Interpretation of MALDI-MS/MS-spectra

A manual for working with Bruker MALDI-MS/MS software, tools and for the use of sophisticated techniques to get more information from your samples
Contributors

Dr. Marcus Macht, Bruker Daltonik GmbH, Applications
Sören-Oliver Deininger, Bruker Daltonik GmbH, Applications
Dr. Arndt Asperger, Bruker Daltonik GmbH, Applications
Dr. Franz-Josef Mayer-Posner, Bruker Daltonik GmbH, Applications
Dr. Detlev Suckau, Bruker Daltonik GmbH, Applications development
Dr. Anja Resemann, Bruker Daltonik GmbH, Applications development
# Content

1. **Introduction**  
   1.1 **Definitions**  
   1.2 **Understanding peptide fragmentation**  
      1.2.1 Charge directed fragmentation versus charge remote fragmentation  
      1.2.2 The influence of specific amino acid residues  
      1.2.3 Special fragment ions in MALDI-MS/MS spectra  
         1.2.3.1 Immonium and immonium related ions  
         1.2.3.2 The origin of $b_x + H_2O$ ions  
      1.2.4 Low energy fragmentation versus high energy fragmentation  
   1.3 **Negative mode MS/MS**  
   1.4 **Tips on acquiring and handling PSD spectra**  
   1.5 **Sodium adducts of peptides – how to identify and why to avoid**  
   1.6 **Mixing LID and CID to improve immonium ion information**  
   1.7 **Using the PLMS for labile compounds**  

2. **Handling of peptide MS/MS spectra**  
   2.1 The easiest case: Database searching using uninterpreted MS/MS spectra  
   2.2 The use of combined MS/MS data sets for Database searches  
   2.3 The use of XMass macros for data interpretation  
   2.4 FlexAnalysis macros for interpretation assistance  
   2.5 The use of the MASCOT sequence search for complex database searches  
   2.6 **De-novo-sequencing**  
      2.6.1 Evaluation of de-novo results  
      2.6.2 Strategy for doing a manual de-novo sequencing  
      2.6.3 LID versus CID for L/I differentiation  
      2.6.4 Using MS-BLAST for the identification of proteins  
      2.6.5 **Labelling techniques in MS/MS analysis**  
         2.6.5.1 $^{18}$O-labelling  
         2.6.5.2 Charge tagging ("CAF")  
      2.6.6 **Identification of posttranslational modifications by MS/MS**  
         2.6.6.1 Methionine oxidation  
         2.6.6.2 Serine- and Threonine-phosphorylation  
         2.6.6.3 N-terminal acetylation  
         2.6.6.4 Methylation of acidic moieties
2.6.6.5 N-methylation of Lysine side chains 77
2.6.7 A practical example 78

3 In-source decay (ISD) 84
3.1 Acquisition of ISD spectra 84
3.2 Interpretation of ISD spectra 86
3.3 Terminus specific fragmentation (\(T^3\)) 87

4 Appendix 90
4.1 Useful tables 90
4.2 Recipes 94
  4.2.1 Gel electrophoresis (according to Laemmli et al.) 94
  4.2.2 Gel stainings 95
  4.2.3 Enzymatic digestion protocols 96
4.3 Suggested readings 99
  4.3.1 Books 99
  4.3.2 Articles 99
  4.3.3 Internet resources 101
    4.3.3.1 Database search engines 101
    4.3.3.2 Helpful Tools 101
    4.3.3.3 Databases 101
1 INTRODUCTION

1.1 Definitions

metastable decay, LID and PSD

Due to the fact that large thermo labile molecules such as proteins can be desorbed and detected as expected by applying MALDI, this technique was denoted “soft ionization technique”. As MALDI was just introduced mainly ions formed intactely were considered and some fragments were considered only as an exception. On evaluating the mass spectra, first only those additional signals corresponding to fragment ions were recognized, which were formed during the desorption process inside the source (in-source decay, ISD). Further experiments showed that a decay of molecular ions later on during the acceleration process or when passing the drift region takes place spontaneously as well. The process of metastable decomposition in the area of the drift region was called PSD (Post Source Decay). Usually, the analysis of PSD ions is used to acquire information concerning the structure of the original molecular ion.

Metastable decomposition of the analyte ions happens by absorbing energy during MALDI by collision excitation processes inside the matrix cloud immediately after the laser pulse has hit the sample. The efficiency of the metastable decomposition depends on four major factors:

a) the density of the matrix cloud after the laser shot and the height of the kinetic energy making analyte ions and matrix molecules collide. The density of the matrix cloud depends on the laser intensity and the temperature the matrix supplies.
b) the laser fluence – higher laser power yields more fragmentation.
c) the acceleration voltage applied defines the collision energy. During metastable decomposition a charged fragment ion as well as a neutral fragment molecule are formed as shown in Figure 1 and
d) the matrix itself: measurements of the initial velocities of MALDI matrices have led to a classification of matrices into “hard” and “soft” matrices. “Hard” matrices like \( \alpha \)-cyanocinnamic acid usually give more fragmentation than “soft” matrices like DHB or sDHB. Sinapinic acid and ferulic acid are somewhere in between.
Figure 1: The graphics shows the dependance of the formation of different particles on the laser power. As you can see, low laser power (close to the threshold for ion formation and observation) produces primarily neutral particles and singly charged ions. When the laser power is raised, the fraction of metastable ions rises so that we can detect the fragment ions produced by metastable decay.

The time scale where the decomposition process takes place is the most important criterion to answer the question, whether fragment ions are detectable with a TOF mass spectrometer or not. Prompt fragmentation, right before the acceleration takes place, delivers signals associated with the correct mass of the fragment (see the paragraph about ISD in this chapter as well as chapter 3). Fragment ions formed inside the acceleration region absorb different kinetic energies depending on the place of decomposition thus deteriorating the S/N ratio of the spectrum.

Figure 2: Schematic view of the locations for the different fragmentation processes. Please note that the CID process is actually only induced within the collision cell, the actual fragmentation might happen at any time/place between the collision cell and the LIFT.
Figure 3: Metastable decay of molecule ions along the drift region of a linear TOF.

Inside a linear ion flight tube, charged fragments ($F^+$) and neutral ones ($F_n$) strike the detector simultaneously with the intact stable molecular ion ($P^+$). This is due to the fact, that PSD ions formed in the drift region continue the flight at the same velocity as the original molecule ion and are, therefore, in principle not distinguished by a linear TOF (Figure 3). As the modification of the mass of the fragment ion comes along with a modification of the kinetic energy ($E_{kin} = \frac{1}{2} m v^2$) the signal splits as soon as an electrical field, which slows down the particles, is applied to a precursor ion and its fragments. Therefore, charged fragments and the intact molecular ions are shifted towards higher m/z values when the spectrum is acquired in reflector mode on a TOF MS instrument. Spengler and Kaufmann described this phenomenon first. Their and other related experiments let estimate that about 1–5 % of the original molecular ions decompose on their way through the flight tube of a TOF MS. The timescale of this, so called post-source decay, is quite slow, usually in the lower ìs range. Results from other researchers [UDE paper] obtained on QIT-TOF instruments indicate that this process might even reach into the ms-timescale.

PSD fragment ions can be separated in mass using a reflector device (Figure 4). When operating the reflector in “normal” mode, only that portion of the fragment ions exhibiting approximately not less than 70% of the mass of the precursor is guided to the reflector detector. PSD ions of less than 70% of the precursor’s mass are guided beside the reflector detector because their energy is too low to be reflected at the correct angle in order to hit the detector. Also, PSD ions striking the detector are displayed at wrong m/z values, since the reflector is not capable to compensate for their decreased energy level. As a consequence of this, PSD ions are registered as signals of comparatively low resolution in a specific, but non-correct mass distance to their precursor ions (Figure 7). The greater the mass difference to the precursor ion, the broader is the signal shape.
To obtain a complete set of fragment ions, spectra are recorded at stepwise-adapted (reduced) reflector voltages. By means of this reduction of the reflector voltage, ions that did not strike the detector at full reflector voltage, will now be guided onto the correct flight curve to get collected on the detector plate. In the primary result of a PSD analysis, 10 to 20 single “segment” spectra (the number of spectra depends on the absolute mass number of the precursor ion, which defines the number of steps when reducing the reflector voltage) are obtained containing fragment ions of differing mass ranges, which have been recorded at stepwise reduced reflector voltages. Finally, these segments are pasted (computer aided) into one single, mass linearized, MS/MS spectrum.

*Figure 4: Metastable decay of molecule ions in a reflector TOF.*
Figure 5: typical PSD segment spectra (upper panels) and the resulting pasted and mass linearized “PSD-spectrum”. The panels show the upper 5 of a total of 19 segments. As you can clearly see from the figure, the PSD segment spectra have two characteristic features: a) they are overlapping and b) the higher the segment number (which means the lower the reflector voltage) is, the smaller is the actual mass range covered by this segment.

Figure 6 shows a typical arrangement of a PSD analysis using a two-stage reflector. Molecular ions $P_1^+$ being decomposed during the flight in the drift region can be detected by reducing the deflection voltage stepwise in an adequate manner. The PCIS-electrode (PreCursor Ion Selector) allows the selection of a specific molecule ion (for further details please refer to the FlexControl 2.0 user manual).
Figure 6: Principle of the PSD mass analysis.

**metastable side chain fragmentation**

The most commonly observed metastable ions in the MALDI spectrum result from neutral losses of some side chains. Frequently observed is the neutral loss of the oxidized Methionine side chain as shown in Figure 7.

The mass assignment in the MALDI mass spectrometer is only correct for ions which pass the reflector with the full energy of the accelerating field. Since the fragment ions travel at the same velocity as the precursor ion it follows that the kinetic energy of the fragment ion is decreased after the decay equivalent to the ratio of their masses $m_F/m_P$. Thus the fragment ions are not detected at their real mass, but at a somewhat higher apparent mass. Typically the apparent mass difference between a precursor ion and the fragment after neutral loss of the oxidized methionine side chain is about 55 Da (rather than the real mass difference of 64 Da).
Figure 7: metastable fragmentation of an oxidized peptide. The unmodified peptide has a mass of 1651.8 Da while the oxidized form appears at 1667.7 Da. You can clearly recognize the significantly less well resolved metastable fragment at an apparent mass of 1612.4 Da. The actual mass loss from the oxidized peptide is 64 Da instead of the apparent 55 Da as shown on the right hand side.

Another modification that may be assigned by metastable ions is a phosphorylation. Phosphorylated peptides can also undergo a neutral loss of phosphoric acid. It is yet unlikely to observe phosphorylated peptides in a peptide mixture unless they are especially purified due to the fact that physiological phosphorylation levels for a specific residue are usually low (~1-5%). However, there are cases where high phosphorylation levels are observed and β-Casein is such a case with a 100% phosphorylation at one residue. Figure 8 shows a part of a spectrum of a tryptic digest of β-Casein. Within this spectrum you can find the two most commonly observed metastable fragmentations: from oxidized Methionine as well as from phosphorylated Serine or Threonine (please note that Phosphotyrosin is much more stable under MALDI conditions and will usually not show intensive neutral loss).
Figure 8: Portion of a MALDI mass spectrum of β-Casein tryptic digest. One can clearly see three different metastable loss signals: a side chain loss from an oxidized Methionine (-CH$_3$SOH), a loss of phosphoric acid (-H$_3$PO$_4$) as well as the associated loss of phosphonic acid (-HPO$_3$).

The implications of this observation are, that metastable ions should not be considered for database searches, since the apparent mass is not the real mass. On the other hand, if one finds a metastable signal in the spectrum one can check whether there is a signal of the precursor ion at approximately +55 Da. If the spectrum is used for database searches, one can then confirm if the matched sequence contains an oxidized Methionine. Since the Methionine residues are often only partially oxidized, one may also find a signal corresponding to the unoxidized peptide.

When comparing peaklists with your spectrum and wondering that metastable peaks might not appear in the peaklist, please keep in mind that later versions of Xmass as well as FlexAnalysis contains a so called “metastable filter”, which eliminates broad peaks from the peaklist, as shown in Figure 9. This will automatically remove peaks with a FWHM larger than the specified value from the peaklist.
#additional new SNAP parameters may be modified here
# FWidth: 0.75 is default; i.e. Fragments broader
# dm >0.75 Da are removed
# RmFrag: 1 is default; i.e. Fragments are removed
# change 0.75 to something else here
# and remove hashes
storepar "FWidth", 0.75, 0, R
storepar "FWidth", 0.75, 0, S
# remove hashes if you don't want fragments being removed
#storepar "RmFrag", 0, 0, R
#storepar "RmFrag", 0, 0, S

Figure 9: Left hand side shows the “metastable filter” part of the Annotate.aura macro of Xmass. On the right hand side you see the “Mass List Find” section of the FlexAnalysis Method parameter with the part which removes the metastable peaks from the peaklist (in this example deactivated, activation would force the software to remove peaks with a FWHM of more than 0.75 Th).

Note that the resolution of the metastable ions does not always have to be that much lower as in the shown example. Sometimes (depending on the experimental conditions, such as laser power) the metastables have only a slightly higher half width than normal signals.

**CID – Collision induced dissociation**

This process occurs within and after a defined, gas filled, collision cell. The ions collide within the cell with atoms or molecules of the collision gas (usually He or Ar at around 6x10^-6 mbar measured in the ion source region, but also N₂, Kr, Xe or reactive gases like NH₃ or CH₄) and are thereby collecting energy. After collisional induced activation the ions will relax into their energy by rotation, vibration or fragmentation. In the classical CID process, the fragmentation event is usually a single collision decay. However, one has to be aware that during the passage of the ions through the collision cell, multiple collisions with gas molecules and therefore multiple fragmentation events may take place. Due to the fact that the ions are accelerated using a potential of several kV, the CID process in MALDI-MS instruments is always a high energy process. This has significant consequences on the appearance of the MS/MS fragment ion spectrum. In classical CID on QqQ, QTof or ion trap type instruments you will usually observe y- or b-ions and their respective neutral losses (H₂O, NH₃). When looking at high energy CID spectra, additional d- and w-type ions are observed which originate from side chain fragmentations. These ions can be used e.g. to distinguish between Leucine and Isoleucine (see chapter 2.6.3 on this issue). However, Figure 10 shows a comparison of data illustrating that using CID has not only advantages but also
disadvantages. The problem is, that while gaining a significant number of fragment ion signals more than compared to LID, only very few of them can be used to solve an analytical problem. The majority of the signals only present in CID can not be assigned to sequence specific fragments in a meaningful way. This means that CID is a helpful tool to answer questions like “is it a Leucine or an Isoleucine at position 7?” but is not helpful for automated identification of proteins based on database searches using peptide MS/MS spectra.

It actually may even be counterproductive since these unassignable signals might lead to false positive identifications in the worst case. Therefore we strongly recommend to use LID for identification work and to acquire a second CID spectrum to answer specifically the questions which deserve the use of this technique. Also, the comparison of both spectra allows for an easier distinction of ions which have specifically been formed during the CID process.

<table>
<thead>
<tr>
<th></th>
<th>LID</th>
<th>CID</th>
<th>CID-LID</th>
</tr>
</thead>
<tbody>
<tr>
<td>total</td>
<td>102</td>
<td>138</td>
<td>36</td>
</tr>
<tr>
<td>backbone</td>
<td>21</td>
<td>20</td>
<td>-1</td>
</tr>
<tr>
<td>neutral loss</td>
<td>12</td>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>internal</td>
<td>17</td>
<td>17</td>
<td>0</td>
</tr>
<tr>
<td>immonium</td>
<td>4</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>side chain</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>assigned</td>
<td>54</td>
<td>55</td>
<td>1</td>
</tr>
<tr>
<td>not assigned</td>
<td>48</td>
<td>83</td>
<td>35</td>
</tr>
</tbody>
</table>

Figure 10: A study illustrating the difference between LID and CID on the number of fragment ions in MALDI-MS/MS spectra. The upper study shows that we gain approximately 40% more signals in the CID spectrum (total of 36 signals), but the majority of those can not easily be assigned to sequence specific fragmentation (35 of the 36). A comparison of a competitor gave essentially the same results with almost identical numbers. The increase in peak numbers was approximately 40% (98 to 141), but 32 of them were internal and 10 were high energy fragments of undefined structure. The corresponding mass spectrum showed only one d- and one w-ion among those.

