Ab-initio Semi-Empirical Mass Spectra Predictions with Galaxy

Wudmir R, Helge H, Zargham A, Elliot P, Jana K

helge.hecht@recetox.muni.cz

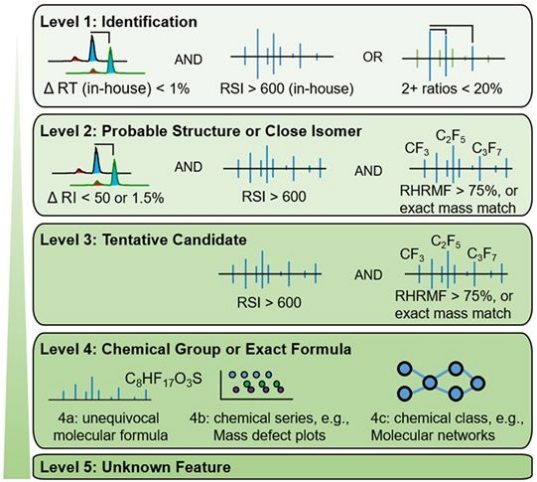
MUNI | RECETOX

Motivation

- MS data annotation poses a universal bottleneck in research.
- *In silico* spectra prediction using machine learning or quantum chemistry is a promising technique for annotation of unknown compounds.
- QCxMS offers reasonably accurate in silico annotation, especially for organic molecules.
- The complexity of quantum chemistry predictions presents challenges for non-HPC experts.
- Integrating QCxMS into Galaxy provides valuable molecular insights.

Our Goal: Make semi-empirical Quantum Chemistry (QC)-based predictions accessible without advanced computational skills.

DOI: 10.1093/exposome/osac007



90 analytes spiked in serum
18 ppb
61 annotated
- 54% annotated wrongly at Level 2
- 60% annotated wrongly at Level 3

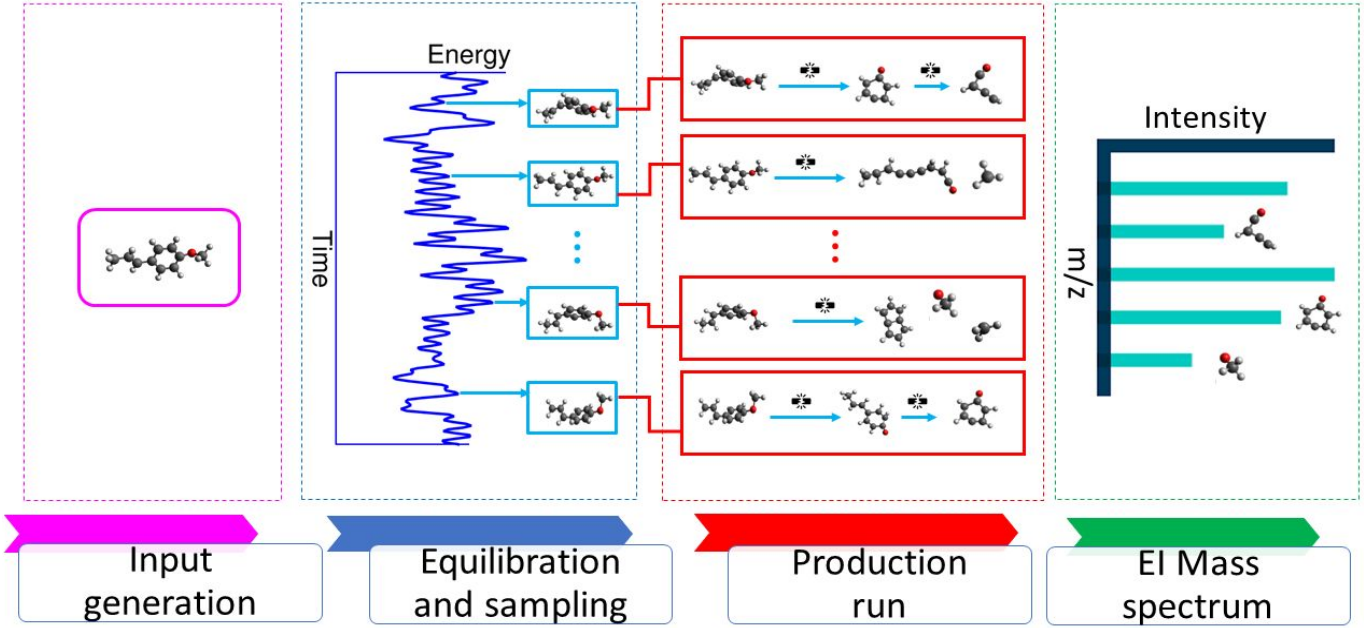
QCxMS Spectra Prediction - Method

Sampling:
- Molecular dynamics (MD):
- T=500 K
- Microcanonical NVE assemble

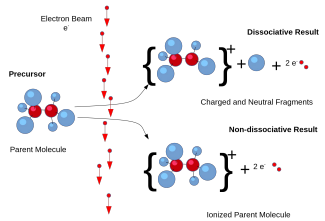
Ionization and Heating phase
- Remove one electron
- Internal conversion (IC)
- Internal Excess Energy (IEE)

