Functional Genomics and Proteomics
National Centre for Biomolecular Research
Faculty of Science Masaryk Univers

CEITEC

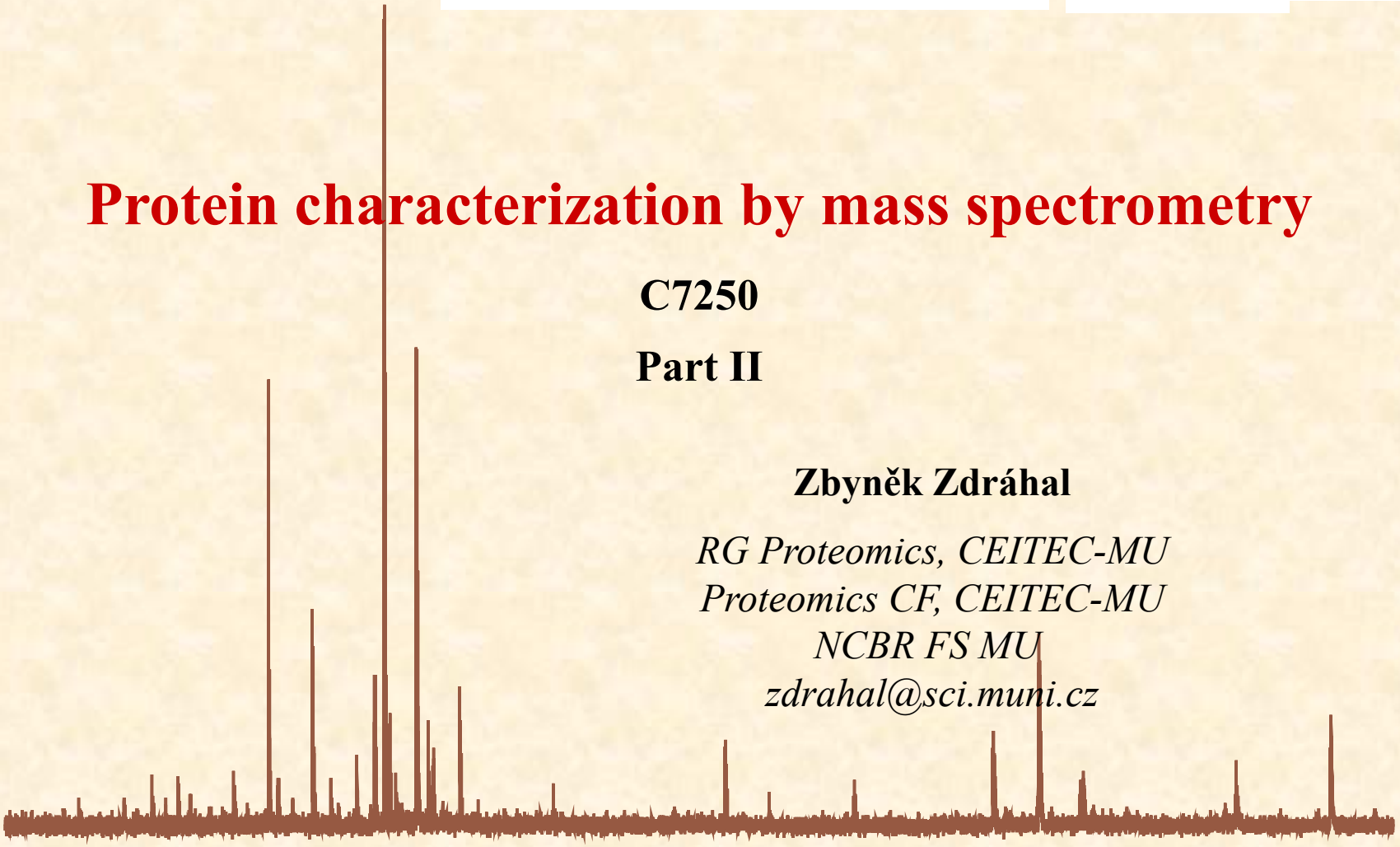
Protein characterization by mass spectrometry

