TECHNISCHE UNIVERSITÄT DRESDEN

Zentrum für Informationsdienste und Hochleistungsrechnen (ZIH)

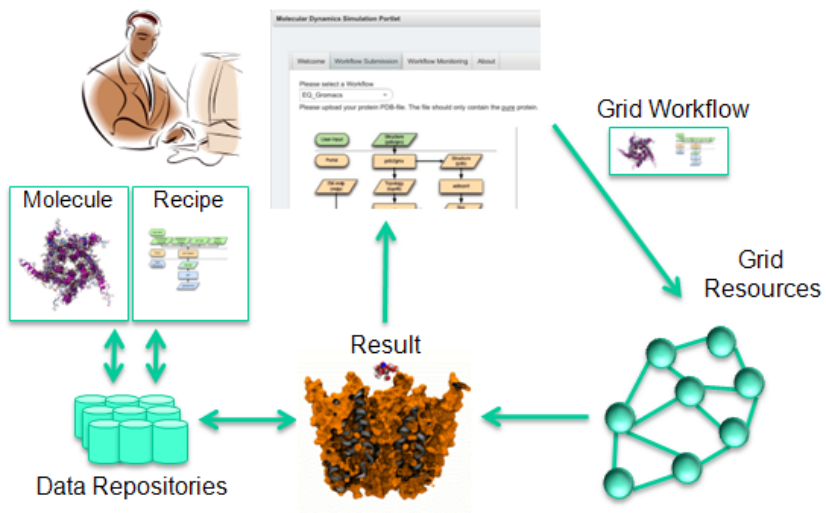
A Data Driven Science Gateway for Computational Workflows

Richard Grunzke (richard.grunzke@tu-dresden.de)

ZIH
Zentrum für Informationsdienste und Hochleistungsrechnen

Introduction

- MoSGrid Science Gateway - Simple and efficient use of complex chemical applications via a portal integrated grid infrastructure



- More simple access to applications via a user friendly **Portal**
- More simple use of complex molecular simulation via **Unified Integration**
- Support work of chemists via linking of applications with **Grid Workflow Management**
- Support re-use of data and workflows via **Data Repository**

Academic Partners

Tübingen, Bioinformatics

- Oliver Kohlbacher
- Jens Krüger
- Sandra Gesing

Berlin, Zuse Institut

- Thomas Steinke
- Patrick Schäfer

Dortmund & Munich, Inorganic Chemistry

- Sonja Herres-Pawlis
- Ines Dos Santos Vieira

Köln, RRZ

- Lars Packschies
- Martin Kruse
- Klaus Warzecha

Dresden, ZIH

- Ralf Müller-Pfefferkorn
- Richard Grunzke

Paderborn, PC2

- André Brinkmann
- Georg Birkenheuer
- Johannes Schuster

Köln, Organic Chemistry

- Dirk Blunk
- Sebastian Breuers

... and many associated partners.

