

# Chapter 2

## In-Depth Protein Characterization by Mass Spectrometry

Daniel Chamrad, Gerhard Körting, and Martin Blüggel

### Abstract

Within this chapter, various techniques and instructions for characterizing primary structure of proteins are presented, whereas the focus lies on obtaining as much complete sequence information of single proteins as possible. Especially, in the area of protein production, mass spectrometry-based detailed protein characterization plays an increasing important role for quality control. In comparison to typical proteomics applications, wherein it is mostly sufficient to identify proteins by few peptides, several complementary techniques have to be applied to maximize primary structure information and analysis steps have to be specifically adopted. Starting from sample preparation down to mass spectrometry analysis and finally to data analysis, some of the techniques typically applied are outlined here in a summarizing and introductory manner.

---

### 1. Introduction

The field of Proteomics has been very successful in identifying the quantification of large sets of proteins (protein mixtures), for example, from whole organelles or cell lysates. Nowadays, hundreds of proteins within a complex sample can be easily identified by mass spectrometry, whereas only few peptides per protein are usually detected (1). This allows elucidating the name of the protein via searching protein sequence databases. In addition to analyzing complex protein mixtures, at least equally challenging is the art of in-depth characterization of individual proteins, or in other words, gaining as much primary structure information (including posttranslational modifications) as possible from a protein of interest.

In-depth protein characterization is of great importance, as it increases the chance to detect posttranslational modification (PTM), which modulates the activity of most eukaryote proteins. Also validating and distinguishing protein isoforms within a sample

demands detailed elucidation of the protein sequence. Especially, therapeutic protein products require thorough characterization, for example, during protein engineering, protein production, and for first in men studies throughout routine testing.

Mass spectrometry (MS) is an excellent tool for this purpose as it allows deducing the primary structure of proteins, including PTM by measuring mass per charge ratios ( $m/z$ ) of peptide ions and corresponding peptide fragment ions in a high-throughput manner (2). Especially, the technology advances in recent years, including the increase in accuracy (today at ppm for peptides and peptide fragments), sensitivity (femtomol) and acquisition speed (more than 10,000 spectra/h) has turned MS into the most valuable analysis tool for detailed characterization of complex molecules like proteins.

While high-throughput protein identification from peptide fragmentation (MS/MS) has become a standard in modern MS-based protein analytics, complete primary structure elucidation, including PTM is still a challenge due to various reasons:

- (a) Masses measured by MS are generally not unique, i.e., different amino acid sequences, including PTM may have identical or similar mass values, making them hard to distinguish.
- (b) Protein and peptide modifications can be induced by sample preparation and these must therefore be carefully distinguished from original in vivo PTM.
- (c) Some protein sequence segments may be hard to monitor by MS, e.g., some peptides are hard to ionize or show poor fragmentation.
- (d) Protein modifications may not be homogenous, and due to numerous gene products caused by alternative splicing and combinations of modifications the protein mixture can be very complex.
- (e) Sample preparation methods have to be individually developed as low protein concentration and interfering small molecules like salt, detergent, and stabilizers in formulation are limiting or even preventing mass spectrometric analysis.

In this chapter, we explore various current methods for complementary primary structure elucidation via mass spectrometry. We also focus on sample preparation as this is an essential prerequisite to enable and improve primary structure discovery.

---

## 2. Methods

### 2.1. Sample Preparation

Sample preparation methods for in-depth protein characterization by MS have to be developed to fulfill two aspects. On the one hand, sample preparation has to be performed to enable mass

spectrometric analysis. On the other hand, it has to be designed in a way to minimize the risk of primary structure change due to the sample preparation.

#### *2.1.1. Enabling Mass Spectrometric Analysis*

Adjuvants and contaminants, such as salt, detergent, or stabilizers, have the potential to prevent or reduce the results of mass spectrometric analysis. In case of liquid chromatography coupled to electrospray ionization mass spectrometry, salts in millimolar concentrations and even low detergent concentrations can be removed online within the HPLC setup (e.g., guard column or dedicated trapping column). For higher concentrations and for MALDI-MS applications, spinning columns (e.g., 3.5-kDa cutoff), dialysis (also available as microdialysis) or precipitation are the methods which are mostly applied. Additionally, separation techniques with high resolving power, such as reverse phase-HPLC or the combination of SDS-PAGE (1D or 2D) with protein digestion, are also well suited to move to an MS compatible buffer, with salts like ammonia carbonate, solvents like water, acetonitril, methanol, and acids like formic or trifluoroacetic acid.

