Molecular Simulations using Workflows and Science Gateways

Jens Krüger

October 9th 2015

Sino - German Workshop on
Multiscale spatial computational systems biology

Applied Bioinformatics Group
EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Outline

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

- Molecular Simulations
  - Quantum Simulations
  - Molecular Dynamics
  - Docking
- Workflows
  - Protocols
  - Abstract Representation as Graphs
  - Concrete Workflows
- Science Gateways
  - Architecture
  - Computing
  - Data Management
  - Metadata Annotation
- Teaching
  - Ease of usage
  - Focus on science
- Conclusion

Molecular Simulations

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

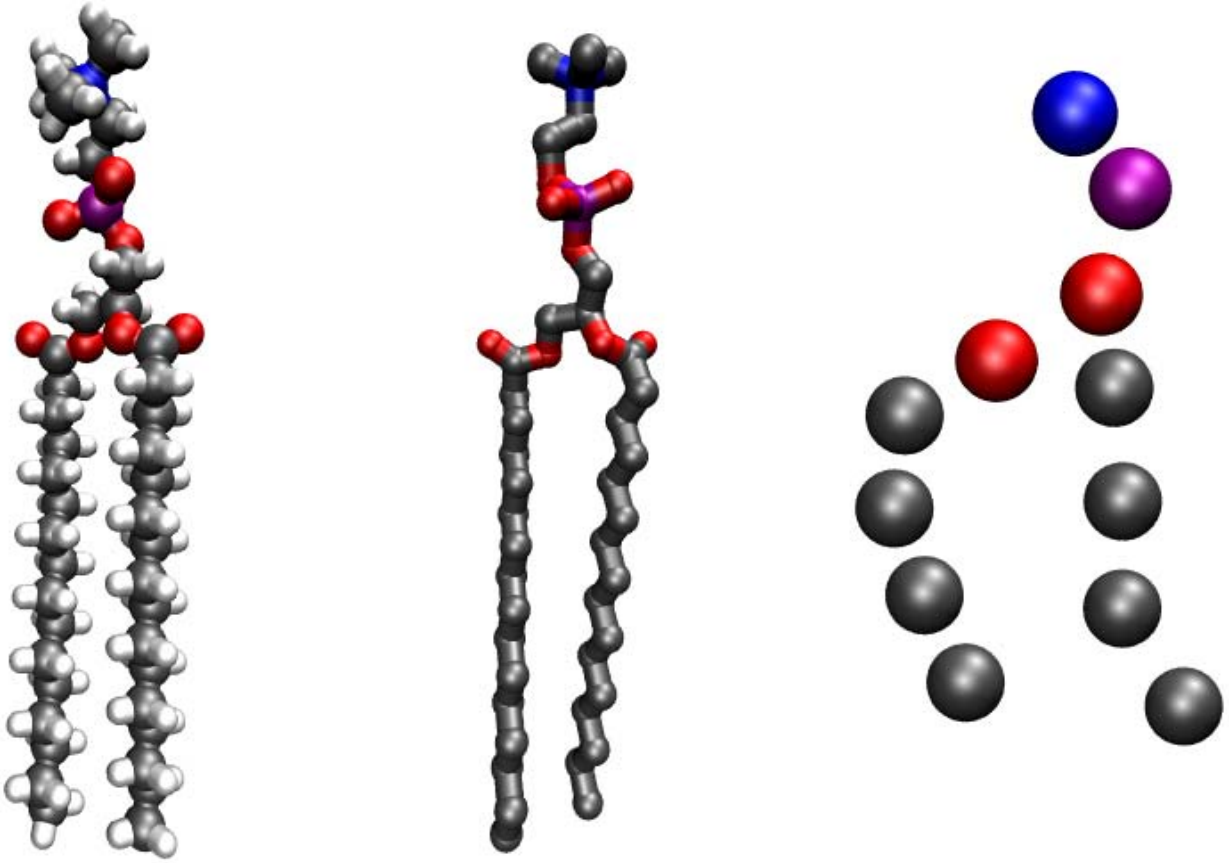
- Processes in living organisms cover multiple orders of magnitude regarding time and space
- Biochemical mechanisms of function can rely on something as small as a single electron or something as big as oligomerization states of membrane proteins
- Hence, different methods, strategies and protocols are required to handle individual scientific challenges

Dimensions and Time Scales

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

	Dimensions	Time Scales
Quantum Mechanics	pico meter [10^{-12}]	pico seconds [10^{-12}]
Molecular Dynamics	nano meter [10^{-9}]	nano seconds [10^{-9}]
Coarse Grained MD	micro meter [10^{-6}]	micro seconds [10^{-6}]
Brownian Dynamics	milli meter [10^{-3}]	milli seconds [10^{-3}]
Continuum Simulations	variable [10^{-12} - 10^{0}]	variable [10^{-12} - 10^{0}]