**ISD – In-source decay**

ISD is a process occurring within the ion source. This fragmentation is somehow similar to ECD (electron capture dissociation) in FTMS and yields primarily c- and y-ions as well as some a-type ions (see Chapter 3 on this). The timescale for this process is very short (<200 ns).

**LIFT**

LIFT is a special technique patented and delivered by Bruker in the new Autoflex- and UltraFlex-TOF/TOF instruments that allows the acquisition of complete fragment ion spectra as a result of a single laser shot and replaces traditional time and sample consuming measurements of segmented FAST spectra with corresponding reduced reflector potentials.
The basic idea of LIFT is to lift the potential of both, precursor and fragment ions right after fragmentation has taken place. That means to add kinetic energy to precursor ions ($P_s^+$) as well as to fragment ions ($F^+$) in parallel to such a high value that the energy difference between parent and smallest fragment does not exceed 30% anymore. As long as this requirement is being met, all fragments are detected simultaneously with the parent ion since they can be focussed on the detector by the reflector in this case.

![Diagram of Ultraflex TOF/TOF mass spectrometer during the LIFT process](image)

**Figure 11:** Functioning of the Ultraflex TOF/TOF mass spectrometer during the LIFT process.

**precursor ion (also known as parent ion)**

The ion which is selected and fragments into pieces during the MS/MS process.

**product ion (also known as fragment ion or daughter ion)**

The fragments coming from a precursor ion. In LID and CID they usually originate from unimolecular decay events, while in CID (high or low energy), they may also originate from multiple collisions.

### 1.2 Understanding peptide fragmentation
1.2.1 Charge directed fragmentation versus charge remote fragmentation

An important issue in discussing peptide fragmentation is the understanding of the different fragmentation mechanisms which might occur. This has been thoroughly investigated during the last years and led to a sound understanding of what is happening during peptide fragmentation in the gas phase. The general understanding is that the fragmentation of peptides under low energy conditions is initiated by migration of the proton from the initial site of peptide protonation (in tryptic peptides usually the N-terminal amino group or the C-terminal side chain) to an amide carbonyl oxygen along the peptide backbone. This is usually termed as the “mobile proton” hypothesis. The cleavage of the peptide bond then occurs by the attack of a neighboring group from an adjacent nucleophilic amide carbonyl group to yield the b-type fragments, y-type fragments or both of them.

$pK_a$ (amino terminus) $\sim$ 9
$pK_a$ (Lys side chain) $\sim$ 10.8
$pK_a$ (Arg side chain) $\sim$ 12.5

Figure 12: Scheme of the “mobile proton” concept. The figure shows a doubly protonated peptide with the protons located at the N-terminal amino group as well as a C-terminal arginine. Due to the higher $pK_a$ value of the Arg side chain the mobile proton is usually the N-terminal one. This proton can move along the peptide chain, destabilize the amide bond and induce fragmentation along the chain.
Figure 13: The figure shows a singly charged tryptic peptide with a proton localized on the C-terminal basic amino acid residue.
Figure 14: Mechanism of the attack of the carbonyl oxygen on the neighbouring carbonyl carbon in the oxonium or alternatively, the quaternary ammonium form. It is unclear whether both or only one options exist. However, the oxonium intermediate is the most likely form to be attached as demonstrated by several studies.

1.2.2 The influence of specific amino acid residues

Figure 15: The cyclic Proline residue hinders the attack of the carbonyl oxygen on the neighbouring carbonyl carbon atom. This explains the low abundance of the fragment ions signals on the C-terminal side of Proline.
Figure 16: Part of the MS/MS-spectrum of the peptide LLYEQPVLPVR. While \( b_5 \) and \( y_7 \), originating from the cleavage of the peptide chain N-terminal to Proline, have high intensities, the corresponding fragment ions \( b_6 \) and \( y_6 \), originating from a cleavage C-terminal to Proline have very low intensities indicating the favourization of the N-terminal cleavage at Proline residues.

Figure 17: Mechanism of acid catalysis of the peptide fragmentation on the C-terminal side of Asp- (and to minor extent Glu-) residues. The cleavage results in a cyclic C-terminal structure for the \( b \)-type fragment.
Figure 18: Part of the mass spectrum of the peptide pyroEGVNDNEEGFFSAR covering the fragment ions of the acidic amino acid residues. As it can be clearly seen, the relative intensity of the fragment ions originating from a cleavage on the C-terminal side of the acidic residue is higher than on the N-terminal side. This is especially true for aspartic acid residues, but also for glutamic acid residues.
Figure 19: Mechanism of the Histidine-assisted cleavage of a peptide yielding a b-type ion with a bicyclic C-terminal structure.

1.2.3 Special fragment ions in MALDI-MS/MS spectra

1.2.3.1 Immonium and immonium related ions
Figure 20: Zoomed view of the immonium ion region of an Arg-containing peptide. In the LID spectrum (a), apart from the immonium ions of Pro, Leu and Lys (residue mass-27 Da), there is a immonium related ion of Gln observed (84 Da). The CID spectrum (b) furthermore shows two Arg-related ions originating from the actual immonium ion (129 Da) according to the reaction scheme. The appearance of those two ions is CID specific and originates from multiple collisions with the collision gas.

1.2.3.2 The origin of b_{x}+H_{2}O ions
Figure 21: $b_n+H_2O$ ions are typically observed in LID spectra (a), but not in CID spectra (b). They originate from a rearrangement of the peptide C-terminus, not from collisions with water e.g. from the crystallization as shown in the reaction scheme. This could be proven by a) $^{18}O$-labelling of the peptide [Gonzalez J, Beasda V, Garay H, Reyes O, Padron G, Tambara Y, Takao T, Shimonishi Y. J. Mass Spectrom 1996: 31; 150-158] and b) the use of $H_2^{18}O$ for the preparation of the sample [M. Macht, unpublished results]. The first experiment led to a 2 Da mass shift of the $b_n+H_2O$ fragment ion, while the latter one did not.

1.2.4 Low energy fragmentation versus high energy fragmentation

Figure 22: MS/MS spectra of ACTH(1-10). The spectra have been acquired on a FAB four-sector magnetic deflection instrument. Spectrum (A) has been acquired at a collision energy of 7 keV using helium as collision gas [high energy CID], while spectrum (B) has been acquired at 45 eV using Argon collision gas [low energy CID]. The figure has been taken from “The interpretation of collision-induced dissociation tandem mass spectra of peptides” from Ioannis A. Papyannopoulos, Mass spectrom.
Figure 23: Tandem mass spectra of ACTH(1-10) acquired on an Esquire ion trap (low energy CID using helium as collision gas), and on an ultraflexTOF/TOF using LID fragmentation (low energy) and CID (high energy fragmentation using argon as collision gas). Please note the agreement of the ions between the 7 keV FAB spectrum and the high energy CID MALDI-TOF/TOF spectrum. This agreement clearly demonstrates that high energy conditions in a TOF/TOF instrument are similar to those in sector field instruments.

1.3 Negative mode MS/MS

The acquisition of negative mode MS/MS spectra is almost identical to the acquisition of po-
sitive mode MS/MS spectra. The major difference is that all ions have a molecular weight of 2 Da less than in positive mode (M-H instead of M+H, so it is actually 2.0156 Da difference). Another difference is the appearance of fragment ion types usually not appearing in positive mode MS/MS spectra. This is especially true for c-type ions as shown in Figure 24.

![Figure 24](image-url)

**Figure 24:** Negative mode LIFT spectrum of Substance P. As in normal positive mode, a-, b- and y-ions are usually observed. However, as can be seen from this figure, in negative mode c-ions are commonly observed as well.

When working with negative mode MS/MS spectra, please make sure that your BioTools param.dat file (in the directory ...\Bruker\BioTools\) contains the following section:

```
[CID-MSTOOL: LIFT negative]
Series 1=a,-COOH3,N,
Series 2=b,-OH3,N,
Series 3=c,-O+N,N,
Series 4=i,-COOH3,I,
Series 5=y,-H,C,
Series 6=y-17,-OH2,C,
```

For further detailed information, you can also refer to the following articles about negative mode MS/MS on MALDI instruments:


J. Jai-nhuknan, C.J. Cassady, „Negative ion matrix-assisted laser desorption/ionization time-
Table 1: Applicability matrix for Brukers MALDI mass spectrometry instrumentation. ° while CID in APEX-Q instruments is done within the collision hexapole region, in APEX systems it is performed within the FT-ICR cell. This difference results in a dramatic increase of the timescale for the MS/MS experiments (about 1-10 sec. on APEX-Q compared to a couple of minutes on APEX).

<table>
<thead>
<tr>
<th></th>
<th>LID</th>
<th>high energy CID</th>
<th>PSD</th>
<th>LIFT-TOF/TOF</th>
<th>ISD</th>
<th>low energy CID</th>
<th>IRMPD</th>
<th>positive mode</th>
<th>negative mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>Omniflex (without LT)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Autoflex</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>AutoflexTOF/TOF</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Ultraflex</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>UltraflexTOF/TOF</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>MALDI-APLEX</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>MALDI-APLEX-Q</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>AP-MALDI-BioTOF-Q</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>AP-MALDI-Esquire</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

1.4 Tips on acquiring and handling PSD spectra

When acquiring PSD segment spectra, you are always acquiring independent data sets. The acquisition of segment number 5 has therefore no influence on the acquisition and appearance of segment number 6 nor number 4. As a consequence, this means that you can always optimize the quality of your segment spectra independently. You don’t have to care about shot numbers, different S/N etc. Acquire as many spectra as needed to get a reasonable spectrum quality, in some cases (especially when doing manual calibration in XMass or FlexControl versions <2.0), this might be more than 1000 per segment! Another specific feature of PSD spectra is also caused by the fact that the final linearized spectrum consists of independently calibrated segments. This calibration is usually very accurate in the middle of a segment, but
might be less accurate at the borders of an individual segment. This has the consequence, that at the seam between two segments sudden jumps in the mass accuracy may appear. This is different in LIFT spectra, where the complete spectrum is acquired at once, the calibration is smooth and therefore sudden jumps must not appear!

Specific tips for the users of the older XMass software are:

a) You should take care that there is always ONLY THE CURRENTLY VALID psd calibration file available on your system. This has to fit to your current PSD method. The most common error when handling PSD spectra is to use the wrong calibration file which causes wrong mass assignments afterwards. The “psd file name”-section in Figure 25 is a good example how it should not like.

b) When performing the FAST segment pasting, usually the option “no overlapping low->high expno” gives the best results, because the resolution on the upper border of a segment is usually better than on the lower border of the previous segment. This is due to the fact that ions with higher energy can be focused more easy than ions with lower energy.

---

Figure 25: XMass’ “FastPaste arguments” window. The best results are usually obtained using the “no overlapping ranges high->low expno” option. (Attention: as shown in the Figure, the “no overlapping ranges (50 % rule)” is activated by default!)

---

1.5 Sodium adducts of peptides – how to identify and why to avoid
When acquiring MS spectra of peptides or in gel digests, on commonly observed adduct is usually the exchange of hydrogen versus sodium ([M+Na]$^+$ instead of [M+H]$^+$), causing a mass shift of +21.982 Da. When there is a lot of sodium in the sample, you might even not be able to observe the protonated peptide. In some cases, washing the sample with 0.1% TFA solution or 50 mM NH$_4$CH$_3$COO (pH 4-4.5) might help to get rid of the sodium adducts. Nevertheless, in some cases you might not be aware of looking at an sodium adduct and trying to acquire a MS/MS spectrum of such a sodiated peptide.

Figure 26: A shows the MALDI mass spectrum of the peptide Tyr(SO$_3$H)-Hirudin (55-65). The expected sequence was DFEEIPETH with a corresponding molecular mass of 1492 Da for the protonated peptide. The observed mass was 1433.6 Da, corresponding to the non-sulfated sodium adduct of the peptide. Please note that absolutely no indication is visible for the presence of the protonated pep-
tide, therefore an identification of the sodium adduct by looking on mass shifts failed. B shows the MALDI-MS/MS spectrum of mass 1433.6 with sequence assignments for one of the sodiated forms (the sodium could be localized at the C-terminus or any of the five acidic residues). This is probably not the only form present but it shows that the sodium can nicely explain the appearance of the spectrum. Please also note the large, unexplained signal slightly below 800 Da.

A typical thing observed with sodiated peptides, which can also be used to get an idea what is going on with the sample, is the appearance of metastable losses resulting in broad unresolved peaks in the spectrum.

![Figure 27: Comparison of MS/MS spectra of a protonated peptide (A) and the same peptide in sodiated form (B). As you can see, the appearance of the spectra is completely different and a way to get an idea about sodium adduct formation is to look at the resolution of the peaks (see the inserts). Due to metastable loss of the sodium, the peaks in the lower spectrum are much broader than in the upper spectrum. Also, the masses determined for the metastable peaks are not correct.](image)

1.6 Mixing LID and CID to improve immonium ion information

A tip to improve especially S/N in the low mass region of a MALDI-MS/MS spectrum is the following procedure:

When acquiring a LID MS/MS spectrum, you might observe that the intensity of the low mass ions (e.g. immonium ions) is rather low. This is especially the case when the vacuum in your MALDI source is very good (low $10^{-7}$ mbar or below). To improve the S/N values in that region of the spectrum, you can, after acquiring a number of laser shots in LID mode, switch on
the CID valve to bring the CID cell to elevated pressure. After switching on the CID valve the
gas pressure will probably go above the pressure threshold, which will cause your instrument
to switch off the high voltage. After a couple of seconds, you can switch on the high voltage
again and wait until the instrument is in “Ready” status again. When you will acquire further
spectra now, this will give you actually a mixture of LID and CID spectra in the sum buffer.
The advantage of this is, that usually in CID mode, the low mass part of the spectrum is more
pronounced thus giving you a better S/N in this region, especially for the immonium ions.

1.7 Using the PLMS for labile compounds

Another trick you can use on TOF/TOF instruments to improve spectra quality is to move the
timing of the so called post-LIFT metastable suppressor (PLMS). The purpose of this device
is to remove precursor ions after the LIFT, which are still metastably activated and could un-
dergo fragmentation on their further way towards the reflector. This would cause the same
effect as described for the metastable side chain fragmentation described in chapter 1.1 (broad
peaks with wrong masses essentially contributing to the spectrums noise). Sometimes you
might observe an effect as shown in Figure 28 with one strong fragment ion signal and low
S/N elsewhere in the spectrum. This happens when a part of the peptide is extremely labile
(pThr, pSer or sTyr but also MetOx) and undergoes a neutral loss. One possibility to increase
the S/N is to increase the number of shots, however, like in NMR, the S/N unfortunately only
improves with the square root of the number of shots, which means that to gain a factor of 10
in S/N, you would have to acquire 100 times more laser shots (e.g. 30000 instead of 300).
Figure 28: Negative mode LID spectrum of Caerulein, an extremely labile peptide with the sequence $p$EQDY(SO$_3$H)TGWMDF-NH$_2$, which looses easily SO$_3$ under MS/MS conditions as can be seen in the spectrum.