Evolution of ion:
- MD on ion at steps of 0.5 fs
- Track secondary fragmentations recursively.
- Choose largest charge part post-dissociation.
- Start new trajectories without extra heating.

Counting:
- 50-100 counts in base peak.
- Record neutral loss.

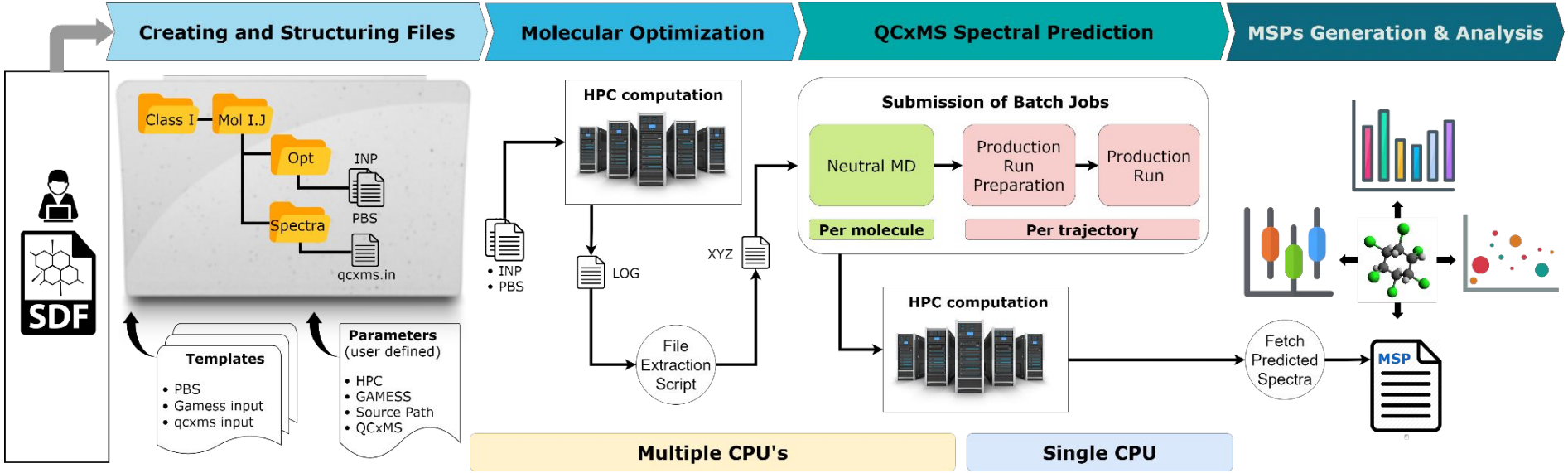


Electron Impact process



Zenodo. https://doi.org/10.5281/zenodo.6791061

