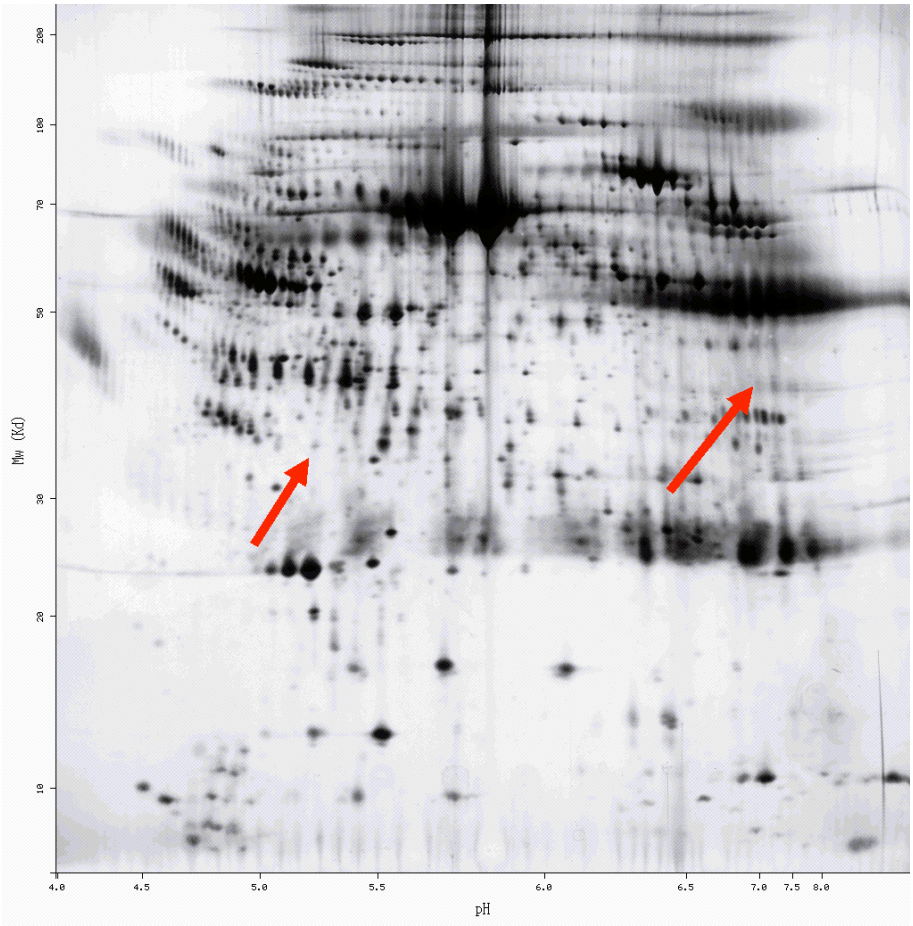
Exercise 2: Protein identification from two-dimensional gel electrophoresis and tandem mass spectrometry

The figure below shows a two-dimensional electrophoresis map of the human plasma proteome:



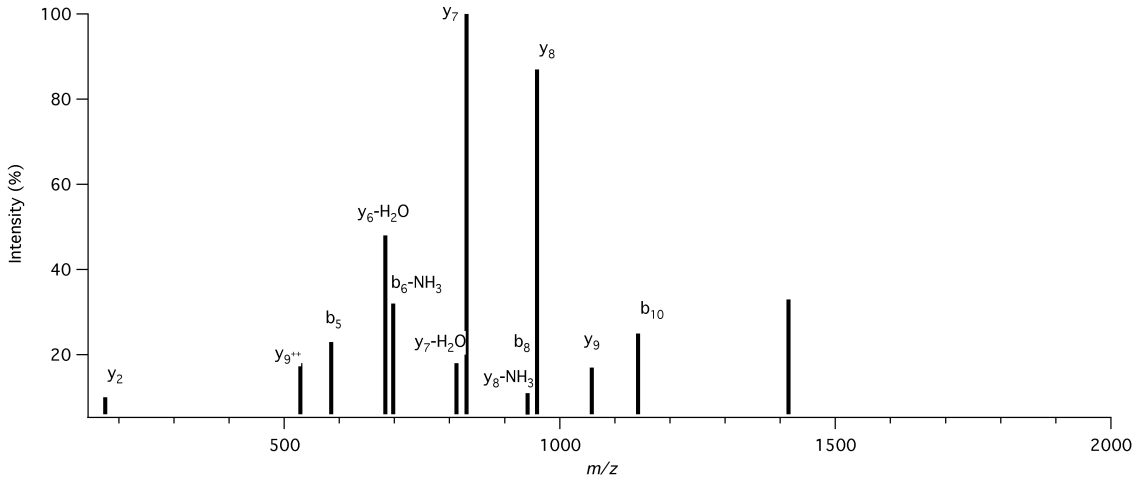
Question 1: Explain what plasma is, compared to blood and serum.

Question 2: Using the Sequence Retrieval System, indicate how many proteins from Homo Sapiens have been documented to date in SwissProt? In TrEMBL? Explain the difference between the two databases.

We want to identify the two spots highlighted in the gel image. The first one has the following properties measured from the gel:
Molecular weight: 34200 Da
Isoelectric point: 5.5

Question 3: From the gel image above, estimate the uncertainty on molecular weight and isoelectric point. Using the tool TagIdent from Expasy, give the number of possible proteins from SwissProt and/or TrEMBL with these molecular weight and isoelectric point, depending on the uncertainties you estimated.

The protein spot is excised, incubated with trypsin, and resulting peptides are eluted, desalted and submitted to ESI-MS/MS. A peptide is measured at 1414,58 Da, and fragmented, yielding the following spectrum:



Point	'Fragment ion'	Mass	Intensity
0	y2	175.12	10
1	y9++	529.3	18
2	b5	585.32	23
3	y6-H2O	683.38	48
4	b6-NH3	697.34	32
5	y7-H2O	812.43	18
6	y7	830.44	100
7	b8	941.49	6
8	y8-NH3	941.51	11
9	y8	958.53	87
10	y9	1057.6	17
11	b10	1141.57	25
12	parent	1414.54	33

Question 4: Using the document provided in annex, and knowing that the peptide sequence is LDEVKEQVAEVR , draw the chemical structure of all the fragments observed in the spectrum (amino acid residues can be represented in 1 or 3 letter code).

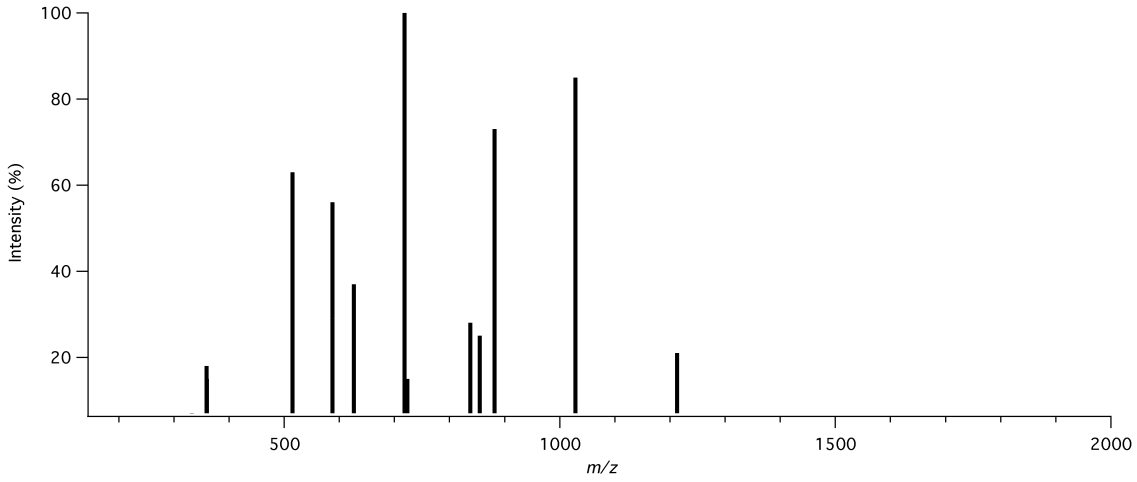
Question 5: The peptide whose MS/MS spectrum is shown above has the sequence LDEVKEQVAEVR. Use TagIdent to identify the protein, knowing:

1. its approximate molecular weight
2. its approximate isoelectric point
3. the sequence of one of its tryptic peptides

As TagIdent allows 6 amino acid tags, estimate if all subsets of 6 amino acids within this peptide yield the same identification e.g. LDEVKE versus DEVKEQ). Conversely, if shorter tags are used, can the protein be identified (e.g. with EQV as tag)? What is the limit of this approach (especially with such low sequence coverage)?

The same procedure is followed for the other spot, whose characteristics are the following:
Molecular weight: 42000 Da
Isoelectric point: 6.9

The protein is digested with trypsin, and one peptide at mass 1212.41 is subjected to tandem mass spectrometry (collision-induced dissociation in an ion trap), yielding the following spectrum:



Point	'Ion type'	Mass	Intensity
0	b3	332.2	7
1	y6++	359.67	15
2	y3	359.2	18
3	y8++	514.74	63
4	y5	587.29	56
5	b5	626.3	37
6	b6-NH3	723.32	15
7	y6	718.33	100
8	b7-NH3	837.36	28
9	b7	854.39	25
10	y7	881.39	73
11	y8	1028.46	85
12	parent	1212.41	21

Question 6: For the ease of exercise, peaks have already been assigned to corresponding fragments. From the peak list above, reconstruct the peptide sequence (remind that MS/MS interpretation relies on mass differences).

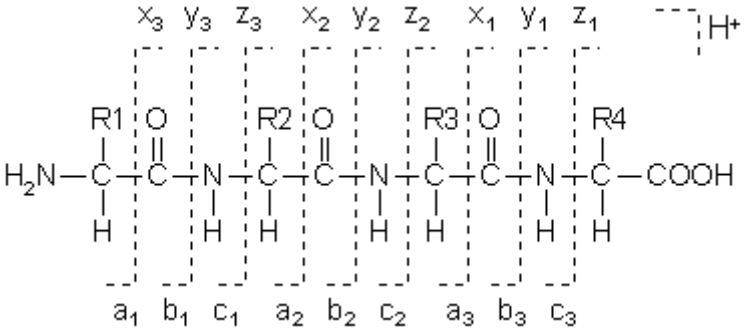
Question 7: From the sequence you determined and using TagIdent as in question 5, identify the protein.