Industrial Partners

GETLIG & TAR

COSMOlogic GmbH & Co. KG

Computational Chemistry and Fluid Thermodynamics

ORACLE®

BioSolveIT

Bayer Technology Services

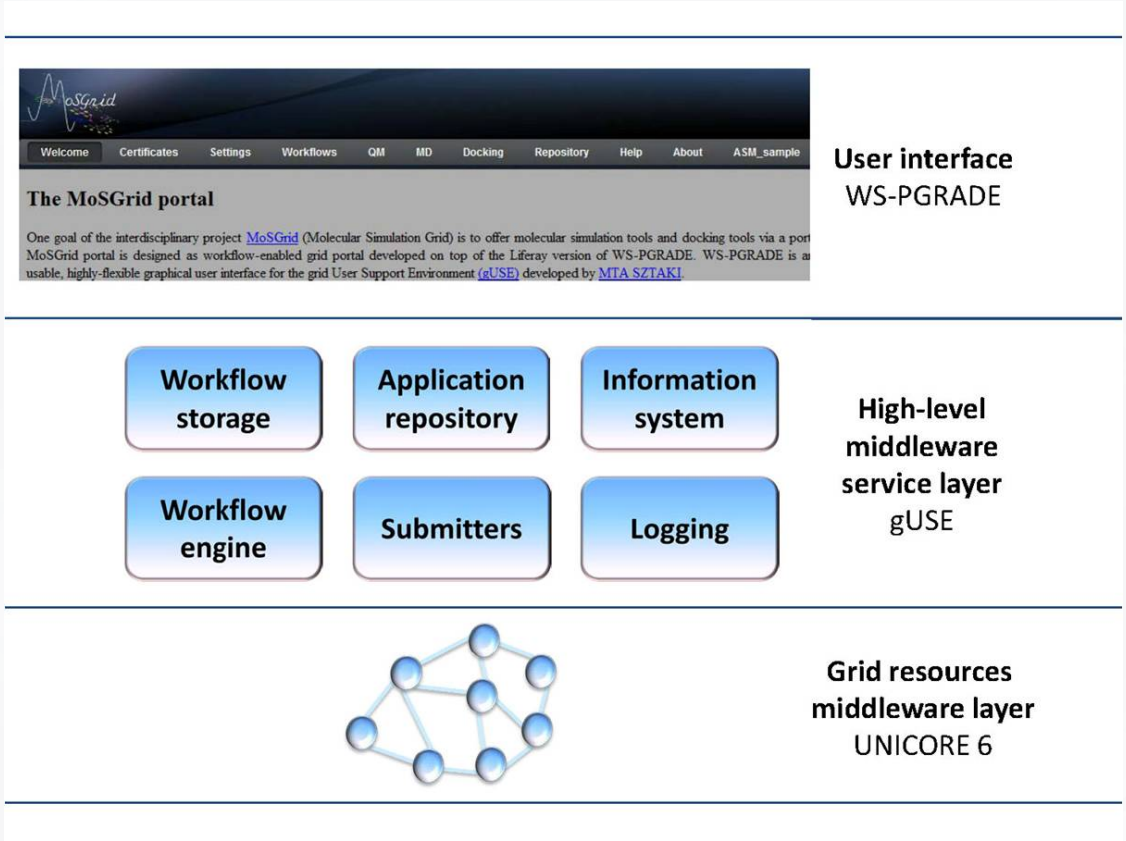
Partners

Background - Liferay

- Open source portal framework
- Widely used in grid and cloud projects
- Supports JSR168 and JSR286 standards
- Highly flexible
- Apache Tomcat as application server

Background – WS-PGRADE/gUSE

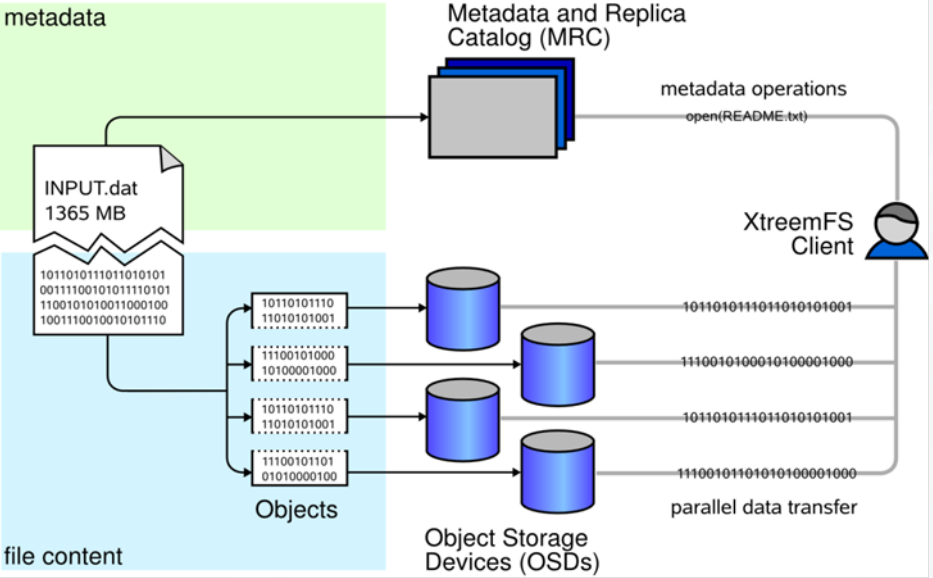
- gUSE (grid User Support Environment)
  - Services providing a collaborative application development environment for DCIs
- WS-PGRADE (Web services Parallel Grid Runtime and Developer Environment)
  - Highly flexible graphical user interface of gUSE