#### *2.1.2. Minimizing Risk of Primary Structure Change*

Oxidation of, for example, Methionine, deamidation of Asparagine, or truncation may occur under conditions of sample preparation. Additionally existing modifications (e.g., phosphorylation) may be removed (e.g., by contact to iron in not inert HPLC systems).

Therefore, the sample preparation steps have to be limited to the minimum steps needed. Harsh conditions have to be avoided (e.g., 4 h, 37°C protein digestion method instead of 24 h, 37°C to avoid deamidation).

There are no universal protocols as the methods have to be adopted and altered to meet several aspects:

- (a) Aim of analysis and intended MS technique.
- (b) Starting protein concentration and nature of buffer content.
- (c) Final protein amount and concentration needed.

Additionally, protein specific aspects like hydrophobicity, tertiary structure, or modification often result in a need for protein-specific method development.

Some general rules provide a guideline to method development:

- (a) Avoid any unnecessary step (e.g., multiple concentration, buffer changes).
- (b) Work at high protein concentrations so that only a minor fraction of the analyzed proteins is lost due to unspecific adsorption and reduce unfavorable adjuvant to protein ratios.
- (c) Minimize harsh stress conditions like high temperature or RT for longer time, freeze/thaw cycle, extreme pH, lyophilization steps; oxidative stress.
- (d) Do not introduce any adjuvants where not needed.

## **2.2. Primary Structure Elucidation by Mass Spectrometry**

The primary structure of a biological molecule is the exact specification of its atomic composition and the chemical bonds connecting those atoms. For a high molecular weight protein like an antibody with approximately 20,000 atoms, the information of its primary structure is very complex. Fortunately, a good portion of this information can be reduced to the amino acid sequence.

However, for proteins the primary structure is not only covering the exact amino acid sequence, but also cross-links like disulfide bridges and modifications. Microheterogeneity will add another level of complexity into sample characterization as it is present in many highly purified recombinant proteins as well.

During the last 20 years, a huge number of mass spectrometric methods were developed to analyze the primary structure in detail. A full molecular weight determination by MS can provide a good insight for the verification of primary sequence and detection of modification. MALDI-TOF-MS is robust in sample preparation and salt concentration and can give you accuracy with as low as a few Daltons for midsized proteins. With this accuracy, information on N-/C-terminal truncation or modifications like glycosylation or phosphorylation can be obtained. However, for modifications like deamidation, disulfid linkage, or even oxidation a higher accuracy may be needed. The ability of Electro Spray Ionization to measure the molecular weight of multiple highly charged ions in parallel results in a much better accuracy. For ESI-FT-MS measurement, these molecular weight determination can be in a sub-Dalton range.

For a more detailed primary characterization, the protein has to be cleaved into subunits or peptides which are then measured by mass spectrometry.

The “MALDI In Source Decay” method fragments a full intact protein within the mass spectrometer and enables here a direct sequencing of the N- and C-terminal sequence area.

A sample preparation with a highly specific enzymatic digestion (e.g., Trypsin, Glu-C, Asp-N, etc.) will result into peptides which can be measured in a mixture (e.g., by MALDI-MS) or separated and analyzed by online LC-ESI-MS. With today's instruments, these peptides can be measured with high sensitivity (fmol) and with highest mass accuracy (low to even sub-ppm level). In the same experiment, these peptides can be fragmented within the mass spectrometer and the resulting peptide fragment pattern will be recorded also with highest mass accuracy and sensitivity.

With this ability and lab automation, it is possible to resolve also very complex primary structures and microheterogeneity of low abundant sequence variants.

However, data analysis becomes increasingly important to unravel the full potential and latest improvements of mass spectrometry.

### 2.3. Signal Extraction

Signal extraction and calibration are the most common first steps in the MS data interpretation process. Most software tools for MS-based protein analysis accept so-called peak lists, which are a collection of signals of a mass spectrum. Peak extraction is a complex task due to signal resolution, noise, signal overlapping, and the need for deisotoping.

In case of ESI-MS, peptides and proteins are typically detected in various charge states ( $z$ ), e.g., with  $z=1-4$  for peptides,  $z=5-100$  for proteins and complexes). In order to determine the exact molecular weight of a peptide or protein, the spectrum has to be deconvoluted (calculate  $M$  or  $MH^+$  from  $m/z$  values). The information of the charge state can be derived directly from the given isotopic  $m/z$  signal pattern using software tools (3, 4). However, one should be aware that the applied software may fail to assign the correct charge state. In case of proteins, molecular mass is derived from  $m/z$  mass peaks of multiple charge states of the same protein. In case of time of flight (TOF) measurements calibration of the spectra is essential to obtain sufficient mass accuracy. Calibration can be done internally (e.g., using theoretical  $m/z$  values of known peptides within the dataset, or by injecting substances in the MS instrument with each spectrum (“lock mass”)), or externally (using the calibration constants of an earlier run, which contains spectra of a known substance).

