session.

Acknowledgments

This work takes place at the Service de Conformations des Macromolécules Biologiques et de Bioinformatique. (SCMBB) of the Université libre de Bruxelles, has been supported by the Région de Bruxelles-Capitale (Belgium) and is supported the Région Wallonne (Belgium).

1 BioPAX: The Biological PAthways eXchange project: http://www.biopax.org/