A different possibility is to increase the laser power. In order to avoid saturation of the digitizer, you can alter the timing of the PLMS and deflect the abundant signals in the upper mass range of the spectrum. As shown in Figure 29, the timing can be easily altered within FlexControl (please note that this screenshot is taken from FlexControl 1.2). To do so, just alter the value after “PLMS” and press enter. The new value will automatically become valid.
Figure 29: Schematic drawing of the UltraflexTOF/TOF to illustrate the position of the PLMS. Normally, the timing of the PLMS is set in a way that only the metastable activated precursor is deflected before entering the reflector. However, the timing can be adjusted in a way that also prominent fragments can be deflected to avoid their appearance in the fragment spectrum. You can achieve this by lowering the value for the PLMS timing (the time after which the PLMS starts to deflect the fragments – in this example 25660 ns after the laser pulse). The lower this value is, the lower the masses which are deflected.

In the newer FlexControl version 2.0 this is easily possible as well, yet slightly more hidden. Figure 30 shows how to reach this value setting in version 2.0. In the LIFT tab just click on “Calibrate” (for the PLMS, NOT the ion selector or LIFT, this is very important!)
Figure 30: Altering the PLMS timing within FlexControl 2.0. In this version, the timing setting is hidden in the PLMS Calibration setup. However, the function and possibility for altering this value is identical to version 1.2.

Figure 31 shows the result of this step. As before, we still see our precursor since this is acquired independent from the fragments. We also see the $y_{10}$-SO$_3$ neutral loss as before, but now not anymore as the dominant signal in the spectrum. However, in contrast to before, we are also observing now a lot more fragment ion signal which were not at all visible before or had a very small S/N. What we did is the following: We acquired the precursor as usual, then switched into the fragment mode. We acquired some laser shots, which were dominated by the neutral loss as before and summed them up. Subsequently, we lowered the time of the PLMS start to deflect the neutral loss signal and acquired some more laser shots. In these last spectra, the neutral loss was completely eliminated hence the spectrum could be easily optimized for a good fragment signal quality.
**Figure 31:** Spectrum resulting after deflecting the prominent neutral loss of SO$_3$. As you can see, it is still present in the spectrum by the overall S/N, especially for the smaller fragments, is dramatically enhanced allowing to achieve much more MS/MS information as compared to the spectrum in **Figure 28**.

Another example is shown in **Figure 32**. This spectrum shows the MS/MS spectrum of a quadruply phosphorylated peptide of mass 3122 Da, which had been isolated by Fe-IMAC (immobilized metal chelate affinity chromatography using Fe$^{3+}$ as metal bound to an immobilized NTA chelate ligand). The peptide has then been analyzed by LID-LIFT and the typical intensive neutral loss of phosphate has been observed. The PLMS timing had been decreased to a level, where the phosphate losses were deflected and subsequently the spectrum could be acquired. This resulted in a spectrum which could be easily used for automatically or manually deducing a sequence tag for further searches.

**Figure 32:** LID-MS/MS spectrum of a quadruply phosphorylated peptide with the sequence RELEELNPGEVEpSLpSpSpSEESITR. To acquire the spectrum, the PLMS was set to a value which caused deflection of masses above approximately 2650 Da. This completely excluded the four possible neutral losses of the phosphate side chain modifications from the acquisition (the residual signal in the spectrum originates from the spectra acquired before shifting the PLMS timing).
2 HANDLING OF PEPTIDE MS/MS SPECTRA

While offering a wealth of information, the interpretation of MS/MS spectra generated either by laser induced dissociation (LID) or collisional induced dissociation (CID) and subsequent metastable decay is sometimes tedious due to different types of fragment ions occurring (i, a, b, y and corresponding neutral losses). This document shall assist in the use of different tools (software, sample prep, knowledge) to simplify the interpretation and to evaluate the meaningfulness of results obtained by manual or automated de-novo sequencing. Some of the techniques described in here are supported by software tools but do not necessarily depend on them and can also be used for manual interpretation.

2.1 The easiest case: Database searching using uninterpreted MS/MS spectra

The most easy way to identify a protein is by using uninterpreted MS/MS spectra information for the database search. This is also the most common way nowadays since it is simple and does not require extensive user interaction which makes it easy to automate. The way this is done is detecting the peaks in the spectra to create a peaklist and to submit this peaklist into a database search engine (e.g. MASCOT or Sonar). Please keep in mind that not all search engines support all types of searches (Sonar and Sequest only support MS/MS searches while ProFound only supports PMF searches). After creating the peaklist using Xmass or flexAnalysis, the spectrum can be transferred into BioTools by clicking on the BioTools icon in the postprocessing software or via the BioTools entry in the Tools menu of flexAnalysis. This submits the raw spectrum data as well as the peaklist (red bars, see Figure 33).
Figure 33: uninterpreted raw MS/MS spectrum displayed by BioTools.

By clicking on the “MSMS” button in the menu line of BioTools the database search window for MASCOT searches is opened (PROWL and EMBL search available via the Search menu are only for MS fingerprint searches) (Figure 34). Here you have to enter your search parameters. Name and email are mandatory when you want to do the search at Matrixscience MASCOT server via the internet, if you are doing the search on a local server they are not needed. Taxonomy and enzyme depend on your experiments as well as the fixed modifications do. Common fixed and variable modifications are mentioned in the figure. Please be careful when taking variable modifications into account, the search times rises exponentially with the number of variable modifications (whereas fixed modifications have no influence on the search time). You should usually enter the alkylating reagent used for the reduction and alkylation of the cysteins (usually 4-vinylpyridine or iodoactamide resulting in a S-pyridylethyl or Carbamidomethyl modification) as fixed modification. When the sample originates from a 1D-gel experiment, please also consider residual monomeric acrylamide as alkylating reagent. In this case you could either use this as fixed modification or in combination with Carbamidomethyl as variable modifications). The most common variable modification is Methionine oxidation and also should be the only one to be considered in an initial search.

Another parameter which is important to look at is “Missing cleavages max.”. Let’s assume you have a peptide LISFKHAATMR with one internal potential tryptic cleavage site (K). This would result from an incomplete cleavage at the Lysine residue. Incomplete cleavage is well known for three cases: When the basic residues form a cluster like ...KKRK..., when the
basic residue is followed by Proline or when the basic residue is in the direct neighborhood of an acidic or phosphorylated residue like ...EKX... or ...XRpS...(X stands for any amino acid). A reasonable value for this parameter is 1, however, sometimes 2 might also be appropriate; especially with 2D-gel samples processed using a short digestion protocol.

**Typical fixed modifications:**
carbamidomethyl (alkylation of Cys with Iodoacetamide), carboxymethyl (alkylation of Cys with Iodoacetic acid), S-pyridylethyl (alkylation of Cys with Vinylpyridine), ICAT labels

**Frequent variable modifications:**
phosphorylation, oxidation (M), acetylation or formylation of protein N-terminus (that’s a tricky one, often DB entries from DNA sequencing projects contain a preceding Met which is missing in the mature protein; in that case the acetylation is located on the second residue (e.g. in adseverin [AcAT...] and SMP30 [AcSS...] from mouse)), carbamylation at Lys (artifical from heating in urea solutions), propionamide (artifical alkylation at Cys from residual monomeric acrylamide in selfmade 1D-gels; can be identified by +14 Da differences to the carbamidomethylated peptides), methylation (from Coomassie staining)

---

**Figure 34:** MS/MS search interface. Typical fixed and variable modifications are mentioned.

A very nice feature of BioTools is the automatic detection of some modifications (phosphorylation and Methionine oxidation). BioTools does this by looking for the presence of peaks in the distance of $-98$ Da and $-80$ Da (phosphorylation, neutral loss of $H_3PO_4$ and $HPO_3$ respectively) and $-64$ Da (Methionin oxidation, neutral loss of $CH_3SOH$ from oxidized side chain) from the precursor mass. The detection of such a modification will be indicated as a small icon in the upper left corner of the sequence window of the the BioTools GUI as shown in **Figure 35**. You can use this information to get an idea whether these modifications might be present or not.
Figure 35: Automatic detection and indication of peptide modifications in BioTools.

The mass accuracy values depend on your acquisition mode and should be set to reasonable values (which may vary from 500 ppm in linear mode using a single calibration spot on the target down to less than 10 ppm when you are sure about your calibration status and maybe also using the HPC feature and/or an internal calibration).

**Excursus: Mass accuracy and the use of mass accuracy windows in database search engines**

When doing database searches, one of the most important parameters is the mass accuracy. Unfortunately, not every search engine uses these values in the same way. This paragraph shall explain the use of different calibration techniques available to the user as well as how mass accuracies are handled by MASCOT as well as ProFound/Sonar.

When calibrating a mass spectrometer using a set of different calibrant masses, usually either a linear or (preferably) a quadratic calibration curve is used. Depending on the specific features of your instrument, this usually gives mass accuracies better than 50 ppm in reflector mode which can be further enhanced to better than 10 ppm by using a subsequent internal calibration. However, due to ion physics, mass spectrometers sometimes don’t behave in an ideal way which will lead to a curve as shown on left hand side of **Figure 36**. To overcome these instrument specific deviations from an ideal calibration curve, Bruker has introduced a
so called high performance calibration (HPC) with the release of FlexControl 2.0. For this, a higher order calibration curve is calculated and (if specified within FlexAnalysis) applied to every spectrum acquired on the instrument using the specific, calibrated parameter set (since the actual shape of the curve also depends on the current parameter settings). The effect on the mass deviations after this correction can be seen on the right hand side of Figure 36.

Figure 36: Effect of Bruker’s high performance calibration HPC.

While this described procedure is merely a hardware issue, there are also other types of calibration tools available from within FlexAnalysis. One is the so called “statistical calibration”. In 1995, Matthias Mann has found that peptides will always appear at masses which are very similar for all peptides having the same nominal integer mass. Figure 37 illustrates this fact. The red bars show the number of peptides appearing at a specific exact mass, while the green lines give the “most likely” mass for a peptide with this nominal integer mass. These “most likely” masses as well as the mass spread which covers 95% of all peptides at this nominal mass can be calculated. What is done then is to calculate the RMS deviations of all peaks in the spectrum and to calculate a calibration which generates the minimum error for all peaks taking into account that the peaks are originating from peptides (this is the only restriction to the algorithm).
Figure 37: Distribution of accurate peptide masses around a “most likely” peptide mass. You can easily see that peptides can not have “any” mass but always appear at distinct masses. E.g. a singly charged peptide could never appear at a mass of 602.8 Da but only around 602.3 or 603.3 Da.

Figure 38 shows the practical effect of this calibration procedure on the mass deviations of a MS spectrum. While the mass errors before the calibration are in the range between +200 and -200 ppm (A), after the calibration the mass errors are between –50 and –50 ppm and now clearly show the instrument specific deviations mentioned before. Associated with this is also a significant reduction of the RMS error.

Figure 38: Effect of the statistical calibration on the mass errors of a spectrum. Panel A shows the original spectrum which has been used for the search. The calibration of the spectrum has been deteriorated on purpose before, to illustrate the effect. B shows the search result using the spectrum without statistical recalibration at a used mass accuracy of 100 ppm while C shows the result of the same
spectrum after the statistical recalibration. In D, the distribution of the mass errors along the mass axis is shown (please note that the mass accuracy had to be lowered down to 250 ppm to get a useful hit from which this distribution could be taken). The RMS error is around 150 ppm. After the statistical recalibration the RMS error is down to 36 ppm (when you look carefully on the mass distribution in E, you will recognize the same curve shape as in Figure 36. Therefore, additional use of the HPC would give you a further gain in mass accuracy).

An approach which had been suggested by E. Mørzt et al. in 1996 is a “quasi internal calibration”. This is done e.g. by the IntelliCal algorithm of a competitor as well as by ProFound/Sonar. The algorithm does the following: When a database search is performed in a first cycle a relatively low mass accuracy is used which creates a decent number of hits. All hits from this list are taken, an internal calibration using some masses of the hits is performed and then a re-scoring of the hits is done using a much more stringent mass accuracy. The purpose is the following: when a calibration is wrong, a wide mass accuracy might still pick up a potential hit. When this one is correct, after an internal recalibration using, let’s say three masses, the remaining ones should fit much better than before, now giving better scores even at significantly lower mass accuracies. The danger with this approach lies in the fact described in Figure 37. Peptides always cluster around certain masses! The only thing you gain this way is to correct a non-statistical error like with the statistical calibration. The difference between “quasi-internal” calibration and the statistical calibration is that the first needs a reasonable hit as a starting point while the second just relies on an intrinsic peptide characteristics.

After all parameters have been set properly, just press the “Start”-button and after the search is finished you will get the result. This may either be a page saying “Wot? No result?” (Figure 82) and some more or less helpful suggestions what to do now, or alternatively (fortunately significantly more often) a list beginning with some search specific information, a list of top candidates followed by a graphics displayed in Figure 39 and a detailed information for each of the found hits.
Excursus: Database searches, probabilities and result interpretation

There are several different possibilities to identify proteins using mass spectral data in a statistical way from databases. Common methods are:

- Probability based MOWSE score (Mascot)
- Bayesian probability (ProFound, Sonar)
- Cross correlation (MS-Fit, Sequest)

Each of these methods has advantages and disadvantages but they always give you a result which states that there is a certain likelihood that the given protein name/accession number is the major component of your analyte.

The search engines MASCOT and Profound usually report their results as shown in Figure 40:

---

**Figure 39:** Search result as displayed by MASCOT. The green area denotes the area where confidence in the result is less than 95%. For a single MS/MS spectrum, this would be a very reasonable result.

<table>
<thead>
<tr>
<th>Accession</th>
<th>M seq</th>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q91884</td>
<td></td>
<td>16601</td>
<td>epimysin - giant panda (fragment)</td>
</tr>
<tr>
<td>P15526</td>
<td>15094</td>
<td></td>
<td>epimysin - Borneo river otter</td>
</tr>
<tr>
<td>P15525</td>
<td>15079</td>
<td></td>
<td>epimysin - rabbit (tentative sequence)</td>
</tr>
<tr>
<td>P15524</td>
<td>15040</td>
<td></td>
<td>epimysin [validated] - bovine</td>
</tr>
<tr>
<td>P15523</td>
<td>15040</td>
<td></td>
<td>epimysin - European beaver</td>
</tr>
</tbody>
</table>

---

A

C

Protein Summary Report

To create a bookmark for this report, right-click this link: Protein Summary Report (data20190107/204469.dat)

Index

<table>
<thead>
<tr>
<th>Accession</th>
<th>M Arab</th>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.72E+7</td>
<td></td>
<td>(re-Other-Mammalia) TL:2.10E+6</td>
</tr>
<tr>
<td></td>
<td>1.72E+7</td>
<td></td>
<td>(re-Other-Mammalia) TL:2.10E+6</td>
</tr>
<tr>
<td></td>
<td>1.72E+7</td>
<td></td>
<td>(re-Other-Mammalia) TL:2.10E+6</td>
</tr>
<tr>
<td></td>
<td>1.72E+7</td>
<td></td>
<td>(re-Other-Mammalia) TL:2.10E+6</td>
</tr>
<tr>
<td></td>
<td>1.72E+7</td>
<td></td>
<td>(re-Other-Mammalia) TL:2.10E+6</td>
</tr>
</tbody>
</table>
A major task of the database search engines is to differentiate between correct identifications and false positive hits. The better this differentiation, the easier is the task to identify your protein. Unfortunately, not all search engines are equally good in that regard. Figure 41 shows a study conducted by a German proteomics company involving analysis of approximately 1600 MALDI data sets in which roughly 10000 database searches were performed. The outcome of the study was that while ProFound/Sonar as well as MASCOT have a good separation of correct hits (red) and false positives (grey), MS-Fit almost completely lacks this differentiation. Therefore, the score of a Profound or MASCOT search is a good indicator whether a hit is really correct or not, while in MS-Fit you always have to evaluate the hit independent of the score since this is not a helpful discrimination criterion.