After calibration, modern MS instruments can achieve a mass accuracy of few ppm.

### 2.4. Peptide Fragmentation Fingerprinting

Fragmentation mass spectra of peptides can be correlated to protein sequences in a database in an automatic manner (5, 6). This can be done by dedicated protein sequence database search software (see Table 1). It is advantageous that this method does not require any a-priori knowledge about the analyzed proteins, and therefore it is often used as an initial step to identify all major protein components in a sample.

**Table 1**  
**Overview on commonly used peptide fragmentation fingerprinting software**

Mascot	<a href="http://www.matrixscience.com/">http://www.matrixscience.com/</a>
MS-Seq	<a href="http://prospector.ucsf.edu/">http://prospector.ucsf.edu/</a>
Phenyx	<a href="http://www.genebio.com/products/phenyx/">http://www.genebio.com/products/phenyx/</a>
Popitam	<a href="http://www.expasy.org/tools/popitam/">http://www.expasy.org/tools/popitam/</a>
SEQUEST	<a href="http://fields.scripps.edu/sequest/">http://fields.scripps.edu/sequest/</a>
SpectrumMill	<a href="http://www.home.agilent.com/">http://www.home.agilent.com/</a>
X! Tandem	<a href="http://prowl.rockefeller.edu/prowl/">http://prowl.rockefeller.edu/prowl/</a>

Initially, the user has to define various input parameters carefully, such as the specificity of the applied proteolysis enzyme, maximum allowed mass errors for peptide parent ion and fragment masses and the protein sequence database to be searched. Then, the software generates theoretical spectra by theoretical fragmentation of peptides obtained from *in silico* digestion of the searched database proteins. The obtained theoretical spectra are compared to the measured spectra and the result is a list of matching peptides and proteins. Commonly, the reported proteins and peptides are sorted by a specific search score that relates to the significance of the found database match.

Protein and peptide modifications can be elucidated with this approach to some extent as typical database search engines that allow searching up to three different variable modifications (each amino acid in question is tested whether it is modified or not) and also fixed modifications (every amino acid is treated to be modified). Also regarding enzyme nonspecificity, missed cleavage sites and even peak picking errors (e.g., failure to detect the correct monoisotopic peptide signal from overlapping isotopic distributions) can be searched but generally applying these search strategies may lead to a drop in sensitivity. Therefore, it is advisable regarding only experimentally induced modifications (e.g., methionine-oxidation) and a maximum of one or two missed cleavages and no unspecific cleavage. In case of in-depth protein characterization, primary structure elucidation beyond this scope should be addressed by dedicated second round search engines (see below).

Mass accuracy is crucial to obtain unambiguous results. The maximum allowed mass error parameters within the search should be set to at least two standard deviations (assuming a normal distribution, about 95% of the measurement errors fall in two times standard deviation). The standard deviation for mass measurements can be determined within routine MS-instrument calibration.

Peptide masses determined by MS are generally not unique and each measured mass can randomly match a peptide from a sequence database. Therefore, a certain risk to obtain false positive results remains. Assessing the correctness of a possible identification is a challenging task. In fact, the probability that the match in question is correct cannot be calculated; however, most reported search scores relate to the probability that the observed peptide match is a pure random event (7, 8). In case of in-depth protein characterization, evaluation of sequence database search results is frequently not done automatically, but remains the task of an expert who manually inspects spectra matching to the protein of interest.

Usually, the primary structure detectable by a single database search is limited and must be extended by further experiments such as using a different cleavage enzyme, or using dedicated second round search engines.

## **2.5. Second Round Searches**

Standard database searches which can be seen as “first round” searches are limited in the elucidation of posttranslational modifications, unspecific, and missed cleavages products, sequence errors, amino acid substitutions, and unsuspected mass shifts. For example, taking more than 200 described posttranslational modifications for all protein sequences of an organism into account would lead to an amount of peptides to be tested that impedes a brute force approach. Apart from the huge time exposure, simply the huge number of possible combinations leads to randomly matching sequences. To overcome this problem, second round searches have been developed, which work similar to peptide fragmentation fingerprinting described above but instead of searching a complete protein sequence database, only few selected protein sequences are regarded (9).

