



# Efficient Mass Spectra Prediction through Container Orchestration with a Scientific Workflow

Maximilian Hanussek  
High Performance and Cloud Computing Group  
Zentrum für Datenverarbeitung  
Eberhard Karls Universität Tübingen

20.06.17  
Poznań



- Introduction
- Key-technologies
- UNICORE workflow for QCEIMS
- Results & Discussion
- Conclusion/Outlook



- Many useful tools are available
  - Hard to install
  - Difficult to use
- Small user community
- Use existing technologies to simplify installation and usage
- Broaden user community

Goal: Evaluation of Docker and UNICORE

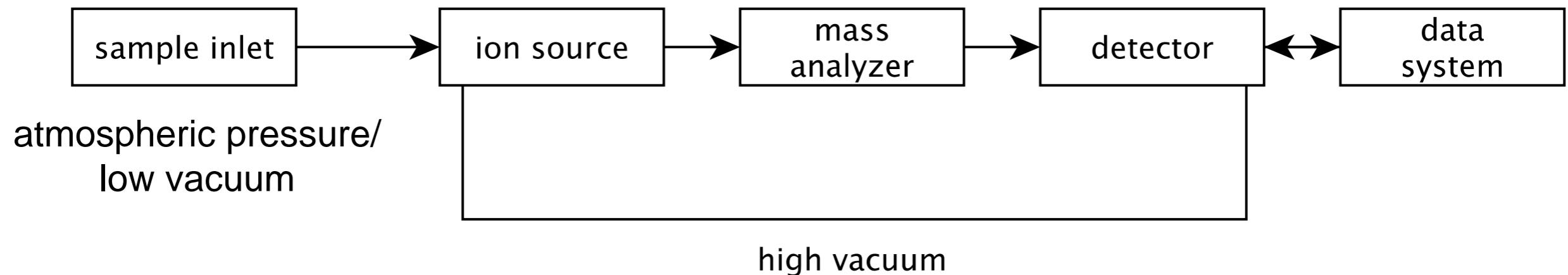


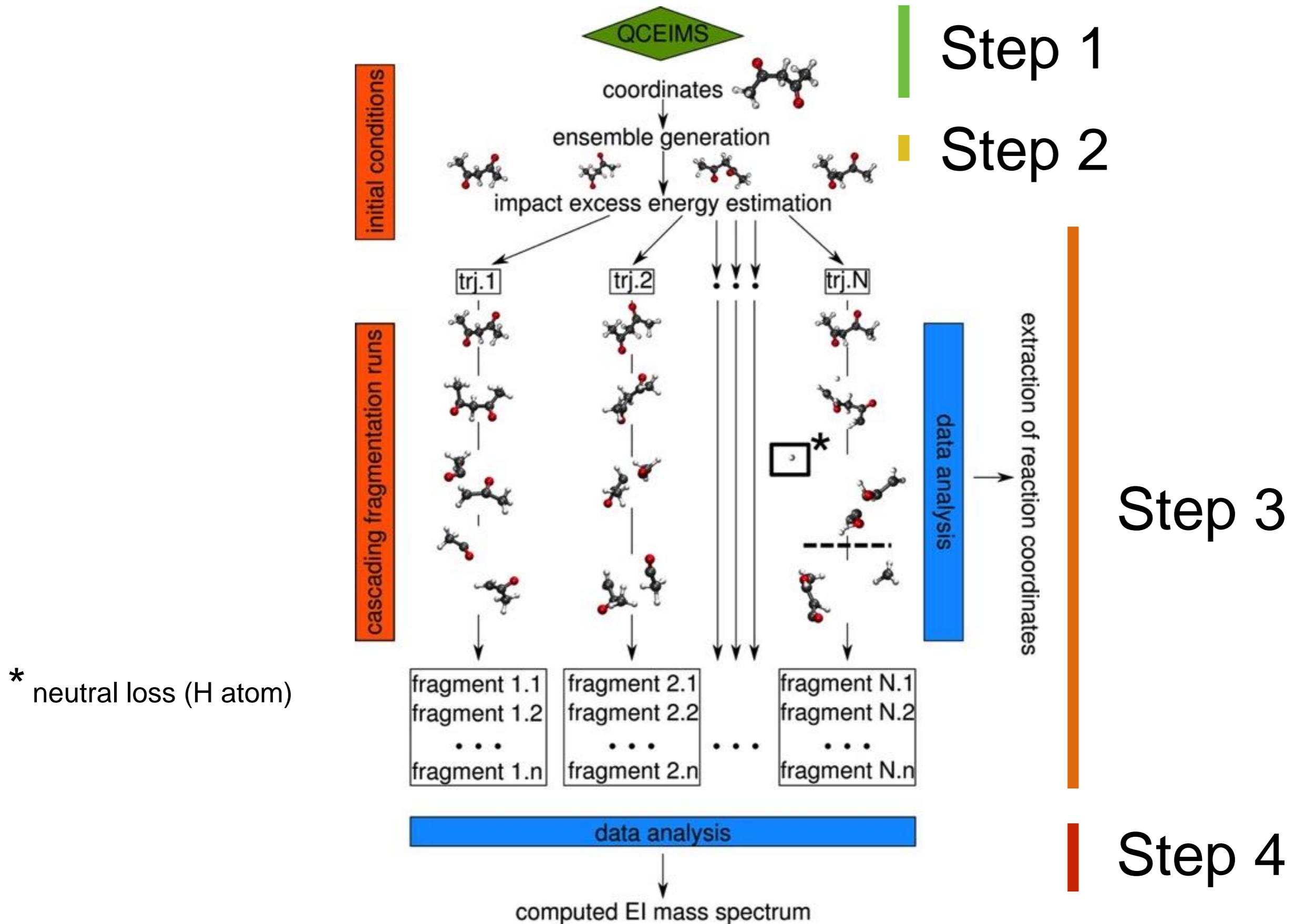
- Quantum Chemical Electron Ionization Mass Spectrometry (QCEIMS)
  - Mass spectra prediction tool (Stephan Grimme et. al)<sup>1</sup>
  - Complex execution procedure
  - Many required additional tools and software packages
  - HPC
- Nonetheless well working tool with good results