Background - UNICORE

- Standards-based grid middleware for secure access to federated compute and data resources
- Used in MoSGrid for
  - Job submission and management
  - Data management
  - Metadata management
- Metadata management based on Apache Lucene
- Metadata stored on same storage in JSON format
- Lucene used to index and search metadata

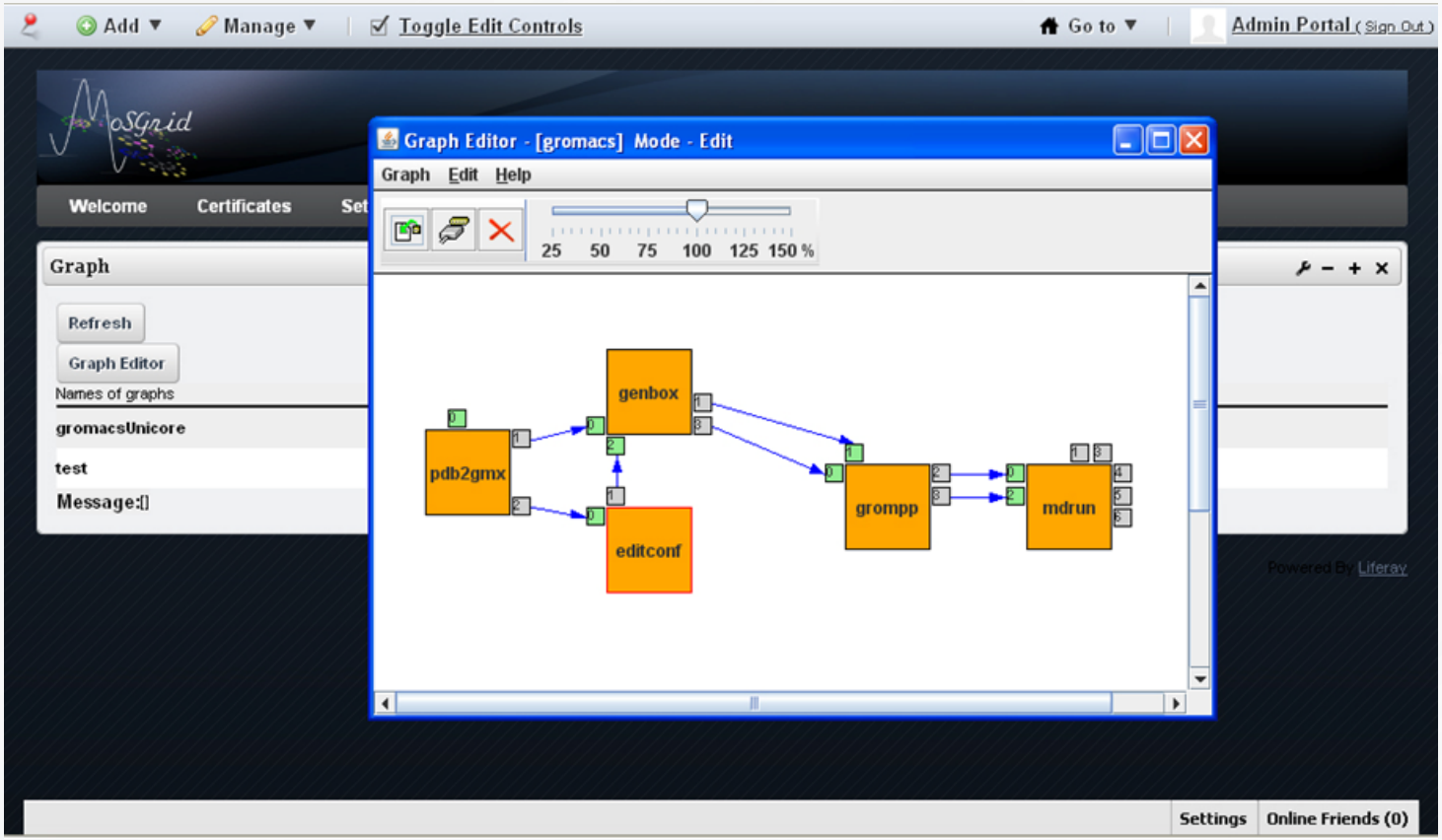
Background - XtreemFS