Typically, protein identification is done in the first step using standard search algorithms. Second round searches are then used in the second step to elucidate previously unexplained spectra. In case of the software tools Mascot and Phenyx, the second round search feature is directly integrated, and can be triggered after the first round search. There is also a dedicated second round search tool named Modiro™ (<http://www.modiro.com>) available. In case of Modiro™, the user can enter own protein sequences, which is of, for example, special interest in case of therapeutic protein products from biotechnology. During the second round, search batches of unidentified spectra (e.g., whole LC-MS/MS runs) are screened in a sequential manner for various different posttranslational modifications, unknown mass shifts, unspecific cleavages, and sequence errors in one single step. A typical search result obtained by using Modiro™ is shown in Fig. 1.

## **2.6. De Novo Sequencing**

As genome sequencing capabilities have increased dramatically during the last decades, many organisms are sequenced today and sequences are available to the public community. However, genome sequence information is still lacking for many organisms at the same time while some of them are of interest in industrial or biochemical research.

As MS/MS spectra of peptides are generated by fragmentation within the backbone of the peptide, the mass difference between two fragment ions directly provide information on the amino acid at a given peptide position. As a result, de novo sequencing is feasible for a peptide and partly also for proteins. However, each fragmentation is highly sequence dependent, and the intensity of the different ions differs a lot for each fragment ion. Therefore, some positions may not be resolved. Additionally, a mass difference may be explained by more than one amino acid combination leading to inconclusive sequences. As additional fragmentation (e.g., from internal fragments, side cleavage, doubly charged ions) may occur and overlay the ion series, the manual interpretation is quite laborious.