All-Atom to Coarse-Grained

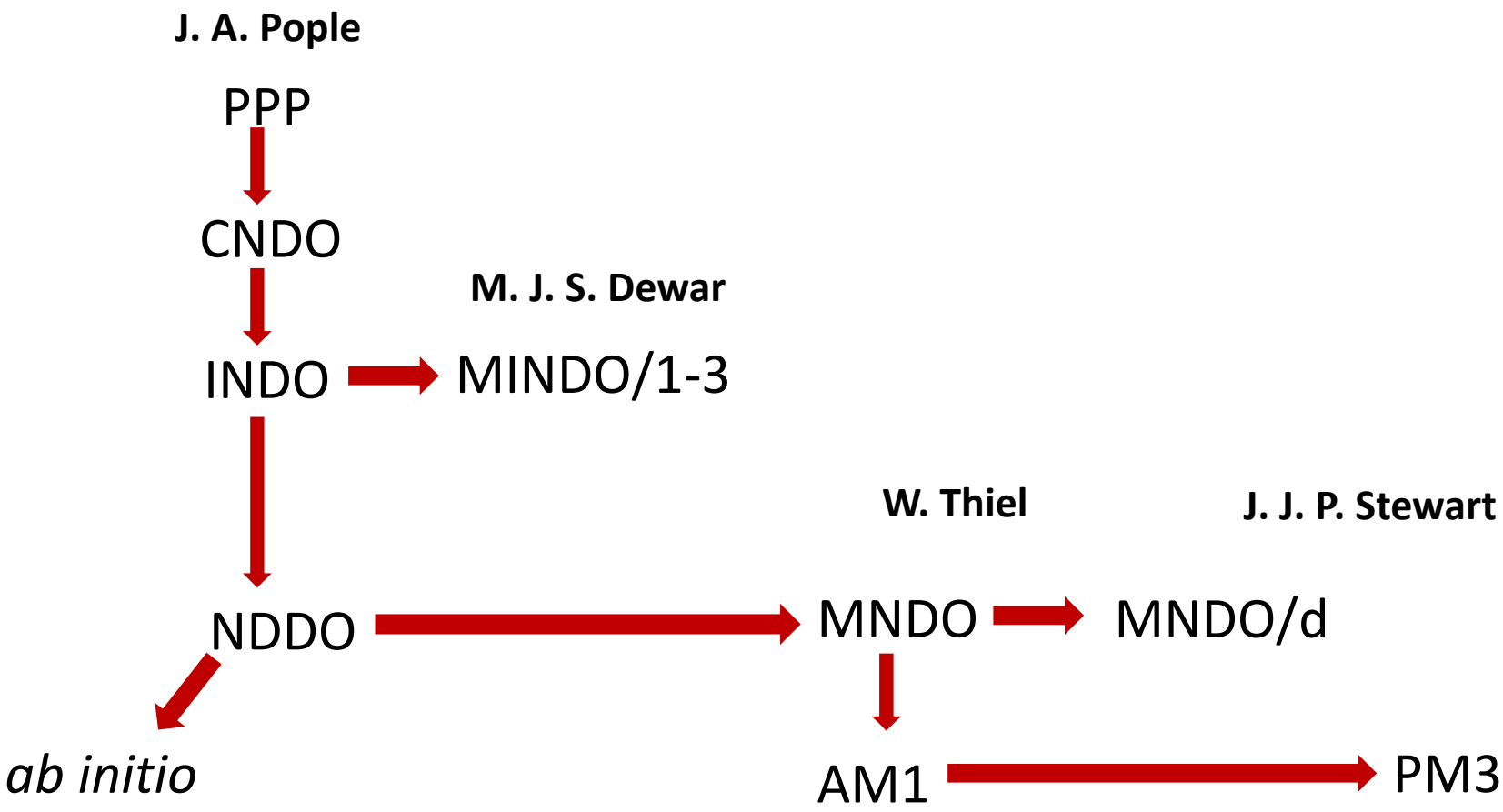


Quantum Calculations

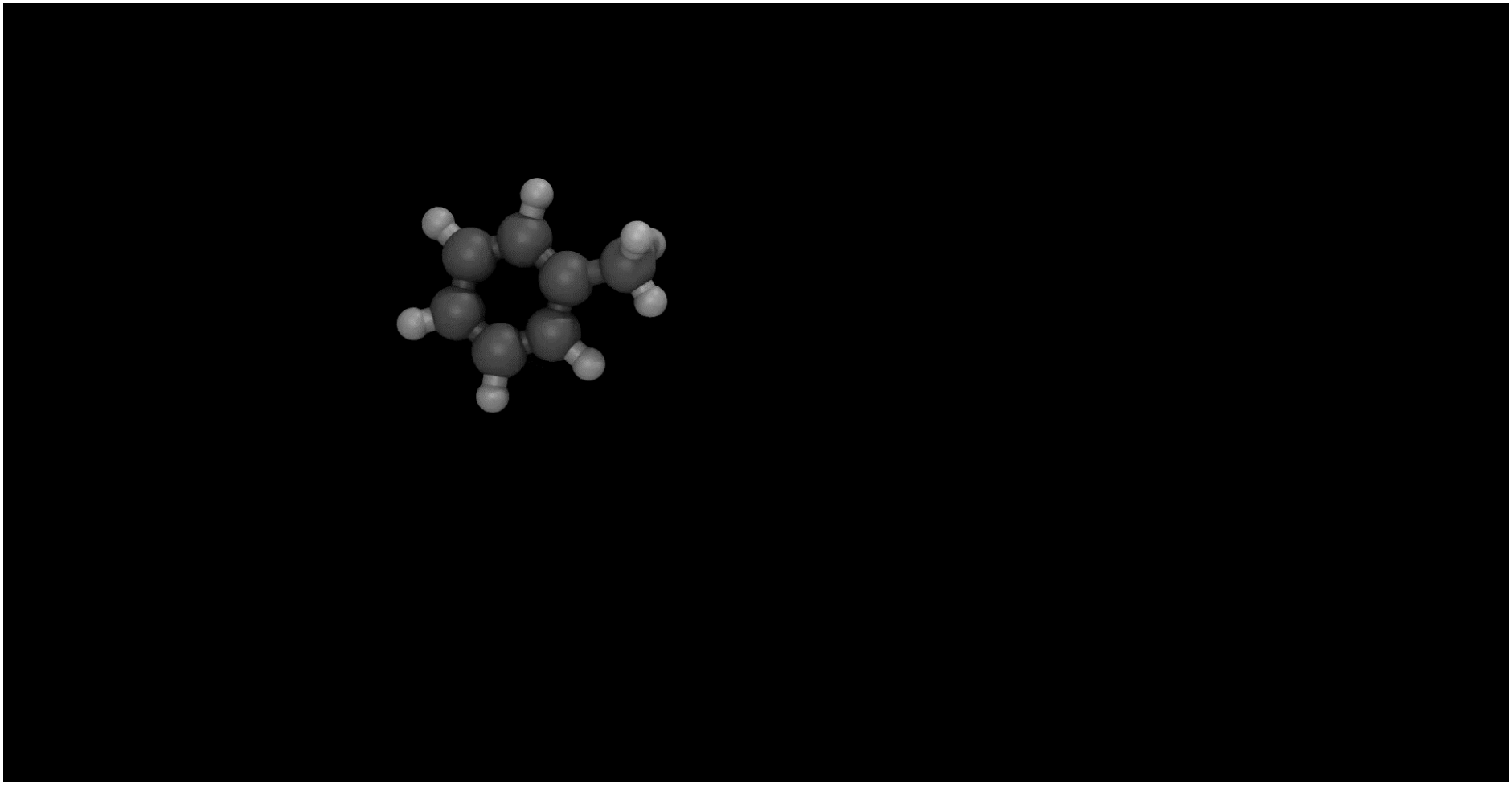
EBERHARD KARLS UNIVERSITÄT TÜBINGEN

- Solving the Schrödinger equation for the electronic Hamiltonian of a molecule determines its electronic structure
- Implicitly all chemical properties are related to the electron structure of a molecule
- The Schrödinger equation can only be solved precisely for the hydrogen atom. For all other molecular systems only approximations can be obtained

Semiempiric approaches

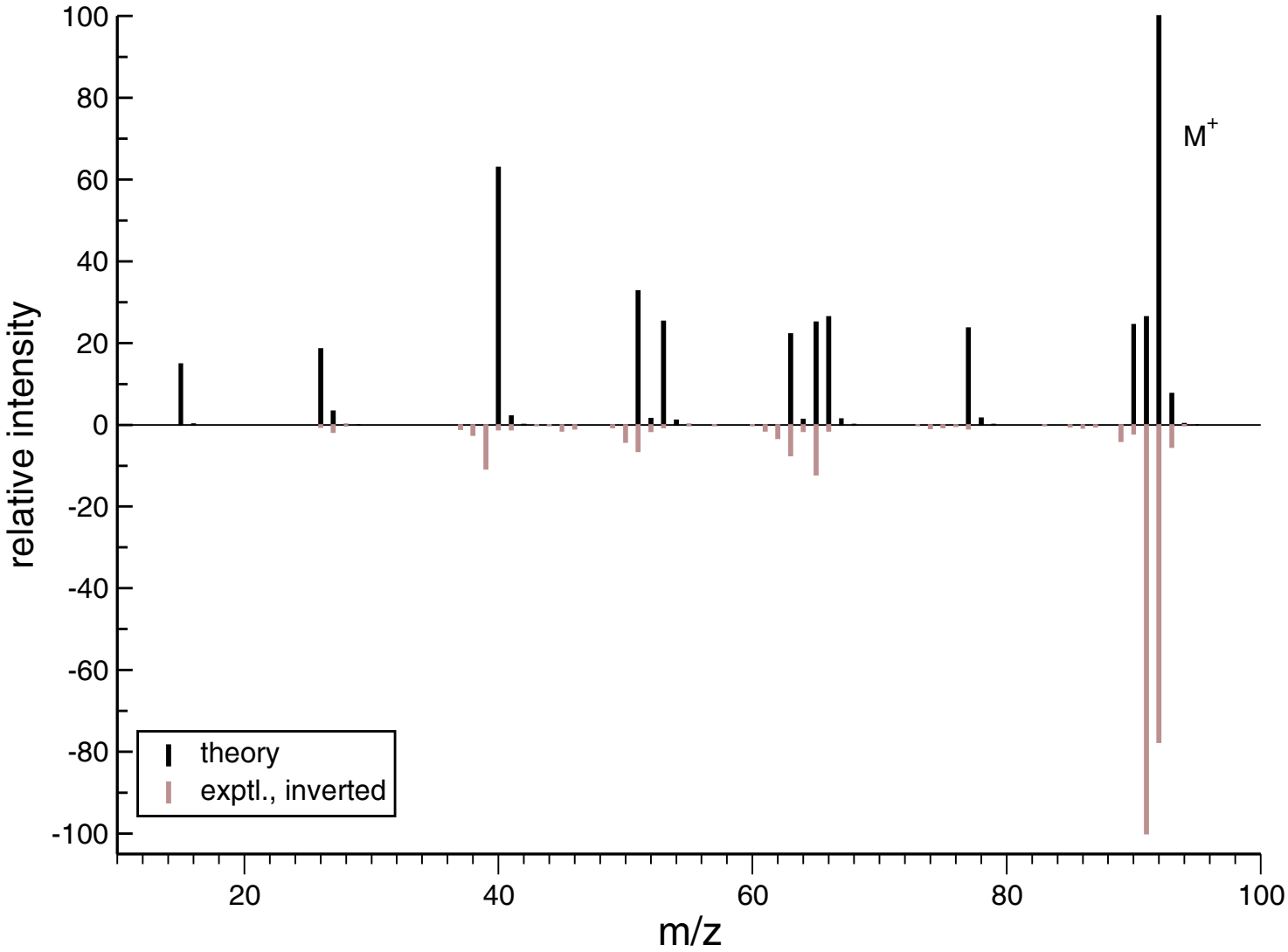


Mass Spectroscopic Fragmentation



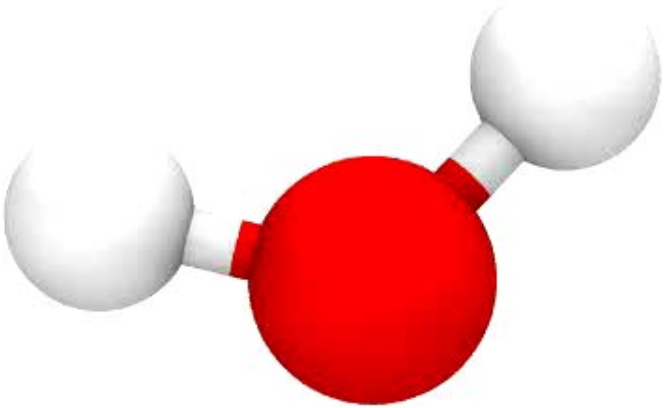
Method of C. A. Bauer and S. Grimme, see Angew. Chem. Int. Ed. 2013, 52, 6306-6312