<sup>1</sup>Towards First Principles Calculation of Electron Impact Mass Spectra of Molecules. Angew. Chem. Int. Ed., 52(24):6306–6312, 2013.



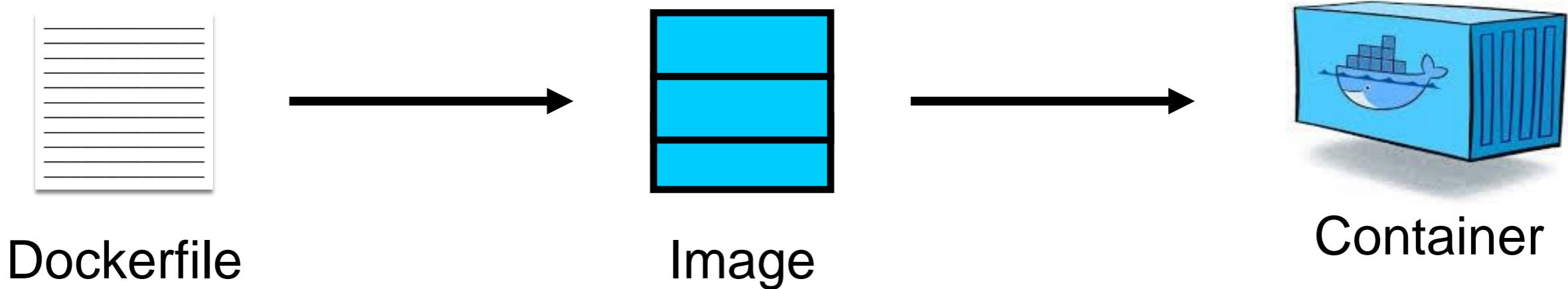
- Quantum Chemical Electron Ionization Mass Spectrometry (QCEIMS)
- Docker (container virtualization technology)
- Uniform Interface to Computing Resources (UNICORE)







# Key-technologies (Docker)



- Built Docker Image from Dockerfile
- Most tools included (MNDO99, DFTB+, ORCA, PubChemPy, ...)
- Stable computing environment
- Installation with root permissions



