Evaluation of LC-MS data for the absolute quantitative analysis of marker proteins

Nathanaël Delmotte[1], Bettina Mayr[1], Andreas Leinenbach[1], Knut Reinert[2], Oliver Kohlbacher[3], Christoph Klein[4], and Christian Huber[1]

[1] Instrumentelle Analytik und Bioanalytik, Universität des Saarlandes, Postfach 15 11 50, 66041 Saarbruecken, Germany
n.delmotte@mx.uni-saarland.de
[2] FB Mathematik und Informatik, AG Algorithmische Bioinformatik, Freie Universität Berlin, Takustrasse 9, 14195 Berlin, Germany
[3] Wilhelm-Schickard-Institut für Informatik, Universität Tübingen, 72076 Tübingen, Germany
[4] IRMM / IHCP-ECVAM, European Commission-Joint Research Centre, Ispra, Italy

Abstract. The potential of mass spectrometric peptide identification in complex mixtures by means of peptide mass fingerprinting (PMF) and peptide fragment fingerprinting (PFF) was evaluated and compared utilizing synthetic mixtures of commercially available proteins and electrospray-ion trap- or electrospray time-of-flight mass spectrometers. While identification of peptides by PFF is fully supported by automated spectrum interpretation and database search routines, reliable identification by PMF still requires substantial efforts of manual calibration and careful data evaluation in order to avoid false positives. Quantitation of the identified peptides, however, is preferentially performed utilizing full-scan mass spectral data typical of PMF. For the absolute quantitation of serum proteins, we have developed an analytical scheme based on first-dimension separation of the intact proteins by anion-exchange high-performance liquid chromatography (HPLC), followed by proteolytic digestion and second-dimension separation of the tryptic peptides by reversed-phase HPLC in combination with electrospray ionization mass spectrometry (ESI-MS).

1 Mass spectrometric peptide identification

1.1 Peptide Mass Fingerprinting

In the Peptide Mass Fingerprinting (PMF) approach, the complex protein mixture one wants to analyze is generally enzymatically digested with a specific enzyme such as trypsin. After protein digestion, the resulting complex peptide mixture is usually separated by two-dimensional high-performance liquid chromatography (2D-HPLC). The separated peptides are then ionized with soft ionization methods such as matrix-assisted laser desorption/ionization (MALDI) or electrospray ionization (ESI) and

Dagstuhl Seminar Proceedings 05471
Computational Proteomics
http://drops.dagstuhl.de/opus/volltexte/2006/539

analyzed by mass spectrometry (MS). In this approach, a peptide is identified by measuring its mass over charge ratio (m/z) and by comparing it to a list of m/z calculated from a list of peptides generated from in silico digestion of the proteins in a database. Protein identification occurs when one or more peptides matching a part of the protein sequence are identified. In this approach, the identification of a peptide is only based on a single m/z measurement. To get high confidence results, it is of importance to get very accurate m/z values (3-5 ppm mass deviation) and to use high-resolution mass analyzers such as time-of-flight (TOF) or Fourier Transform Ion Cyclotron Resonance (FT-ICR) being able to differentiate lysine (m/z 128.095) from glutamine (m/z 128.059). Anyways, to get the best mass accuracy it is recommended to perform internal mass recalibration. This procedure is, however, time consuming for the operator and seldomly automated.

1.2 Peptide Fragment Fingerprinting

In the Peptide Fragment Fingerprinting (PFF) approach, the complex protein mixture one to be analyzed is digested and separated as in the previously described PMF approach. The difference appears in the mass spectrometric detection. After MALDI or ESI ionization, peptides are fragmented in the gas phase. The fragments obtained in the gas phase are finally separated, detected, and compared to sets of theoretical fragments. Theoretical fragments are obtained from in silico fragmentation of peptides, according to fragmentation rules and the type of instrument used for the fragmentation. In this approach, a peptide is identified with by the mass of the intact precursor peptide ion, but also by the m/z of each fragment observed in the spectrum. For this reason it is possible to get unambiguous peptide and protein identification without high accuracy mass instruments.

1.3 Peptide Mass Fingerprinting *vs.* Peptide Fragment Fingerprinting

23 commercially available proteins were used to prepare synthetic mixtures of proteins. These mixtures were proteolytically digested with trypsin after denaturation with urea and cystein carboxymethylation. The resulting peptide mixtures were individually separated on a 200 µm i.d. poly-(styrene-divinylbenzene) (PS-DVB) monolithic column coupled to an ESI ion-trap mass spectrometer performing data-dependent auto-MS/MS analysis of the most intensive ions. The same mixtures were also separated on the PS-DVB column coupled to an ESI time-of-flight mass spectrometer. Mass spectra of both HPLC run sets were computed for identification with MASCOT (http://www.matrixscience.com). In the PMF approach (MS with time-of-flight analyzer) 438 peptides were identified as only 226 were found significant with the PFF approach (MS/MS with ion-trap analyzer). The results sum up the advantages and the disadvantages of the two methods. With PMF, more peptides are identified but despite mass spectra recalibration and high accuracy mass measurement, false positive hits might be reported. This approach is also practically difficult to analyze complex mixtures of unknown compounds. The number of undistinguishable peptides is substantially increasing with the size of the database. On

the contrary, PFF delivers high confidence identification data but at the cost of a smaller number of identifications.