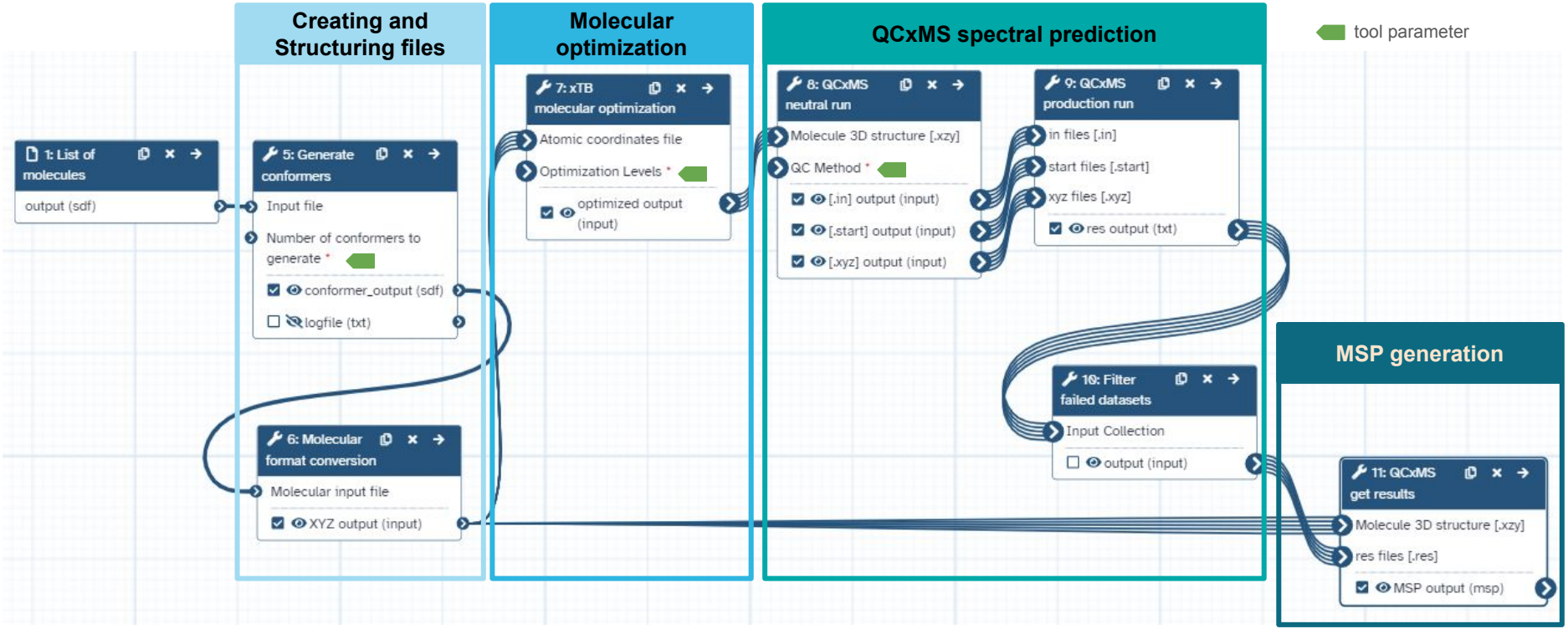
QCxMS Spectra Prediction - HPC Workflow



HPC workflow: https://doi.org/10.5281/zenodo.10853686

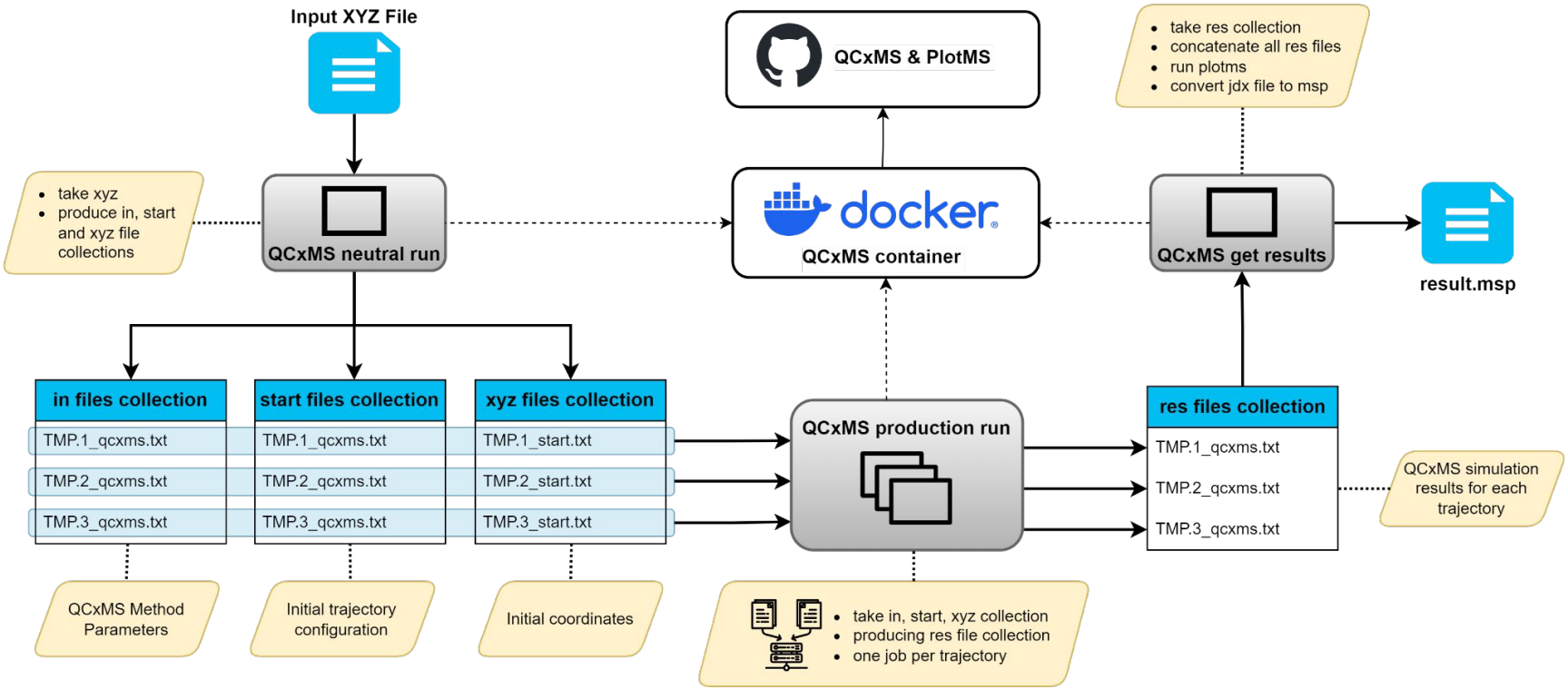
Hecht, et. al. Quantum chemistry based prediction of electron ionization mass spectra for environmental chemicals. ChemRxiv. 10.26434/chemrxiv-2024-2ngwq-v2