C7250

Part II

Zbyněk Zdráhal

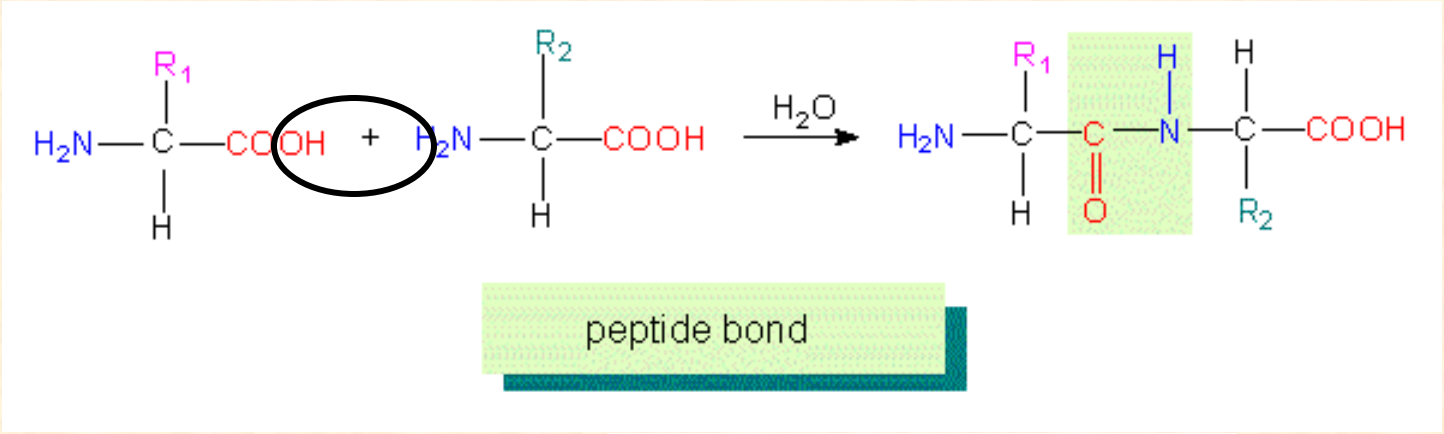
RG Proteomics, CEITEC-MU
Proteomics CF, CEITEC-MU
NCBR FS MU
zdrahal@sci.muni.cz



Functional Genomics and Proteomics
National Centre for Biomolecular Research
Faculty of Science Masaryk University

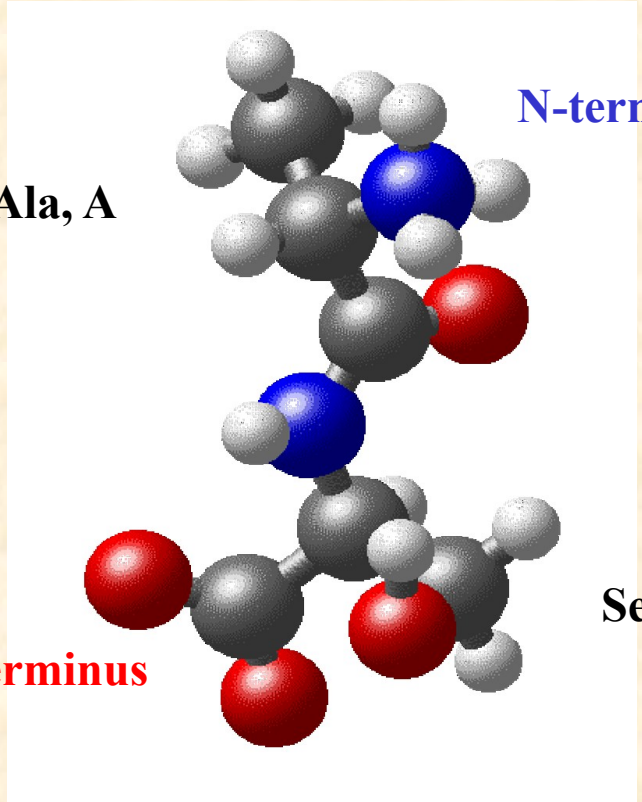
Basic methods of protein characterization by mass spectrometry