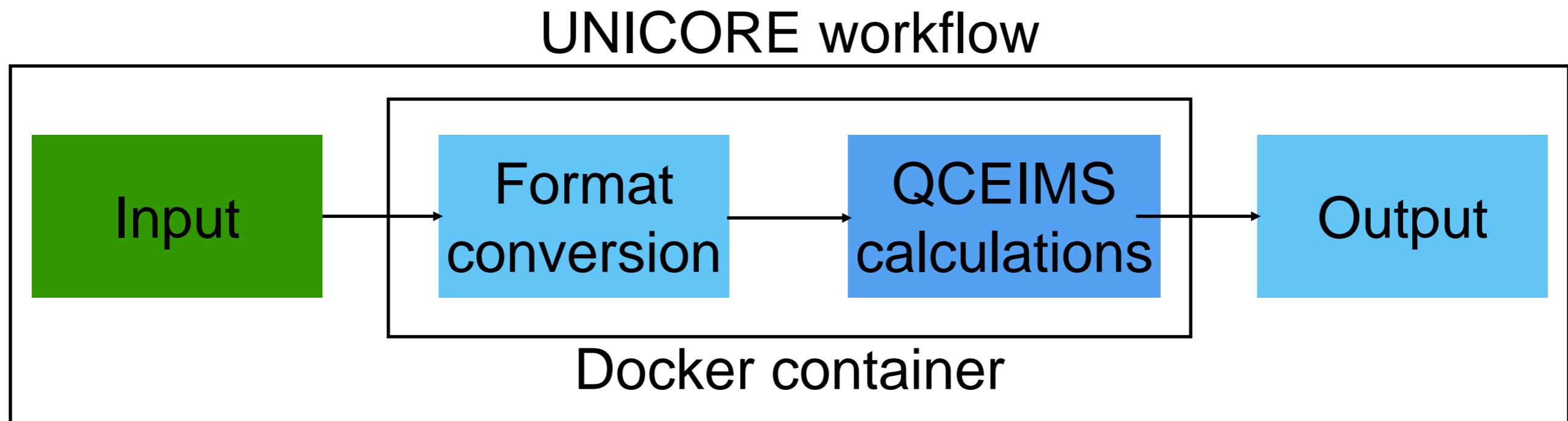
- Middleware software for distributed computing systems
- Developed at the research center Jülich and by further partners
- Offering different components:

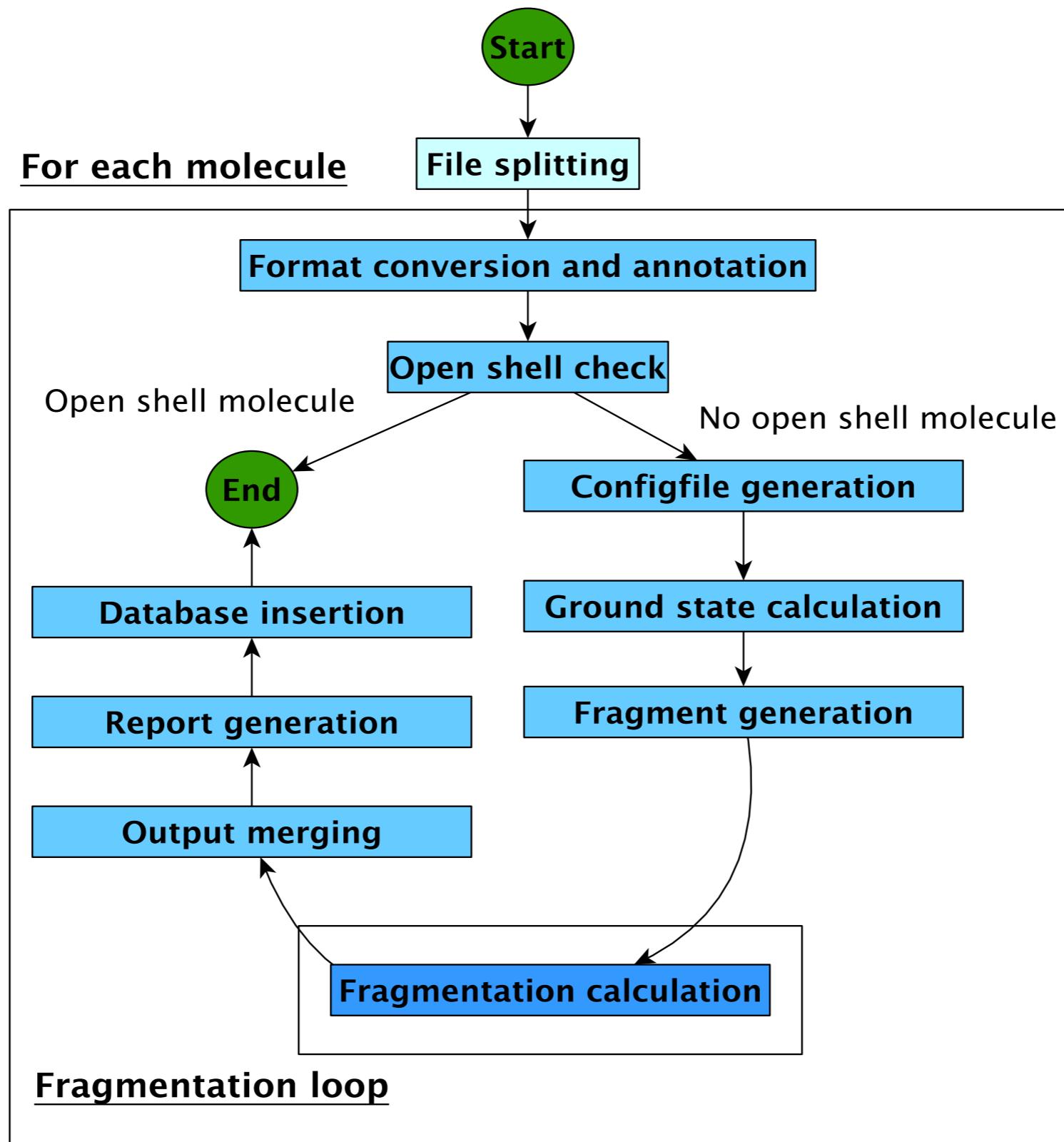
UNICORE Server	UNICORE Workflows	UNICORE Portal
<ul style="list-style-type: none"><li>• Job submission</li><li>• Data handling</li><li>• Resource allocation</li></ul>	<ul style="list-style-type: none"><li>• Workflow engine</li><li>• Connected to UNICORE Server</li></ul>	<ul style="list-style-type: none"><li>• Web access</li><li>• Use existing workflows</li></ul>

Benedyczak and Schuller. Unicore 7—middleware services for distributed and federated computing. In High Performance Computing & Simulation (HPCS), 2016 International Conference on, pages 613–620. IEEE, 2016.



- Combined usage of Docker and UNICORE
- Automating the QCEIMS calculations
- Most software already preinstalled in Docker image



