Open Computing GRID for Molecular Science and Engineering
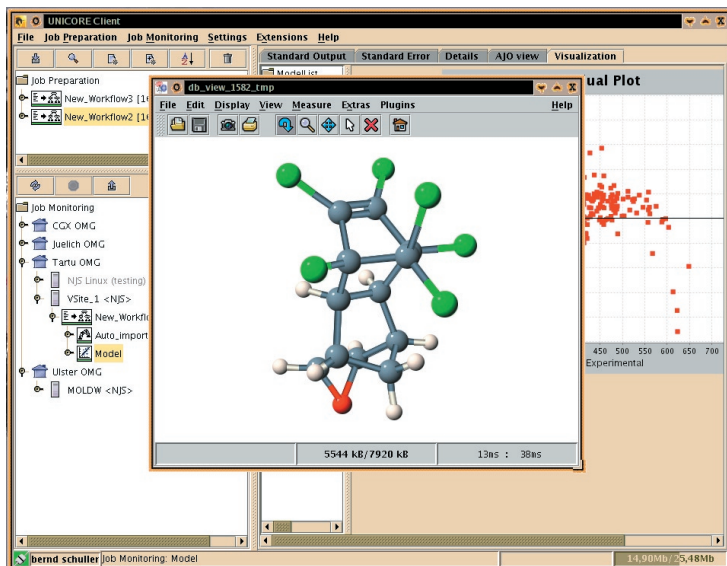
OpenMolGRID

Scope

The OpenMolGRID project has followed a Grid approach to molecular engineering problems. The project builds an automated, human controlled drug design pipeline. The methodology that has been used relies on the Quantitative Structure Property/Activity Relationship (QSPR/QSAR) models.

OpenMolGRID used the UNICORE Middleware to integrate the various applications that are needed for automated and distributed molecular engineering. These include databases and molecular engineering and prediction modules. Building on the basic Grid Middleware, substantial functionality has been added to the client and the abstraction layers that are used to hide the system complexity from the user. UNICORE Middleware has been extended with new features to support some required functionalities.

Innovation and Functionality

Grid-enabled data warehousing: The OpenMolGRID data warehousing repository provides pre-computed data in order to improve the efficiency and effectiveness of subsequent automatic and semiautomatic data analysis and data mining operations. By providing cached computations of frequently used and computationally expensive data processing tasks, it offers value for all OpenMolGRID users. In order to support the data warehouse's data transformation and pre-computation tasks, a command-line interface and an application programming interface have been developed for UNICORE.

Contract number
IST-2001-37238

Type of project
Cost-shared Research and Technological Development Action

Project coordinator
Forschungszentrum Juelich GmbH

Contact person
Mathilde Romberg
Forschungszentrum Juelich GmbH
ZAM
52425 Jülich
Germany

m.romberg@fz-juelich.de

Project website
http://www.openmolgrid.org/

Maximum Community contribution to project
EUR 1990 000

Project start date
1 September 2002

Duration
30 months

Information Society

European Commission

Application-neutral workflow support: The UNICORE client is extended by a workflow component which uses application metadata from both the server- and client-sides, to automatically generate UNICORE jobs from a workflow description. This includes the splitting of data-parallel tasks onto multiple target systems and assigning appropriate resources for each step. Molecular engineering is the first application to make use of this: QSPR/QSAR analysis uses data mining methods to develop predictive models by analysing relationships between properties/activities and molecular structures. The model development process is carried out in a workflow that involves several time-consuming computation steps.

Integrated tool for chemists, drug developers, and toxicologists: OpenMolGRID also provides design principles for the next-generation of molecular engineering tools. Molecular engineering is the design of molecular compounds and materials with predefined target properties. The design of molecular compounds involves a large number of calculations in a single design task and, thus, the Grid approach becomes vital.

Positioning

Within the Grid related aspects of OpenMolGRID, data and compute services have been developed for building an environment for solving molecular design and engineering problems. These use UNICORE as the underlying Grid Middleware in a distributed, collaborative environment.

Target Users and User Benefits

The OpenMolGRID project is one of the first realisations of technology in the field of drug design and development. Property prediction, specifically the modelling of toxicity, which is a part of the drug design pipeline, is used by chemical modelling and environmental scientists. Besides this, the project will lead to shortened time-to-solution for drug design and molecular engineering tasks. In addition it will shorten the times needed for the decision making involved in the design of drugs and the development pipeline. The results of the project will lead to an integrated and secure Grid solution for the pharmaceutical industry.

Maturity and Availability of Tools

The developed system is fully functional and is used in real-life test cases within the project. The specification documents are available for download from the project's web site. The developed software, mainly the extensions to UNICORE and the abstract application/data interfaces, will be available as open source (under the BSD, GPL, and LGPL licenses) at: http://unicore.sourceforge.net, after the end of the project. Discussion fora for users and administrators are supported by Sourceforge and are already used for UNICORE. The integrated applications

(e.g. descriptor calculation, structure enumeration, structure optimisation) that are owned by third parties, are available under their specific licenses.

Compliance with Standards

OpenMolGRID is based on, and all its developments are fully compliant with, the UNICORE Middleware. The defined XML workflow description schema reflects the capabilities of UNICORE's Abstract Job Object (AJO) architecture.

Interoperability

The OpenMolGRID architecture is developed on top of the UNICORE infrastructure, which is based on a client-server model. A UNICORE client plugin is developed for each specific task and application wrappers are developed for available software packages that can carry out these task.

Value-Added Services and Next Generation Development

An important aspect in the project is the dissemination and exploitation of the OpenMolGRID results to support drug design and other complex application pipelines in science and industry.

The Grid-related achievements, in the form of additions to the UNICORE software, will be taken up by new European Grid projects such as the DEISA and UniGridS projects. The European "Quality of Life" projects DEMETRA and FATEALLCHEM have or are currently exploiting the OpenMolGRID software to develop dose-response relationships and predictive models.

Project Partners

Organisation name and country

COMGENEX	HU
FORSCHUNGSZENTRUM JUELICH GMBH	DE
ISTITUTO DI RICERCHE FARMACOLOGICHE "MARIO NEGRI"	IT
UNIVERSITY OF TARTU	EE
UNIVERSITY OF ULSTER	UK