Please keep always in mind that these results are statistical results! This is never the definite truth but always only an answer like “You can be 99.99% sure that this is the protein”. Generally MASCOT displays an area which is marked with green stripes. Everything which is within this area has a 5% chance to be a false positive. This means, everything with a higher
MOWSE score than the confidence area has a less than 5% chance to be a false positive, the higher the score, the better. Reasonable values for a single MS/MS spectrum start at approximately 30-40, significant hits are usually in the range of 50-100, very good data sets of combined MS/MS data can provide values up to more than 1000 (in this case you can be pretty sure that it is the correct protein, but still not necessarily the correct species!). A reasonable result with a less than 5% chance of a false positive when the search has been run in ProFound has a Z-score value around or above 1.6.

To evaluate MOWSE score data you should also know that an increase of the MOWSE score by 10 units means an increase in the confidence in the result of the search of a factor of 10. Therefore a hit with a MOWSE score of 120 is 1000 times more confident than one with a score of 90.

Typical observations are also that MASCOT gives higher scores to smaller proteins so that fragments of a protein appear higher in the rank list than the intact protein while ProFound or Sonar prefers to report proteins with higher masses. So when you identify a protein, also check whether this could be only a fragment of a larger one or if the top hit is an unprocessed precursor protein.

Please be also careful in this regard in the use of min. and max. masses for searches. Most search engines use these strict limits which has the effect that a protein, which does not fall within the given limits is excluded from the search prior to the search. A possible result could be the following: You observe a spot on the gel with an apparent mass of 42 kDa and you enter 30 and 50 kDa as lower and upper limit in the search engine (always use at least 10, better 25% +/- for the borders!). In case that the observed spot is a fragment of a larger protein (in this real example of a 66 kDa protein), you will not be able to identify it with many search engines because it will fall out of the limits. MASCOT uses a different approach. MASCOT has only one value which is used as a window sliding over all sequences in the database. This has the effect that no protein sequence is excluded a priori and the fragment would be identified as part of a larger protein (in fact, in the given example, the spot at 42 kDa was a N-terminal fragment of mouse serum albumine (66 kDa) appearing on a 2D gel of mouse liver cell lysate and could be clearly identified by ESI-MS/MS in conjunction with MASCOT search).

## 2.2 The use of combined MS/MS data sets for Database searches
To improve the confidence in the identification as well as to have a better chance to identify a protein at all, one can acquire several MS/MS spectra. While using classical PSD this is a tedious and time consuming process, MALDI-TOF/TOF with its fast data acquisition offers a convenient access to this and has therefore promoted it to a standard procedure especially in automatic data acquisition. With the data sets, one has two principal options: Search each data set independent or combine all data sets into one and search with this combined data set. The latter way has the advantage that it is faster, you don’t have to combine the results manually and the significance of the result is usually better. To do this, you open all the MS/MS spectra as well as the associated MS spectrum in BioTools in parallel. By clicking on “Combine multiple LIFT spectra” (Figure 42) under the File menu a window will open which allows the selection of MS/MS spectra to combine them with a MS spectrum into one single file (Figure 43). The number of MS/MS spectra is not limited, however one MS spectrum must always be in the selection.
Figure 42: The File menu with the “Combine multiple LIFT Spectra” option.

![File menu with Combine multiple LIFT Spectra option highlighted](image)

Figure 43: The “Combine spectra” window. In this case two MS/MS and one MS spectrum are available for combination.

![Combine spectra window](image)

After selecting the spectra and clicking “OK”, a new spectrum is generated. This appears as a stick spectrum in BioTools and looks like the MS spectrum of the combined data set. In parallel, this file is automatically saved to disk as a mgf (mascot generic format)-formatted textfile which can be used for BioTools-independent searches as well. For the newly created mgf-file we now have the MS as well as the MS/MS search option active in the BioTools icon bar as shown in Figure 44.
Figure 44: The combined data set as it appears in BioTools. Please also note that all three search options for BioTools’ MASCOT searches are now active, indicating that this data set contains MS as well as MS/MS data.

Running the search now with this combined data set should usually significantly improve the result of the MASCOT search as it is shown in Figure 45.

Figure 45: Search result for a combined MS/MS data set using several MS/MS spectra.

Please note that the search result comes up automatically with the peptide summary report of MASCOT. BioTools however, is hardcoded to the Protein summary report of MASCOT which might give you a different result in this case as shown in Figure 46. The difference results from the fact that for the scoring of the Protein view also MS data is taken into account while this is not the case for the Peptide view where only the MS/MS information is used for the scoring.
Figure 46: Comparison of Peptide summary report (left) and Protein summary report (right). Note the score difference of almost 240 units between the both results coming from the same data set. The search has been carried out with a combination of 1 MS and 21 MS/MS spectra and resulted in a combined score of nearly 1000 for this particular data set.

2.3 The use of XMass macros for data interpretation

Sometimes the search using uninterpreted MS/MS data does not lead to a result or the user wants to get more information from his samples in the form of amino acid sequence stretches etc. For those of you, which use the older Xmass software, this software provides you with two simple macros which can assist you in the interpretation of peptide MS/MS spectra. Lets assume we are starting from a MS/MS spectrum like the one shown in Figure 47:
The first step would be to look for peak pairs which add up to the precursor mass +1 Da ([M+2H]^+). These pairs are formed by one b-ion and one y-ion. This means that two signals forming a B-Y pair cannot belong to the same ion series. To get this pair information click on “Annotate B-Y Ions” in the AuraCmds menu of the Xmass menu bar. Subsequently, a window will pop up which allows you to specify a mass tolerance as well as a precursor mass (Figure 49). Reasonable settings for the mass tolerance depend on the type of spectrum. Usually 0.3-0.5 Da should be used for PSD-, 0.2-0.3 Da for LIFT-MS/MS spectra. After setting the parameters click on “OK”.

After the execution of the macro additional numbers should appear in the spectrum as shown in Figure 50. Those numbers indicate the pairs like pair number 3:

\[1064.37 + 561.15 = 1625.52\]
In case you observe an ion belonging to a pair with the mass 175.12 or 147.12 you can use this as a starting point for a y-ion series (175.12 is the $y_1$-ion of an Arg-terminated tryptic peptide, 147.12 the corresponding one of a Lys-terminated peptide). You can then search for ions belonging to pairs which have mass distances corresponding to amino acid masses.

**Figure 50:** Added number pairs after running the Aura command “Annotate B-Y ions” on the annotated raw spectrum. The command looks for b-/y-ion pairs which add up to the parent ions mass within the given mass accuracy limit. In this case, pair one corresponds to $b_{10}/y_{10}$, pair 2 to $b_9/y_9$, pair 3 to $b_8/y_8$, pair 4 to $b_7/y_5$ and pair 5 to $b_6/y_6$. Please note that in pair 4 the y-ion is lighter than the b-ion while in all other cases the situation is vice-a-versa. This pairing does not give you any information about certain fragment series.

To look for the mass differences, you can use a second macro within the **AuraCmds** menu called “Annotate Mass Diffs” (Figure 51). By clicking on the menu entry the “LabelDelta-Arguments” window comes up where you have to enter a mass tolerance, a distance mass list and to specify whether mass errors shall be displayed or not (Figure 52). The mass tolerance here can be smaller than in the “BYpairs arguments” window because the relative mass differences are usually smaller than the absolute mass errors, especially for LIFT-MS/MS spectra. In PSD this is not always the case due to the segmented nature of PSD spectra which does not allow a smooth calibration curve. For peptides, use the “aminomasses” mass list, there are also other mass lists available, e.g. for glycans or nucleic acids. These mass lists can
be found and edited to your particular needs in the directory “...\Xmass\prog\5.1\macros”. The mass errors should usually not be displayed since this is mainly confusing the display. However it might be helpful when you are using relatively large mass tolerances to identify sudden jumps in the mass tolerance indicating a wrong assignment especially in LIFT-MS/MS. In PSD mode, sudden jumps are to be expected at the segment borders.

![Figure 51](image1.png)  
**Figure 51:** Choose “Annotate Mass Diffs” in the AuraCmds menu

![Figure 52](image2.png)  
**Figure 52:** Enter the appropriate values in the window (e.g. 0.2-0.5 Da for PSD or 0.1-0.3 Da for LIFT spectra) and choose the mass list. Set “print mass errors” to “no”.

After clicking “OK” the macro is executed and results in a picture like the one shown in Figure 53. Potential amino acid mass differences within the defined mass error tolerances are displayed. The letters indicate the amino acid in single letter code corresponding to that particular distance while the arrows indicate the start and the end of the distance. L/I and K/Q are not differentiated since they can not be distinguished by PSD or LID-LIFT. Please note that these annotations can also be caused incidentally like in some cases in Figure 53 as well. The correct assignments in this case are shown in the red and green circles. The red circles indicate mass differences belonging to a y-ion series while the green indicate those belonging to a b-ion series. This would already allow us to assign the sequences [LI]GSF[LI]Y and GSF[LI] from the two fragment series to the peptide. As you can see in this case, the fragment series are overlapping, indicating that they are originating from a middle part of the sequence. Usually (but not always) fragment ions in the high mass part of the spectrum originate from y-ions while those in the lower mass part originate from b-ions. As already stated, 175.12 and 147.12 are good starting points for analyzing a y-ion series. b-ions can be analyzed by looking for a/b-ion pairs with a mass difference of 28 Da resulting from CO neutral loss. They are usually especially prominent for a2/b2 and a3/b3 (for the masses of b2-ions refer to the look-up Table 3 in Chapter 4). The middle part (300-700 Th) of the spectrum is often very crowded by the appearance of internal fragment ions resulting from double cleavages.
Figure 53: Amino acid difference annotation after using the AURA-command “Annotate MassDiffs”. The circled amino acids show correct annotations. The command simply annotates mass differences according to the distance mass list specified in the LabelDelta window. Therefore, in peptide MS/MS, no fragment ion series information can be obtained. However, one can often identify one or two consecutive series of amino acids, which correspond to y- (red) and b-ion series (green). The correct peptide sequence in this example is DAFLGSFLYEYSR, of which the underlined part can be determined from the figure.

2.4 FlexAnalysis macros for interpretation assistance

For FlexAnalysis, a new macro has been created which has essentially the same features like the “LabelDelta” macro from XMass. Its name is “MassListSearch” and it can be found on the disk associated with this manual. Basically, you enter the mass tolerance in ppm or Da, select the appropriate mass control file (in this case one containing the amino acid masses) and then click on calculate. Unlike in XMass, the differences are not displayed in the spectrum but only as a list as shown in the example below. The macro has to be copied into ...
\Methods\FlexAnalysisMacroModules and the mass control files into ...
\Methods\MassControlFiles.
2.5 The use of the MASCOT sequence search for complex database searches

Apart from MS and MS/MS searches, MASCOT offers a third option called „sequence search“ accessible via the „SEQ“ button in the BioTools menu bar. This option allows you to setup much more complex searches using all information gathered so far and also to include information you may have collected using other techniques (e.g. Edman sequencing). The principal syntax is like this:
1234.56 seq(...) ions(...) comp(...)

The first number corresponds to the precursor mass ([M+H]^+) followed by information about the sequence, the ions appearing in the spectrum and/or compositional information. For the exact options and syntax please refer to the MASCOT help available online on every MASCOT server (e.g. http://www.matrixscience.com). An example as the one from Figure 53 could look like this:

1624.56 seq(*-GSF[LI]) comp(*[Y]*[F]) ions(y-175.12)

This would correspond to a precursor ion of mass 1624.56 Da containing a sequence stretch GSF[L or I] in unknown direction (C→N-terminus or vice a versa), containing at least one Tyrosine and at least one Phenylalanine (this information might come e.g. from the observed immonium ions) and having a y1 fragment at mass 175.12 Da (please note that this would also be equivalent to the statement seq(c-R), seq(y-R) or comp(*[R]) in the MASCOT sequence search syntax). When Edman sequencing data are available, these could be included as seq(n-XYZ) indicating that this particular sequence stretch originates from the N-terminus of the peptide.
2.6 De-novo-sequencing

By far the most demanding task when dealing with MS/MS data is the de-novo sequencing attempt. This means to deduce more or less the complete amino acid sequence from the spectrum without any further knowledge.

In principal, there are two different approaches for this task:

- the “mass difference” approach and
- “real” de-novo sequencing

In chapter 2.3 you have seen how you can use macros to deduce sequence stretches from your spectra. A similar function is available within BioTools. You can find this in the Analyze menu as “De-novo sequencing” (note that this is not identical with “Full de-novo sequencing” available for users with a RapiDeNovo license which will be explained later in this chapter). Figure 57 shows the output of such a de-novo sequencing attempt.
Figure 57: Result of a the BioTools de-novo sequencing tool. The assigned upper sequence in this case is absolutely correct, however, the orientation is still unknown (actually only y-ions have been assigned and therefore the sequence has to be read backwards).

While this technique works quite well for in-source decay (ISD) (see chapter 3) spectra of intact proteins as well as for very simple spectra like those resulting from charge tagged peptides (see chapter 2.6.5.2 for this), usually MALDI-PSD or –LIFT-MS/MS are too complex to use this tool efficiently.

To overcome this problem, a new de-novo sequencing algorithm is available for BioTools (versions greater 2.2) which is based on a method from Ishikawa, Niwa and Sasagawa. The algorithm is invoked by clicking on the entry “Full de-novo sequencing” in the Analyze menu of BioTools (which is only active after entering the license key). For a detailed description please refer to the BioTools manual, which contains an extensive description as well as a tutorial for the use of the algorithm.

In the following, the procedure shall be described shortly and its use for the generation of MS-BLAST input files for subsequent homology search shall be described. Figure 58 shows the setup window of the algorithm which appears after pressing “Full de-novo sequencing” in its expanded view (after clicking on “Expand” in the lower left corner of the window.)
**Figure 58:** DeNovo settings window of the RapiDeNovo algorithm. The four lists on the top contain various modifications which can be set as fixed or variable modifications. The exact availability of specific modifications depends on the used BioTools modification list (see top line). As described in chapter 1.1, you can use BioTools' information to get an idea about some modifications. In the middle part, the experimental parameters are entered and in the lower part sequence information obtained from the spectrum (automatically by clicking on “Low mass ion info” or manually) or e.g. Edman data (as N-terminal hint) can be entered.

For the modifications, except if you don’t have other indications for their presence (see Chapter 2.1 for potential variable modifications, fixed modifications depend on the sample pretreatment) nothing should be entered for a first search. In the middle part the precursor mass is taken automatically from the spectrum information. The mass tolerances should be set according to the calibration and mass accuracy of your instrument (the more accurate, the better). The “Stringency of calculation” defines how fast sequences should be taken out of consideration, the higher the value, the more stringent the analysis. Reasonable values usually range from 2-10. By pressing the “Low mass ion info” button, an automatic analysis of the low mass ion region containing immonium ions as well as y1-ions is performed. According to the masses detected there, entries are automatically made in the fields “C-Terminal:”, “Absent amino acids:” and “Present amino acids:”. As described earlier, 175.12 Da and 147.12 Da are indicators for an Arg-y1 or Lys-y1 respectively of tryptic peptides. When nothing is detected but you know the cleavage reagent, you can use this information to give a hint here (e.g. “K
R” for trypsin, “E” for Endo-GluC/V8-Protease, “F W Y” for chymotrypsin or “M” for CNBr [attention: in this case also consider the appropriate C-terminal modification!]).

Automatic detection of absent and present amino acids works on the search for indicative immonium ions. Usually the aromatic amino acids (F, W, Y) and Histidine (110 Da) are very reliable (Caveat: when Histidine is detected because of a peak at 110 Th, this should be a really significant one, normally much more intensive than most other immonium ions and definitely more intensive than 112 Da). Critical candidates are L/I (86 Da) and V (72 Da), when these are not clear peaks, they should neither set present nor absent. See Figure 59 for examples.