Quantum Calculations



Molecular Dynamics

- Molecule connected with spri
- The force l as their time dep th Newton's

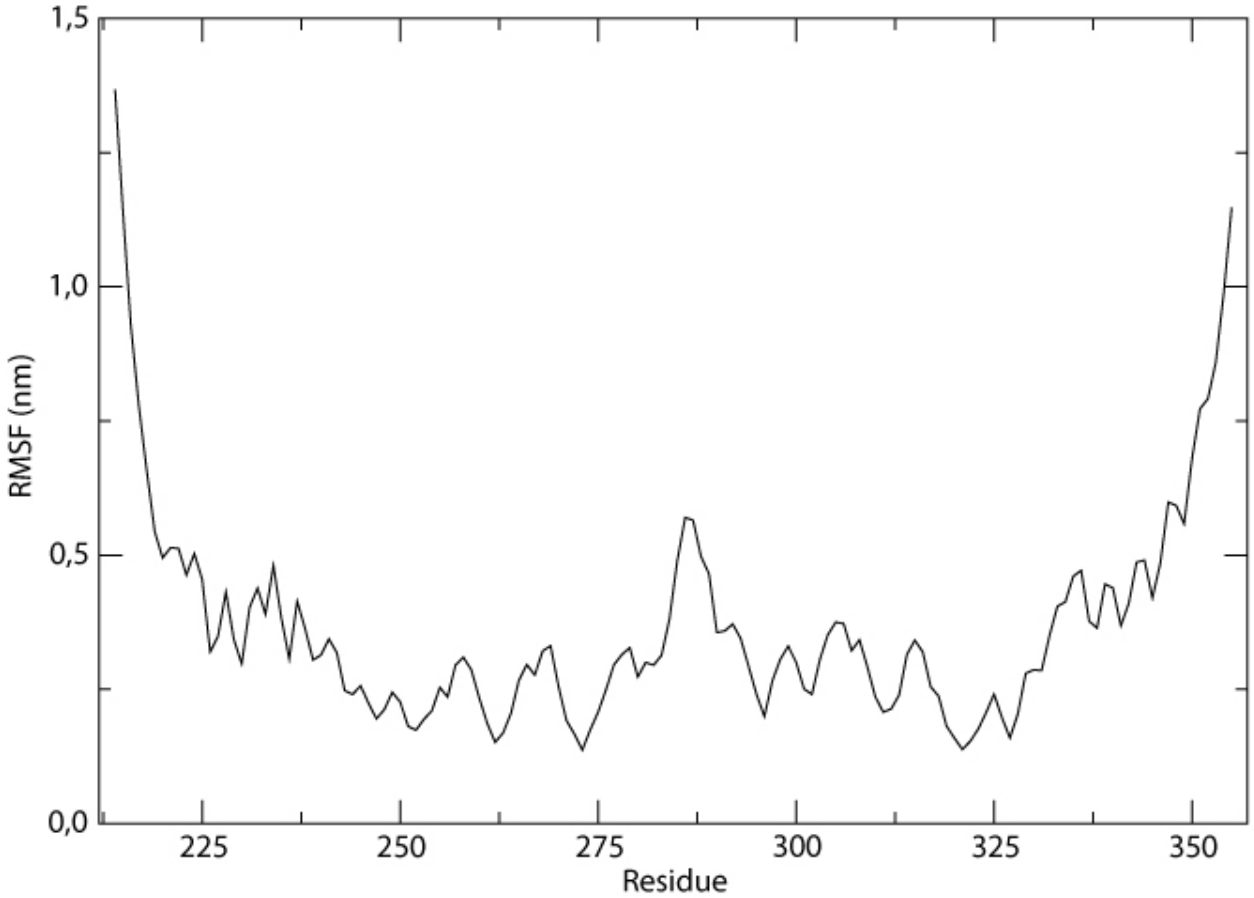


$$F_i = m_i a$$

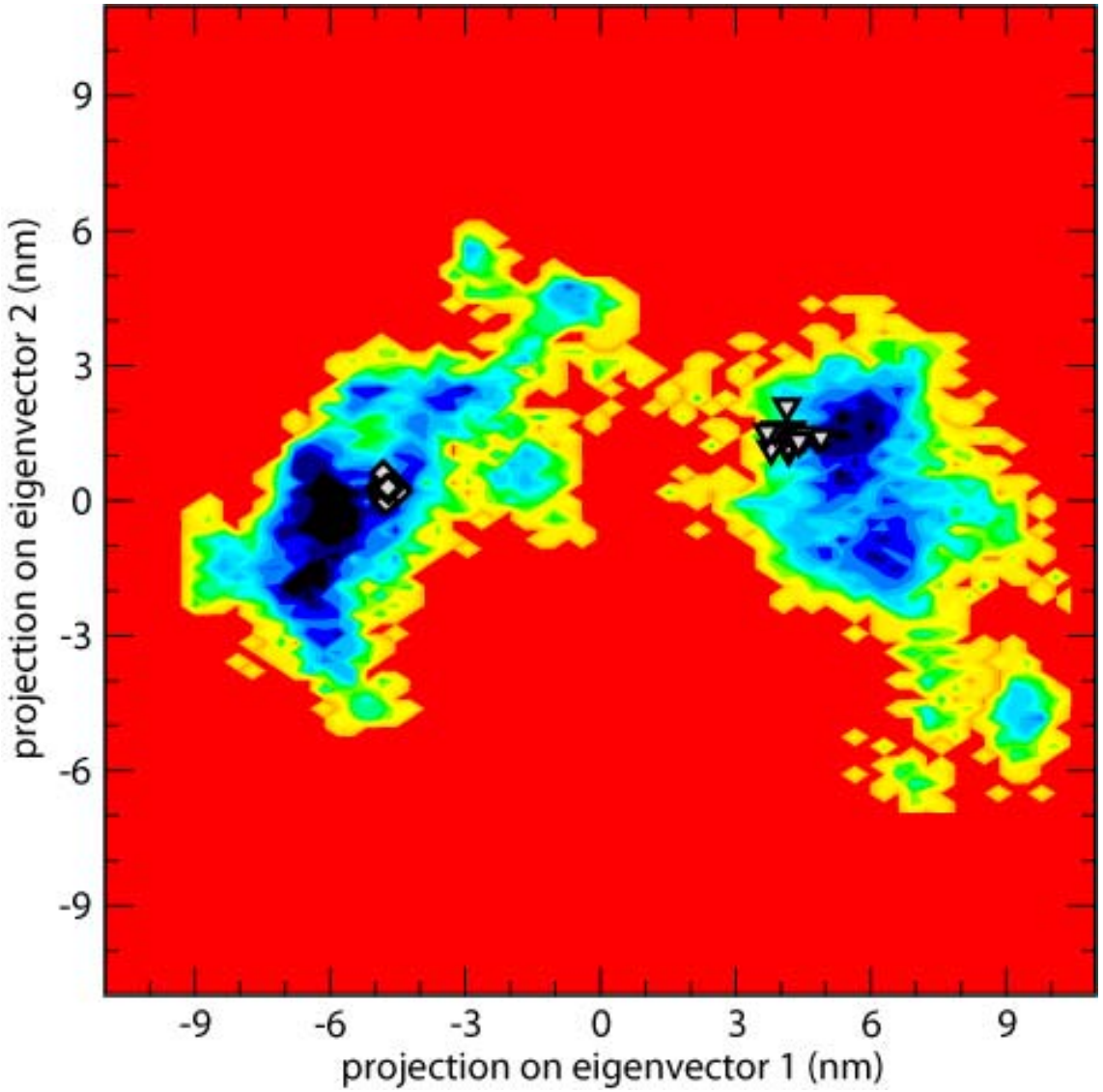
$$dr_i$$

$$m_i \frac{d^2 r_i}{dt^2}$$

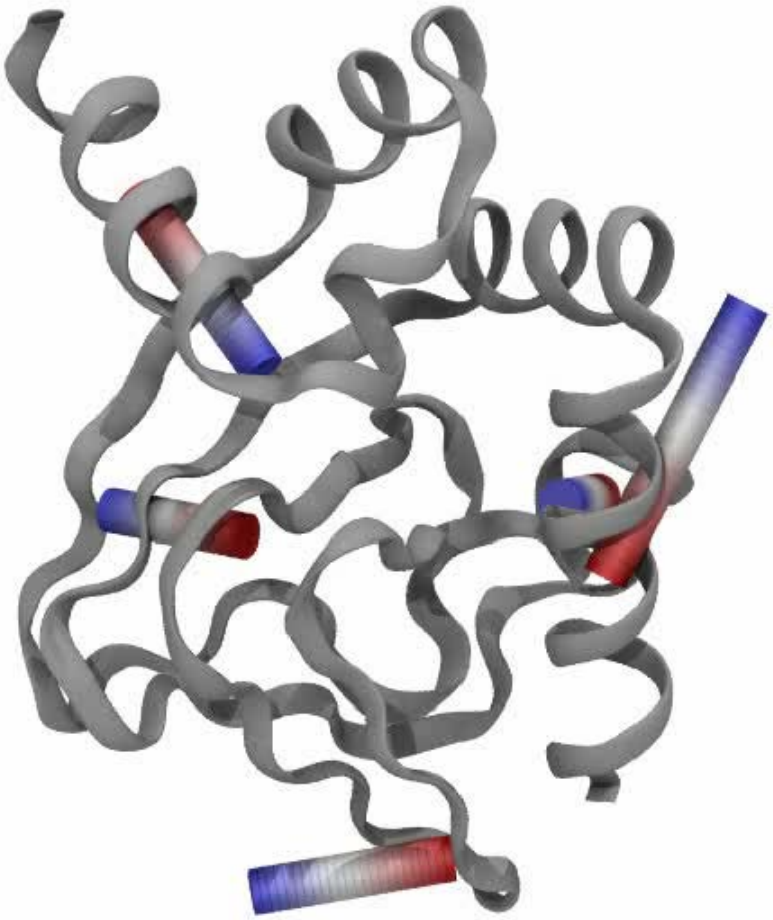
Fluctuations



Free Energy Map



Amplitude of Movement



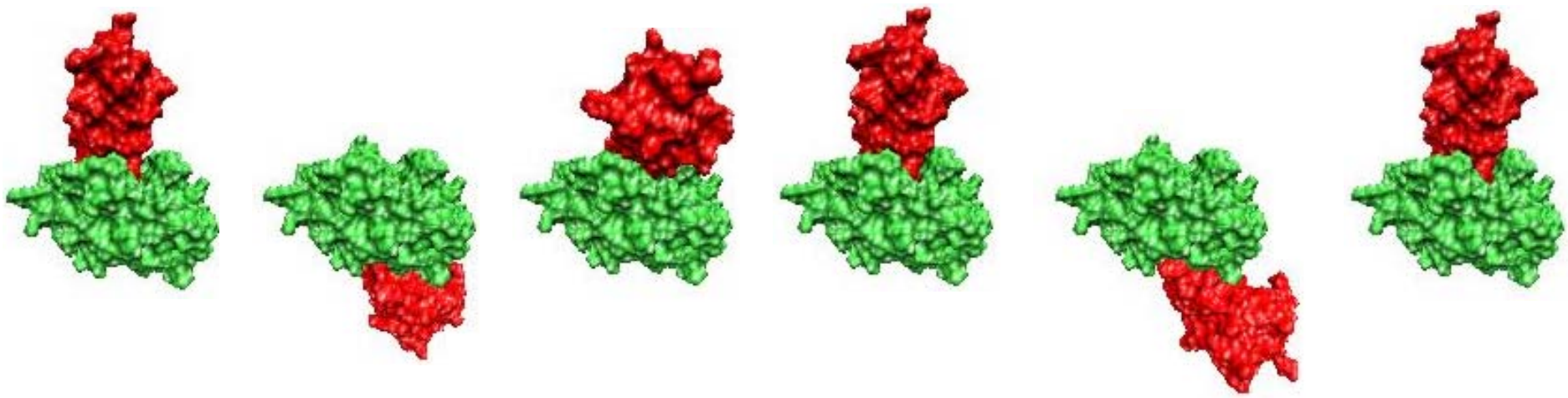
Docking

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Base Algorithm

Structure Generation

- Create a large number of plausible complex conformations of *A* and *B*
