OpenMS - A Software Platform for Shotgun Proteomics

Marc Sturm[*], Eva Lange[†], Clemens Gröpl[†], Ole Schulz-Trieglaff[†],
Nico Pfeifer[*], Oliver Kohlbacher[*], Knut Reinert[†]

Over the last few years numerous experimental techniques for the differential analysis of proteome expression levels have been developed and HPLC-MS has become one of the standard techniques in high-throughput shotgun proteomics. The acceptance as well as the impact of an experimental technique is however limited by the ability to efficiently handle and analyze the large amount of resulting data. Therefore the development of software platforms for mass spectrometry-based proteomics is currently a very active field of research.
Our *OpenMS* C++ framework offers data structures and algorithms for the major steps in the analysis of HPLC-MS experimental data, which are: preprocessing, peak picking, protein identification, quantitation, alignment of samples, and statistical analysis. In contrast to other proteomics open source software projects ([1],[2],[3]) OpenMS offers a broader range of functionalities in a consistent package.
OpenMS ensures interoperability with MS vendor software and other proteomics tools through standard formats (mzData, mzXML, ANDI-MS, AnalysisXML) and several other file formats. File handling is complemented by the OpenMS database schema (SQL) which is based upon the PEDRO/MIAPE model (HUPO-PSI).

The OpenMS Proteomics Pipeline (TOPP) [4] packages the major analysis algorithms of OpenMS in a set of convenient computational tools. These tools range from simple utilities (file format conversion, peak picking, visualization) over wrapper applications for known applications (e.g. Mascot, Sequest) to completely new algorithmic techniques for data reduction and data analysis. All the components can be used both individually as command line tools or easily chained into linear or more complex pipelines. OpenMS is fully documented and comes with an elaborate tutorial and various examples.

In our software demonstration we will show the design of the OpenMS framework and give a short introduction how to implement a proteomics application using OpenMS classes. Furthermore we will demonstrate a number of example proteomics workflows based on TOPP components.

OpenMS and TOPP are available as open-source software under the lesser GNU public license (LGPL). Source code is available from the project web site at `www.OpenMS.de`.

References

[1] K. C. Leptos et al. MapQuant: open-source software for large-scale protein quantification. *Proteomics*, 6(6):1770–1782, 2006.

[2] M. Bellew et al. A suite of algorithms for the comprehensive analysis of complex protein mixtures using high-resolution LC-MS. *Bioinformatics*, Advance Access published June 9, 2006.

[3] A. Keller et al. A uniform proteomics MS/MS analysis platform utilizing open XML file formats. *Mol. Syst. Biol.*, 1:1744–4292, 2005.

[4] O. Kohlbacher et al. TOPP - The OpenMS proteomics pipeline. *Bioinformatics*, 2006. to appear.

[*] Eberhard Karls University Tübingen, Simulation of Biological Systems, Tübingen, Germany
[†] Free University Berlin, Algorithmic Bioinformatics, Berlin, Germany