Figure 59: Examples of low mass regions of various peptide MS/MS spectra. A shows an example where the 110 Da signal is a typical strong signal indicating the presence of Histidine while in B, albeit Histidine is present in the peptide, the signal is significantly smaller in intensity. C shows a peptide where you can clearly recognize the 86 Da signal indicating Leucine or Isoleucine. In D, albeit two Leucines are present, the signal can not be observed. E shows a spectrum clearly showing the 72 Da immonium ion of Valine while in F it is not visible despite Valine is present in the sequence.
Proline usually forms an intensive immonium ion at 70 Th, however, this mass can also originate from other amino acids and is therefore not used for the analysis. As you can see, a proper peak labeling is the key in case of a de-novo analysis. After all necessary settings have been made, click on “Calculate”. Depending on the speed of your processor and the complexity of the spectrum, after a while a window as displayed in Figure 60 pops up.

Figure 60: DeNovo result window.

The top part contains partial sequences identified by the algorithm, the lower part contains sequences which are “filled up” to the precursor mass of the peptide. Every sequence is assigned with a score for the quality of the assignment. This is composed e.g. by the length of sequence stretches which can explain observed ions. You can manually check these sequences by clicking on a specific sequence in the lower window and the assignment for this sequence will automatically be transferred into the spectrum and used for the sequence annotation. By browsing through the sequences (using up and down keys) you can easily visually inspect all the sequences and also zoom out and concentrate on dubious regions.

2.6.1 Evaluation of de-novo results

- Always be aware that the rules for interpretation of MS/MS spectra are based on human experience. What does this mean? The vast majority of de-novo sequenced peptides originates from tryptic digests. Therefore, general rules and spectral appearance are well known for these type of samples. Other might behave completely different! E.g. peptic
peptides usually might form intensive b-ion series, which can be misinterpreted as y-ion series which results in a wrong direction of a sequence readout.

- Look for residue specific fragmentation patterns: one should observe intensive fragment ion signals on the C-terminal side of Asp and on the N-terminal side of Pro (Figure 61), Gly-Gly usually only gives a very poor fragmentation yield, cleavages in between are hard to find.

![Image of a spectrum with annotated y-ion series]

**Figure 61:** Example of a peptide containing two acidic Asp as well as a Pro residue in the sequence. One can clearly see that intensive fragmentation takes place on the N-terminal side of Pro and on the C-terminal side of Asp (the annotated ion series is a y-ion series, therefore the sequence has to be read from right to left).

- Be careful with isobaric amino acids (I=L, K=Q, F=M\textsubscript{Ox}) and dipeptide combinations (TT=CV=DS=202 Da, for further ambiguous masses consult Table 5 in the tables section)

- Look for the immonium ions in your spectrum. Aromatic ions are quite reliable (F, Y, W). His must show a really intensive ion when present, a small signal might originate from other sources, a signal smaller than the 112 Da signal usually indicates absence of His. Val is sometimes hard to see (72 Da), Ile/Leu (86 Da) should normally be visible but might be not very intensive. Pro usually yields a very intensive signal at 70 Da.

- Look for a/b-ion combinations, they show a mass difference of 28 Da. This is especially a good way to locate and verify the b\textsubscript{2} ion in the 200-300 Da area. To identify the b\textsubscript{2}-ions you can use Table 3 in the tables section.
• look for the \( y_1 \) ions. When working with tryptic peptides, you will find either 175.12 Da (Arg-\( y_1 \)) or 147.12 Da (Lys-\( y_1 \)) (except for the C-terminal peptide). When you don’t observe any of those two, the terminus will most likely consist of Lys for tryptic peptides.

• look for neutral losses which indicate specific residues (-64 Da from the loss of CH\(_3\)SOH from oxidized Met, -98 Da from the loss of H\(_3\)PO\(_4\) from phosphorylated residues [primarily from Ser and Thr, Tyr is much more stable], -80 Da from the loss of SO\(_3\) from Tyr-OSO\(_3\)H – especially this can be very intensive, see Figure 28 for an example). You can also use these losses to localize the modified residue.

### 2.6.2 Strategy for doing a manual de-novo sequencing

• Calculation rules for ion masses:
  
  \[
  \begin{align*}
  b_1 &= AA + 1 \\
  b_{n-1} &= [M+H]^+ - 18 - AA \\
  y_1 &= AA + 19 \\
  y_{n-1} &= [M+H]^+ - AA
  \end{align*}
  \]

  where AA is always an amino acid residue mass

• In the first step you should inspect the low mass region of the spectrum for the identification of immonium ions to get some hints on the amino acid compositions. Amino acid immonium ion masses are given in Table 2. Characteristic ions originate from the aromatic amino acids and Leucine/Isoleucine. Also Proline and Histidine show very prominent immonium ions, but these have to be really intensive to clearly indicate those amino acids.

• Look in the low mass region for the \( b_2 \)-ion. You can use Table 3 to look for possible fragment ion masses in that region. Often the \( b_2 \)-ion is accompanied by the corresponding \( a_2 \)-ion, which is 28 Da less than the \( b_2 \)-ions. In FlexAnalysis or XMass you can use macros to look for this 28 Da difference. When you have identified the \( b_2 \)-ion, you can look in the high mass region for the \( y_{n-2} \) ion (\( y_{n-2} = [M+H]^+ - b_2 + 1 \)).

• Again, look in the low mass region. This time we are heading for the \( y_1 \)-ion. As described earlier, tryptic peptides usually (the only exception is the original protein C-terminus) show either a \( y_1 \)-ion at 147.12 Da (for Lysine terminated peptides) or at 175.12 Da (for Arginine terminated peptides). When a \( y_1 \)-ion can not be detected the peptide is either the C-terminal peptide or, more likely, terminated by Lysine which is usually not as
prominent as the Arg-y1-ion. After the identification of the y1-ion, you can use this to calculate the mass of the b_{n-1} ion in the high mass region (b_{n-1} = [M+H]^+ - y_1 + 1).

- Look in the high mass region for the y_{n-1} ion. After the identification of the b_2 ion, the number of possible amino acids is limited by the possible combinations from the b_2 ion.

- Since you have already identified y_n (= [M+H]^+), y_{n-1} and y_{n-2}, the next step is to extend the y-ion series towards the N-terminus. Check the fragment ion signals for mass differences corresponding to amino acid residual masses (Table 2) starting from the y_{n-2}-ion. You can also calculate the corresponding b-ion masses and check for those starting from b_2 towards higher masses.

- After the identification of the peptide sequence, calculate the peptide mass from the sum of the residual masses plus water plus a proton. This mass should be identical to [M+H]^+.

- Check the deduced amino acid sequence versus the observed immonium ions. When you identified Proline or Histidine, is there an intensive immonium ion indicating those amino acids? When you identified Asp or Pro in the sequence, are the neighboring fragment ions intensive? (refer to Figure 61)

- Try to identify the remaining ions in the spectrum. For neutral losses you can use Table 7 as a lookup table. For MALDI-MS/MS spectra, in the region between 250 and 700 Da there are usually many internal fragments originating from double cleavages. These are a bit more complicate to calculate because of the number of combinations, however their mass generally corresponds to (sum of residual masses + 1 Da). When you have identified a sequence for the peptide you can also use the tool MS-Product of the Protein Prospector suite (accessible via the WWW at http://prospector.ucsf.edu) to calculate the internal as well as all other possible sequence specific fragments.

- One simple way to check your result would be the following: when you identified an acidic residues in your peptide (Asp or Glu), you can selectively modify those by esterification using methanolic hydrochloric acid (this can be done by repeated lyophilization of the peptide/digest mixture with commercially available methanolic hydrochloric acid). The mass of the intact peptide should shift towards higher masses by this modification by (n+1)*15 Da (+1 because the C-terminus is modified as well). When doing MS/MS, every fragment containing Asp or Glu should also show the appropriate mass shift in comparison to the unmodified peptide (see Chapter 2.6.6.4).
2.6.3 LID versus CID for L/I differentiation

Sometimes you may face the problem that you have to make a definitive statement whether a certain position is a Leucine or an Isoleucine. This is usually the case when molecular biologists want to generate a so-called “degenerated” primer for fishing a certain gene out of a huge pool of DNA for sequencing or cloning of this gene. To reduce the complexity of this task (which is anyway very complex due to the high possible number of nucleic acid combinations), they usually want to know exactly which amino acids are present. In the past this has traditionally been done by Edman sequencing. With a TOF/TOF or PSD instrument, you can use high energy CID to differentiate between Ile and Leu. Lys and Gln can not be differentiated by CID! Figure 62 shows a comparison of a LID and a CID spectrum. You can clearly see the additional peaks appearing under CID conditions. These side chain fragmentations solely observed under CID conditions are called d- or w-ions, depending on their chemical structure.

![Figure 62: Comparison of LID- and CID-MS/MS data. You can clearly see the w-ions appearing only in the CID spectrum which allow the unequivocal assignment of the Leucines and Isoleucines on the respective positions.](image)

To get an idea, which side chain fragmentation-originating ions to expect, you should look into your spectrum and analyze which ions are present under LID conditions. When you observe intensive y-ions for the positions of interest, look for the corresponding w ions, when observing a and b ions, look for d as shown in Figure 63.
Figure 63: Examples of high energy derived side chain fragmentation. The upper panel shows w-ions originating from the corresponding y ions while the lower panel shows a d-ion originating from the corresponding a-ion.

A general comment on this: the side chain fragmentation usually appears under CID conditions, but this is not a must. Sometimes even under CID conditions, no side chain fragmentations are visible. A very nice peptide to observe side chain fragmentations (and also to check your CID conditions in general) is Substance P. The best way to clearly observe the differences is to used DHB as matrix since this will not give you too much PSD fragmentation interfering with the CID fragmentation.

2.6.4 Using MS-BLAST for the identification of proteins

A relatively new approach in comparison to the use of accurate MS/MS data for database
searches is the idea of using the BLAST homologous sequence search tool for protein identification using multiple de-novo sequenced peptides (see Shevchenko et al. for reference). The basic idea is to create multiple sequences which are not necessarily identical, but more or less similar to the actual peptide sequence for each of the peptide MS/MS spectra. This can be easily achieved by using the RapiDenovo-Tool within BioTools as shown in Figure 64. After sequences have been found, you can select each which seems to be reasonable to you and then click on “Copy/MS Blast”. This copies all the selected sequences into the clipboard in a MS-Blast compatible format (starting with a B, K and Q replaced by Z, L or I represented by L and all sequences separated by a dash). Subsequent clicking on “Open MS Blast” starts your web browser and directs him directly to the MS-Blast homepage at the European molecular biology laboratory (EMBL) in Heidelberg (Germany). By pressing Ctrl-V you can paste the sequences from the clipboard into the browser as shown in Figure 65.

Figure 64: BioTools’ de-novo result window. By clicking on “Copy/MS Blast” all selected sequences are copied into the clipboard in a MS-BLAST compatible format (see Figure 65) and subsequent clicking on “Open MS Blast” starts your internet browser and directs him to the MS-Blast homepage (http://dove.embl-heidelberg.de/Blast2/msblast.html) at the EMBL.
Figure 65: Example of an MS-BLAST input. Each block represents the results of an individual de-novo sequencing attempt on a distinct spectrum (so three spectra are displayed here). Please note that every sequence starts with a B, which represents a basic amino acid (K or R) due to the tryptic cleavage. Q and K are represented by Z while L and I are represented by L only.

The basic principle of MS-Blast is the following: the software searches for sequence stretches in the database which are homologous to your different entered short sequences. For normal BLAST searches you would normally need long sequence stretches to get meaningful results. MS-Blast overcomes this problem by using multiple sequences. For example, you will probably have one hit with a weak similarity to each of the sequences in the sequence set of peptide A. But when you look on three different peptides, the chance that you will find three weakly similar hits, one for each of the three peptides, all fitting to one common protein sequence, will be rather small. The better the fit is (which obviously also depends on the quality of the de-novo sequencing), the smaller will be the chance for a random hit. Also, taking more independent peptides into account will raise the chance to get useful results. Figure 66 gives an schematic illustration on the underlying principle of MS-Blast.

Figure 66: Schematic view of the function of MS-Blast.

A typical result is shown in Figure 67. You can see the hits being ordered in three blocks (and colors). The top red one gives you the hits, where the software assumes from the statistics that
these are very clear hits. The second, green one are the borderline hits while the black ones could potentially also be random hits. The green ones should always be checked manually whether they make sense or not and the same is of course true for the black ones as well. In the given example you can see that even the fourth (already a borderline) hit has the same protein name as the first three. However, you will also note a significant drop in the scoring values from above 700 to about 120. The reason is that the first hit is from mouse, the second one from Drosophila, the third from a fungus related to Aspergillus and the fourth from the slime mold Dictyostelium. By the way: the sample was from none of those mentioned, it originated from the African claw frog Xenopus. From this you can see that even very remotely related sequences of completely different organisms can be identified by this technique. Especially the power to identify proteins from organisms which have not been sequenced so far (and this is currently about 99.995% of all living beings) is the major strength of this method. You will have serious problems to get comparable results with database search engines using exact mass matches like MASCOT or ProFound/Sonar.

**Figure 67:** Example of a MS-BLAST search result. The red hits represent clear hits with significant similarities to the sequences used as input. Borderline hits are shown in green while critical hits which require manual inspection are shown in black.
2.6.5 Labelling techniques in MS/MS analysis

2.6.5.1 $^{18}$O-labelling

A relatively easy way to differentiate between fragment ions originating from the N-terminus and the C-terminus of a peptide is to introduce a stable isotope label in the form of Oxygen-18. This can be done very easy during digestion of an intact protein. When a proteolytic enzyme cleaves a protein into peptides, one water molecule is added to the newly formed peptide termini. This water is taken from the solvent. When the solvent consists of a mixture of water containing Oxygen-16 and Oxygen-18, this results in a statistical incorporation of both atom types into the peptide. By this statistical incorporation every peptide appears as a signal pair separated by 2 Da as shown in Figure 68.

![Figure 68: MALDI mass spectrum of a proteolytic peptide mixture after enzymatic digestion in a mixture of $^{16}$O and $^{18}$O water. The inset shows a zoomed view to illustrate the isotopic peak pattern with the signal pairs resulting from the incorporation of the isotopic mixture.](image)

The isotopic label is always incorporated within the newly formed carboxylic group at the C-terminus of the peptide (therefore there is always on unlabeled peptide in the mixture, which is the original C-terminal peptide of the protein). Upon fragmentation of these peptides, the resulting MS/MS spectrum will show two types of signals: single signals and signal pairs. The first class consists of fragment ions which do not contain the C-terminus (usually i-, a-, or b-type ions) while the latter one consists of fragment ions containing the C-terminus (usually y-
type ions). Figure 69a shows a typical spectrum obtained from an isotopically labeled peptide while in b the $y_1$-ion of a Arginine terminated tryptic peptide is shown, clearly demonstrating the incorporation of the $^{16}\text{O}/^{18}\text{O}$ mixture and showing the typical isotopic pattern.

Figure 69: A) MALDI-MS/MS spectrum of a tryptic peptide. You can see that the annotated pairs always belong to $y$-type ions, while singlet peaks are either $b$-type, $a$-type or internal fragment ions. B) zoomed view on the $y_1$-ion of a tryptic peptide with an Arginine at the C-terminus. You can clearly recognize the doublet structure with 2Da separation for the $^{16}\text{O}$ and the $^{18}\text{O}$ species. You can also recognize a small third peak resulting from a double incorporation of $^{18}\text{O}$ (from reentering of an already cleaved peptide into the active site of the enzyme and a second oxygen exchange by the enzyme). This effect is the more pronounced the longer the digestion takes place and can not be completely avoided.
2.6.5.1.1 Tryptic digest in $^{18}$O-water

The digest in H$_2^{18}$O is fairly simple. All steps are performed as usual, only the actual digestion is done in a separate buffer. This usually consists of a 1:1 mixture of H$_2^{16}$O and H$_2^{18}$O (by diluting enriched $^{18}$O water with e.g. Milli-Q water). The actual mixing ratio depends on the degree of enrichment of the $^{18}$O water (usually 70, 90, 95 or 99%, the latter two can be mixed just 1:1).