- Requirement: *A* in contact with *B*
- Only implicit structure generation in some algorithms
- Some of these complexes will be energetically unfavorable



Docking

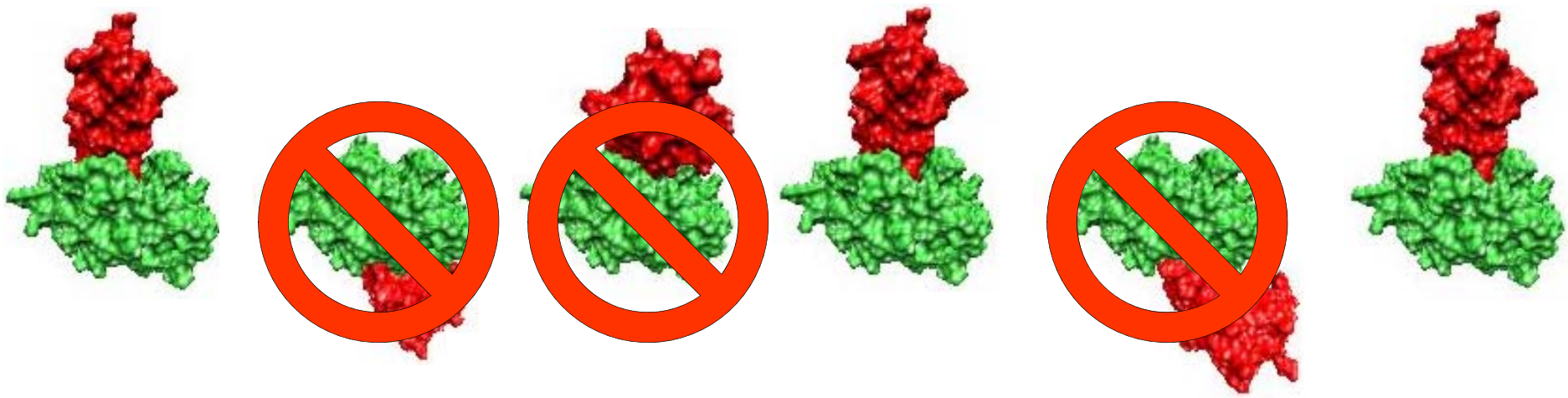
EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Base Algorithm

Structure Generation

Filtering

- Remove complex structures that are geometrically or energetically unfavorable
- Since we filter a large number of candidates, this evaluation has to be simple and computationally efficient