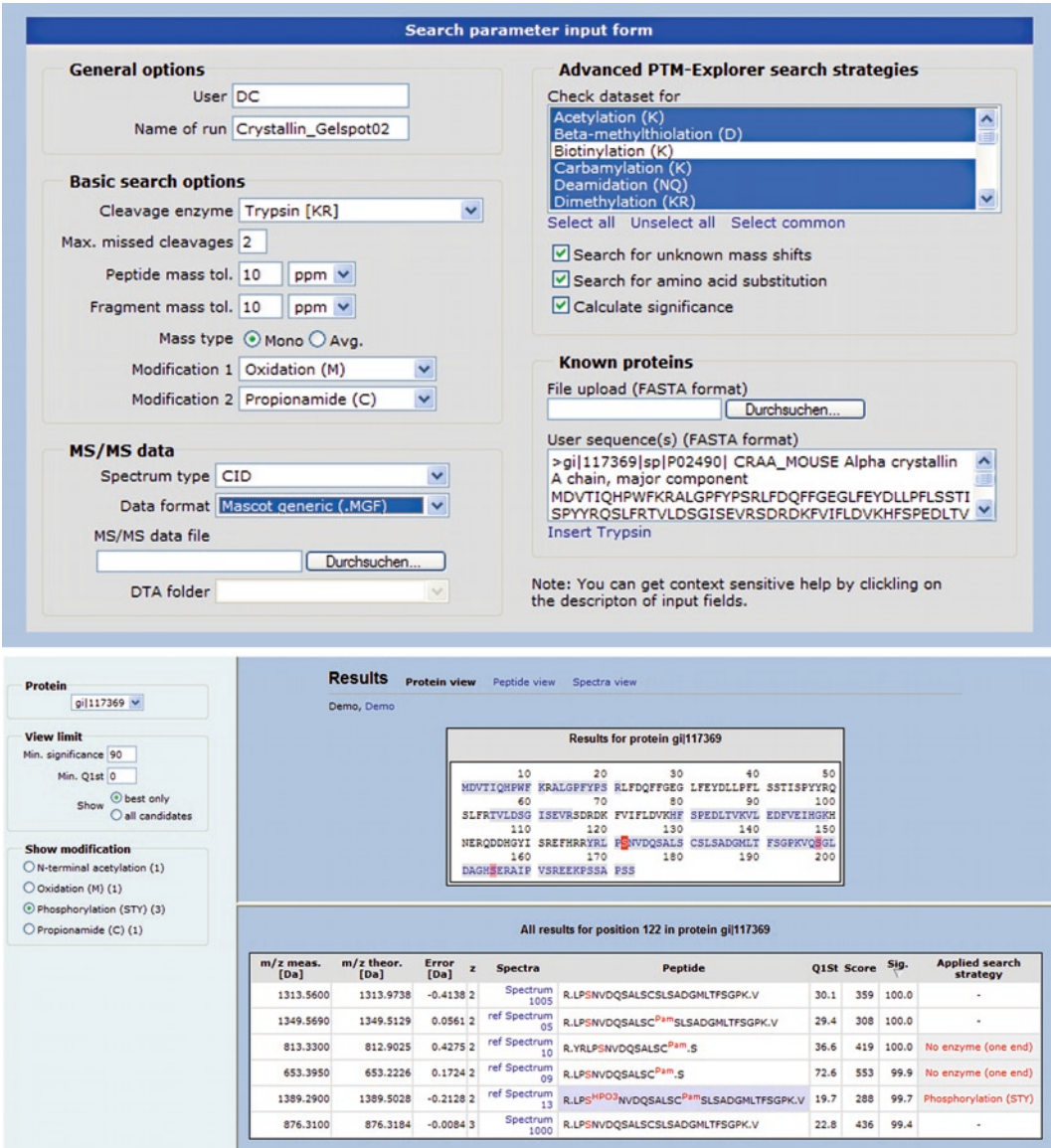


Fig. 1. Screenshots of the Modiro™ Software showing search parameter input and the obtained result page, including detected protein modifications in MS/MS datasets.

Several software solutions were developed to perform an automated de novo sequencing (e.g., PEAKS (10), PepNovo (11), Lutefisk (12)). They provide the best guess of the sequence, at least a sequence tag. The accuracy of this prediction highly depends on the quality of the fragmentation spectra. Resulting peptide candidates can be easily searched for homology against sequence databases. MS-BLAST (13) is a dedicated alignment tool for this purpose.



Additionally, MS instrument providers deliver software packages where either a full de novo algorithm is incorporated or sequence tag generation is supported by interactive annotation of a resulting MS/MS spectrum (e.g., BioTools, Bruker Daltonik GmbH).

Although knowing that a given protein is derived from a non-sequenced organism, its MS/MS data should be analyzed in the first round by a search engine (see Subheading 2.4) with no or broad taxonomy restriction. For some peptides, the homology might be sufficient to pick up the homolog protein from another already sequenced organism, which reduces the workload for de novo sequencing.

For isolated unknown proteins from an unsequenced organism internal protein sequence parts are needed, in order to construct nucleotidic degenerative primers for PCR and subsequent DNA sequencing. For this purpose, high quality sequence information ideally from the C-terminal region and long (minimum 7, best 15 amino acids) stretches are best suited.

## 2.7. Combination of Results (see Fig. 2)

In-depth characterization of protein requires the identification of the complete protein sequence. Usually, within a single MS analysis, some sequence areas are not identified or confirmed, as some peptides are outside the mass range detectable with a specific MS instrument, or have poor fragmentation. Therefore, it is advisable to make several MS runs, using different enzymes (or enzyme combinations) for proteolysis, or to apply other sample preparation techniques. Ideally, missing sequence areas will be different for the different runs and applied techniques, yielding more complete sequence coverage after the combination of the found

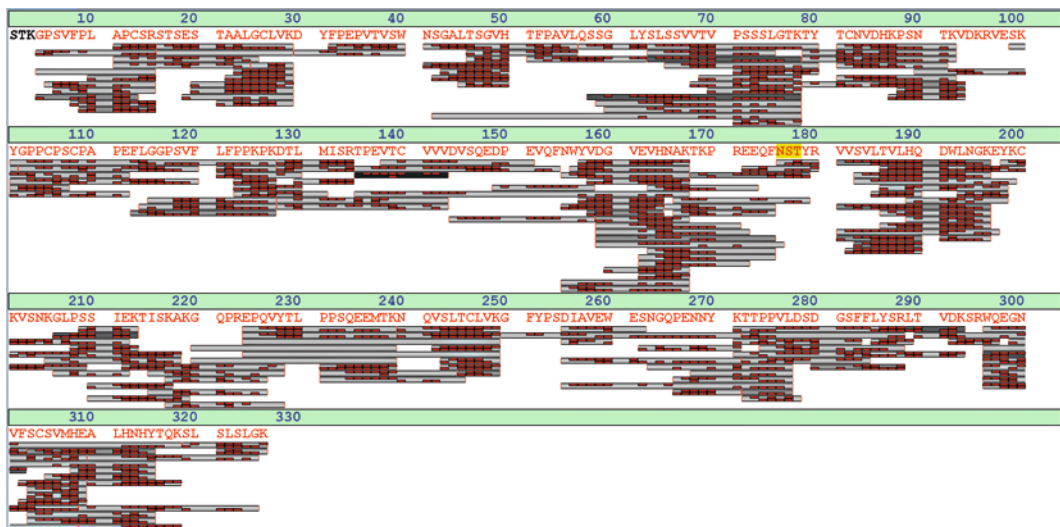


Fig. 2. Combining search results of MS/MS runs with several cleavage enzymes to get nearly complete sequence coverage.

peptides. Equally, analyzing the sample with differing MS instrumentation (e.g., MALDI-MS and LC-ESI-MS/MS) will give a complementary dataset.