- Distributed object-based Grid und Cloud filesystem
- Replication for availability, locality, bandwidth, and latency
- GSI and SAML support
- Easy integration in heterogenous environments
- FUSE client and Java API

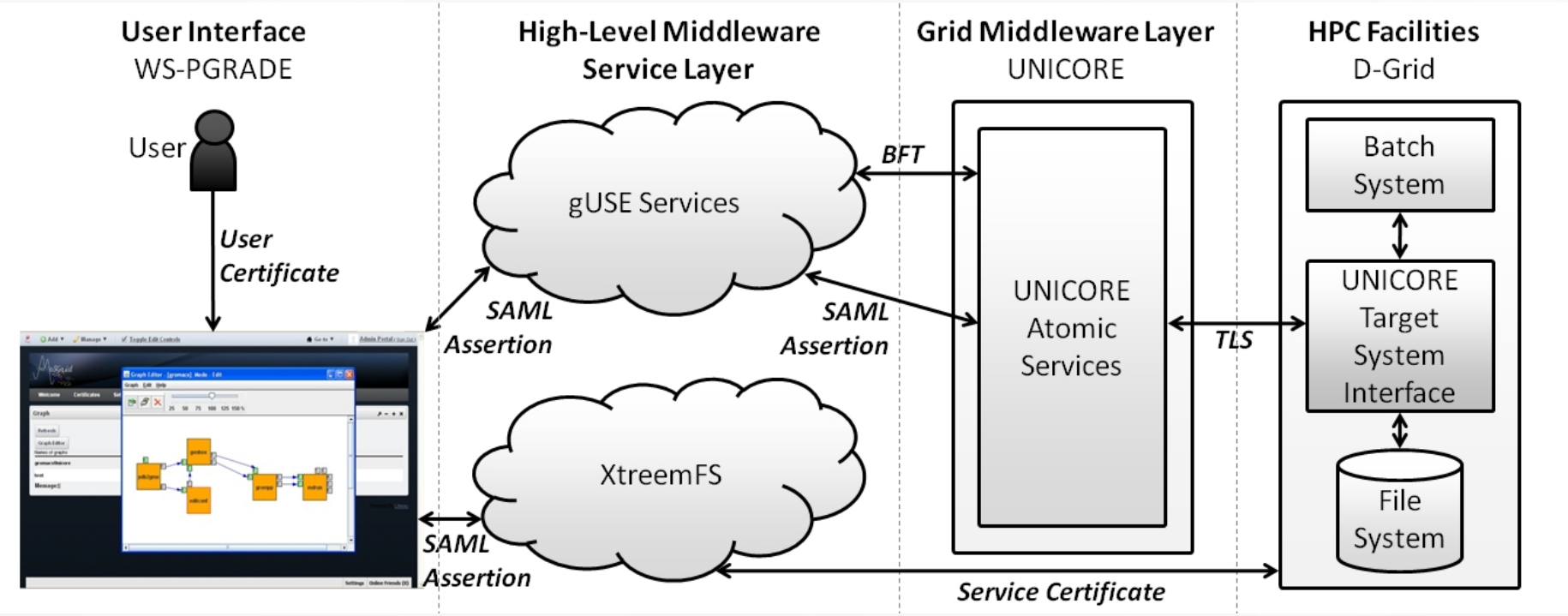
MoSGrid Software Structure - GUI

- Liferay using Tomcat, WS-PGRADE with gUSE, and domain portlets as the GUI
- WS-PGRADE workflow editor:



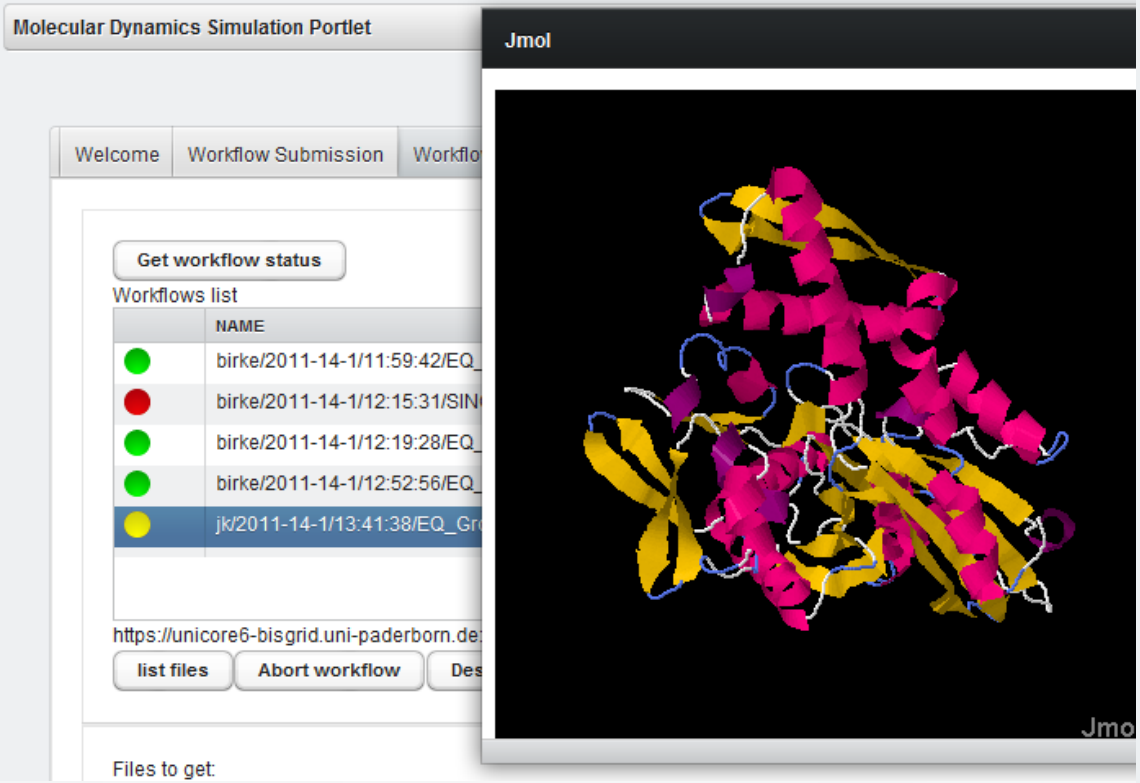
MoSGrid Software Structure - Security

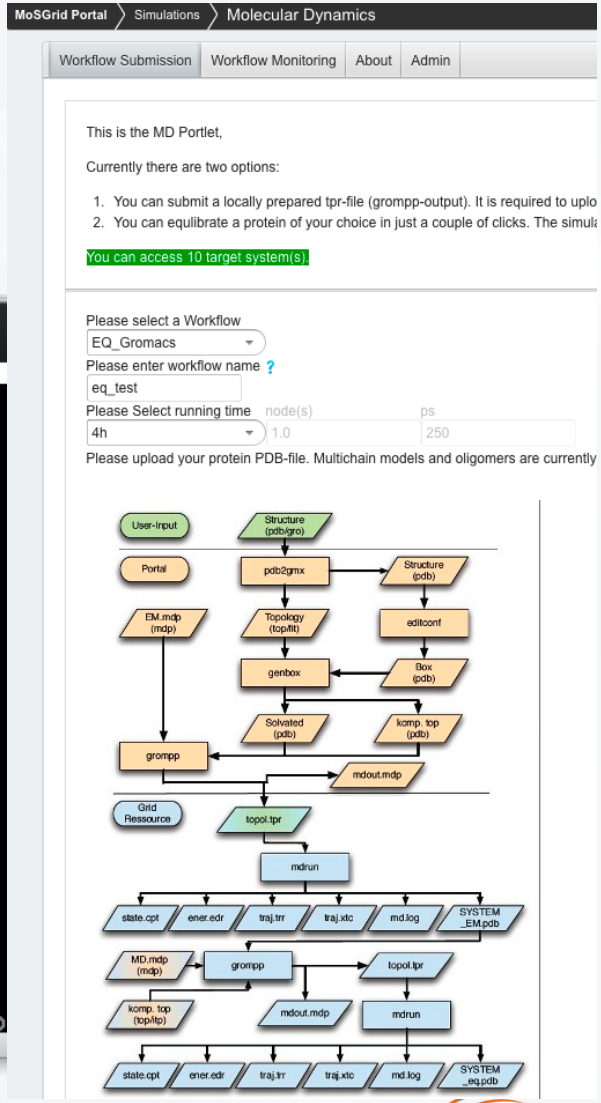
- User certificate for creating a SAML assertion
- SAML assertion for authentication to gUSE, XtreemFS, and UNICORE



