Transhumus: Integrated software solution for interpretation of FT-ICR mass spectra of natural organic matter

Anton Grigoryev[1, 2], Alexey Kononikhin[2, 3], Irina Perminova[4], Eugene Nikolaev[2, 3]
[1]ansgri@gmail.com
[2]Institute for Energy Problems of Chemical Physics, Moscow, Russia
[3]Institute of Biochemical Physics, Moscow, Russia
[4]Lomonosov Moscow State University, Moscow, Russia

Natural organic matter (NOM) and humic substances in particular are important research objects as they play important roles in the movement of nutrients in ecosystems and are one of the largest carbon reservoirs. NOM is very complex, consisting of thousands of chemically distinct compounds. Ultrahigh-resolution Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS) allows characterizing the compositional and structural diversity of NOM. However, there still are problems with interpretation of mass spectra of NOM. Particularly notable is the lack of specialized software tools for automated analysis of NOM mass spectra. The vast array of existing mass spectrometry software was created for proteomics, genomics, forensics, and pharmaceutical applications and is not suitable for specific needs of NOM analysis. We present software *Transhumus*, specifically designed for automated analysis of FT-ICR mass spectra of NOM.

The software aims to incorporate the most recent developments in interpretation of ultrahigh resolution mass spectra and can serve as a platform for further algorithm development. Main functionality is illustrated by Figure 1: you start with a peak list (m/z values), perform preliminary identification to assess mass measurement accuracy and number of compounds, recalibrate the mass spectrum using unambiguous identifications, and then you can explore the structure of the spectrum using the mass difference statistics. In particular, there is an efficient algorithm for charge state determination which doesn't depend on the presence of ^{13}C-isotopologues of compounds in the spectrum. The neutral mass list can then be used to obtain a final set of compound formulas, using both brute force and "formula extension" approaches.

In addition, the software includes means of result visualization (e.g. van Krevelen plots) and is rather fast, which enables its application in large-scale studies.

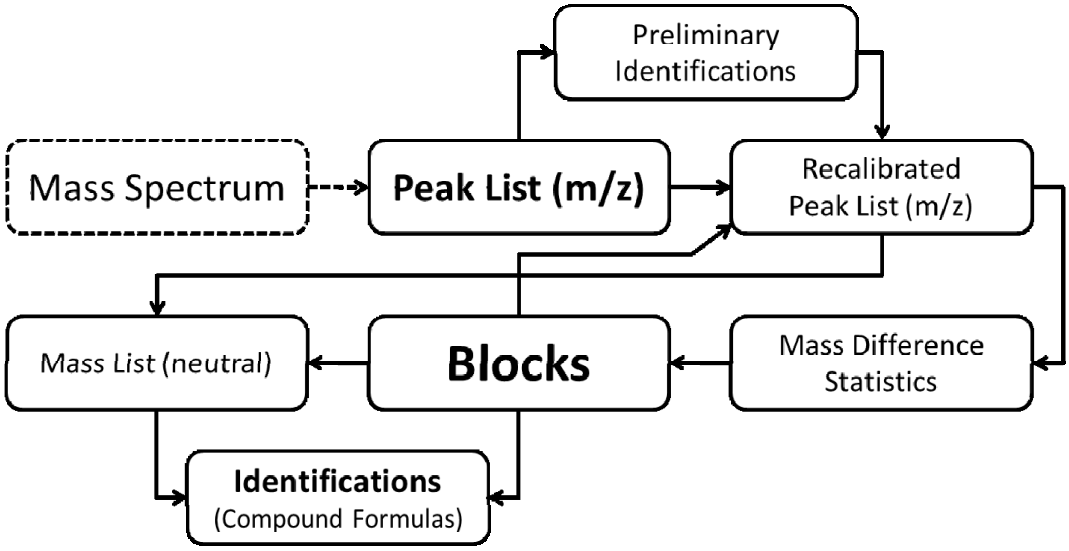

Figure 1. Spectrum interpretation workflow using *Transhumus* software. Boxes show data tables, arrows represent the relation "is needed to obtain" between data. The "Mass Spectrum" box is dashed because peak picking means are not included in *Transhumus*.