Detailed report

In the present OTKA project we have developed and used novel theoretical approaches to describe mass spectrometry experiments commonly used in biochemistry. Most of these relate to the energetics of fragmentation, whose knowledge is key for optimal peptide/protein sequencing using mass spectrometry. This research was based on the principal investigator's experience in fundamental mass spectrometry of small molecules (the MassKinetics model). The aim was to extend this model to macromolecules (especially towards proteomics), to ionization methods suitable for macromolecules (ESI and MALDI) and to the MS/MS fragmentation of peptides. This strategy was successful. Fundamentals developed for small molecules were successfully applied to the macromolecular field, and we have obtained a good theoretical basis of various methodologies used in proteomics workflows.

We have used the results of this basic research as a starting point to innovation: Using our theoretical models we have identified weak points in current mainstream mass spectrometry and proteomics workflows, and designed new experimental strategies to improve them. We use these improvements in our practical workflows, extending our studies beyond the state of art, and gaining a competitive advantage in industrial collaborations.

Below we describe our main achievements. Most of these results have already been published in leading international journals; some are in preparation.

1) Fundamentals and the use of internal energy distributions in biomolecular mass spectrometry

We have adapted our theoretical framework (MassKinetics) and the corresponding software for determining internal energy distributions in ESI and MALDI ionization. The internal energy distribution of thermometer ions that cross the desolvation region of an electrospray source was simulated using this software. The desolvation region is considered as a collision zone and a partially elastic multiple-collision model is used to account for the accumulation of internal energy in the gaseous ions. The ion survival yields of the theoretical mass spectra calculated by the MassKinetics software, were fitted with the experimental ion survival yield of substituted benzylpyridinium cations, obtained in the case of a PE SCIEX ESI source interfaced with a quadrupole mass spectrometer. This model allowed development of a new evaluation method for internal energy distribution of ions in various parts of an ESI source. The internal energy distributions were found to be similar to that of a thermal distribution at "characteristic temperature" (T_{char}) of between 1020 and 1550 K. In addition, this study also provided evidence for a linear correlation between T_{char} and the mean internal energy. The characteristic temperature is a useful feature of the mass spectrometry experiment; much better than the collision (or skimmer-cone) voltage, commonly used in proteomics. Using T_{char} instead of collision energy to set up mass spectrometry experiments is likely to lead to better reproducibility and much easier method transfer in proteomics.

Internal energy effects on ESI ionization have been studied on various instrument types, on QTOF, ion trap, QQQ and Orbitrap type mass spectrometers. Tandem mass spectra of synthetic peptides have been measured on 3 different instruments (QTOF, QQQ, IT); a new technique was established to determine activation parameters for ESI ionization. Using the developed method, activation parameters for leucine enkephalin have been determined.

We have developed a novel method to perform single collision experiments on a quadrupole type instrument, which was thought to be impossible before. We have decreased the collision gas pressure by 2-3 orders of magnitude in a QQQ type instrument, which is considered well out of the working range. This led to conditions, when the selected ion, on average, collides less than once in an MS/MS experiment. Under such conditions it is easier, more reliable, and requires less modeling assumptions to determine the internal energy distribution of the selected ion. This in turn leads to more reliable activation parameters than previous techniques, and requires simple, widely available instrumentation. This development not only leads to better description of energetics, but also offers a better understanding of various (MSMS) processes within mass spectrometers.

While we had managed to get major advances describing fundamentals in ESI and MS/MS experiments, those planned for MALDI and BIRD were less successful. We have determined the internal energy distribution in MALDI spectra and compared different matrices, as outlined in the workplan. However, this line of research did not lead to a major breakthrough, and was abandoned. It was not possible to establish collaboration using the BIRD technique (available in only 2-3 labs worldwide), so this line of research was not pursued.