QCxMS Spectra Prediction - Galaxy Workflow



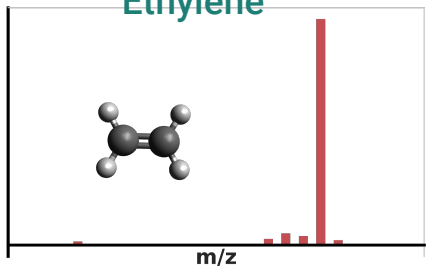
Galaxy workflow: https://doi.org/10.48546/WORKFLOWHUB.WORKFLOW.897.1

QCxMS Galaxy Tool Structure



Runtime Performance Metrics

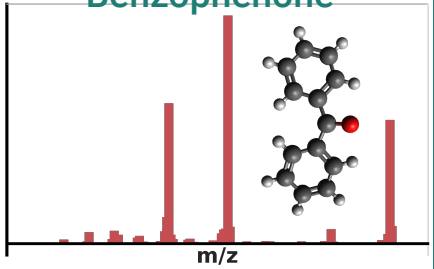
Ethylene



6 atoms
Elements: C, H

Slots: 155
Job Runtime (s): 34624
CPU usage time (s): 2325007517
CPU user time (s): 1716160059
CPU system time (s): 608847386
Memory allocated (TB): 0.58

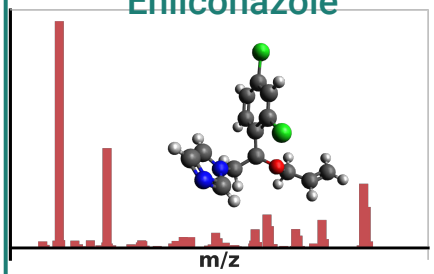
Benzophenone



24 atoms
Elements: C, H, O

Slots: 605
Job Runtime (s): 679037
CPU usage time (s):10185690473
CPU user time (s): 7641539889
CPU system time (s): 2544150283
Memory allocated (TB): 2.25

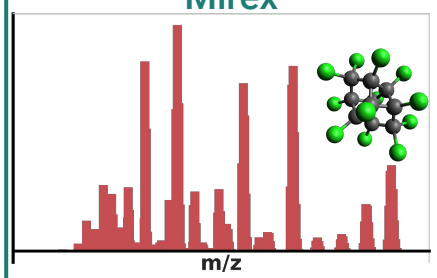
Enilconazole



33 atoms
Elements: C, H, O, N, Cl

Slots: 830
Job Runtime (s): 1720209
CPU usage time (s):13987695689
CPU user time (s): 10219951769
CPU system time (s): 3767743519
Memory allocated (TB): 3.08