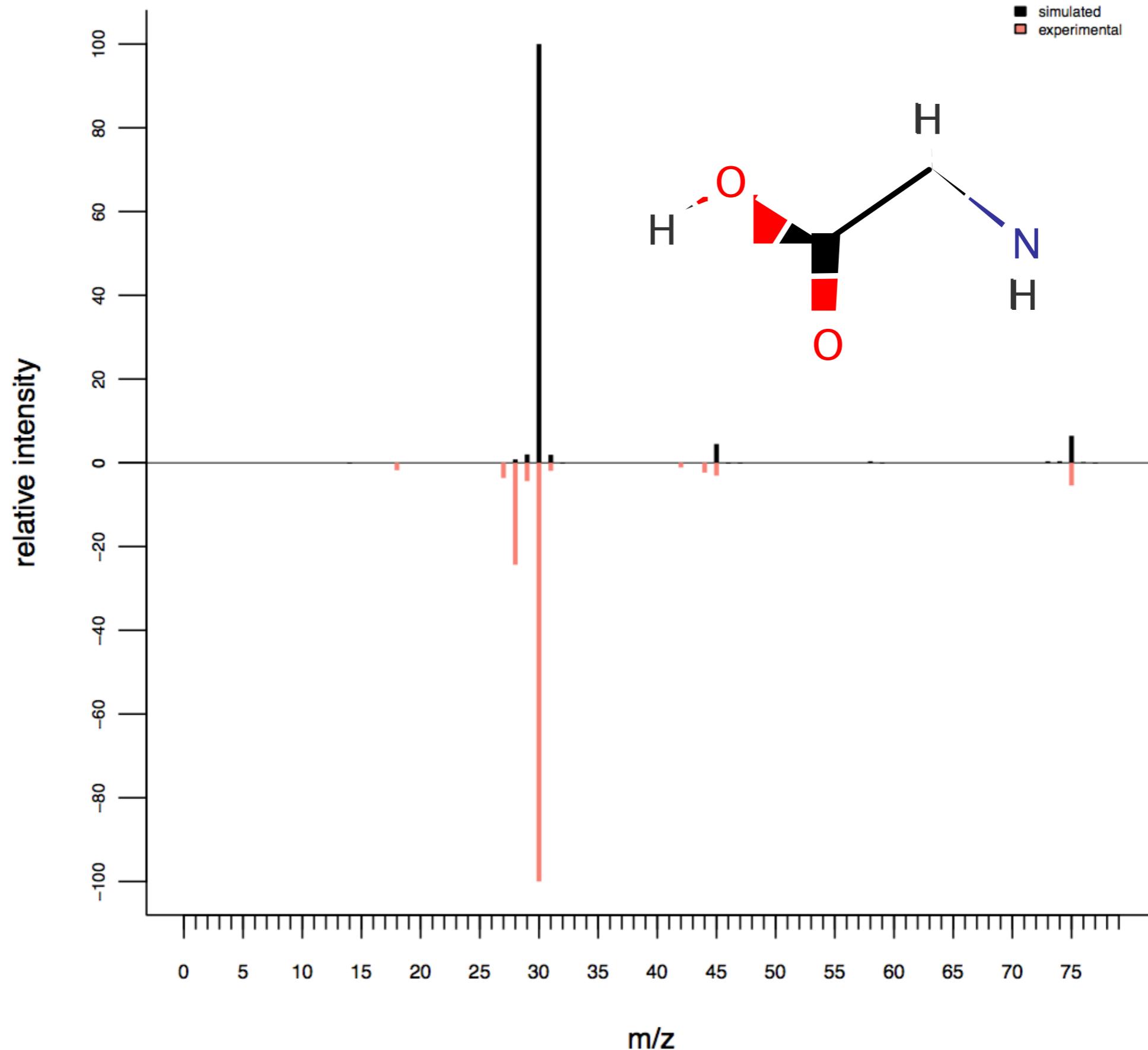




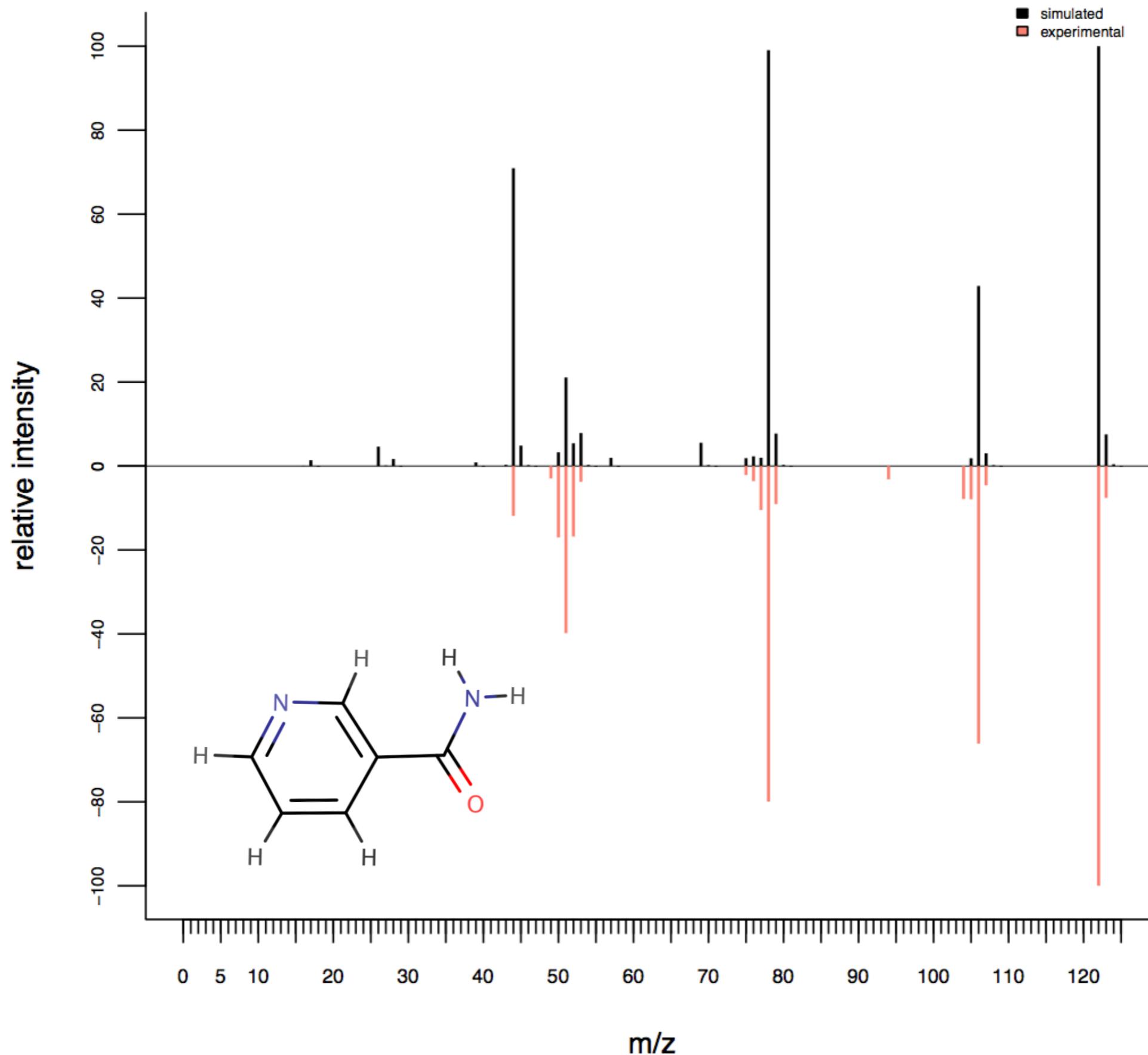
- 11 molecules
- 2 QC tools (MNDO99, DFTB+)
- OM2-D3 (MNDO99)
- DFTB3-D3 (DFTB+)
- Absolute value distance<sup>2</sup>:

$$D_A(I_E, I_S) = \left| 1 + \frac{I_E - I_S}{I_S} \right|^{-1}$$

<sup>2</sup> Stein and Scott. Optimization and testing of mass spectral library search algorithms for compound identification. Journal of the American Society for Mass Spectrometry, 5(9):859–866, 1994.