Dedicated software is required to combine the outcome of the database searches, as a combined search with, e.g., different cleavage rules or mass spectrometric methods is not possible using currently available sequence database search software. In ProteinScape (Bruker Daltonik GmbH and Protagen AG), which is a Proteomics Bioinformatics Platform (14, 15), an algorithm for this task is integrated. Within ProteinScape, a new protein list is built, combining all peptides from all searches. Additionally, only the best matching sequence for each spectrum is annotated.

## 2.8. Differential EIC

For complete protein characterization of therapeutic proteins, it is necessary to show that the amino acid sequence, including modifications such as glycosylation meets the expected patterns. Second round searches with tools like Modiro<sup>TM</sup> can help to analyze existing modifications.

In case of LC-ESI data, the level of a specific modification can be validated by the visualization of Extracted Ion Chromatograms (EIC) of the modified and unmodified peptide. An EIC shows the mass spectrometric signal intensity of a specific  $m/z$  value over the retention time. With an overlay of two EICs, showing the  $m/z$  of the modified and the unmodified peptide, the level of modification can be detected (Figs. 3 and 4). If both signals are visible, there should be a retention time shift between them.

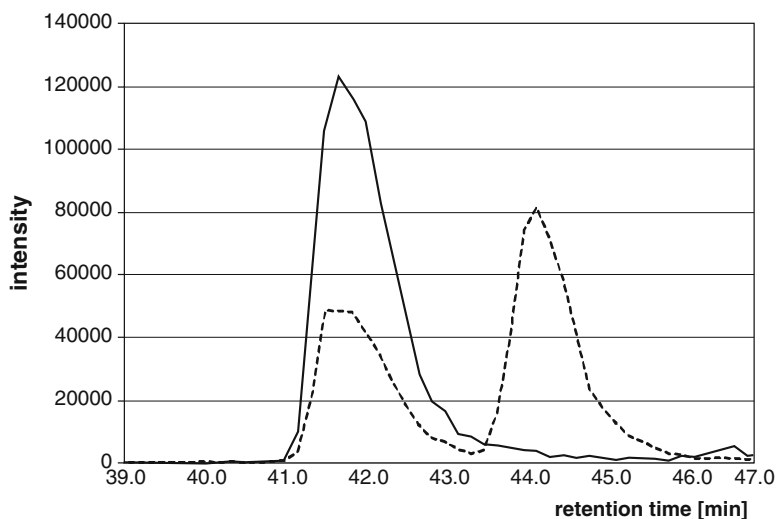


Fig. 3. Overlay of the extracted ion chromatograms of the unmodified and deamidated peptide W.LNGKEY.K. The peak at 42.0 min is the unmodified peptide ( $m/z=723.3672$ ), the peak at 44.5 min the deamidated peptide ( $m/z=724.3512$ ). By comparing the peak intensities or areas a medium deamidation can be estimated. The lower signal at 42.0 min is the second isotope of the unmodified peptide which has nearly the same  $m/z$  as the deamidated peptide.

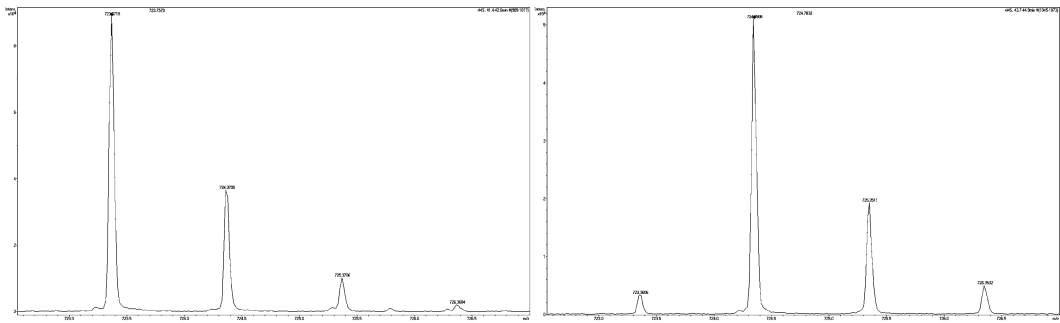


Fig. 4. MS spectra of the deamidation of Fig. 3. The first spectrum is the unmodified peptide at 42.0 min, the second spectrum the deamidated peptide at 44.5 min.