To perform an in-gel digest please lyophilize the gel plug to complete dryness before adding the $^{16}$O/$^{18}$O water containing buffer. Otherwise the actual $^{18}$O concentration will be lower than 50% resulting in different peak patterns. Also, when the digest shall be performed in solution, carry out all necessary preparation steps (reduction, alkylation, etc.) as usual, then desalt your protein (e.g. by dialysis or Zip-Tip) and lyophilize it to complete dryness before adding the digestion buffer.

Please note that this technique is not only available to tryptic digestion. As cleaving enzymes also chymotrypsine, endo-LysC, endo-GluC (V8-Protease), endo-AspN or endo-ArgC can be used. Carboxypeptidases could be used in principle as well, however with those you can get the sequence information also by running a digestion time course and looking at the masses of the resulting peptides.

2.6.5.1.2 Use of a FlexAnalysis macro for $^{18}$O-assisted de-novo sequencing

For the users of FlexAnalysis, a new macro has been created which is called “MassListSearch2Da”. The function is similar to that of “MassListSearch”. The major difference is that
the macro MassListSearch2Da first looks for pairs with 2 Da differences and in a second step only takes the first peak of those pairs into account for the search for sequences.

**Figure 71:** Result of the “MassListSearch2Da” macro. The red circles show the correct assignments starting from the $y_1$-ion indication Arginine as C-terminal residue. The C-terminal sequence is therefore ...GEVR.

Based on the evaluation of these data and further manual work, a sequence could be deduced from the spectrum within less than half an hour, which had been used for a MS-Blast homology search which led to the unequivocal identification of the sample as a eukaryotic initiation factor as shown in Figure 72.
Figure 72: BioTools ion series annotation of the sequence DDFQLIGIQDGYLSLNEDSGEVR from eukaryotic initiation factor 5A (eIF 5A). Again in this case you can clearly observe the typical high signal intensity on the C-terminal side of aspartic acid.

2.6.5.1.3 Sources for $^{18}$O-water

Cambridge stable isotopes, Andover (MA), USA
Promochem GmbH, Wesel, Germany
Sigma-Aldrich Chemie GmbH, Taufkirchen, Germany
Sigma-Aldrich Handels GmbH, Wien, Austria

2.6.5.2 Charge tagging (“CAF”)

Another way to simplify MS/MS spectra is to suppress specific ion types. One way to do this has been described by Keough and Lacey and is now commercially available as CAF (chemically assisted fragmentation) kit from Amersham Pharmacia Biotech. The basic technique is to introduce a charge tag on the N-terminus of a peptide which results in the suppression of the N-terminal fragment ions (b-type ions) and enhancement of the C-terminal y-type ions as shown in Figure 73.
Figure 73: Example of a N-terminally modified fibrinopeptide A (Sequence ADSGEGDFLAEGGGVR) with the “CAF” charge tagging kit (Amersham Pharmacia Biotech). The spectrum shows a clear series of y-type fragment ions which is very easy to interpret.

Figure 74 shows the basic principle of the reaction. The charge tag usually consists of a highly acidic group with a pKₐ-value below 3 (e.g. a sulfonic acid) and a reactive group targeting to free amino groups. The commercially available reagent is NHS activated Propane-1-carboxylic acid-3-sulfonic acid. This reagent can be used in aqueous solutions. Instead of this, alternative (commercially available) reagents are e.g. cyclo-2-Sulfobenzoic acid anhydride or Chlorsulfonyl-acetic acid chloride. Both reagents also introduce a negative charge at the N-terminus.

One major problem with this technique is the fact that the major side reaction is the modification of lysine side chains. This results in negative charge tags on both (N-terminal and C-terminal) sides of a tryptic peptide which yields to an almost complete loss of all fragment ions in MS/MS. One way to overcome this is, to introduce a modification on the lysine side chain before modifying the peptide N-terminus. This can be done selectively by
guanylation of the lysine side chain, using either N-methylisourea (the sulfate is more expensive than the hemisulfate but better suited) or pyrazole-1-carboxamidine hydrochloride, converting the lysine into a homoarginine. Apart from the protection against the charge tag, this also increases the gas phase basicity of the former lysine side chain resulting in a more efficient desorption/ionization of the peptide as well as a better stabilization of the charge at the C-terminal side chain. Together with the highly negative tag at the N-terminus this results in MS/MS spectra which show almost exclusively y-type ions as shown in Figure 73.

For detailed descriptions of the modification reactions see:

Charge Tagging:
Keough, T.W., Youngquist, R.S., Lacey, M.P., Anal Chem. 2003;75(7):156A-165A

Guanylation:

Also refer to: Bruker Daltonics Application note MT-53 “N-terminal Derivatization of Tryptic Peptides for de-novo Sequencing by PSD: A Powerful New Tool in Proteomics”

2.6.6 Identification of posttranslational modifications by MS/MS

De-novo sequencing of peptides becomes even more complicated than it is already when modifications of the peptide have to be taken into account. These modifications can be divided in several subgroups, e.g. those which are due to biochemical processes (e.g. phosphorylation) or those which are due to preparation artifacts (e.g. oxidation). Another differentiation can be made by dividing them into modifications, which can not be easily
detected by MS/MS or those which can be detected by MS/MS. Examples for the first group are cystein alkylations, which can only be identified by the mass difference of the corresponding modified amino acid in the MS/MS ion series. The following chapter will deal with those modifications which will give you hints upon their presence due to the appearance of the MS/MS spectrum, e.g. by characteristic neutral loss signals. The modifications will by discussed roughly in the order of likeliness of their appearance and also some information on their origin and biological role will be mentioned.

2.6.6.1 Methionine oxidation

Oxidation of methionine residues is probably the most common artifical modification appearing in MALDI mass spectra. It can usually always be considered as a partial modification in PMF database searches. The oxidation takes places in solution, e.g. during pipetting or sample preparation and can have various extends which also depends on the particular peptide sequence. There are various known factors effecting the extend of oxidations in your samples:

a) the target surface seems to have an influence. Usually, we are observing more oxidation on Brukers’s AnchorChip targets than on conventional steel targets. This is especially very well known for the peptide Bombesin,

b) the sample composition has some influence on this process. Scavengers like n-octylglucopyranoside (a MALDI compatible, sugar-based detergent) can almost completely prevent the formation of oxidations, even on AnchorChip targets,

c) Studies have also shown that the composition of the athmosphere has a significant influence on this process. It has been reported that the extend of oxidation products correlated with the activity on a photocopier machine next to the labs door (probably due to higher ozone levels in the laboratory air).

Usually, methionine oxidation can be easily recognized in the MS/MS spectrum by the neutral loss of CH$_3$SOH from the precursor, which is in many cases the base peak of the spectrum. Figure 75 shows this neutral loss:
2.6.6.2 Serine- and Threonine-phosphorylation

Phosphorylation is the biochemically most important process for the transmission of information. Molecular switches are set putting adding or removing a phosphate group to certain serine or threonine residues (also to tyrosine, however, this is chemically much more stable and does not show a proper neutral loss in MS/MS). Usually, biological phosphorylation levels are in the range of 1-5%. However, there are also cases, where these levels are significantly higher. This means, that without taking precautions or especially looking for those modifications, you will usually not easily detect them. Analysis of phosphorylations is also not only tricky due to the low amounts of available modified material, but also to its „sticky“ behaviour to metal, especially iron, surfaces. On the other hand, Fe-IMAC or Ga-IMAC chromatography is one way how these peptides can be selectively enriched for further analysis as shown in Figure 76.

Figure 75: Neutral loss of CH₃SOH (-64 Da) from a methionine oxidized peptide.
Figure 76: MALDI mass spectra of bovine β-Casein (A) after a tryptic digest without further processing and (B) after isolation of the phosphopeptides by chromatography using Fe$^{3+}$ loaded Millipore ZipTip MC$^{	ext{TM}}$. As you can see in the lower panel, the two peptides, singly phosphorylated at mass 2061 Da and quadruply phosphorylated at mass 3122 Da could be selectively enriched by this technique. The signal at 1977 Da corresponds to the metastable neutral loss of H$_3$PO$_4$.

Another feature, which can be observed in spectra with good intensities is a series of neutral losses from the b-ion series (b-phosphate, usually shortened as b-Pi). This is available in the BioTools annotation options. shows an example of the above mentioned 2061 Da peptide from Casein, which nicely demonstrates the appearance of this series.
2.6.6.3 N-terminal acetylation

A relatively common modification is the acetylation of the N-terminus of a peptide. Often you can already easily recognize them from the MS spectrum since N-acetylated peptides are often significantly more intensive than the other peptides in the spectrum. The origin of the acetylation is from the peptide synthesis in the organism. Usually, N-formyl- or N-acetylmethionin is used to initiate the peptide synthesis. Often, this is removed after the synthesis of the complete peptide, but sometimes it remains on the protein, especially when the organism has better things to do like in phases of hunger, heat shock or something else. It may also occur that the initial Methionine is cleaved of and the second residue is acetylated, in this case a formylation is usually never observed. shows a MS/MS spectrum of a pure, N-acetylated peptide, showing the neutral loss of the acetyl group indicating the acetylation of the N-terminus.

Figure 77: MS/MS spectrum of the peptide FQpSEEQQTEDLQDK from bovine β-Casein.
Figure 78: MALDI-MS/MS spectrum of Nα-AcRRPYIL. The neutral loss of the actetyl group (-42 Da) clearly indicates the acetylation of this peptide. Due to the fact that y₅ shows (in comparison to y₆/the precursor ion) no corresponding loss, the modification can be easily assigned to the N-terminus of the peptide.

The BioTools annotation in Figure 78 has been generated using

[CID-MSTOOL: Acetyl-peptides]
Series 1=a,-COOH,N,
Series 2=b,-OH,N,
Series 3=y-Ac,-CHCO,C,
Series 4=y,+H,C,
Series 5=i,-COOH,I,
Series 6=b+18,+H,N,

in the CID-Tools section of BioTools param.dat file.

- 1-2 reale Beispiele MS (intensives, nicht zugeordnetes Fragment) und MS/MS des acetylierten N-Terminus

2.6.6.4 Methylation of acidic moieties

Methylation of acidic moieties can occur for several different reasons. The two most common ones are staining artifacts, especially in colloidal Coomassie staining solutions like the one given in the recipes chapter, or due to modification of these moieties on purpose. This can be done for several reasons:

a) to enhance the specificity to discriminate acidic peptides in IMAC mediated isolation of phosphorylated peptides or

b) to get an idea about the number of acidic moieties in a particular peptide. Each acidic residue adds one methyl group (+CH₃-H resulting in a 14 Da mass shift) to the sum formula of a peptide.
Shows an example of a peptide which had been modified to count the number of acidic groups. In total, 2 methyl groups had been added to the peptide by the modification using methanolic hydrochloric acid, which brings us the information that one acidic residue is within the peptide (the second methyl group is attached to the peptides C-terminus).

Figure 79: LIFT spectra of the peptides LGEYGFQNALIVR (A) and LGE(OMe)YGFQNALIVR-OMe (B). The arrows indicate the 14 Da shifts for the individual fragment ions containing the acidic moieties. Please note especially the shift of the y$_{11}$-ion. The b$_2$ ion has not been shifted and the y$_{13}$ would be shifted by 28 Da, but is not visible in the spectrum.
2.6.6.5 N-methylation of Lysine side chains

N-methylation is a modification occurring mainly in two cases: as a natural modification especially in the basic histone protein family to enhance basicity of the side chains for a stronger interaction with the DNA (histones are proteins which play a role in the structural organization of the DNA in the chromosomes) or, alternatively, as an artificial modification in trypsines to enhance their stability towards autoproteolysis. **Figure 80** shows an example of the latter kind.

![MALDI-MS/MS spectrum of the peptide LGEHNIDVLEGNEQFINAAK(Me2). This peptide originates from modified trypsines e.g. the Promega trypsine. You can clearly see the neutral losses of the two methyl groups from the y-ions while the b-ion series remains unchanged.](image)

Methylation can occur as singly, doubly and triple modification (the last one generating a charged, quarternary ammonium ion). Twofold methylated Lysine can be distinguished from Arginine (residual masses 158.1419 Da versus 156.1011 Da) by its neutral loss of the methyl groups from the side chain as shown in **Figure 80**. The annotation in BioTools can be done by adding the section

```
[CID-MSTOOL: LysCH3-peptides]
Series 1=a,-COOH,N,
Series 2=b,-OH,N,
Series 3=y-CH3,-CH,C,
Series 4=y-(CH3)2,-C2H3,C,
Series 5=y,+H,C,
Series 6=i,-COOH,I,
```
into your BioTools param.dat file.

### 2.6.7 A practical example

This chapter shall give you an idea on how to handle an unknown sample in the course of an analysis. The example is a practical example, which ideally illustrates which steps you can perform to achieve a result on an unknown, tricky sample.

The sample has been sent to us by a customer to evaluate the machines performance. The sample was a tryptic digest of a protein dissolved in a solution of 6M urea, sodium phosphate, sodium chloride and potassium phosphate. Since the same customer sent a couple more samples and all of them turned out to be either cytochrome C or BSA, we assumed that it would be either of them as well.

From the sample a peptide mass fingerprint as well as four different MS/MS spectra had been acquired (the latter ones are shown in Figure 81).

![Figure 81: Four LIFT-MS/MS acquired from the sample after fingerprint analysis.](image)

Searches performed independently and also with the combined MS and MS/MS spectra gave no results as shown in Figure 82. Also searches with less restriction regarding species, more missed cleavage sites, etc. were unsuccessful.
Due to the fact that no result had been obtained by simple searches, one of the MS/MS spectra of a precursor with mass 2159 Da was chosen for a de-novo-sequencing attempt. As you can see from Figure 83, this resulted in a nice, high scoring sequence tag.

The potential sequences were selected, copied to the clipboard and used for a MS-Blast search. As shown in Figure 84, this resulted in an excellent hit with identical sequence. As already assumed, the protein turned out to be most likely BSA.
Figure 84: Result of a MS-Blast search using the de-novo results of the above shown sequence. The result shows a clear hit (red) and an excellent fit between the query and the sequence of bovine serum albumine.

The thereby identified sequence tag from BSA was then used to locate its position in the BSA’s original, intact sequence to get an idea why it could not be identified using simple fingerprint of MS/MS search.

Figure 85: Sequence search in BioTools for the sequence tag which had been identified to belong to BSA by MS-Blast.

A potential tryptic peptide sequence containing the tag as well as one missed cleavage site was then copied in a separate window of the BioTools sequence editor. Its molecular mass turned out to be 43 Da less than that of the analyzed precursor. By looking into the “DeltaMass”-list, a potential candidate for 43 Da is carbamylation. This can happen by heating up proteins in solutions containing high urea concentrations. The modification side is the amino terminus or lysine side chains.
Figure 86: A closer examination of the sequence reveals that the tryptic peptide containing the identified sequence has a molecular weight of 43 Da less than the observed one. By looking into the "DeltaMass"-list (http://www.abrf.org/index.cfm/dm.home) reveals carbamylation as a potential modification, an artificial modification well known to be formed in high concentration urea solution especially upon heating.

Since BioTools can handle optional modifications, both were entered as such and the sequence was submitted from the sequence editor into BioTools. During the submission process, the sequence editor asks for the precursor mass to calculate the minimum and maximum number of potential variable modifications. Since this turned out to be one, two sequences were sent to BioTools, each containing one modification at different sites.
Figure 87: Comparison of the two potential carbamylation sites (N-terminal or Lysine side chain) clearly shows that the carbamylation had occurred at the Lysine side chain. This also indicates that the solution has been heated up before the tryptic digestion since otherwise the modification would also have taken place at the N-terminus of the peptide.