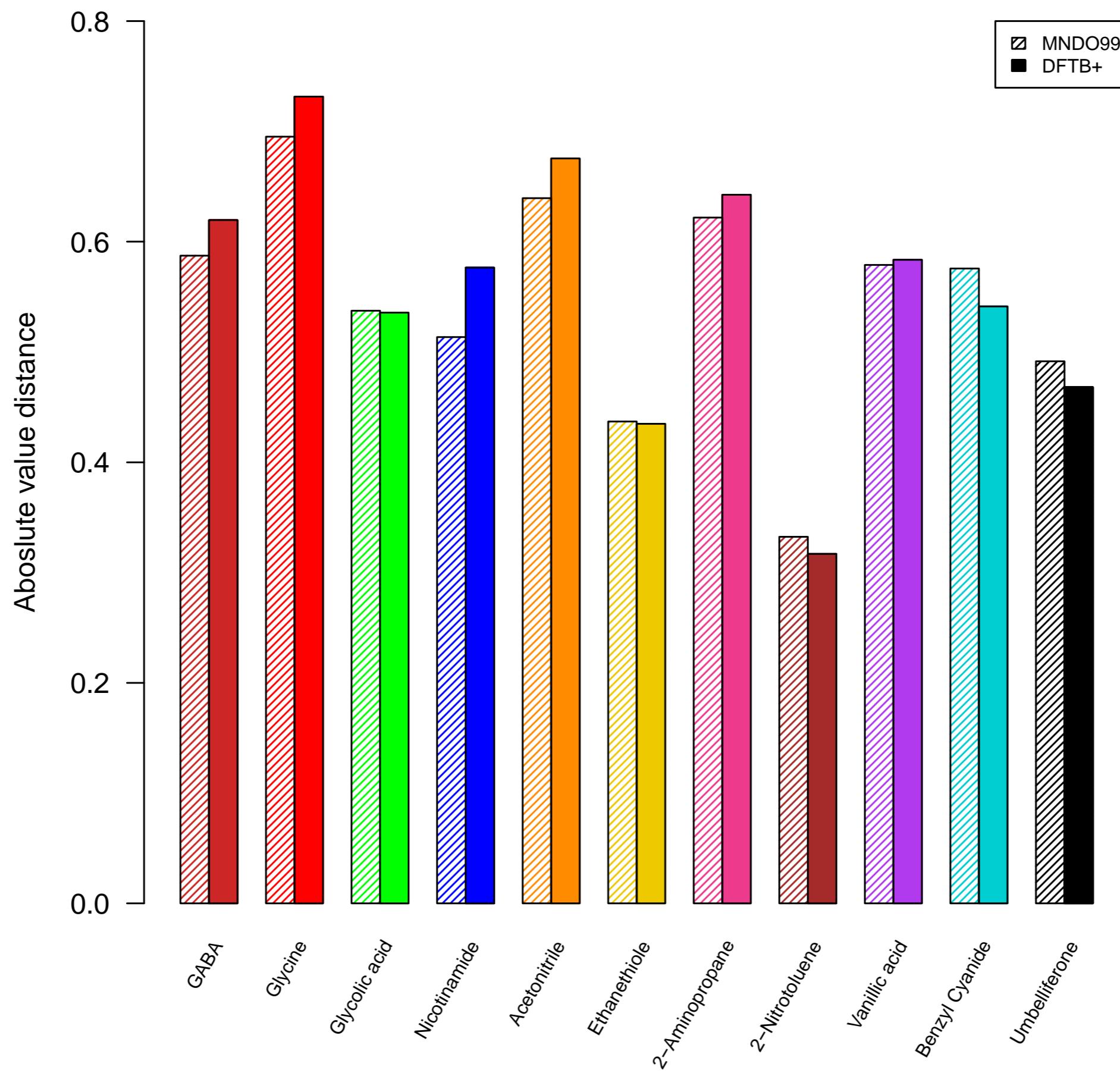


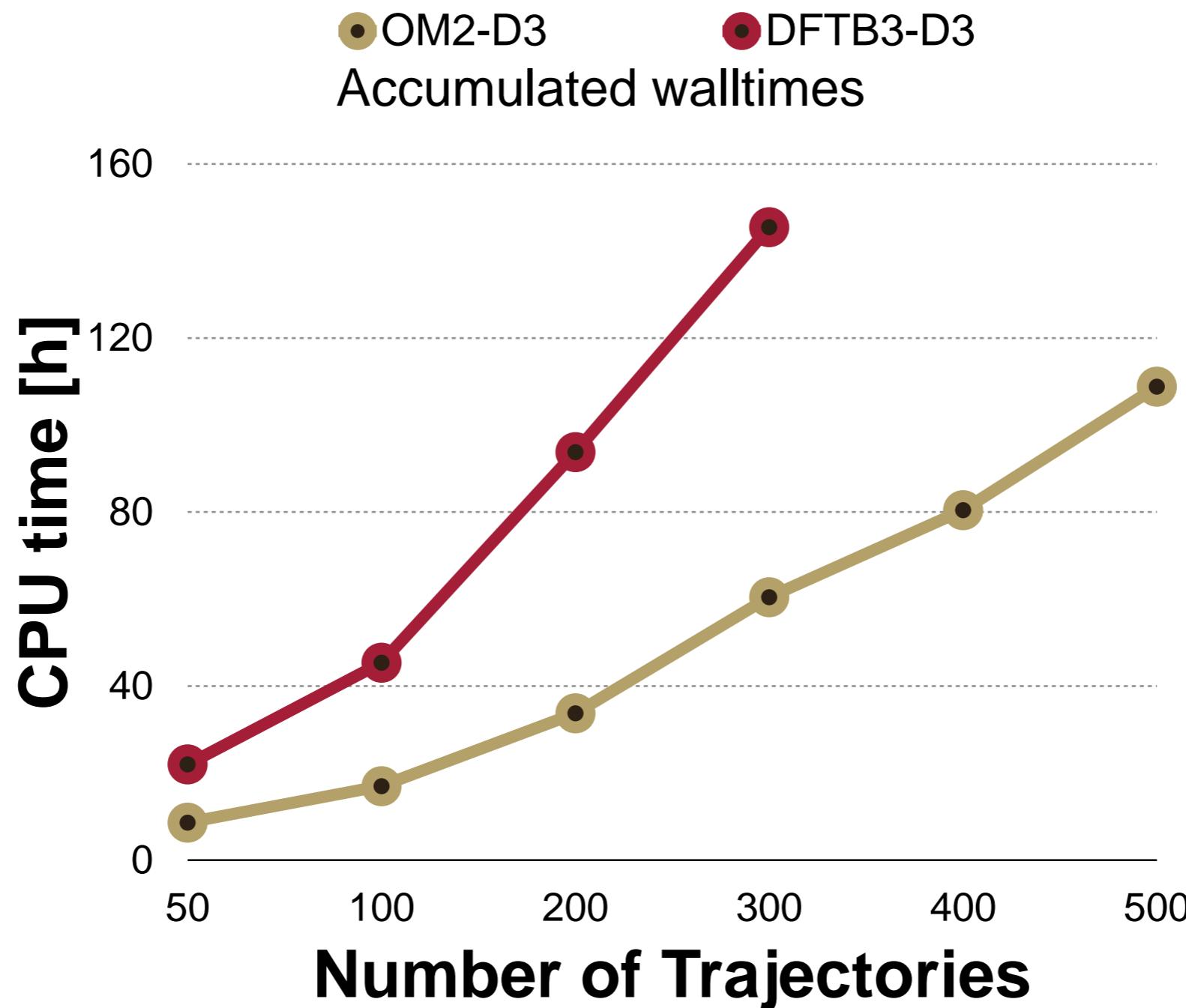
Glycine (DFTB3-D3, 300 trj.,  $D_A$ : 0.752)