Peptide Fragmentation

Sequence Ions

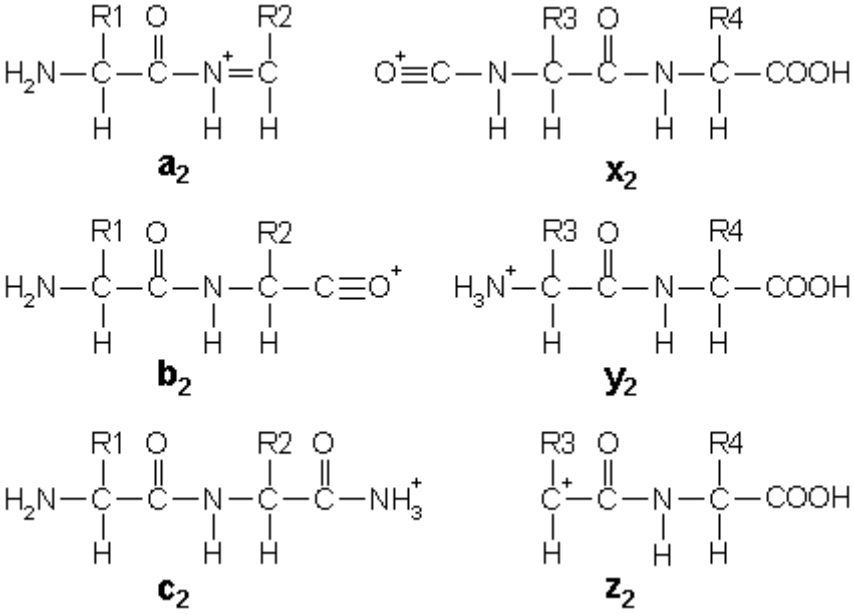
Bibliography

The types of fragment ions observed in an MS/MS spectrum depend on many factors including primary sequence, the amount of internal energy, how the energy was introduced, charge state, etc. The accepted nomenclature for fragment ions was first proposed by Roepstorff and Fohlman [Roepstorff, 1984 #672], and subsequently modified by Johnson *et. al.* [Johnson, 1987 #680].



Fragments will only be detected if they carry at least one charge. If this charge is retained on the N terminal fragment, the ion is classed as either *a*, *b* or *c*. If the charge is retained on the C terminal, the ion type is either *x*, *y* or *z*. A subscript indicates the number of residues in the fragment.

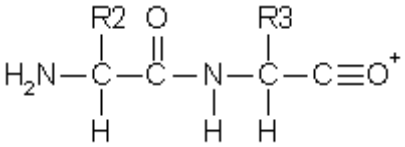
In addition to the proton(s) carrying the charge, *c* ions and *y* ions abstract an additional proton from the precursor peptide. Thus, the structures of the six singly charged sequence ion are:



Note that these structures include a single charge carrying proton. In electrospray ionisation, tryptic peptides generally carry two or more charges, so that fragment ions may carry more than one proton.

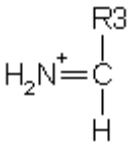
Internal Cleavage Ions

Double backbone cleavage gives rise to internal fragments. Usually, these are formed by a combination of *b* type and *y* type cleavage to produce the illustrated structure, an amino-acylium ion. Somethimes, internal cleavage ions can be formed by a combination of *a* type and *y* type cleavage, an amino-immonium ion. Internal fragments are labelled with their 1 letter amino acid code.



Immonium Ions

An internal fragment with just a single side chain formed by a combination of *a* type and *y* type cleavage is called an immonium ion. These ions are labelled with the 1 letter code for the corresponding amino acid.