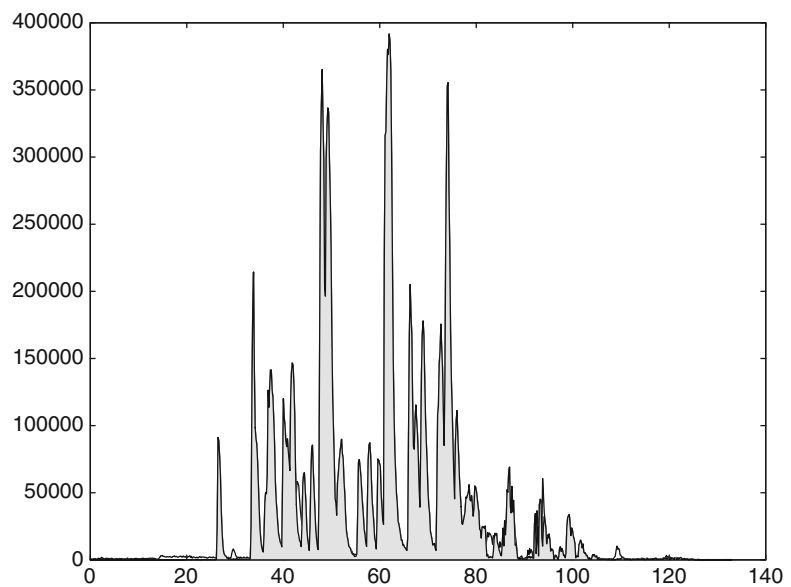


Fig. 5. Base peak chromatogram with identified peaks colored. Most of the MS run is explained. The remaining peak at 42 min was assigned to a peptide containing glycan, but the MS/MS fragmentation was not sufficient for identification.

Another way of assuring that there are no major signals left unexplained can be done by coloring identified peptides in a base peak chromatogram (Fig. 5). Ideally, there should be no peaks left unexplained. If major signals are still unexplained, the corresponding MS and MS/MS spectra must be analyzed further.

### 3. Conclusions

In-depth protein characterization by MS is significantly different from the task to identify proteins from simple or complex mixtures. The whole analysis process from sample preparation to MS acquisition

and data interpretation has to be specifically adopted to the analyzed protein samples in order to increase the amount of elucidatable primary structure information. Therefore, in-depth protein characterization is not standardized and remains an expert task. The major keys to successful primary structure characterization are firstly, sample preparation for the isolation and enrichment of the proteins to be analyzed, and secondly, the combination of several analysis methods to maximize the protein sequence coverage.

Significantly, more material is needed compared to protein identification approaches which require mapping of only a few peptides of each identified protein. The focus lies more on the enrichment or isolation of structural variants, including product impurities which are in low concentration. Chromatographic, electrophoretic separations or immunoaffinity purification are usable methods to isolate suitable amounts of the protein to be analyzed. Often milligrams of proteins are isolated to enable in-depth protein characterization.

Applying different complementary analysis methods is required. An example is increasing sequence coverage by using different proteolysis enzymes or combinations to make more protein sequence segments accessible to the MS measurement. Also the combination of various software tools for analyzing mass spectrometric data maximizes the primary structure yield contained in the acquired data.

As much as possible MS data has to be collected and must be evaluated in a combinatorial approach. However, MS data interpretation can only be partly automated by software. Laborious manual evaluation of mass spectra and primary structure assignment is still required.

---

## 4. Notes

1. Due to computational reasons, MS spectrum identification via software (Peptide Fragmentation Fingerprinting, De Novo Sequencing) works on peak lists rather than the originally acquired raw spectra. The preceding automatic peak picking procedures are not flawless and not lossless. Deconvolution and deisotoping is not always correct. Additionally, signals with low signal to noise ratio may be missed. For that purpose, it can be very helpful to validate a critical peptide match in question manually, using raw spectra. MS instrument providers usually deliver suitable software for manual raw spectrum annotation.
2. Especially, in the area of quality control for protein production, it is very important to elucidate sample preparation and MS-induced artifacts which are not related to the production process itself. Examples are  $\text{Na}^+$  adducts, nonspecific proteolysis,

skimmer nozzle fragmentation, keratin contaminations, pyroglutamate formation from N-Term of internal peptides, etc. As long as these spectra remain unexplained, one cannot be sure about the purity of the product. Here, second round search engines are very helpful as they allow screening a wealth of possible modification in parallel, including also artificially induced ones.