Domain Portlets – Molecular Dynamics

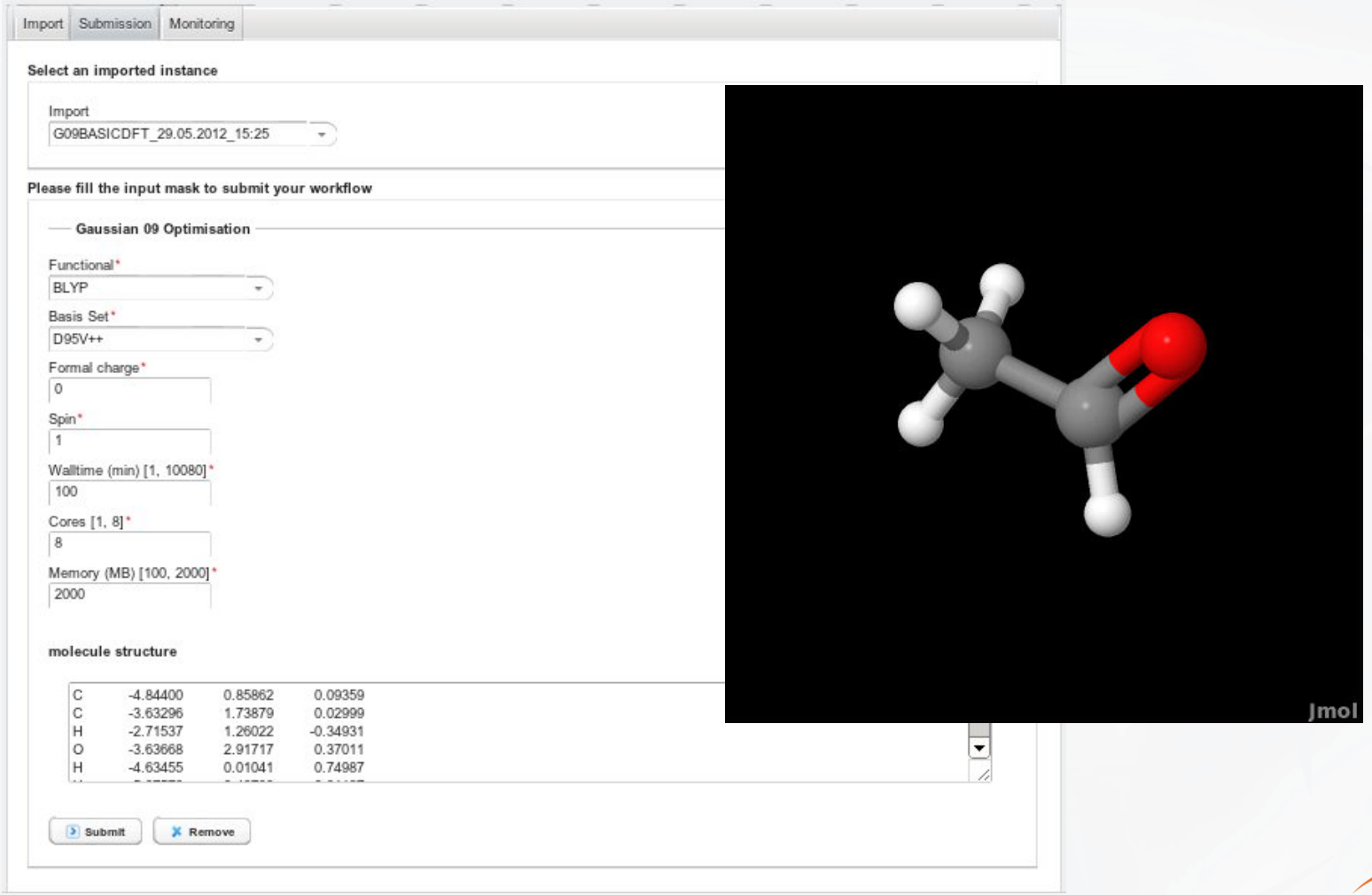
- Applications: Gromacs, Amber, NWChem





Domain Portlets – Quantum Chemistry

- Applications: Gaussian, Turbomole, NWChem



Domain Portlets – Docking

- Applications: CADDSuite, FlexX

Docking Portlet

Import | Submission | Monitoring | About | ?

Select an imported instance

Import

StandardDockingWorkflow_2012-03-30-125439_29.(

Please fill the input mask to submit your workflow

PDBCutter

Filename *

1DX6.pdb

Upload PDB

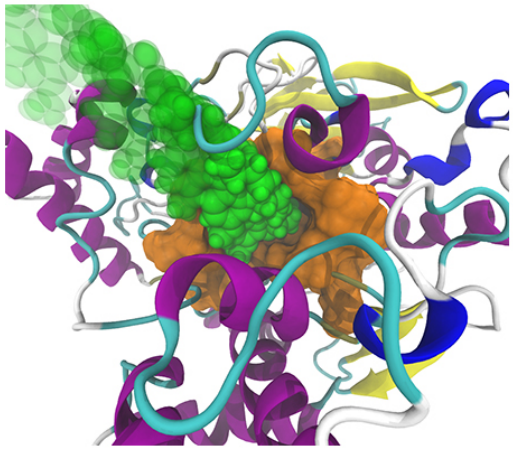
PDB Model *

Model 0

Chain A

Chain S

Select a protein chain from your PDB input file to act as receptor (secondary structure) including the binding pocket (orange).



Chain name of ligand *

A

Name of ligand as stated in pdb file *

GNT

Specifiy a reference ligand (green) by it's three letter code including the corresponding chain. It might be necessary to open the input PDB file with an editor. This information is required for the indentification of the binding site and the calculation of RMSD values.