Docking

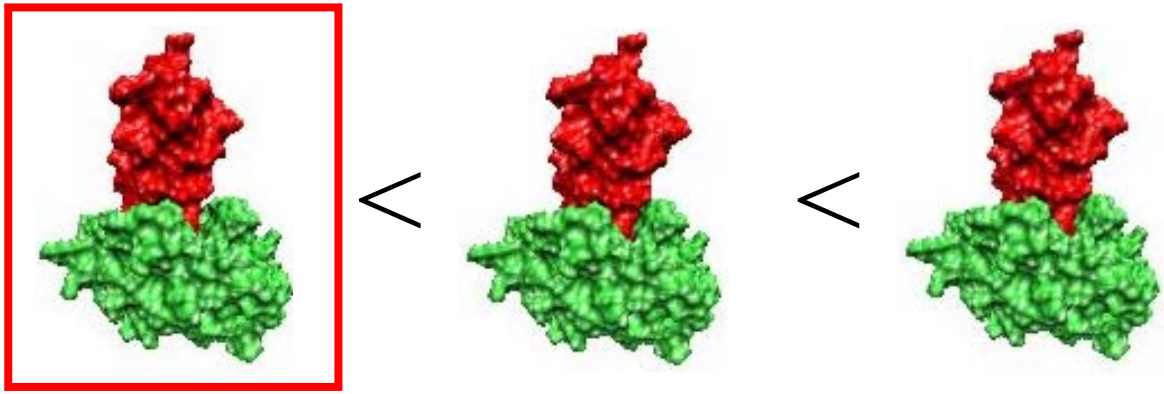
Base Algorithm

Structure Generation

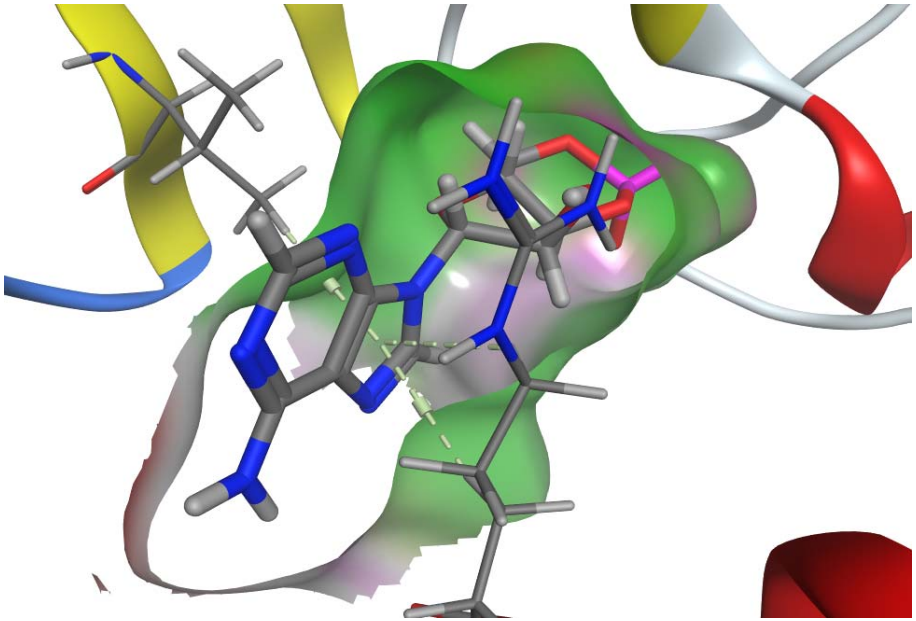
Filtering

Evaluation

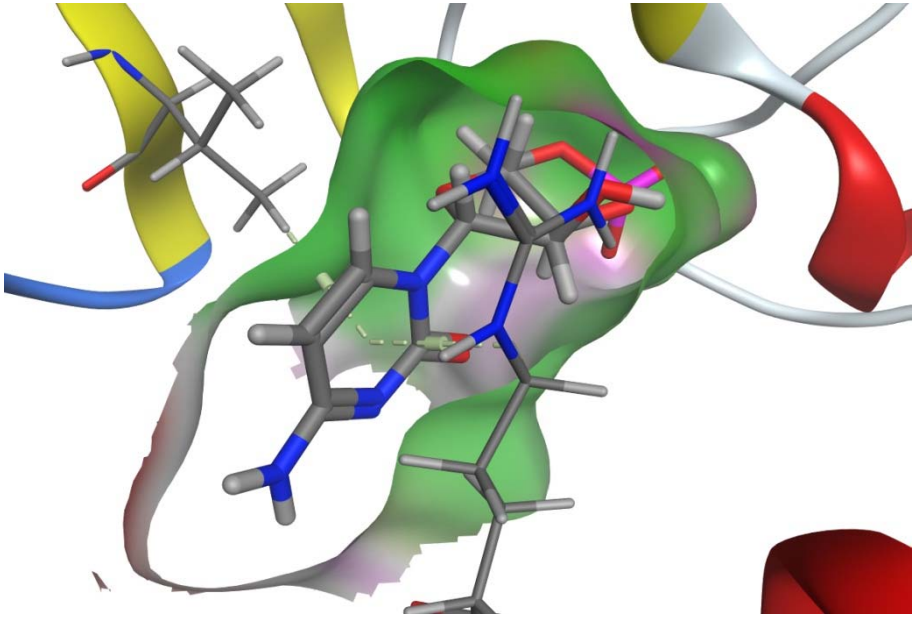
- Estimate ΔG for each candidate
- Many scoring/energy functions are rather simple
- **Result:**
  list of candidates ordered by their ΔG
- **Good result:**
  head of the list close to true structure



Binding Site



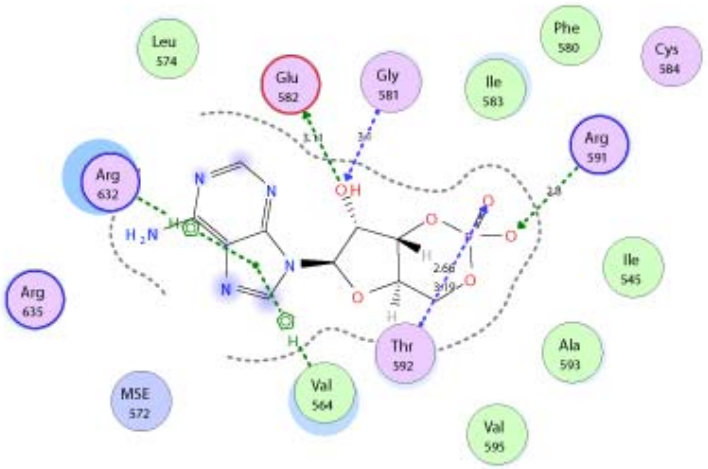
cAMP



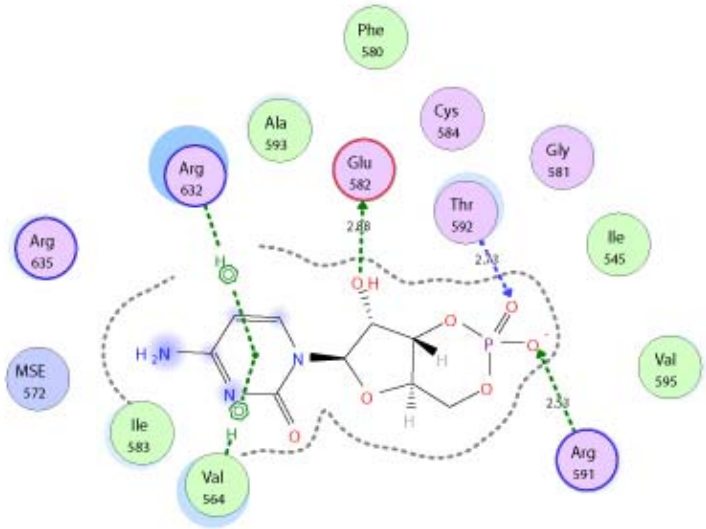
cCMP

X. G. Zong, S. Krause, C. C. Chen, J. Krüger, C. Gruner, X. C. Cao-Ehlker, S. Fenske, C. Wahl-Schott, and M. Biel: *Regulation of HCN channel activity by cyclic cytidine 3', 5'- monophosphate (cCMP)*, Journal of Biological Chemistry, 287(32): 26506-26512, 2012.

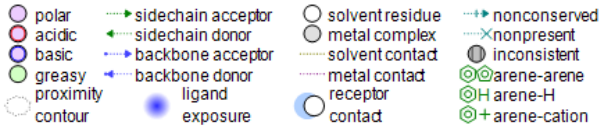
Interactions



cAMP



cCMP



Workflows

- All simulation methods require multiple steps of pre- and post-processing
- Simulations protocols can be represented as acyclic directed graphs
- The nodes correspond to processing steps
- The data flows along the edges of the graph

Protocols

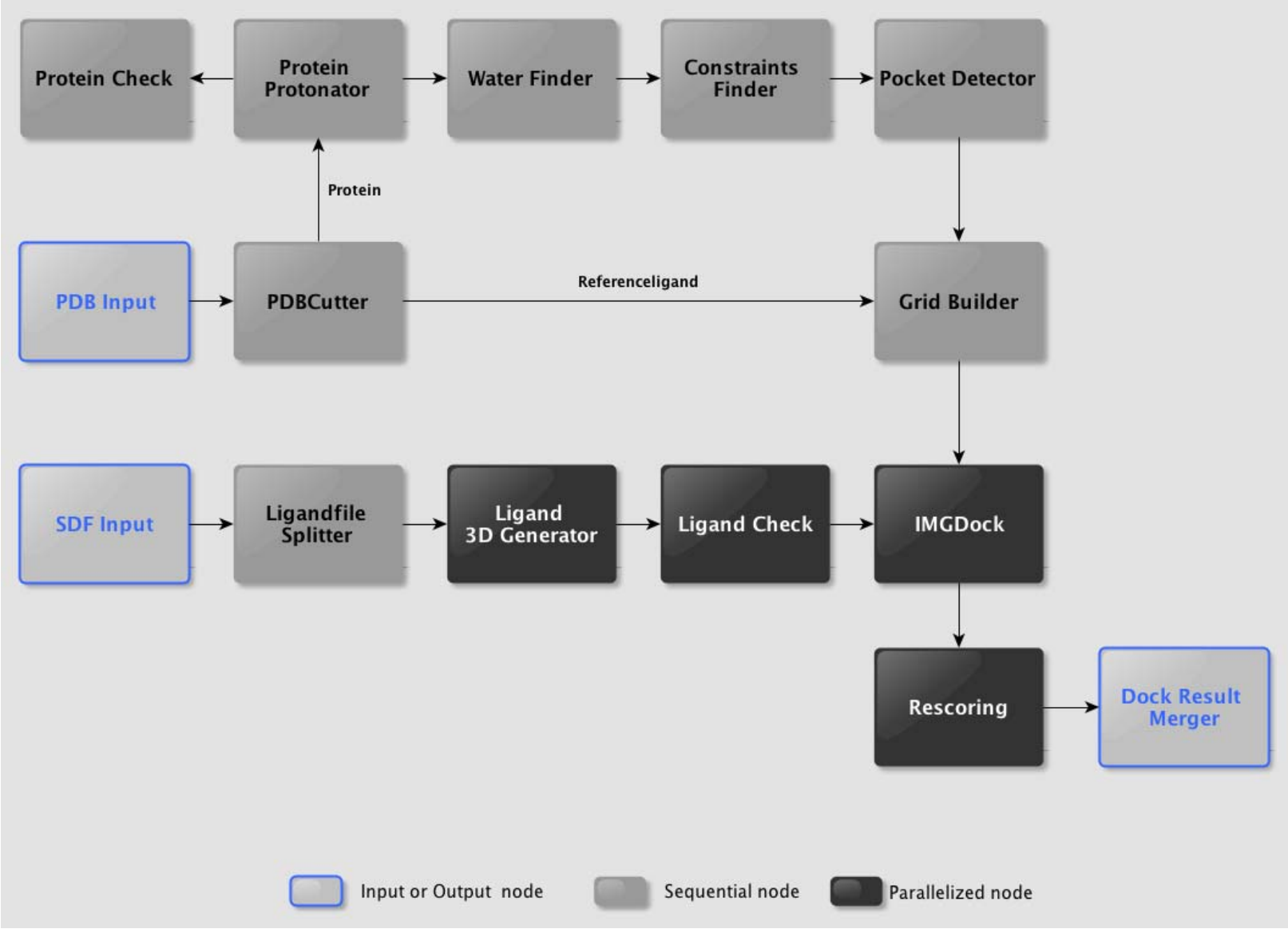
EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Example: Molecular Docking

