HelmholtzZentrum münchen
German Research Center for Environmental Health

CADASTER

CAse studies on the Development and Application of in-Silico Techniques for Environmental hazard and Risk assessment

MOPAC@home - a structural database for small chemical compounds

Stefan Brandmaier, Vijyant Srivastava, Igor V.Tetko, Robert Körner, Wolfram Teetz, Sergii Novotarskyi, Anil Pandey, Iurii Sushko

Institute of Bioinformatics & Systems Biology, 85764 Neuherberg, Germany

Introduction

Molecular descriptors derived from 3D structures, in particular quantum chemistry descriptors, are frequently reported as a useful representation of chemical compounds, that can significantly improve the quality of QSAR models to predict physico-chemical and biological properties, as well as toxicity of chemicals.

Unfortunately, the results of these studies are often not traceable and thereby not assessable, as the 3D structures of the molecules, that were used to derive the models, are not accessible. Another obstacle, that makes it difficult to work with 3D structures is, that computing global minimum energy conformations can be very time expensive, especially for large and therefore flexible molecules.

We introduce an online database http://cadaster.eu/mopac that optimizes molecular conformations with the semi-empirical AM1 algorithm from the MOPAC7 package. For each molecule in the database not only one structure is available, but a whole set of them, derived from different initial conformations.

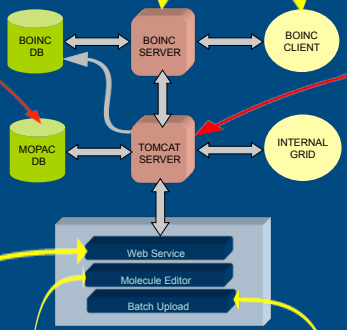
BOINC Server

BOINC (Berkeley Open Infrastructure for Network Computing) is an open-source software platform for computing using volunteered resources.

- allows grid computing
- contains of a a server and client(s)
- computational problems, that are paralelizable can be distributed to several computers.

BOINC Client

- enables the connection to multiple BOINC projects
- users with internet access share their computational power
- download, processing and upload of tasks is automatized



Workflow

Six parallel cronjobs

- **Processing of user submitted conformations**
  - one conformation calculated immediately
- **Additional conformations**
  - utilizing OpenBabel, Obgen, Balloon
  - comparison with included structures to avoid redundancies
- **Submitting to BOINC server**
  - Boinc workunits are created and distributed to clients
  - binaries, wrappers and tasks are automatically sent to the clients
  - Computation of optimized structures uscng MOPAC AM1 algorithm
- **Collection of results**
  - automatic recalculation of tasks failing for known errors
  - optimized conformations for each molecule are stored to database
- **Calculation of quantum chemical descriptors**
  - Mopac descriptors are calculated and available for model building
- **Validation of Work**
  - check for stereochemical consistency (utilizing InChI key)
  - comparison of total energy

OCHEM

- 245398 measured values from literature
- Implementation of many machine learning algorithms
- Calculation of 20 different descriptor sets
- Various filtering options
- Application of peer reviewed published models
- All experimental conditions can be added to a record (unique feature)

Web interface

- pload of molecules or specific conformations
- Browsing for:
  - conformations with global minimum energy
  - local minimum energy from user provided conformation
- Support of stereo-chemistry

WebService

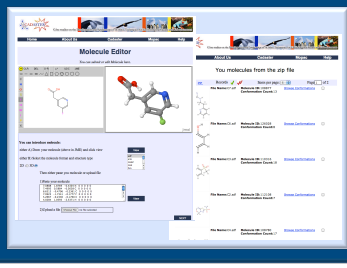
- Soap based web service
- Client is available for download
- Providing:
  - SDF file with optimized structure and partial charges
  - MOPAC output file with quantum chemistry descriptors

Molecule Editor

- Introduction via:
  - Drawing in JME
  - Pasting into text box
  - Uploading files
- Supports all molecule formats recognized by OpenBabel
- 3D-depiction in Jmol
- Browsing of stereoisomers
- Export in several formats

Batch Upload

- Upload multiple molecules for optimization
- Support for:
  - Compression formats (zip, gzip, tar)
  - Multi molecule files or folders
- friendly interface



Results

Features

- Many ways to access data
- Free access
- Revision of local minimum energy conformations
- Optimization of several initial conformations
- User upload of new molecules or conformations
- Support for stereo-chemistry
- Sophisticated algorithm to optimize structures
- Quantum chemistry descriptors

Numbers

MOPAC@home contains:

- 125.000 different molecules
- 240.000 different stereoisomers
- 290.000 user uploaded conformations
- 1.270.000 optimized conformations

Conclusion

MOPAC@home is developed to be a database of optimized conformations derived by using the MOPAC7 tool and utilizing distributed volunteer computing using BOINC software.

The provided WebService is a convenient way to integrate the structures in any application (eg. model building). Both the structures and the implicitly provided quantum chemistry descriptors can be an important contribution to advance environmental toxicity prediction and drug discovery. Furthermore, the optimized conformations produced by this database could be used as a reference for QSAR/QSPR models, which depend on 3D descriptors.

HELMHOLTZ ASSOCIATION