Protein formation



Alanin, Ala, A

N-terminus



Dipeptide Ala-Ser

Serin, Ser, S

C-terminus

Protein structure

and mass spectrometry

Beta-glucosidase (maize), 566 aminoacids, 8280 atoms, 128 474 Da

Primary structure (aminoacids ladder):

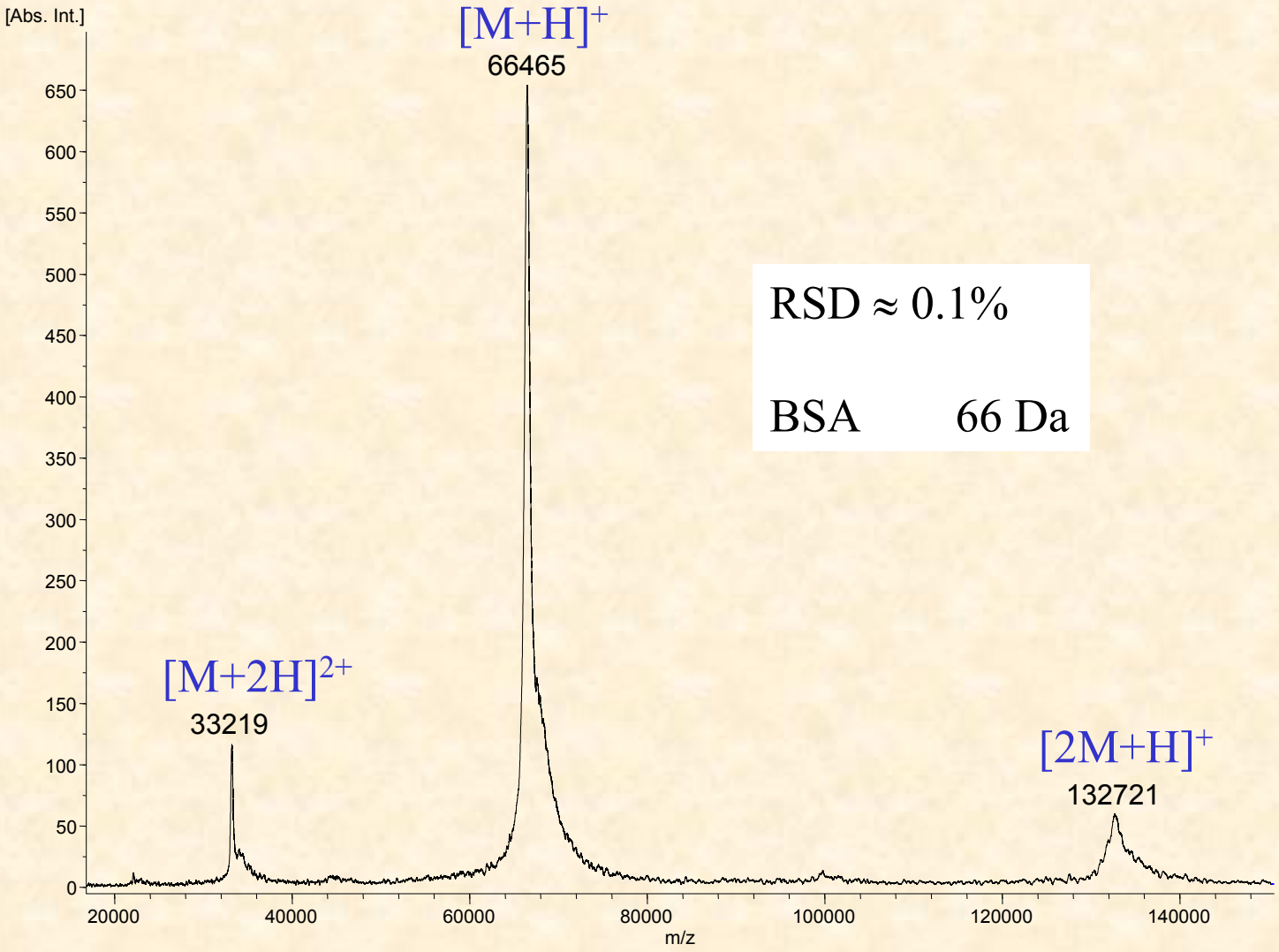
MAPLLAAAMNHAAAHPGLRSH PQSSKRRCNLSFTTRSAR
VGSQNGVQMLSPSEIPQRDWE WNEDGKGESNWDHFCHNH
PERILDGSNSDIGANSYHMYK SWPRILPKGTKEGGINPD
GIKYYRNLINLLLENGIEPYV LDKSHKSIVEDYTYFAKV
CFDNFGDKVKNWLTFNEPQTF LDCAYPTGNSLVEPYTAG
HNILLAHAEAVDLYNKHYKRD SFLDKQAEERSWDINLGW
FLEPVVRGDYPFSMRSLAREF LGLNYYTSRFSKNIDISP
NYSPVLNTDDAYASQEVNGPD GLKDLLMIMKNKYGNPPI
YITENGIGDVDTKETPLPMEA KESIDLGSNVQGYFAWSL
LDNFEWFAGFTERYGIVYVDRNNNCTRYMKESAKWLKEFNTAKKPSKKILTPA