Results described above have been published in two papers, while some (single collision experiments and activation energy determination) are in preparation:

1) D. Rondeau; L. Drahos; K. Vekey, Internal energy distribution in electrospray ionization: Towards the evaluation of a thermal-like distribution from the multiple-collision model, ***Rapid. Commun. Mass Spectrom.***, 28, 1273-1284, 2014

2) Bazso, FL; Ozohanics, O; Schlosser, G; Ludanyi, K; Vekey, K; Drahos, L, Quantitative Comparison of Tandem Mass Spectra Obtained on Various Instruments, **J. Am. Soc. Mass Spectrom.**, 27, 1357-1365 (2016);

3) Gömöry Á., Bazsó F. L., Vékey K., Drahos L., Novel Single Collision Experiments to Determine Activation Energy, *J. Mass. Spectrom.* in preparation.

2) Innovations: Improving reproducibility and transferability of *HPLC-MS*

One of the major problems in HPLC-MS and tandem MS based methods is reproducibility and transferability (i.e., transferring methods from one lab to another or one instrument to another). These are especially critical for biomolecular studies, where long experiment series and long HPLC runs are typical. In the pharma industry it is especially important to be able to transfer methods from one instrument to another, but this often fails, due to intricacies of the MS/MS. We have used our understanding of the fundamental processes occurring in ESI to

identify temporal variations in signal intensity. This led to an empirical model and to a simple algorithm to compensate for these fluctuations. As an end-result, we have developed a method which significantly improves reproducibility in long HPLC-MS runs. We have also used our understanding of MS/MS processes to develop methods to compare energy and collision gas pressure variations in various instrument types. This study indicated that MS/MS spectra taken on various QQQ and QTOF instruments can be adjusted to be nearly identical, adjusting the collision energy and collision gas pressure. Between other instruments, like QQQ vs. ion traps, it is impossible. These results are essential to determine method transfer from one laboratory to another.

During the determination of mass spectra of model peptides we have found that signal intensities in series of experiments often vary and may cause bias in the results. It was found that the sensitivity of various components change differently; in our case variability is in the order of 20-40%; and it is due to changing conditions in ESI ionization. The change in signal intensities (peak areas) was described by a polynomial function; using a 4th order polynomial proved best. This correction improved reproducibility up to 4 times.

The similarity between two mass spectra, which were measured on different instruments, were compared quantitatively using the similarity index (SI); defined as the dot product of the square root of peak intensities in the respective spectra. This function was found to be useful for comparing energy dependent tandem mass spectra obtained on various instruments. We have compared energy dependent mass spectra on various instruments. The aim was to determine, if, how, and to what degree it is possible to get a tandem mass spectrum, which is closely similar to another one taken on a different instrument. As it is well known, the collision energy is the most important single parameter, which influences the appearance of tandem mass spectra. The similarity index between spectra taken on two instruments, at all combination of collision energies was calculated. Experiments were centered on leucine enkephalin, as it is an often used standard in mass spectrometry, has various fragments in a wide energy range, and its spectra have well-described energy dependence. All other studied compounds yielded analogous results.

We have found that the similarity index (square root dot product) was a good way to compare spectra taken under different conditions. If long term reproducibility is an issue, we suggest scanning the collision energy, and determining the optimum using the similarity index (comparing the new, energy dependent spectra to an old reference spectrum). We also suggest using YGGFL as a quality control standard for energy resolved studies; determining the tandem MS spectrum at the selected collision energy. This spectrum may be used later as a reference spectrum. In future experiments the mass spectrometer should be tuned using YGGFL (by varying the collision energy) to get the best similarity index with the reference spectrum. This will result in better reproducibility than e.g. using the same tuning file on an instrument; because it takes into account e.g. possible deposits in the ion source, misalignment of or impurities on the quadrupole rods etc., which may vary in time. The use of YGGFL is advantageous, as it has fragment ions in a wide energy range, and its fragmentation characteristics are well known.

