

Supplementary Material on peptide validation to

The abc's (and xyz's) of peptide sequencing

by Hanno Steen and Matthias Mann

Validation of peptide hits

A challenging, but critical step is the verification of a retrieved sequence by comparing its calculated fragment ion pattern to the experimental MS/MS data that only provide borderline identification scores and where the peptide is the only one identifying the protein. In order to evaluate those ambiguous peptide identifications, several rules of thumb can be used to assess whether a particular protein identification is reasonable or not. Some of these rules are described using the example of a tandem mass spectrum that was acquired using a quadrupole–time-of-flight instrument and the doubly protonated tryptic peptide FLYM(Ox)WPNAR (mass-to-charge value (m/z) = 607.32; the precursor ion is marked with an asterisk):

- It is often not possible to rationalize all the fragment ions that are observed in a tandem mass spectrum. However, in the case of doubly charged tryptic peptides, the majority of the most abundant peaks in the m/z range above and around the precursor ion should be indicative of a (short) continuous series of y-type fragment ions (see the red fragment ions that are assigning the sequence tag -L-Y-M(Ox)-W-). b-type fragment ions of lower intensity are expected to be present when ion traps have been used for the analysis, or if the peptide comprises an internal basic amino-acid residue.
- Peptide bonds amino-terminal to proline, and carboxy-terminal of aspartate residues are particularly labile — that is, more abundant fragment ions are observed for this cleavage compared to those for the cleavage of the preceding and subsequent peptide bonds. In fact, cleavage carboxy-terminal of proline, and amino-terminal to aspartate, is energetically unfavored. Fragment ions that are derived from the labile cleavage should therefore be much more abundant than those derived from the hampered cleavage (see, for example, the signal intensities

at the m/z value of 457.26 versus 359.19 in the figure).

- If a side-chain modification — such as serine/threonine phosphorylation, glycosylation and/or methionine oxidation — is present, fragment ions that comprise this modification can be accompanied by so-called ‘satellite ions’. This is a result of the ready loss of modification-specific fragments (for example, phosphoric acid (98 Da) for phosphorylated species or CH_3SOH (64 Da) for oxidized methionine; see the green m/z values in the figure). Depending on how facile this loss is, the satellite ions can be more abundant than the related fragment ion. (Note: Since mass spectrometry measures the mass to charge ratio also mass differences are measured as mass to charge ratio, that is the m/z difference depends on the charge state: for example ‘satellite’ ions can appear at m/z -49 for doubly charged phosphorylated fragments or m/z -32.66 for triply charged fragments (oxidised Met: m/z -32 for doubly charged and m/z -21.33 for triply charged fragments))
- It is often assumed (by researchers and also by the database-searching algorithms that are used at present) that fragment ions have a lower charge state than the precursor ion from which they derive. However, some intense fragment ions in the m/z range below the precursor ion that have not been accounted for can actually be fragment ions that have the same charge state as the precursor ion (see the fragment ion that is marked with a red arrow; this ion is the doubly protonated form of the fragment ion at the m/z value 953.44).
- b_1 ions are absent (exception: acylated N-termini), and y_2 ions are often of very low intensity; this is contrasted by an intense a_2/b_2 fragment ion pair (see fragment ions at 233.17 and 261.16).
- Fragment ions are often accompanied by the loss of water (-18 Da) or ammonia (-17 Da). Those ‘satellite’ fragment ions can be quite intense if the peptide is rich in Ser, Thr, Glu, Gln, Asp, and/or Asn residues.
- Accuracies: Triple quadrupoles and ion traps provide mass accuracies of ± 0.5 Da; this number can be even greater (up to 1 Da) in the case of weak data, that is, the m/z values of all fragment ions should be within this range. When quadrupole TOF instruments are used, the mass deviation can be as low as ± 0.01 Da,

depending on how good the instrument is tuned and/or calibrated; nevertheless, it should not exceed ± 0.1 Da. In addition, the mass deviation should be consistent.

- Whereas quadrupole ion traps suffer from a so-called low mass cut-off, that is, a limited m/z range with a cut-off at $1/3$ of the m/z value of the precursor fragmented, triple quadrupole and quadrupole TOF instruments provide also information in the low mass region which can be very useful. These are in particular the y_1 -ions of Lys (m/z 147.11) or Arg (m/z 175.12), a strong immonium ion deriving from the N-terminal amino acid (especially in the case of Leu, Ile (both 86.10), His (110.07 Da), Phe (120.08 Da), Tyr (136.08 Da), Val (72.08 Da), and Trp (159.09 Da)) and immonium ions of aromatic amino acid residues within the sequence (see part B in the figure). Irrespective of whether b-type fragment ions are present in the spectrum, an ion pair in the lower m/z range spaced by 28 Da is often quite intense, corresponding to the a_2 and b_2 fragment ions (see above). In addition, the low mass range can help to confirm modified amino acid residues; for example, tyrosine phosphorylated and proline hydroxylated peptides can give rise to a characteristic immonium ion at m/z 216.05 and 86.06, respectively, when high resolution, high accuracy instruments such as quadrupole TOF mass spectrometers are used.
- Tandem mass spectra which are very complex are sometimes caused by fragmentation of two peptides together. This can be ascertained by looking for evidence of two precursor ions in the corresponding mass spectra. Both peptides can often be identified by an iterative search, that is, the database again with 'leftover' peaks from the first identification.