OM-CJF

Functional Genomics and Proteomics
National Centre for Biomolecular Research
Faculty of Science Masaryk University

Determination of protein mass
Intact mass analysis

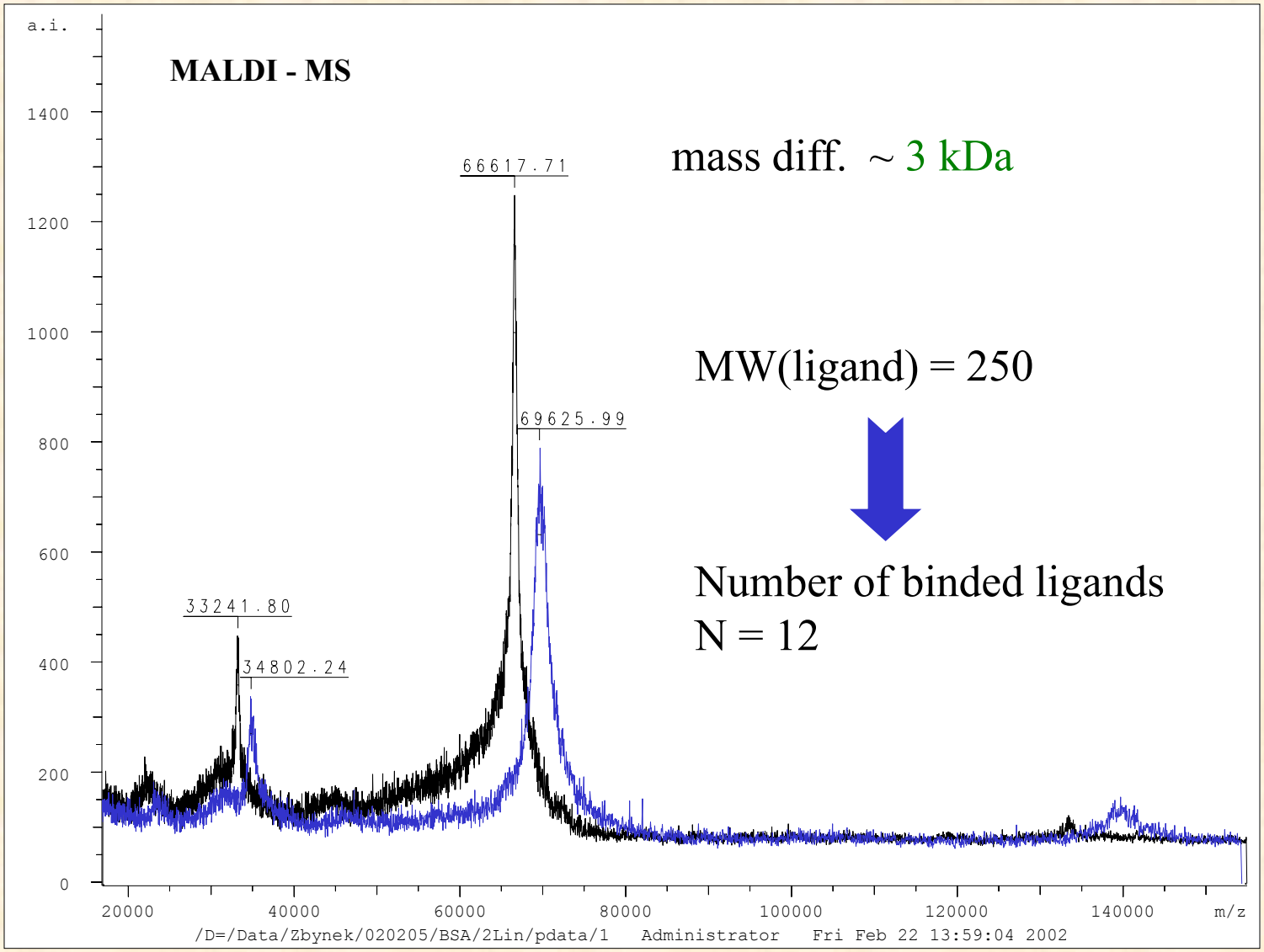
MALDI MS spectrum of protein (BSA ≈ 15 pmol)