3. Characterizing a protein via peptide fragmentation fingerprinting relies on the correctness of the protein sequence which is matched to the spectra. In case of, for example, sequencing errors, elucidation of corresponding fragmentation spectra fails. Sequence errors from single amino acid exchanges can be elucidated by second round searches. Other sequence errors must be elucidated via de novo sequencing.
4. In case of analyzing a specific peptide via an EIC, there may be unrelated signals and other peptides visible in the chromatogram with nearly the same  $m/z$ . Therefore, corresponding MS and MS/MS spectra have to be checked, too. The MS spectrum must show that the signal is a monoisotopic peak, and has the correct charge state, the MS/MS spectra must match to the peptide sequence.
5. To cover the whole amino acid sequence of a protein, LC-MS/MS runs from digests with several enzymes and enzyme combinations are necessary. To minimize the laboratory work, use software tools and theoretical digests to predict which enzymes or enzyme combinations are optimal to get peptides within the  $m/z$  acquisition range of a mass spectrometer.

## References

1. Yates JR 3rd, McCormack AL, Schieltz D, Carmack E, Link A (1997) Direct analysis of protein mixtures by tandem mass spectrometry. *J Protein Chem* 16:495–497
2. Yates JR 3rd (1998) Mass spectrometry and the age of the proteome. *J Mass Spectrom* 33:1–19
3. Coombes KR, Fritsche HA Jr, Clarke C, Chen JN, Baggerly KA, Morris JS, Xiao LC, Hung MC, Kuerer HM (2003) Quality control and peak finding for proteomics data collected from nipple aspirate fluid by surface-enhanced laser desorption and ionization. *Clin Chem* 49:1615–1623
4. Gras R, Müller M, Gasteiger E, Gay S, Binz PA, Bienvenut W, Hoogland C, Sanchez JC, Bairoch A, Hochstrasser DF, Appel RD (1999) Improving protein identification from peptide mass fingerprinting through a parameterized multi-level scoring algorithm and an optimized peak detection. *Electrophoresis* 20:3535–3550
5. Mann M, Wilm M (1994) Error-tolerant identification of peptides in sequence databases by peptide sequence tags. *Anal Chem* 66:4390–4399
6. Eng JK, McCormack AL, Yates JR III (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* 5:976–989
7. Elias JE, Gygi SP (2007) Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods* 4:207–214
8. Keller A, Nesvizhskii AI, Kolker E, Aebersold R (2002) Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal Chem* 74:5383–5392

9. Chamrad DC, Körting G, Schäfer H, Stephan C, Thiele H, Apweiler R, Meyer HE, Marcus K, Blüggel M (2006) Gaining knowledge from previously unexplained spectra-application of the PTM-Explorer software to detect PTM in HUPO BPP MS/MS data. *Proteomics* 6:5048–5058
10. Ma B, Zhang K, Hendrie C, Liang C, Li M, Doherty-Kirby A, Lajoie G (2003) PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid Commun Mass Spectrom* 17:2337–2342
11. Frank A, Pevzner P (2005) PepNovo: de novo peptide sequencing via probabilistic network modeling. *Anal Chem* 77:964–973
12. Taylor JA, Johnson RS (1997) Sequence database searches via de novo peptide sequencing by tandem mass spectrometry. *Rapid Commun Mass Spectrom* 11:1067–1075
13. Shevchenko A, Sunyaev S, Loboda A, Shevchenko A, Bork P, Ens W, Standing KG (2001) Charting the proteomes of organisms with unsequenced genomes by MALDI-quadrupole time-of-flight mass spectrometry and BLAST homology searching. *Anal Chem* 73:1917–1926
14. Blueggel M, Chamrad D, Meyer HE (2004) Bioinformatics in proteomics. *Curr Pharm Biotechnol* 5:79–88
15. Thiele H, Glandorf J, Hufnagel P, Körting G, Blüggel M (2008) Managing proteomics data: from generation and data warehousing to central data repository. *J Proteomics Bioinform* 1:485–507



Data Mining in Proteomics

From Standards to Applications

Hamacher, M.; Eisenacher, M.; Stephan, C. (Eds.)

2011, XII, 461 p., Hardcover

ISBN: 978-1-60761-986-4

A product of Humana Press