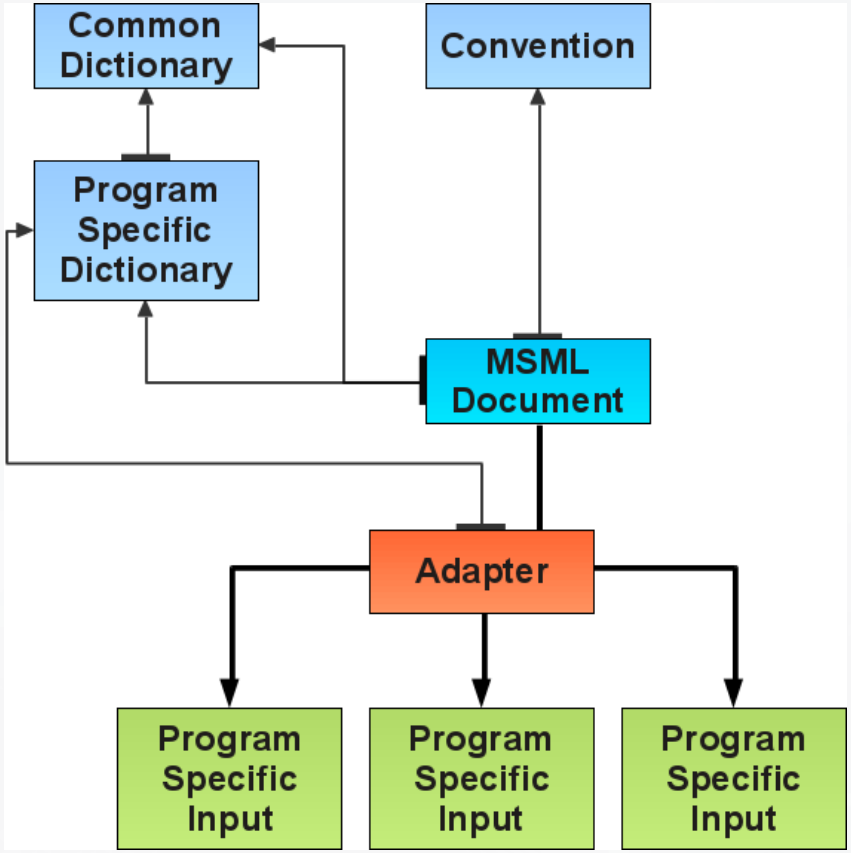
Protein Chains that are to be deleted

Repository - Metadata with MSML

- Repository consists of data and metadata storage
- Development of MSML (Molecular Simulation Mark-up Language)
- Subset and extension of CML (Chemical Mark-up Language)
- Unified data representation for all 3 domains
- Used for storing structures, simulation descriptions, and results
- Parsers and adapters used for conversions to and from MSML

Repository - Parser and Adapter

- Structure parser for input structure formats (PDB, SDF) to MSML
- Parser for application output to MSML
- Adapter for MSML to application input



Repository - Data Management

- Upload of new data
  - Structure parser to convert raw structures to MSML
  - Metadata extractor to convert MSML to JSON
  - UNICORE and Lucene used for indexing JSON files

Repository - Data Management

- Use of data
  - Chooses domain portlet and suitable workflow
  - Adjusts parameters for applications and selects input
  - First step in workflows: MSML to application specific input
  - Results are computed
  - Output to MSML
  - Last step in workflows: MSML to JSON metadata
  - JSON indexed for searchable results

Repository - UNICORE integration

- Integration of XtreemFS in UNICORE with URL scheme (Thanks Bernd)
  - UNICORE extension to make xtreemfs:// available
  - XtreemFS access via TSI local mount point or remote SMS
- Integration of UNICORE in gUSE with submitter
  - Connection between gUSE and UNICORE to submit jobs
  - To index JSON metadata at the end of workflows
- Integration of UNICORE metadata capabilities in portal
  - Metadata extractor to convert MSML to JSON
  - Searching in search portlet and in domain portlets

Outlook

- Updating infrastructure for sustainability
- Finishing data and metadata integration
- Improvements of domain portlets with portlet API
- Automatic login via certificates
- Release MoSGrid as open-source

Follow-up Projects

- ER-flow – EU project to build european workflow sharing community
  - Integration of applications in SHIWA simulation platform
  - Study of data exchange between workflow systems
- SCI-BUS – EU project for semantic search and visualization
  - Search over workflows, results and structures
  - Workflow editor and visualization with WebGL

Questions?