

Einladung zur Vortragsreihe *Algorithmische Bioinformatik*

Herr Dr. Clemens Gröpl, FU Berlin

spricht über

Perspectives in Algorithmic Mass Spectrometry

Datum: Dienstag, 23. Dezember 2008
Zeit: 9:00 Uhr s.t.
Ort: B-IT, Dahlmannstr. 2, Rheinsaal

The combination of liquid chromatography and mass spectrometry is an essential analytical technique in proteomics and metabolomics. The development of new chromatographic separation techniques, precise mass analyzers and experimental protocols is a very active field of research. This leads to more complex experimental setups yielding ever increasing amounts of data. Consequently, analysis of the data is currently often the bottleneck for experimental studies.

Most of my recent work has been centered around OpenMS, our open-source software framework for rapid application development in mass spectrometry. Although software tools for many data analysis tasks are available today, they are often hard to combine with each other or not flexible enough to allow for rapid prototyping of a new analysis workflow. OpenMS has been designed to be portable, easy-to-use and robust while offering a rich functionality ranging from basic data structures to sophisticated algorithms for data analysis. OpenMS is available under the Lesser GNU Public License (LGPL) from the project website at <http://www.openms.de>.

Started in 2003, the project has now reached a fairly mature state and algorithmic solutions are available for most of the essential steps in a typical data processing pipeline. I will give an overview of OpenMS and highlight some newly developed algorithms for peak picking in mass spectra, detection of two-dimensional features in LC-MS data, and the alignment of multiple LC-MS measurements.

I will then move on to perspectives of ongoing and future research. Many aspects of the later stages of LC-MS/MS data processing can best be described using concepts from graph theory. The spectrum graph of a (single) MS/MS spectrum is fundamental for the identification of peptides, especially de novo approaches. The similarity structure between (multiple) MS/MS spectra or identified peptides is captured by so-called spectral networks, which bear great potential to reduce false peptide identifications and can help with protein inference. A related line of research aims to improve our understanding of the genome using peptide mass spectrometry (proteogenomics). The initial assignment of peptides to genomic sites is refined by analysis of the underlying graph. This way, the relevance of alternative splicing forms or single nucleotide polymorphisms can be confirmed on the translational level. Related problems also occur when reads from next-generation deep sequencing technology are mapped back to the genome; ambiguous matches should not be disregarded, and repeats are a major challenge. Indexing data structures are mandatory to speed up the similarity search, but these will be more complicated in the case of peptides due to the nature of their MS/MS spectra.