- Process the protein target structure
  - Separate protein and reference ligand (if present)
  - Remove detergents, unwanted ions, …
  - Adjust protonation states
  - Convert to a suitable format
- Process the ligand library
  - Convert to a suitable format
  - Generate 3D conformations
  - Adjust protonation states
- Define binding site/box
  - Use the reference ligand (if present) to define a binding site
    or use another method to create a binding site description
  - Create a docking grid, interaction points, …
- Dock
  - Place ligands within the binding site
  - Evaluate energies
- Collect and sort results
  - Filter and sort results
- Analysis and Visualization
  - Scatter plots
  - Box plots
  - Enrichment
  - Molecule view

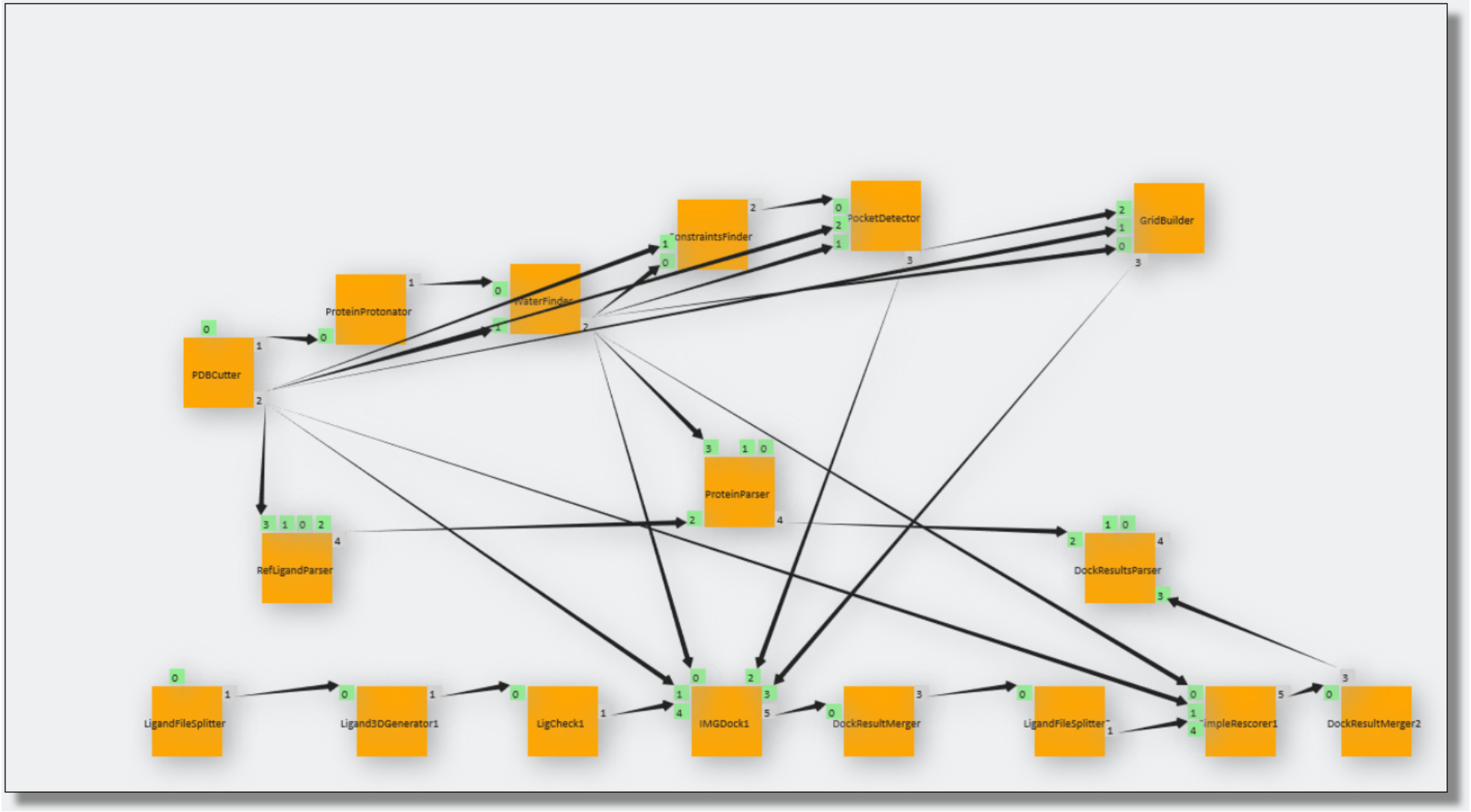
Incomplete and inaccurate description!

Abstract Representation



Concrete Workflows

EBERHARD KARLS UNIVERSITÄT TÜBINGEN



Science Gateways

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

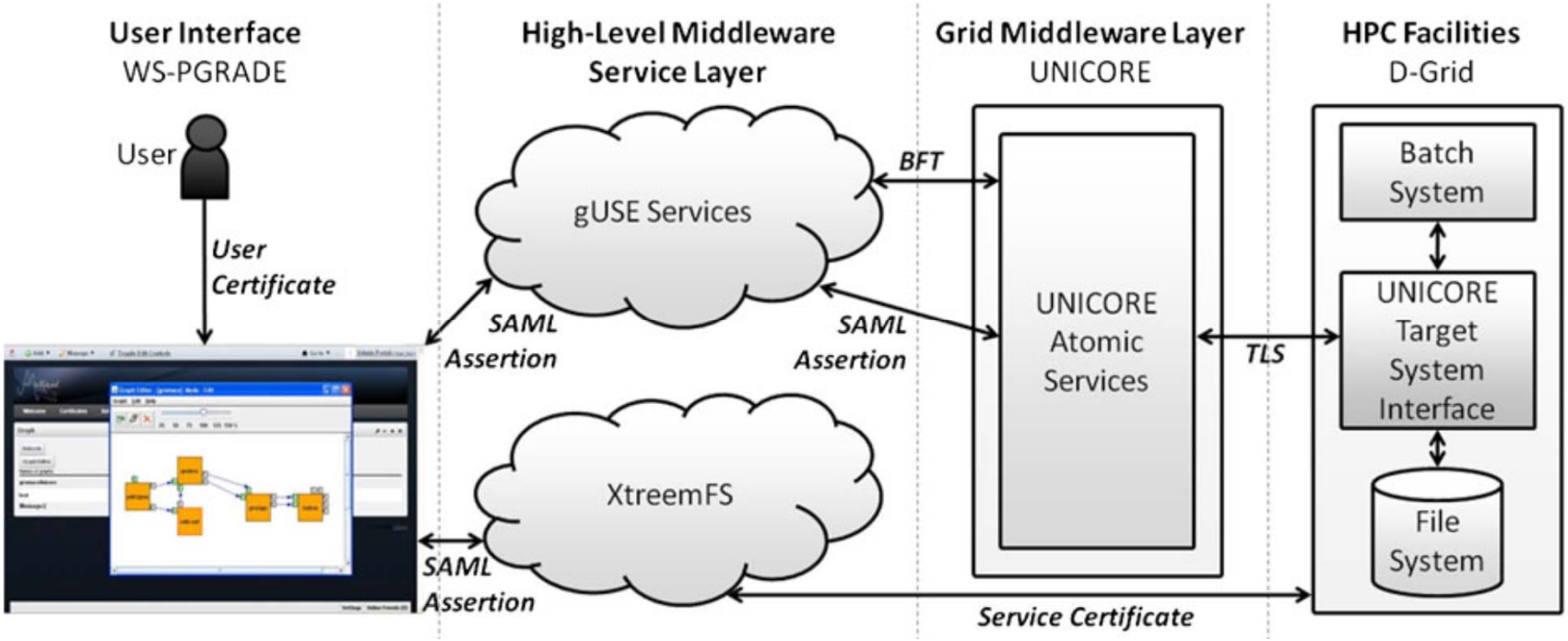
- A portal facilitating the access to data, applications and tools
- Community dependent implementation
- Integration of workflows, compute resources, visualization and data management
- Platform for collaborative working
- Secure and resilient environment