Nicotinamide (DFTB3-D3, 200 trj.,  $D_A$ : 0.660)

## Mean molecule score







- Successful integration as wrapping tool
- Occurred problems
  - Docker group = root
  - User mapping is mandatory
  - Manual garbage collection of exited and dead containers
  - Accumulation of Docker metadata (/var/lib/docker)



- Challenges to overcome
  - Encapsulation of multiple molecules
  - Encapsulation of a molecule for fragmentation
  - Merging generated results by unique identifier
  - Explicit data staging
  - QC tool monitoring



- Unicore-portal workflow adaption
  - Export workflow from UNICORE Rich Client (URC)
  - Integrate input field
  - Set independent path to workflow node scripts



UNICORE portal authentication

[TLS login](#)

You will be logged as: CN=Maximilian Hanussek,OU=Universitaet Tuebingen,O=Gric

[Login with certificate](#)

[Switch to user registration](#)

[Home](#)[Create Job](#)[My Jobs](#)[New Workflow](#)[My Workflows](#)[My Sites](#)[Data Manager](#)

## Job Computation



### New UNICORE Job ×

#### Application

Job name:

Tag:

▼

Select application:

 ▼

Select version:

 ▼

Command line arguments:

### Workflow template parameters

[Select a template](#)

[Editing workflow template file: UNICORE\\_QCEIMS\\_workflow.xml](#)

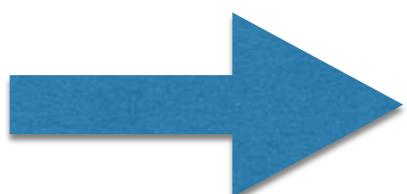
splitSdf\_INPUT:



Submit



- Successful simplification of QCEIMS
  - Default parameter change, QC tools (ORCA)
- Docker: Great potential, serious problems
- UNICORE: Generally applicable to complex tools
  - Login/Registration procedure



Relatively good *in silico* spectra  
is better than no spectrum.



## Acknowledgment

---

Applied Bioinformatics group Tübingen

- Prof. Dr. Oliver Kohlbacher

High performance and cloud computing group Tübingen

- Dr. Jens Krüger

UNICORE-Support

- Dr. Bernd Schuller

QCEIMS

- Christoph Bauer, Prof. Dr. Stefan Grimme