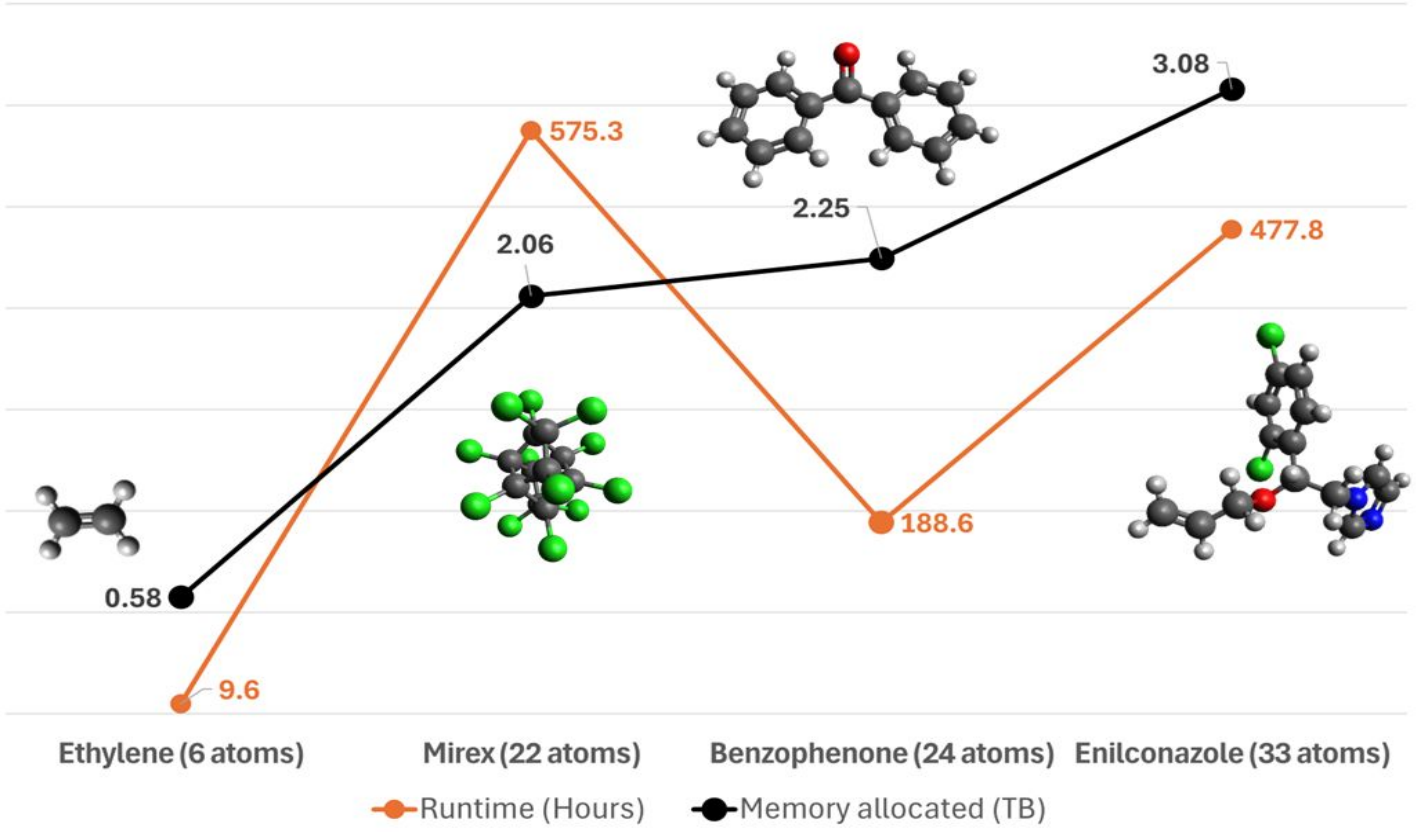
Mirex



22 atoms
Elements: C, Cl

Slots: 555
Job Runtime (s): 2070941
CPU usage time (s): 9914289674
CPU user time (s): 7506592875
CPU system time (s): 2407696515
Memory allocated (TB): 2.06

Runtime Performance Metrics - Examples



Acknowledgements

Human Exposome Research Group

Jana Klánová –Head
Kapil Mandrah – MSCA fellow
Žiga Tkalec – ERA fellow
Helge Hecht – PhD candidate
Akrem Jbebli – PhD candidate
Hana Seličová – PhD candidate
Thomas Contini – PhD candidate

Biomarker Analytical Laboratories RI

Eliška Benešová - Researcher
Kateřina Coufaliková – Senior researcher
Gabriela Přibyl Dovrtělová - Researcher
Štěpán Koudelka – Senior researcher
Veronika Vidová – Senior researcher
Helge Hecht – Senior researcher
Wudmir Yudy Rojas Verastegui – Senior researcher
Zargham Ahmad – Researcher
Kristína Gömöryová – Senior Researcher

Trace Analytical Laboratories RI

Petra Přibylová & team

Data services

Richard Hůlek & team

EIRENE RI coordination

Jan Ostřížek
Shachar Dvir

RECETOX RI coordination

Petra Růžičková

Operations and management

Martin Životek, Šárka Palátová & team

Partners

Institute of computer science at MU
CERIT-SC
CESNET
ELIXIR-CZ
Galaxy Project
EIRENE consortia

MUNI | RECETOX