MoSGrid

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

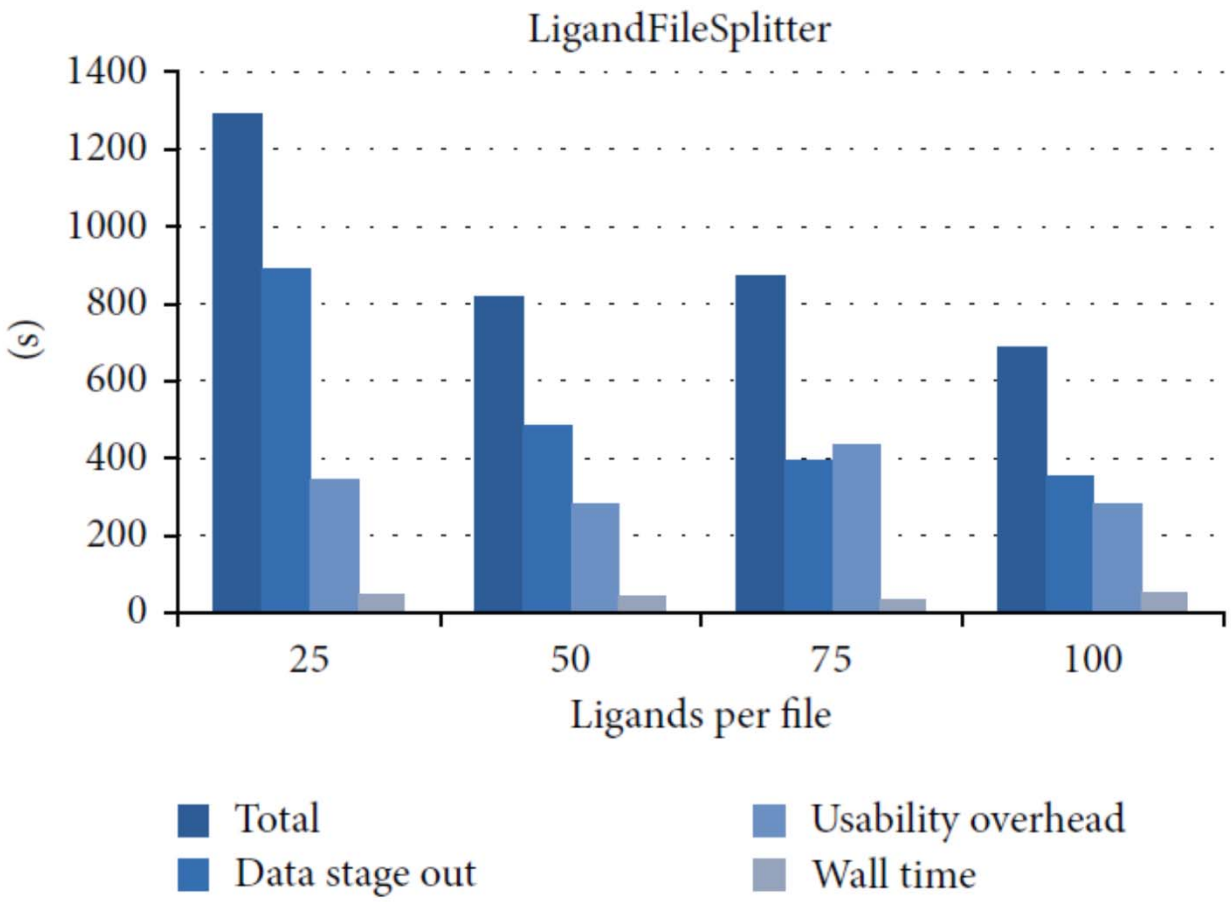
- ~ 400 registered users
- 3 application domains (QC, MD, docking)
- Gaussian, NWChem, GROMACS, eSBM-Tools, CADDSuite, Autodock, FlexX, …
- DCIs on 4+ sites all over Germany
- ~ 7500 cores available 24/7
- 100+ workflows
- ~ 90k+ individual jobs per year

Architecture



R. Grunzke, S. Gesing, J. Krüger, G. Birkenheuer, M. Wewior, P. Schäfer, B. Schuller, J. Schuster, S. Herres-Pawlis, S. Breuers, A. Balaskó, M. Kozlovszky, A. Szikszay Fabri, L. Packschies, P. Kacsuk, D. Blunk, T. Steinke A. Brinkmann, G. Fels, R. Müller-Pfefferkorn, R. Jäkel and O. Kohlbacher: *A Single Sign-On Infrastructure for Science Gateways on a Use Case for Structural Bioinformatics*, Journal of Grid Computing, 10:769-790, 2012.

Computing



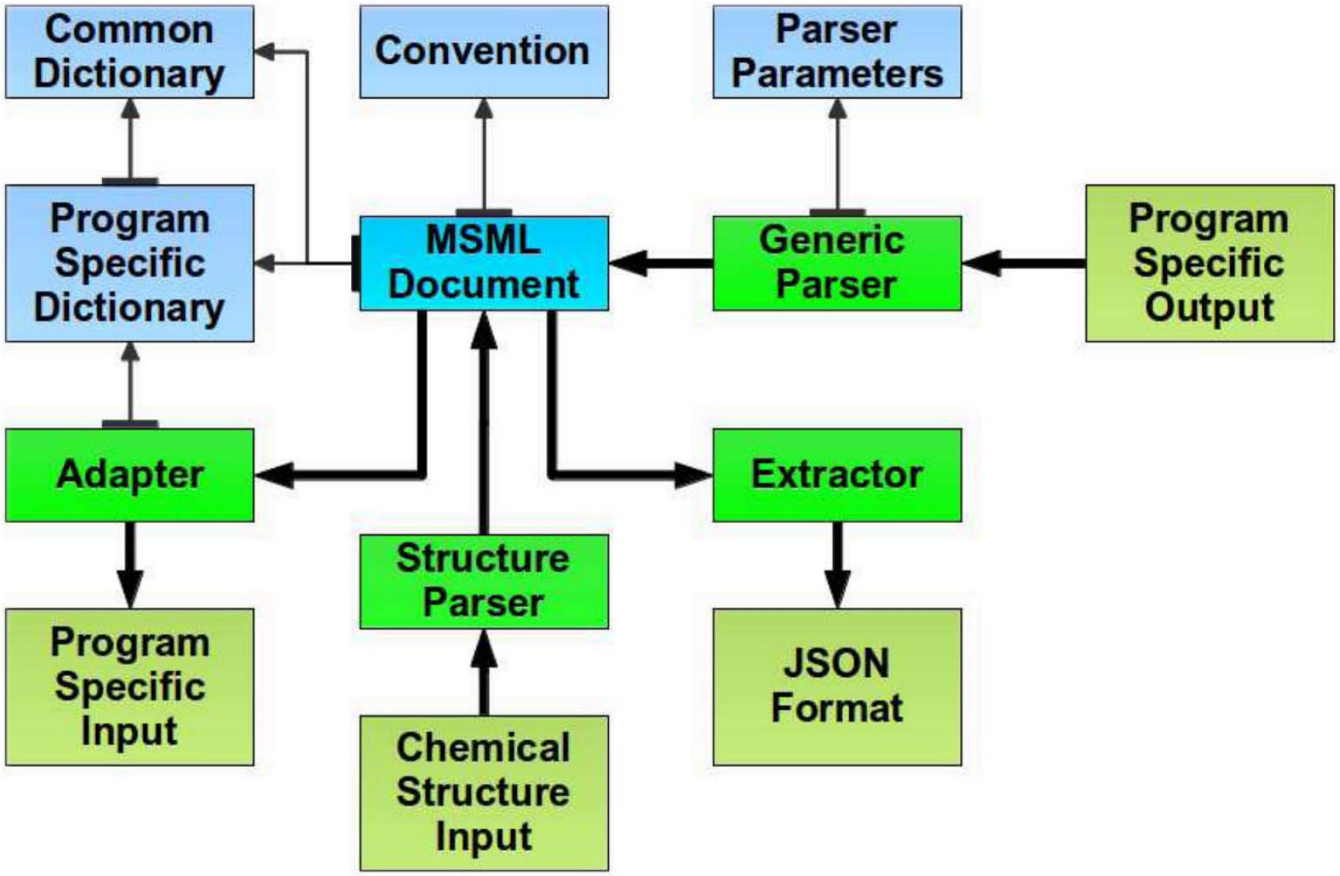
J. Krüger, R. Grunzke, S. Herres-Pawlis, A. Hoffmann, L. de la Garza, O. Kohlbacher, W.E. Nagel and S. Gesing: *Performance Studies on Distributed Virtual Screening*, BioMed Research International BioMed Research International 2014, Article ID 624024, http://dx.doi.org/10.1155/2014/624024.

Data Management

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

- A federated storage based on XtreemFS is used, directly accessible from the portal instance and all compute instances
- All input data is put on the storage upon job assembly
- All intermediate data serving as input or output is put on the storage
- All processed and analyzed data is annotated and stored