The results described above were published in two papers:

1) Tóth E., Hevér H., Ozohanics O., Telekes A., Vékey K., Drahos L.: Simple correction improving long-term reproducibility of HPLC-MS, ***J Mass Spectrom*** 50, 10 1130-1135, (2015)

2) Bazso, FL; Ozohanics, O; Schlosser, G; Ludanyi, K; Vekey, K; Drahos, L, Quantitative Comparison of Tandem Mass Spectra Obtained on Various Instruments, ***J. Am. Soc. Mass Spectrom.***, 27, 1357-1365 (2016);

3) Advancements for protein identification and improving sequence coverage in proteomics and for MAb analysis

One of the key issues in proteomics is protein sequence validation, relying on the evaluation of tandem mass spectra. In an ideal case, the complete peptide sequence can be identified based on the MS/MS fragment ions; but this is rarely the case in practice. The problem is, that the peptides do not show cleavages between all, but only some peptide bonds. Fragmentation highly depends on the collision energy, which is the most important experimental parameter, and needs to be adjusted. Using a special, home-developed software we have analyzed the collision energy dependence of identification scores for 1000s of peptides, allowing us to use a more relevant optimization target, than currently used. Using this optimization process, it was possible to improve both peptide/protein identification and sequence coverage by up to 50%. These results are important not only in proteomics, but also for the pharma industry. We are exploring the use of this optimization algorithm for improving sequence coverage of monoclonal antibodies; which is one of the challenges for analysis of biological medicines.

Correlation between survival yield and information content of fragmentation for various peptides has been studied: energy dependent tandem mass spectra of E. coli lysate and HeLa digest have been measured.

We have integrated this with a general methodology for optimizing proteomics methodologies with respect to energetics. We have compared different charge states, and observed that the commercially developed parameter sets need to be significantly changed for optimal results. We have determined optimal collision energies for peptides, and extended these results to glycopeptides.

We have developed novel software capable of extracting information on MS/MS energetics in large-scale proteomics studies (i.e. tens of thousands of compounds in an HPLC-MS/MS run). This allowed us not only to look at the energy dependence of a few standard compounds, but to derive large-scale statistics. This made it possible to optimize energetics in proteomics experiments; and will allow us to extend it not only to glycopeptides, but also to other post-translational modifications or to selected peptide classes.

We have studied energy dependent fragmentation of unprecedentedly large number of (over 1000) tryptic peptides individually, which allowed obtaining statistically relevant results. We have found various novelties. 1) Characterization of the „optimum” mass spectra may be best

based on the probability based peptide identification score (e.g. Mascot score); 2) Many peptides show two maxima on the Mascot score – collision energy plot – this feature has never been observed before. 3) Combining results obtained at various collision energies results in higher sequence coverage.

Results have been published, and a further publication with Richter Pharmaceuticals is in preparation:

1) Á. Révész, T. A. Rokob, D. J. Dit Fouque, L. Turiák, A. Memboeuf, K. Vékey, L. Drahos: Selection of collision energies in proteomics mass spectrometry experiments for best peptide identification: study of Mascot score energy dependence reveals double optimum, ***Journal of Proteome Research,*** 17, 1898-1906, (2018)

4) Applications:

The results and developments described above are not only important for fundamental research, but can be used in practice as well. They are in use in our laboratory practice, enabling us to resolve problems beyond the state of art. In one example, we were utilizing improved reproducibility for identifying prostate cancer biomarkers. In the other example, we were using our theoretical models for internal energy manipulation (and its effects on MS/MS) to advance structural identification of glycopeptides. Both of these examples are direct application of the methods developed in the course of the OTKA project, and demonstrate the short-term applicability of our research:

1) Turiák, Lilla; Ozohanics, Oliver; Tóth, Gábor; Ács, András; Révész, Ágnes; Vékey, Károly; Telekes, András; Drahos, László: High sensitivity proteomics of prostate cancer tissue microarrays to discriminate between healthy and cancerous tissue, ***Journal of Proteomics*** doi: 10.1016/j.jprot.2018.11.009. [Epub ahead of print]
2) Ács. A,; Ozohanics O. ,Vékey K., Drahos L., Turiak, L.: *Distinguishing Core and Antenna Fucosylated Glycopeptides Based on Low-Energy Tandem Mass Spectra*, ***Anal. Chem.*** 90, 12776-12782 (2018)