RSD ≈ 0.1%

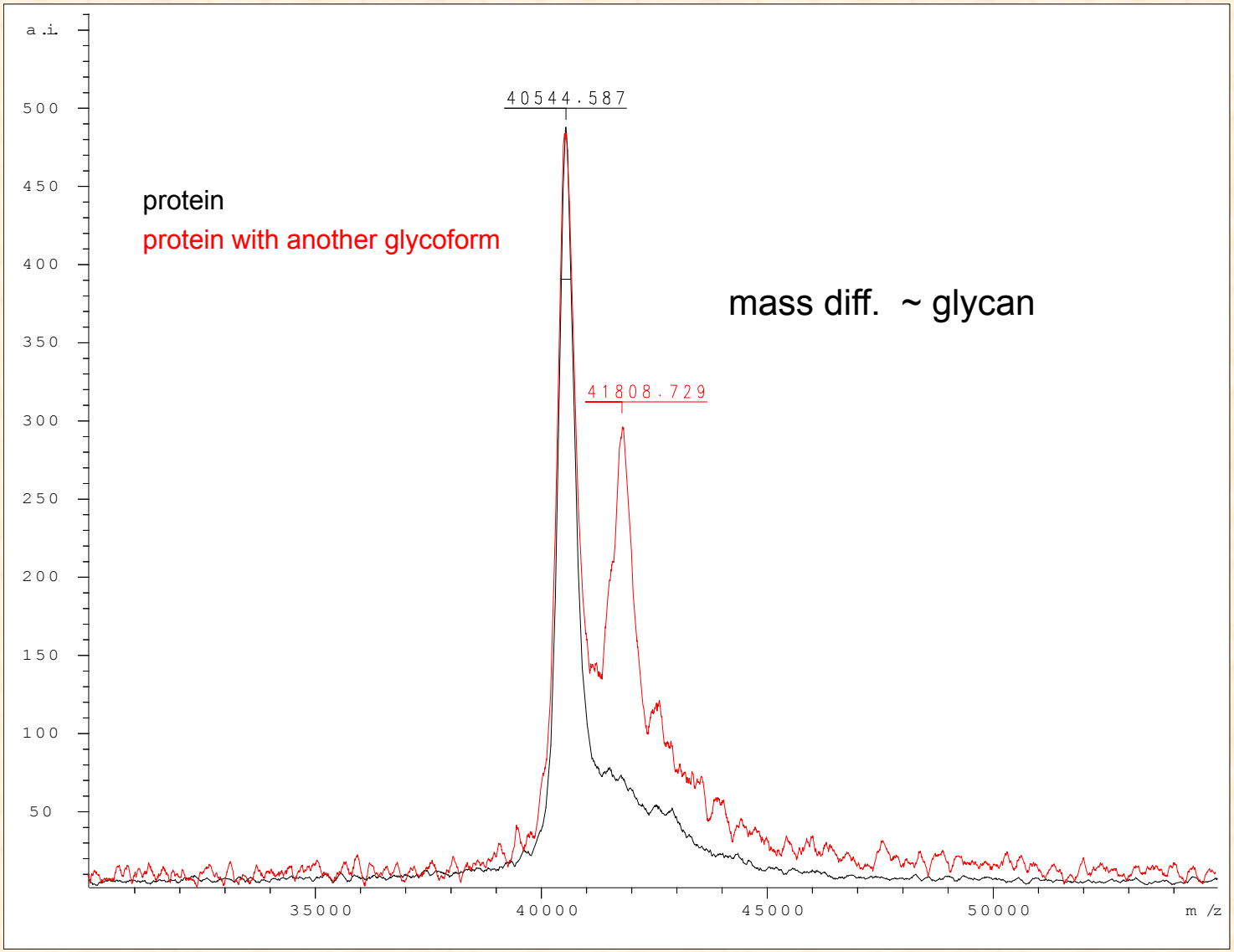
BSA 66 Da

Modified BSA *vs.* std. BSA (≈ 5 pmol)

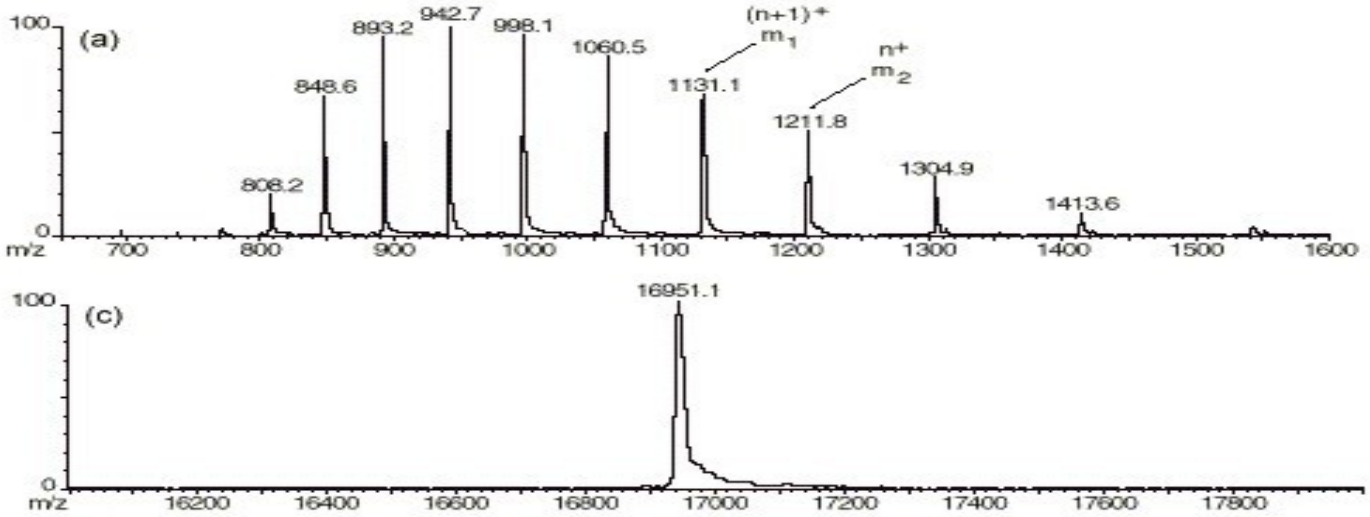


C7250

MALDI-MS spectrum of glycoprotein



ESI spectrum of myoglobin (16 951 Da)



(b)

Mass to Charge Ratio (m/z)	No. of charges (n)	Molecular Mass (RMM)
1542.04	11	16951.40
1413.59	12	16950.95
1304.93	13	16950.94
1211.80	14	16951.11
1131.12	15	16951.62
1060.46	16	16951.26
998.11	17	16950.67
942.75	18	16951.30
893.15	19	16950.71
848.57	20	16951.25
808.21	21	16951.14
771.49	22	16950.72
	Mean	16951.09
	S.D.	±0.30