Metadata Annotation

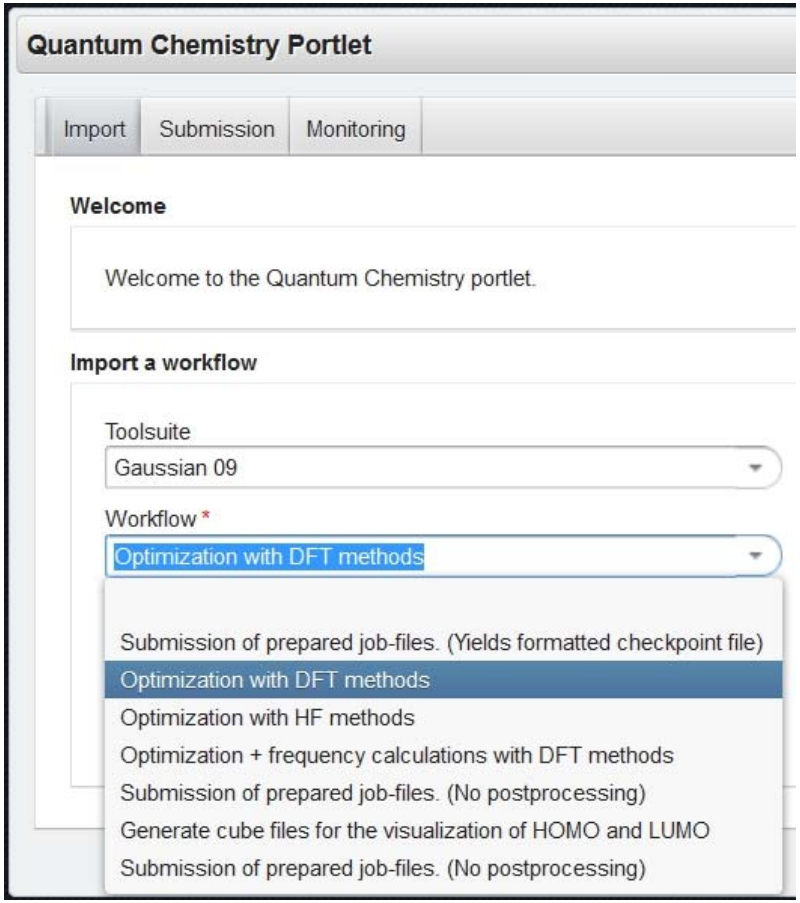


R. Grunzke, D. Blunk, S. Breuers, L. de la Garza, S. Gesing, S. Herres-Pawlis, M. Kruse, L. Packschies, P. Schäfer, C. Schärfe, T. Schlemmer, A. Zink, R. Müller-Pfefferkorn and J. Krüger: *Standards-based Metadata Management for Molecular Simulations*, Concurrency and Computation: Practice and Experience 2013, 26(10):1744-1759, DOI: 10.1002/cpe.3116.

PortletAPI

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

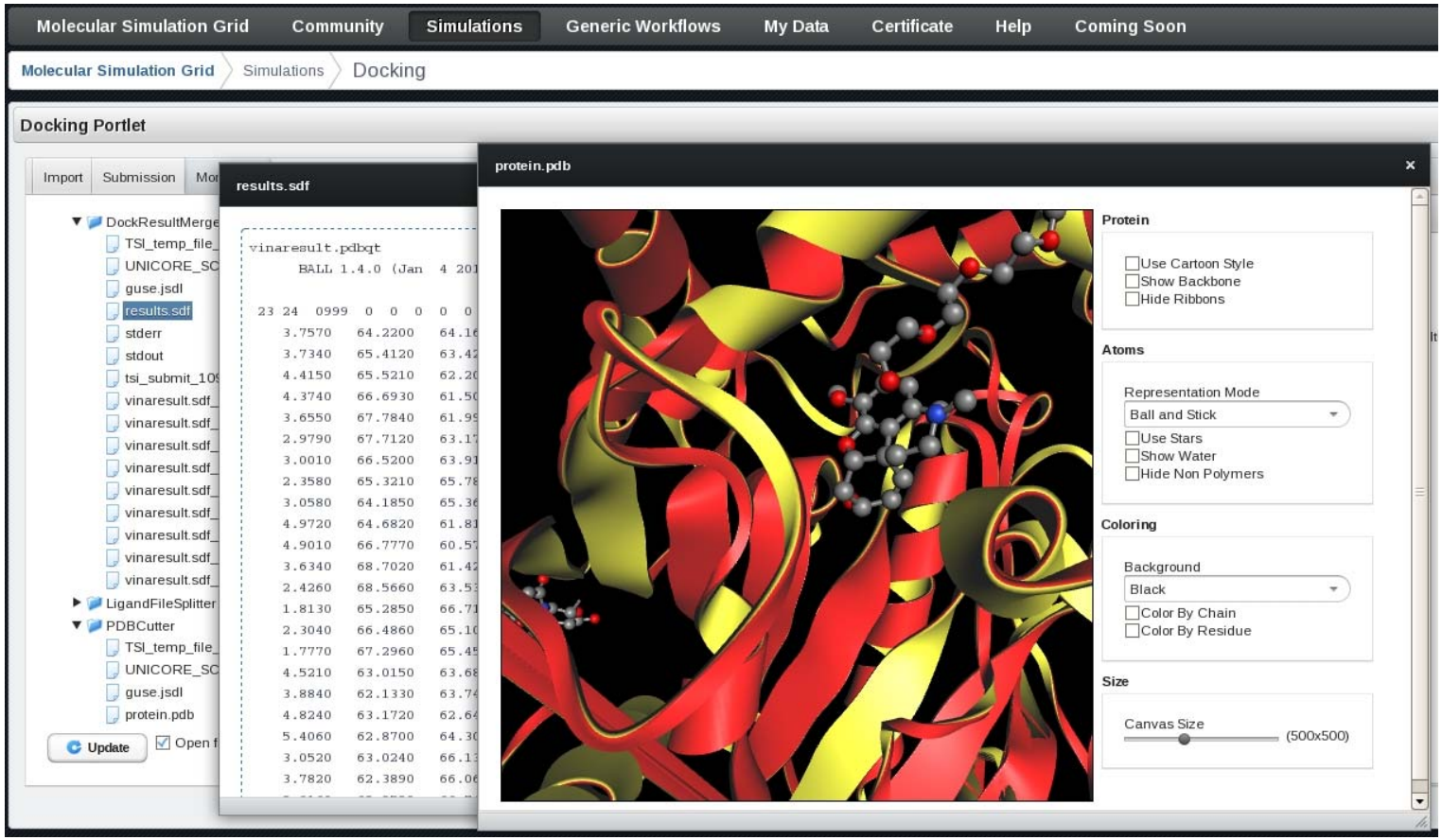
```xml
<module dictRef="compchem:initialization">
    <parameterList>
        <parameter dictRef="g09:loglevel">
            <scalar dataType="xsd:string" units="si:none">p</scalar>
        </parameter>
        <parameter dictRef="g09:jobtype">
            <scalar dataType="xsd:string" units="si:none">opt</scalar>
        </parameter>
        <parameter dictRef="g09:hf.theory" cmlx:editable="true">
            <scalar dataType="xsd:string" units="si:none">hf</scalar>
        </parameter>
        <parameter dictRef="g09:basisset" cmlx:editable="true">
            <scalar dataType="xsd:string" units="si:none">6-31G</scalar>
        </parameter>
        <parameter dictRef="g09:formal.charge" cmlx:editable="true">
            <scalar dataType="xsd:integer" units="si:none">0</scalar>
        </parameter>
        <parameter dictRef="g09:spin" cmlx:editable="true">
            <scalar dataType="xsd:integer" units="si:none">1</scalar>
        </parameter>
        <parameter dictRef="g09:checkpointfile">
            <scalar dataType="xsd:string" units="si:none">job.chk</scalar>
        </parameter>
    </parameterList>
    <cmlx:adapterConfiguration adapterID="g09adap"
        fileExtension="com" portName="job.com"/>
</module>
```



J. Krüger, R. Grunzke, S. Gesing, S. Breuers, A. Brinkmann, L. de la Garza, O. Kohlbacher, M. Kruse, W.E. Nagel, L. Packschies, R. Müller-Pfefferkorn, P. Schäfer, C. Schärfe, T. Steinke, T. Schlemmer, K. Warzecha, A. Zink and S. Herres-Pawlis: *The MoSGrid Science Gateway – A Complete Solution for Molecular Simulations*, Journal of Chemical Theory and Computation 2014, 10 (6), 2232–2245, DOI: 10.1021/ct500159h.

Teaching

EBERHARD KARLS UNIVERSITÄT TÜBINGEN



S. Delgado Olabarriaga, A. Benabdelkader, M. W. A. Caan, M. Mahdi Jaghoori, J. Krüger, L. de la Garza, C. Mohr, B. Schubert, A. Danezi and T. Kiss: *WS-PGRADE/gUSE-Based Science Gateways in Teaching*, in Science Gateways for Distributed Computing Infrastructures, ed. by P. Kaczuk, Springer 2014, ISBN 978-3-319-11267-1

Conclusion

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

- Representing scientific simulations protocols as workflows can be highly useful
- The reusability and reproducibility is largely enhanced
- Science gateways offer convenient access to compute and storage resources
- Collaborative working is improved
- The entrance hurdle is lowered, especially for novice users

Acknowledgment

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

- All group members
- All project partners
- The people in the compute centers supporting us every day

J. Krüger, R. Grunzke, S. Gesing, S. Breuers, A. Brinkmann, L. de la Garza, O. Kohlbacher, M. Kruse, W. Nagel, L. Pachschies, R. Müller-Pfefferkorn, P. Schäfer, C. Schärfe, T. Steinke, T. Schlemmer, K. Warzecha, A. Zink, S. Herres-Pawlis,
The MoSGrid Science Gateway – A Complete Solution for Molecular Simulations,
J. Chem. Theory Comput. 2014, 10 (6), 2232–2245. http://pubs.acs.org/doi/abs/10.1021/ct500159h

SPONSORED BY THE

Federal Ministry of Education and Research

SCI-BUS

BMBF (Förderkennzeichen 01IG09006)

RI 283481

IWSG ’16

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

- International Workshop on Science Gateways
- https://sites.google.com/a/nd.edu/iwsg2016/home
- 8-10 June 2016
- Rome, Italy

IWSG ‘16

8th International Workshop on Science Gateways

8-10 June 2016
Rome, Italy

Thank you!