2 Absolute quantitation of myoglobin in serum

The serum complexity makes the absolute quantitative analysis of medium to low-abundant proteins very challenging. Tens of thousands proteins are present in human serum and dispersed over an extremely wide dynamic range. The reliable identification and quantitation of proteins, which are potential biomarkers of disease, in serum or plasma as matrix still represents one of the most difficult analytical challenges (Chelius, Zhang et al., 2003;Bondarenko, Chelius et al., 2002;Chelius & Bondarenko, 2002). The difficulties arise from the presence of a few, but highly abundant proteins in serum and from the non-availability of isotope-labeled proteins, which serve to calibrate the method and to account for losses during sample preparation. An analytical scheme for the absolute quantitative analysis of myoglobin was developed.

2.1 Myoglobin biological activity

Myoglobin is a 17 kDa protein which permits the oxygen storage of primary aerobic working muscles such as heart. The protein is made of a single polypeptide chain of 153 amino acids. The property of myoglobin to bind oxygen molecules is due to the presence of a heme group. This non-covalent group made of a porphyrin ring is able to complex iron and is responsible for the red color of myoglobin. The dimensions of the protein measured in solution are: 4.5 x 3.5 x 2.5 nm. The tertiary structure of the protein is mostly α-helical; eight α-helical portions are separated by unarranged structures. Two histidine residues inside the natural protein play a decisive role for the binding of oxygen on the heme group. When a muscle (e.g. heart muscle) is injured, myoglobin is released into serum. Such a case occurs during myocardial infarction. After 2 hours, the plasma myoglobin concentration rises, and after 6-9 hours a concentration peak is reached (200 – 1000 ng/mL plasma) (Kilpatrick, Wosornu et al., 1993;Bhayana, Cohoe et al., 1994;Beuerle, Azzazy et al., 2000). After 12-24 hours the myoglobin concentration recovers a value of 3 -65 ng /mL.

2.2 Myoglobin absolute quantitation

In our approach, a human serum sample was spiked with horse myoglobin as internal standard. The sample was also spiked with different concentrations of human myoglobin to perform a quantitation based on the method of standard additions. The sample was injected over a strong-anion-exchange (SAX) chromatography column and a fraction corresponding to myoglobin elution was collected. The fraction containing myoglobin as intact protein was digested with trypsin. The resulting peptides were injected over a 200 µm PS-DVB monolithic column hyphenated to an

ESI ion-trap mass spectrometer monitoring in full-scan mode. Extracted ion chromatograms corresponding to tryptic peptides of human and horse myoglobin were traced and areas of the eluting peaks were computed. It was possible to determine myoglobin concentrations in human serum down to 100-500 ng/mL. Calibration graphs were linear over at least one order of magnitude and the relative standard deviation of the method ranged from 7-15%. Compared to manual result computation, bioinformatic algorithms (OpenMS) led to a significant reduction both of absolute deviation and confidence interval (Gröpl. C., Lange et al., 2005).

3 Conclusion

According to our results for peptide identification and quantitation, algorithmic solutions for PMF that incorporate both recalibration and automated feature finding on the basis of peak elution profiles and isotopic patterns are highly desirable in order to speed up the process of data evaluation and calculation of quantitative results. Absolute quantification of protein in biological matrices is feasible upon calibration using additions of authentic analyte. Computer-based algorithms, based on modelling of peak volumes both in retention-time and m/z dimension result in a significant inmprovement in quantitative accuracy.

References

Part of this work has been published in J. Proteome Res. 5, 2006, 414-421.

Beuerle, J. R., Azzazy, H. M., Styba, G., Duh, S. H.,Christenson, R. H.," Characteristics of myoglobin, carbonic anhydrase III and the myoglobin/carbonic anhydrase III ratio in trauma, exercise, and myocardial infarction patients", Clin. Chim. Acta, 294, 115-28 (2000).

Bhayana, V., Cohoe, S., Pellar, T. G., Jablonsky, G.,Henderson, A. R.," Combination testing for myocardial infarction using myoglobin, creatine kinase-2, and troponin T", Clin. Biochem., 27, 395-406 (1994).

Bondarenko, P. V., Chelius, D.,Shaler, T. A.," Identification and relative quantitation of protein mixtures by enzymatic digestion followed by capillary reversed-phase liquid chromatography-tandem mass spectrometry", Anal. Chem., 74, 4741-9 (2002).

Chelius, D.Bondarenko, P. V.," Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry", J. Proteome Res., 1, 317-23 (2002).

Chelius, D., Zhang, T., Wang, G.,Shen, R.-F.," Global Protein Identification and Quantification Technology Using Two-Dimensional Liquid Chromatography

Nanospray Mass Spectrometry", Anal. Chem., 75, 6658-6665 (2003).

Gröpl. C., Lange, E., Reinert, K., Kohlbacher, O., Sturm, M., Huber, C. G., Mayr, B. M., and Klein, C. (2005). Algorithms for the Automated Absolute Quantification of Diagnostic Markers in Complex Proteomics Samples. *In* "Lecture Notes in Bioinformatics, Proceedings of the First Symposium on Computational Life Sciences (CLS 2005)" (M. Berthold, R. Glen, K. Diederichs , and O. F. I. Kohlbacher, Eds.), Vol. 3695, pp. 151-161. Springer, Heidelberg.

Kilpatrick, W. S., Wosornu, D., McGuinness, J. B.,Glen, A. C.," Early diagnosis of acute myocardial infarction: CK-MB and myoglobin compared", Ann. Clin. Biochem., 30 (Pt 5), 435-8 (1993).