By switching between the two sequences and their respective annotations (as well as by simply looking on the BioTools score for the annotation), it was easy to distinguish between the correct assignment of an carbamylation on the lysine side chain at position 12 and the incorrect assignment of a carbamylation at the amino terminus of the peptide. Finally, a correct, fully annotated sequence could be obtained for the BSA peptide including the clear identification of the modification sites.
Figure 88: Automatic annotation by BioTools of the now completely identified peptide sequence ETYGDMADC*CKQEPER. The underlined part indicates the sequence which has been identified by the automated de-novo sequencing.

Figure 89: General recommended workflow for the analysis of MS/MS data. The green boxes are tasks which can be performed using the BioTools software, while the yellow boxes show tasks which involve other search engines or databases. Generally, you should follow this procedure stepwise from the easiest to the most difficult tasks. A comment to pacify your mind: you will hardly encounter occasions where you will get 100% sequence coverage (I personally had one case so far in eight years which involved combination of MALDI and ESI as ionization techniques). Therefore, having gaps or unclear regions in your sequences is absolutely normal. This is especially the case when you are working with so far unsequenced genomes and have to rely on “similar” entries in the databases.
3 IN-SOURCE DECAY (ISD)

In-source decay is a term introduced by Brown and Lennon in 1995 and the process can be used to acquire MS/MS information on intact, undigested proteins. While these studies had been performed in linear mode, Suckau and Cornett later observed the same phenomenon in reflector spectra as well. In ISD mode, the molecular ion undergoes spontaneous fragmentation within the source region (see Figure 2) and typically produces fragment ions of the c- and y-type series, slightly depending on the matrix used. The ions can be observed typically between 1000 and more than 4000 Da, provided that the sequence contains no disulfide bridges within this region. Therefore, reduction and alkylation is beneficial to improve the sequence coverage in ISD-MS/MS. Also proline residues will produce short gaps in the ion series. In reflector mode, isotopic resolution and mass errors less than 0.1 Da can be expected within this mass range, which allows to obtain sequence information close to the N-terminus even when the N-terminus is blocked and therefore not amenable to Edman sequencing.

3.1 Acquisition of ISD spectra

To acquire protein ISD spectra, two major preconditions must be fulfilled:

a) the sample has to be a more or less purified protein (>80%). This is because ISD is only a pseudo-MS/MS technique. Everything within the sample undergoes fragmentation in the source region. Due to this fact, there is no way for a preselection of a particular precursor and this means that mixtures can not be analyzed by this technique.

b) the sample amount has to be in the range of 10-20 pmol, ideally per µl. If possible, the acetonitrile content should be below 20%.

To prepare the sample, a so called double layer preparation is used. For this, a thin layer of sinapinic acid is prepared from a saturated solution in ethanol. On this thin layer, 0.5-1 µl of a 1:1 mixture of the sample with a saturated solution of sinapinic acid in 0.1% TFA/ 30% acetonitrile is pipetted and allowed to dry at ambient air.

For the spectra acquisition choose a parameter set using an acceleration voltage of 25 kV and a delay time of around 200 ns should be used. To achieve maximum sensitivity, the detector should be set to 1.8-1.9 kV and 100 mV digitizer sensitivity should be used. Deflection should be set to 800. Start with the analysis by acquiring 50 laser shots and increase laser power until you see fragment peaks appearing in the mass range between 1000 and 3500. Usually you have to acquire several hundred laser shots to achieve a sufficient S/N. The calibration can be performed using a normal peptide standard mixture. Figure 90 shows an example of the 12
kDa protein Thioredoxin. In the right hand part you can see the Precursor ion (broad saturated peak), some PSD fragment ions (unresolved metastable peaks between 9000 and 12000 Da) as well as the doubly charged precursor ion at mass 5800 Da. The lower panel shows the zoomed view of the region between 1000 and 4500 Da which contains the ISD fragment ions. As you can see from the annotation, mainly c- and y-ions are observed.

Figure 90: Example of ISD spectra of the protein Thioredoxin.
3.2 Interpretation of ISD spectra

Interpretation of ISD spectra is insofar easier than LID or CID interpretation, that almost exclusively only one or two ion series are appearing which are completely isotopically resolved and have very good mass accuracies. This makes the readout of ion series a rather straightforward task and this can be done using the “DeNovo Sequencing” option in BioTools Analysis menu. For details refer to the section I.4 in the BioTools manual which explains the analysis of samples with known sequence as well as unknown sequence.

Using this technique, detailed information about the terminal parts of an intact protein can be obtained as shown in Figure 91. This example shows a recombinant protein which has been fused to a short sequence of another protein to promote expression in the host organism, which is a typical case in molecular biology and recombinant protein production.
Figure 91: Example of an ISD spectrum of recombinant protein A produced in E. coli. In the spectrum two sequence stretches are encircled. As you can see in the lower part, one can identify two different proteins based on these two short sequence stretches. The first is E. coli β-glucuronidase while the second one is the recombinant protein A from S. aureus. This has been unexpected, but is typical for recombinant proteins. The reason is the following: E. coli does not always want to produce foreign proteins. One way to convince the organism to do so is to make him think he produces one of his own. This has been done here. The cloning has been done in a way that the molecular biologist put the start of an E. coli housekeeping protein into the vector and then proceeded with the foreign sequence. The effect is that E. coli starts with the protein synthesis and when it recognizes that it synthesizes or foreign protein, it’s already too late to stop.

3.3 Terminus specific fragmentation (T³)

A rather new technique which had been presented as a combination of ISD and PSD for the first time on an ASMS poster in 1997 (D.S. Cornett, “Prompt In-source Fragmentation of Biomolecules in a Gridless Reflectron-TOF”) and has been reinvented under the term T³ (Terminus TOF/TOF) by D. Suckau and A. Resemann as a combination of ISD and LID in a TOF/TOF instrument. Figure 92 illustrates the general principle. First, by ISD fragment ions originating from the N- or C-terminal region are formed. In a second step, one of these fragment ions is selected, undergoes metastable PSD fragmentation, the precursor and the fragment are further accelerated in the LIFT unit and a complete MS² spectrum of the fragment ion is obtained. Since the first isolation step is not by MS but by chemical or biochemical isolation to obtain a pure precursor, this technique is a pseudo-MS³ technique.
Figure 92: Scheme of the principle operation of T3-sequencing. By ISD N- and C-terminal fragment ions (usually c- and y-type) are obtained, which are undergoing further metastable fragmentation giving the usually observed ion series which are also obtained by regular LID spectra.

The appearance of the spectra acquired this way is almost identical to MS/MS spectra from classically formed peptides. However, depending on the ion type used for isolation and fragmentation, certain modification might have to be considered:

a) a-type fragment ions have a loss of COOH in comparison to the unmodified peptide,
b) c-type fragment ions behave as peptide amides and
c) y-type fragment ions are identical to normal, unmodified peptides.

When you take these modifications into account, analysis is relatively easy. For automated database searches, c- and y-type ions can be searched without any changes just by allowing amidation on the C-terminus as a variable modification. For the a-type ions, the modification has to be set up on the MASCOT server (this is only possible on in-house servers) by adding the following line into the file "mod_file" in the directory ..\inetpub\MASCOT\config on your local MASCOT server:

* 
Title: ISD a-ion 
Cterm: -43.9898 -44.0098 
* 

Figure 93 shows an example of such a T3-analysis of RNAse B. The ISD spectrum has been acquired from DHB as matrix. From the ISD spectrum, three precursors had been selected for subsequent MS3 experiments of which one was a c-, one a y- and one an a-ion (the spectra of
the first and the last are shown). A database search has been performed using all three MS/MS spectra as a combined data set (see chapter 2.2) and allowing the modifications “Amide (C-term)” as well as the above described “ISD a-ion”. A charming effect of this type of search is that you don’t have to know which of your fragments belongs to which series since this will be recognized by the search engine automatically due to the fitting modification.

**Figure 93:** TSF spectra (ISD-LID) of RNase B in DHB: a) c_{14} and b) a_{12}; upper panel with the Mascot result of the TSF-a_{12} spectrum, lower panel with the result from the search of the combined TSF spectra of a_{12}, c_{14} and y_{12}. c) Sequence coverage map of RNase B precursor resulting from the combined search, identifying the processed N-terminus and the proper C-terminus; upper row of red bricks indicates the matching b-, lower row the matching y-ions in the three peptides. d) LID spectrum of synthetic N-terminal 14mer-amide in HCCA, structure identical to c_{14}.  

89
4 APPENDIX

4.1 Useful tables

Table 2: Amino acid residue masses (important immonium ions are written boldfaced).

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>Single letter code</th>
<th>Residue mass (Da)</th>
<th>Immonium ions (Da)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alanine</td>
<td>A</td>
<td>71.04</td>
<td>44</td>
</tr>
<tr>
<td>Arginine</td>
<td>R</td>
<td>156.10</td>
<td>129, 175 (y1)</td>
</tr>
<tr>
<td>Asparagine</td>
<td>N</td>
<td>114.04</td>
<td>87</td>
</tr>
<tr>
<td>Aspartic acid</td>
<td>D</td>
<td>115.03</td>
<td>88</td>
</tr>
<tr>
<td>Carbamidomethylcysteine C</td>
<td>*</td>
<td>160.03</td>
<td>133</td>
</tr>
<tr>
<td>Carboxamidomethylcysteine</td>
<td></td>
<td>161.03</td>
<td>134</td>
</tr>
<tr>
<td>Cysteine</td>
<td>C</td>
<td>103.01</td>
<td>76</td>
</tr>
<tr>
<td>Glutamic acid</td>
<td>E</td>
<td>129.04</td>
<td>101</td>
</tr>
<tr>
<td>Glutamine</td>
<td>Q</td>
<td>128.06</td>
<td>102</td>
</tr>
<tr>
<td>Glycine</td>
<td>G</td>
<td>57.02</td>
<td>30</td>
</tr>
<tr>
<td>Histidine</td>
<td>H</td>
<td>137.06</td>
<td>110</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>I</td>
<td>113.08</td>
<td>86</td>
</tr>
<tr>
<td>Leucine</td>
<td>L</td>
<td>113.08</td>
<td>86</td>
</tr>
<tr>
<td>Lysine</td>
<td>K</td>
<td>128.09</td>
<td>101, 147 (y1)</td>
</tr>
<tr>
<td>Methionine</td>
<td>M</td>
<td>131.04</td>
<td>104</td>
</tr>
<tr>
<td>Oxidized Methionine M_{Ox}</td>
<td></td>
<td>147.04</td>
<td>120</td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>F</td>
<td>147.07</td>
<td>120</td>
</tr>
<tr>
<td>Proline</td>
<td>P</td>
<td>97.05</td>
<td>70</td>
</tr>
<tr>
<td>Propionamidocysteine C^a</td>
<td></td>
<td>174.04</td>
<td>147</td>
</tr>
<tr>
<td>Serine</td>
<td>S</td>
<td>87.03</td>
<td>60</td>
</tr>
<tr>
<td>Threonine</td>
<td>T</td>
<td>101.05</td>
<td>74</td>
</tr>
<tr>
<td>Tryptophane</td>
<td>W</td>
<td>186.08</td>
<td>159</td>
</tr>
<tr>
<td>Oxidized Tryptophane</td>
<td></td>
<td>218.08</td>
<td>191</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>Y</td>
<td>163.06</td>
<td>136</td>
</tr>
<tr>
<td>Valine</td>
<td>V</td>
<td>99.07</td>
<td>72</td>
</tr>
</tbody>
</table>

Table 3: b₂-ion masses lookup table (refer to Table 2 for modified amino acid residues). Dipeptide masses can be calculated by subtracting 1 Da (see calculation rules in Chapter 2.6.2).

<table>
<thead>
<tr>
<th></th>
<th>G</th>
<th>A</th>
<th>S</th>
<th>P</th>
<th>V</th>
<th>T</th>
<th>C</th>
<th>I/I</th>
<th>N</th>
<th>D</th>
<th>Q/K</th>
<th>E</th>
<th>M</th>
<th>H</th>
<th>M_{Ox}/F</th>
<th>R</th>
<th>C^a</th>
<th>Y</th>
<th>C^a</th>
<th>W</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>115</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>129</td>
<td>143</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>145</td>
<td>159</td>
<td>175</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>155</td>
<td>169</td>
<td>185</td>
<td>195</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>157</td>
<td>171</td>
<td>187</td>
<td>197</td>
<td>199</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>159</td>
<td>173</td>
<td>189</td>
<td>199</td>
<td>201</td>
<td>203</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>161</td>
<td>175</td>
<td>191</td>
<td>201</td>
<td>203</td>
<td>205</td>
<td>207</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L/I</td>
<td>171</td>
<td>185</td>
<td>201</td>
<td>211</td>
<td>213</td>
<td>215</td>
<td>217</td>
<td>227</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>172</td>
<td>186</td>
<td>202</td>
<td>212</td>
<td>214</td>
<td>216</td>
<td>218</td>
<td>228</td>
<td>229</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>173</td>
<td>187</td>
<td>203</td>
<td>213</td>
<td>215</td>
<td>217</td>
<td>219</td>
<td>229</td>
<td>230</td>
<td>231</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q/K</td>
<td>186</td>
<td>200</td>
<td>216</td>
<td>226</td>
<td>228</td>
<td>230</td>
<td>232</td>
<td>242</td>
<td>243</td>
<td>244</td>
<td>245</td>
<td>257</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>187</td>
<td>201</td>
<td>217</td>
<td>227</td>
<td>229</td>
<td>231</td>
<td>233</td>
<td>243</td>
<td>244</td>
<td>245</td>
<td>246</td>
<td>258</td>
<td>259</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>189</td>
<td>203</td>
<td>219</td>
<td>229</td>
<td>231</td>
<td>233</td>
<td>235</td>
<td>245</td>
<td>246</td>
<td>247</td>
<td>248</td>
<td>260</td>
<td>261</td>
<td>263</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>195</td>
<td>209</td>
<td>225</td>
<td>235</td>
<td>237</td>
<td>239</td>
<td>241</td>
<td>251</td>
<td>252</td>
<td>253</td>
<td>266</td>
<td>267</td>
<td>269</td>
<td>275</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M_{Ox}/F</td>
<td>205</td>
<td>219</td>
<td>235</td>
<td>245</td>
<td>247</td>
<td>249</td>
<td>251</td>
<td>261</td>
<td>262</td>
<td>263</td>
<td>276</td>
<td>277</td>
<td>279</td>
<td>285</td>
<td>295</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>214</td>
<td>228</td>
<td>244</td>
<td>254</td>
<td>256</td>
<td>258</td>
<td>260</td>
<td>270</td>
<td>271</td>
<td>285</td>
<td>288</td>
<td>294</td>
<td>304</td>
<td>313</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C^a</td>
<td>218</td>
<td>232</td>
<td>248</td>
<td>258</td>
<td>260</td>
<td>262</td>
<td>264</td>
<td>274</td>
<td>275</td>
<td>276</td>
<td>289</td>
<td>290</td>
<td>292</td>
<td>298</td>
<td>308</td>
<td>321</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y</td>
<td>221</td>
<td>235</td>
<td>251</td>
<td>261</td>
<td>263</td>
<td>265</td>
<td>267</td>
<td>277</td>
<td>278</td>
<td>279</td>
<td>292</td>
<td>293</td>
<td>295</td>
<td>311</td>
<td>320</td>
<td>324</td>
<td>327</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C^a</td>
<td>232</td>
<td>246</td>
<td>262</td>
<td>272</td>
<td>274</td>
<td>276</td>
<td>278</td>
<td>288</td>
<td>289</td>
<td>290</td>
<td>303</td>
<td>304</td>
<td>306</td>
<td>312</td>
<td>322</td>
<td>331</td>
<td>335</td>
<td>338</td>
<td>350</td>
<td>361</td>
</tr>
<tr>
<td>W</td>
<td>244</td>
<td>258</td>
<td>274</td>
<td>284</td>
<td>286</td>
<td>288</td>
<td>290</td>
<td>300</td>
<td>301</td>
<td>302</td>
<td>315</td>
<td>316</td>
<td>318</td>
<td>324</td>
<td>334</td>
<td>343</td>
<td>347</td>
<td>350</td>
<td>361</td>
<td>373</td>
</tr>
</tbody>
</table>
Table 4: Lookup table for $y_2$-ions depending on the terminal residue for tryptic peptides.

