OpenMolGRID: QSAR/QSPR Applications in Grid Environment

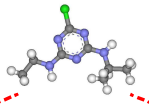
Sulev Sild, Andre Lomaka, Uko Maran
University of Tartu, Estonia

Outline

- QSAR/QSPR
  - Background
  - Challenges
- OpenMolGRID
  - Integrating applications
  - Automating workflows
- Summary

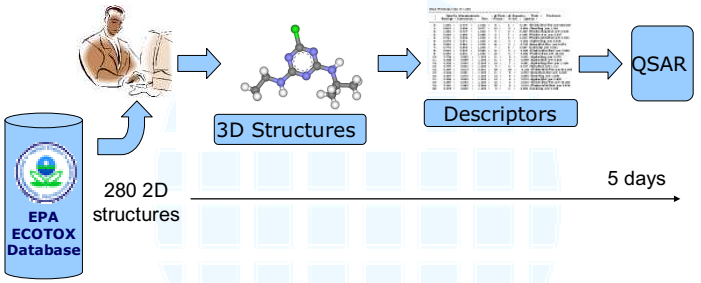
Background

QSAR/QSPR—Same methodology, but different targets



Biological activity	Chemical property
Lethal dose	Boiling point
Skin irritation	Melting point
Carcinogenicity	Viscosity
…	…

Typical workflow



There must be a better way!!!

Challenges in QSAR/QSPR

- Large data sets are computationally demanding
- Manual preparation of input data from multiple data sources (distributed and heterogeneous)
- Multiple programs are involved in a complex workflow
- Integration and automation required to speed up the process

OpenMolGRID

- The Open Computing Grid for Molecular Science and Engineering
- OpenMolGRID aims to provide a standardised environment to help
  - Predict properties of chemicals
  - Predict new chemical compounds that exhibit specific properties (reverse engineering)
- Data warehousing, data mining and the Grid are key technologies

Collaboration

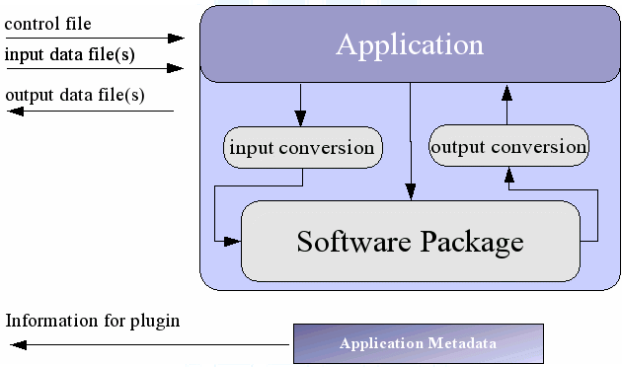
Testbed based on UNICORE

UNiform Interface to COmputing Resources

Extensions to UNICORE

- Seamless integration of legacy applications
- Support for scientific workflows
- Data warehouse for molecular data with distributed data transformations
- Data access tool
- Command line interface

Integrating applications



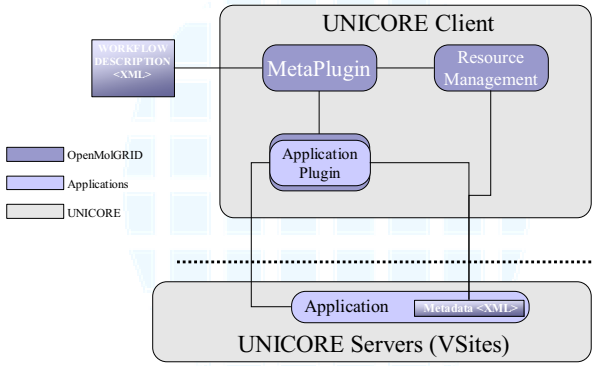
Integrated applications

- 2D to 3D conversion
- Semi-empirical quantum chemical calculations
- Molecular descriptor calculation
- Model Building
- Property/activity prediction
- Structure generation

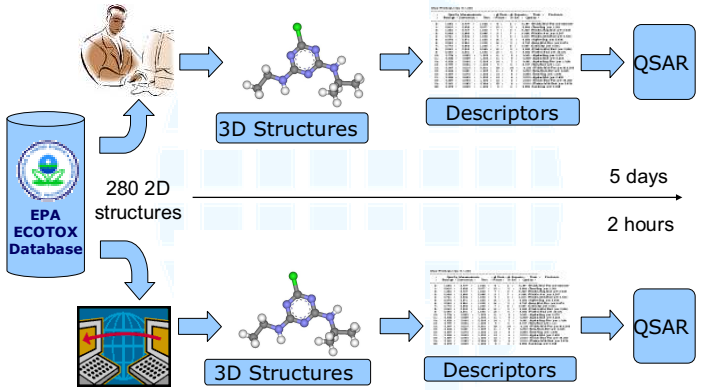
OpenMolGRID workflows

- Generate concrete UNICORE jobs from abstract workflow description
- Resource management
  - Automatic transformation of input/output formats
  - Automatic selection of target sites and application/data resources
  - Distribute tasks to multiple Vsites
  - Automatic transfer of data between Vsites

Workflow support



OpenMolGRID in action



Summary

- Grid computing is very suitable for the QSAR/QSPR applications
- Proper data management and workflow automation has a huge time saving potential
- UNICORE middleware has good potential for standardisation, automated scientific workflows, and integration of legacy applications

The End

Thanks!