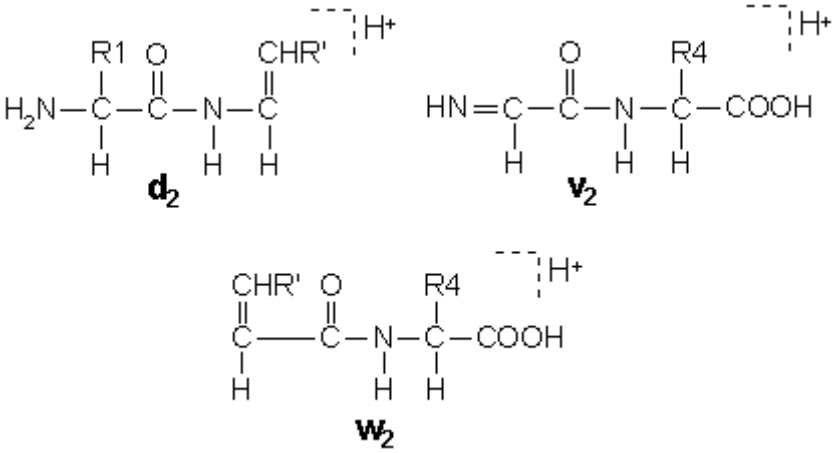


Residue	3-letter code	1-letter code	Immonium ion*	Related ions*
Alanine	Ala	A	44	
Arginine	Arg	R	*129*	59,70,*73*,87,100,112
Asparagine	Asn	N	*87*	*70*
Aspartic acid	Asp	D	*88*	*70*
Cysteine	Cys	C	*76*	
Glutamic acid	Glu	E	*102*	
Glutamine	Gln	Q	101	56,84,*129*
Glycine	Gly	G	30	
Histidine	His	H	**110**	*82*,*121*,*123*,*138*,166
Isoleucine	Ile	I	**86**	44,*72*
Leucine	Leu	L	**86**	44,*72*
Lysine	Lys	K	*101*	*70*,84,112,129
Methionine	Met	M	*104*	61
Phenylalanine	Phe	F	**120**	*91*
Proline	Pro	P	**70**	
Serine	Ser	S	60	
Threonine	Thr	T	74	
Tryptophan	Trp	W	**159**	77,117,**130**,132,**170**,**171**
Tyrosine	Tyr	Y	**136**	*91*,*107*
Valine	Val	V	72	41,55,69

Immonium and related ion masses after Falick, 1993 #690 and Papayannopoulos, 1995 #681 . * Bold face indicates strong signals, italic indicates weak.

Satellite Ions

Collision induced dissociation of ions at keV energies can generates additional ion types due to side chain cleavages, [Johnson, 1988 #679].



In these structures, R' is the substituent, if any, at the beta carbon following the loss of the gamma carbon by side chain cleavage. Isoleucine and threonine are doubly substituted at the beta carbon, so that side chain loss can give rise to two different ion structures. These pairs are designated *d* and *d'* or *w* and *w'*.

A useful feature of *d* and *w* ions is that they enable the isobaric residues leucine and isoleucine to be differentiated.

Low Energy CID

Bibliography

In low energy CID (i.e. collision induced dissociation in a triple quadrupole or an ion trap) a peptide carrying a positive charge fragments mainly along its backbone, generating predominantly *a*, *b* and *y* ions. In addition, peaks are seen for ions which have lost ammonia (-17 Da) denoted *a**, *b** and *y** and water (-18 Da) denoted *a°*, *b°* and *y°*. Satellite ions from side chain cleavage are not observed.

High Energy CID

Bibliography

All of the ion series described above are observed in high energy collision spectra. Relative abundances are composition dependent. Unlike low energy CID, ions do not readily lose ammonia or water.

The characteristic effects of individual amino acid residues on peptide fragmentation behaviour have been reviewed in detail by Papayannopoulos [Papayannopoulos, 1995 #681].

Post Source Decay

Bibliography

The most abundant fragment ion types observed in MALDI-TOF PSD are *a*, *b*, and *y*. If collision gas is used, then the spectra resemble high energy CID. All ion series can be accompanied by composition dependent satellites due to loss of ammonia or water.

When a peptide contains an internal proline, strong ion series due to internal cleavage are observed, extending from the proline in the direction of the C terminus.

Negative Ions

The structures illustrated above are all for ions with a single positive charge. Mascot also includes support for negative ions, and the negative ion types are the same as positive, but with one proton per charge subtracted rather than added. This may be an over-simplification, and we would welcome feedback concerning any additional ion types that should be considered.

Formulae to Calculate Fragment Ion m/z values

[N] is the molecular mass of the neutral N-terminal group, [C] is the molecular mass of the neutral C-terminal group, [M] is molecular mass of the neutral amino acid residues. To obtain m/z values, add or subtract protons as required to obtain the required charge and divide by the number of charges. For example, to get a^{+}, add 1 proton to the M_r value for a. To get a^{--}, subtract 2 protons from the M_r value for a and divide by 2.

Ion Type	Neutral M_r
a	[N]+[M]-CHO
a*	a-NH_3
a°	a-H_2O
b	[N]+[M]-H
b*	b-NH_3
b°	b-H_2O
c	[N]+[M]+NH_2
d	a - partial side chain
v	y - complete side chain
w	z - partial side chain
x	[C]+[M]+CO-H
y	[C]+[M]+H
y*	y-NH_3
y°	y-H_2O
z	[C]+[M]-NH_2