<table>
<thead>
<tr>
<th>Residue</th>
<th>$y_1$=K</th>
<th>$y_1$=R</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>204.141</td>
<td>232.141</td>
</tr>
<tr>
<td>A</td>
<td>218.157</td>
<td>246.157</td>
</tr>
<tr>
<td>S</td>
<td>234.152</td>
<td>262.152</td>
</tr>
<tr>
<td>P</td>
<td>244.173</td>
<td>272.173</td>
</tr>
<tr>
<td>V</td>
<td>246.188</td>
<td>274.188</td>
</tr>
<tr>
<td>T</td>
<td>248.168</td>
<td>276.168</td>
</tr>
<tr>
<td>C</td>
<td>250.129</td>
<td>278.129</td>
</tr>
<tr>
<td>I</td>
<td>260.204</td>
<td>288.204</td>
</tr>
<tr>
<td>L</td>
<td>260.204</td>
<td>288.204</td>
</tr>
<tr>
<td>N</td>
<td>261.163</td>
<td>289.163</td>
</tr>
<tr>
<td>D</td>
<td>262.147</td>
<td>290.147</td>
</tr>
<tr>
<td>Q</td>
<td>275.178</td>
<td>303.178</td>
</tr>
<tr>
<td>K</td>
<td>275.215</td>
<td>303.215</td>
</tr>
<tr>
<td>E</td>
<td>276.163</td>
<td>304.163</td>
</tr>
<tr>
<td>M</td>
<td>278.160</td>
<td>306.160</td>
</tr>
<tr>
<td>H</td>
<td>284.179</td>
<td>312.179</td>
</tr>
<tr>
<td>M$_{Ox}$</td>
<td>294.160</td>
<td>322.160</td>
</tr>
<tr>
<td>F</td>
<td>294.188</td>
<td>322.188</td>
</tr>
<tr>
<td>R</td>
<td>303.221</td>
<td>331.221</td>
</tr>
<tr>
<td>Y</td>
<td>310.183</td>
<td>338.183</td>
</tr>
<tr>
<td>W</td>
<td>333.199</td>
<td>361.199</td>
</tr>
</tbody>
</table>

Table 5: Amino acid combinations which equal to single amino acid masses

<table>
<thead>
<tr>
<th>Amino acid combination</th>
<th>Residue mass</th>
<th>Single amino acid</th>
</tr>
</thead>
<tbody>
<tr>
<td>GG</td>
<td>114</td>
<td>N</td>
</tr>
<tr>
<td>GA</td>
<td>128</td>
<td>K, Q</td>
</tr>
<tr>
<td>GV</td>
<td>156</td>
<td>R</td>
</tr>
<tr>
<td>GE / AD / SV</td>
<td>186</td>
<td>W</td>
</tr>
<tr>
<td>SS</td>
<td>174</td>
<td>C$^a$</td>
</tr>
</tbody>
</table>

Table 6: Dipeptide masses for Pro-containing peptides (if the cleavage in between can not be detected)

<table>
<thead>
<tr>
<th>P</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>154.17</td>
</tr>
<tr>
<td>A</td>
<td>168.20</td>
</tr>
<tr>
<td>S</td>
<td>184.20</td>
</tr>
<tr>
<td>P</td>
<td>194.23</td>
</tr>
<tr>
<td>V</td>
<td>196.25</td>
</tr>
<tr>
<td>T</td>
<td>198.22</td>
</tr>
<tr>
<td>C</td>
<td>200.26</td>
</tr>
<tr>
<td>I</td>
<td>210.28</td>
</tr>
<tr>
<td>L</td>
<td>210.28</td>
</tr>
<tr>
<td>N</td>
<td>211.22</td>
</tr>
<tr>
<td>D</td>
<td>212.21</td>
</tr>
</tbody>
</table>

91
Table 7: Neutral losses observed from peptides containing specific amino acids

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>Neutral loss (Da)</th>
<th>Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>18</td>
<td>H₂O</td>
</tr>
<tr>
<td>T</td>
<td>18</td>
<td>H₂O</td>
</tr>
<tr>
<td>C</td>
<td>34</td>
<td>H₂S</td>
</tr>
<tr>
<td>N</td>
<td>17</td>
<td>NH₃</td>
</tr>
<tr>
<td>D</td>
<td>18</td>
<td>H₂O</td>
</tr>
<tr>
<td>Q</td>
<td>17</td>
<td>NH₃</td>
</tr>
<tr>
<td>K</td>
<td>17</td>
<td>NH₃</td>
</tr>
<tr>
<td>E</td>
<td>18</td>
<td>H₂O</td>
</tr>
<tr>
<td>M</td>
<td>48</td>
<td>CH₂SH</td>
</tr>
<tr>
<td>M₉₉x</td>
<td>64</td>
<td>CH₃SOH</td>
</tr>
<tr>
<td>pS</td>
<td>80, 98</td>
<td>HPO₃, H₃PO₄</td>
</tr>
<tr>
<td>pT</td>
<td>80, 98</td>
<td>HPO₃, H₃PO₄</td>
</tr>
<tr>
<td>sY</td>
<td>80</td>
<td>SO₃</td>
</tr>
<tr>
<td>Nα-AcX</td>
<td>42</td>
<td>CH₂CO</td>
</tr>
</tbody>
</table>

Figure 95: Structures of the fragment ions occurring from sequence specific fragmentation of peptides during MS/MS. i represents an immonium ion of amino acid number 2 while I represents an internal fragment corresponding to \( y_{3b3} \) (see Figure 94).

Figure 96: Structures of d- and w-ions useful for the differentiation between Leucine and Isoleucine. R corresponds either to H (Leucine) or CH\(_3\) (Isoleucine). The mass of a dL-ion is therefore 14 Da higher than that of a dL-ion (the same is true for wL and wL). The numbering scheme is a little bit different from that for the a- and y-ions (figure 25) from which the d- and w-ions originate. Further fragmentation of an a\(_x\)-ion leads to a d\(_{x-1}\)-ion while side chain loss of a y\(_x\)-ion leads to a w\(_{x-1}\)-ion.
4.2 Recipes

4.2.1 Gel electrophoresis (according to Laemmli et al.)

Separating gel

<table>
<thead>
<tr>
<th></th>
<th>20 %, 0.53 %</th>
<th>15 %, 0.4 %</th>
<th>10 %, 0.27 %</th>
<th>5 %, 0.13 %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.875 M Tris(^1)/pH 8.8</td>
<td>1 ml</td>
<td>1 ml</td>
<td>1 ml</td>
<td>1 ml</td>
</tr>
<tr>
<td>40 % Acrylamide</td>
<td>2.5 ml</td>
<td>1.87 ml</td>
<td>1.25 ml</td>
<td>1 ml</td>
</tr>
<tr>
<td>2 % Bisacrylamide</td>
<td>1.3 ml</td>
<td>1 ml</td>
<td>0.67 ml</td>
<td>0.32 ml</td>
</tr>
<tr>
<td>10 % SDS(^2)</td>
<td>50 µl</td>
<td>50 µl</td>
<td>50 µl</td>
<td>50 µl</td>
</tr>
<tr>
<td>10 % APS(^3)</td>
<td>100 µl</td>
<td>100 µl</td>
<td>100 µl</td>
<td>100 µl</td>
</tr>
<tr>
<td>TEMED(^4)</td>
<td>10 µl</td>
<td>10 µl</td>
<td>10 µl</td>
<td>10 µl</td>
</tr>
<tr>
<td>H(_2)O</td>
<td>40 µl</td>
<td>970 µl</td>
<td>1920 µl</td>
<td>2520 µl</td>
</tr>
<tr>
<td>total</td>
<td>5 ml</td>
<td>5 ml</td>
<td>5 ml</td>
<td>5 ml</td>
</tr>
</tbody>
</table>

Collecting gel (4 %, 0.11 %)

<table>
<thead>
<tr>
<th></th>
<th>20 µl</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.6 M Tris(^5)/pH 6.8</td>
<td>20 µl</td>
</tr>
<tr>
<td>40 % Acrylamide</td>
<td>200 µl</td>
</tr>
<tr>
<td>2 % Bisacrylamide</td>
<td>110 µl</td>
</tr>
<tr>
<td>2.8 % SDS(^6)</td>
<td>36 µl</td>
</tr>
<tr>
<td>10 % APS(^7)</td>
<td>10 µl</td>
</tr>
<tr>
<td>H(_2)O</td>
<td>1624 µl</td>
</tr>
<tr>
<td>TEMED(^8)</td>
<td>2 µl</td>
</tr>
<tr>
<td>total</td>
<td>2 ml</td>
</tr>
</tbody>
</table>

\(^1\) Final concentration 0.375 M  
\(^2\) Final concentration 0.1 %  
\(^3\) Final concentration 0.2 %  
\(^4\) Final concentration 6 mM  
\(^5\) Final concentration 50 mM  
\(^6\) Final concentration 0.05 %  
\(^7\) Final concentration 0.1 %  
\(^8\) Final concentration 0.1 %
10 x Elektrophoresis buffer, 1 l
60 g Tris
28.8 g Glycine
10 g SDS
add Milli-Q water to a final volume of 1 l

Sample buffer, 50 ml
5 ml 0.6 M Tris pH 6.8
0.5 g SDS
5 g Sucrose
0.25 ml 2-Mercaptoethanol
5 ml 0.5 % Bromphenolblue

4.2.2 Gel stainings

Coomassie Protein staining solution
0.1 % coomassie R250
50 % methanol
10 % acetic acid

Destaining solution for Coomassie
10 % methanol
7 % acetic acid

Staining is performed overnight while destaining takes place within a couple of hours. For a faster destaining you can also heat up the solution in a microwave oven to about 50°C.

Colloidal Coomassie
0.1 % coomassie G-250
2 % H₃PO₄
10 % (w/v) (NH₄)₂SO₄
20 % methanol
 detection limit ~ 0.3 – 1.0 ng/mm²
Destaining solution for colloidal Coomassie stainings

Milli-Q water

Staining takes 3-5 days while destaining is performed overnight.

Tip: Putting a Kleenex into the destaining solution significantly reduces the destaining time for both stainings since the Coomassie binds to the Kleenex which reduces the Coomassie concentration within the destaining solution.

4.2.3 Enzymatic digestion protocols

Drying and washing Coomassie stained gel bands

- Incubate for 30 min at 37°C 250 µl 20 mM DTT in 50 mM NH₄HCO₃ (3.09 mg/ml, always prepare solution fresh) (*not necessary with 2D gel spots*)

- Incubate for 30 min at RT with 250 µl 50 mM Iodoacetamid in 50 mM NH₄HCO₃ (9.25 mg/ml, always prepare solution fresh) (*not necessary with 2D gel spots*)

- Add 250 µl Acetonitrile to the gel plugs

- Incubate for 15 min

- Remove the supernatant

- Add 250 µl 50 mM NH₄HCO₃

- Incubate for 15 min

- Remove the supernatant

- Add 250 µl Acetonitrile to the gel plugs

- Incubate for 15 min

- Remove the supernatant

- Add 250 µl 50 mM NH₄HCO₃

- Incubate for 15 min

- Remove the supernatant

- Add final 250 µl Acetonitrile
Incubate for further 15 min

Remove the supernatant and dry the gel plugs in a vacuum centrifuge to complete dryness

**Tryptic digest of proteins and elution of resulting tryptic fragments**

- break large gel plugs (>>1 mm²) into small pieces
- Add Trypsine solution (12.5 ng/µl in 50 mM NH₄HCO₃ [Trypsine can be stored as a stock solution in 1 mM HCl; this can be diluted immediately prior to use to a working concentration using 50 mM NH₄HCO₃; we have good experiences with TPCK-treated Trypsine from Sigma (T1426), Trypsine sequencing grade from Roche Diagnostics and Trypsine sequencing grade from Promega]) and let the plug swell for 1 h at 4°C on ice.
- Remove supernatant Trypsin solution and wash twice with 50 mM NH₄HCO₃
- add enough 50 mM NH₄HCO₃ to cover the gel plugs completely
- Incubate for 4-16 h at 37°C and slight shaking (approx. 800 rpm)
- Collect the supernatant in a 0.5 ml microreaction cup and dry it (SpeedVac)
- Add 250 µl 0.1% TFA/ Acetonitril (2:3) to the gel plug
- Incubate for approx. 6 h at 37°C while shaking (approx. 1200 rpm)
- Collect the supernatant and add it to the first one (dry it in the SpeedVac)
- Add 250 µl 0.1% TFA/ Acetonitril (2:3) to the gel plug
- Incubate over night at 37°C while shaking (approx. 1200 rpm)
- Collect the supernatant and add it to the first one (dry it in the SpeedVac)

Subsequently the tryptic fragment peptides can be analyzed by mass spectrometry to identify the proteins.

Note: This is one of many protocols. We have made good experience with it, which does not necessarily mean that it is the best and most sensitive one. When you want to have your results faster, you can reduce the digestion times (see J. Havlis, H. Thomas, M. Sebela and A. Shevchenko, Anal. Chem. (2003) 75(6), 1300-1306 for a rapid 30 min digestion protocol), the elution times and you can also support the elution of the peptides by ultrasonification. Some people say that drying the peptides will lead to losses due to sticking on the wall. This
is probably true. When you want to avoid this, you have two options: a) using siliconized plastics (strongly NOT recommended) or b) avoiding a complete drying of the samples. For the reconstitution prior to mass spectrometric analysis you should keep in mind what you want to do with the samples afterwards.

Recommended reconstitution solutions are:

MALDI-MS: 0.1% TFA/acetonitrile (1:1)
ESI-MS (offline nanoESI): 0.1-1% formic acid/acetonitrile (1:1) or 0.1-1% formic acid/isopropanole (1:1)
ESI-MS (online nanoLC): 0.1-1% formic acid, max. 5% organic modifier (e.g. acetonitrile)
4.3 Suggested readings

4.3.1 Books


Michael Kinter, Nicholas Sherman, “Protein Sequencing and Identification Using Tandem Mass Spectrometry”; Wiley

Peter Snyder, “Interpreting Protein Mass Spectra: A Comprehensive Resource”; ACS

for german speaking readers:
Friedrich Lottspeich, Haralabos Zorbas, “Bioanalytik”; Spektrum Verlag

4.3.2 Articles


K. Gevaert, H. Demol, L. Martens et al., “Protein identification based on matrix assisted laser desorption/ionization-post source decay-mass spectrometry”

ISD


D. Suckau, D.S. Cornett, Analusis (1998), 26(10), M19-M20

Posttranslational modifications


Database searching


**Miscellaneous**


### 4.3.3 Internet resources

#### 4.3.3.1 Database search engines

Protein Prospector: [http://prospector.ucsf.edu/](http://prospector.ucsf.edu/)


MS-Blast: [http://dove.embl-heidelberg.de/Blast2/msblast.html](http://dove.embl-heidelberg.de/Blast2/msblast.html)

#### 4.3.3.2 Helpful Tools

ABRF: [http://www.abrf.org/](http://www.abrf.org/)

DeltaMass list: [http://www.abrf.org/index.cfm/dm.home](http://www.abrf.org/index.cfm/dm.home)

PROWL: [http://prowl.rockefeller.edu/](http://prowl.rockefeller.edu/)


Pedro’s Tools: [http://www.public.iastate.edu/~pedro/research_tools.html](http://www.public.iastate.edu/~pedro/research_tools.html)

Scientific Instruments Services: [http://www.sisweb.com/index/sis/referenc.htm](http://www.sisweb.com/index/sis/referenc.htm)

#### 4.3.3.3 Databases
