The rpm package: aligning LC/MS mass spectra with R

Saussen B.[1], Kirchner M.[1,2,3], Steen H.[2,3]
Steen J.A.J.[2,3], Hamprecht F.A.[1]

[1]Interdisciplinary Center for Scientific Computing,
University of Heidelberg, Germany
[2]Steen & Steen Lab, Children's Hospital Boston, Boston, USA
[3]Systems Biology, Harvard Medical School, Boston, USA

The application of mass spectrometry methods in Systems Biology and particularly in Proteomics is a rapidly evolving and promising field, that provides high-accuracy qualitative and quantitative measurements.

For the LC/MS analysis of complex (e.g. organic) samples, mass spectrometers (MS) are operated in line with liquid chromatography (LC) systems, which uses an organic solvent gradient to separate the sample based on the chemical properties of its constituents before subjecting it to ionization and mass analysis.

Despite the usefulness of the LC separation, this procedure also presents several analytical challenges while processing data. Due to physical properties of the LC process, multiple runs of the same sample as well as comparative runs of different samples suffer from non-linear shifts in the retention time domain, rendering direct comparisons difficult or impossible.

Proper sample registration is also required for samples which do not use the LC separation dimension as one often observes small shifts along the mass/charge domain amongst multiple samples in MS analysis.

We provide an R package called "rpm", which implements the Robust Point Matching Algorithm (RPM) of [Chui and Rangarajan, 2000] which we successfully applied to the registration of real-world MALDI and LC/MS data. With RPM being non-landmark-based, outlier-insensitive and capable of modeling non-linear transformations, we were able to overcome the drawbacks of state-of-the-art methods that are mainly based on piecewise linear alignment of hand-picked landmark peaks and to achieve proper registration of samples.

We consider this work a successful effort to integrate R further into the bioinformatics and proteomics field as a platform for efficient data analysis and prototyping.