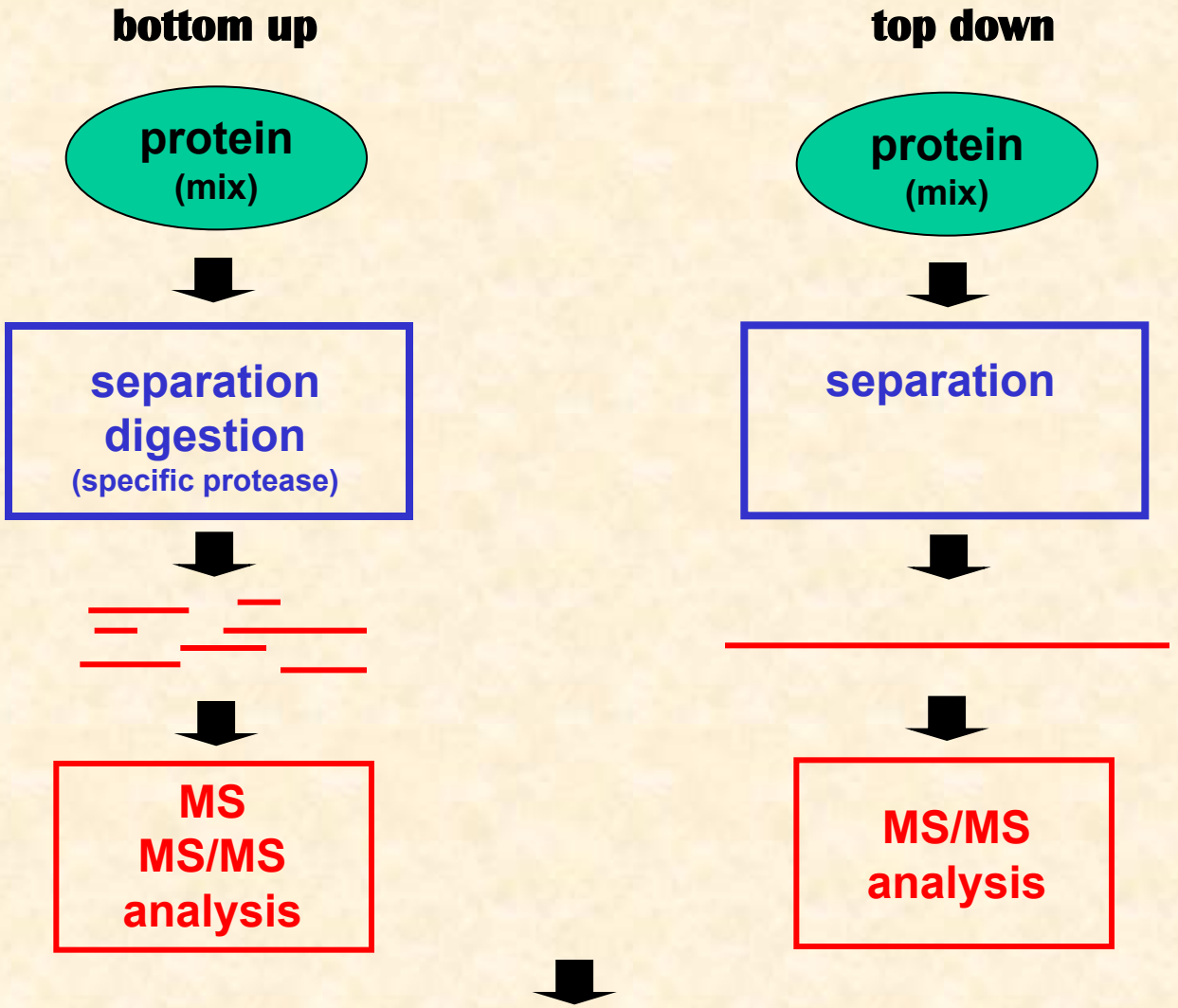
Determination of protein mass

- useful information for initial characterization
- does not enable protein identification and it is not necessary for identification
- in limited way allows characterization of modifications (mainly using high-resolution MS)

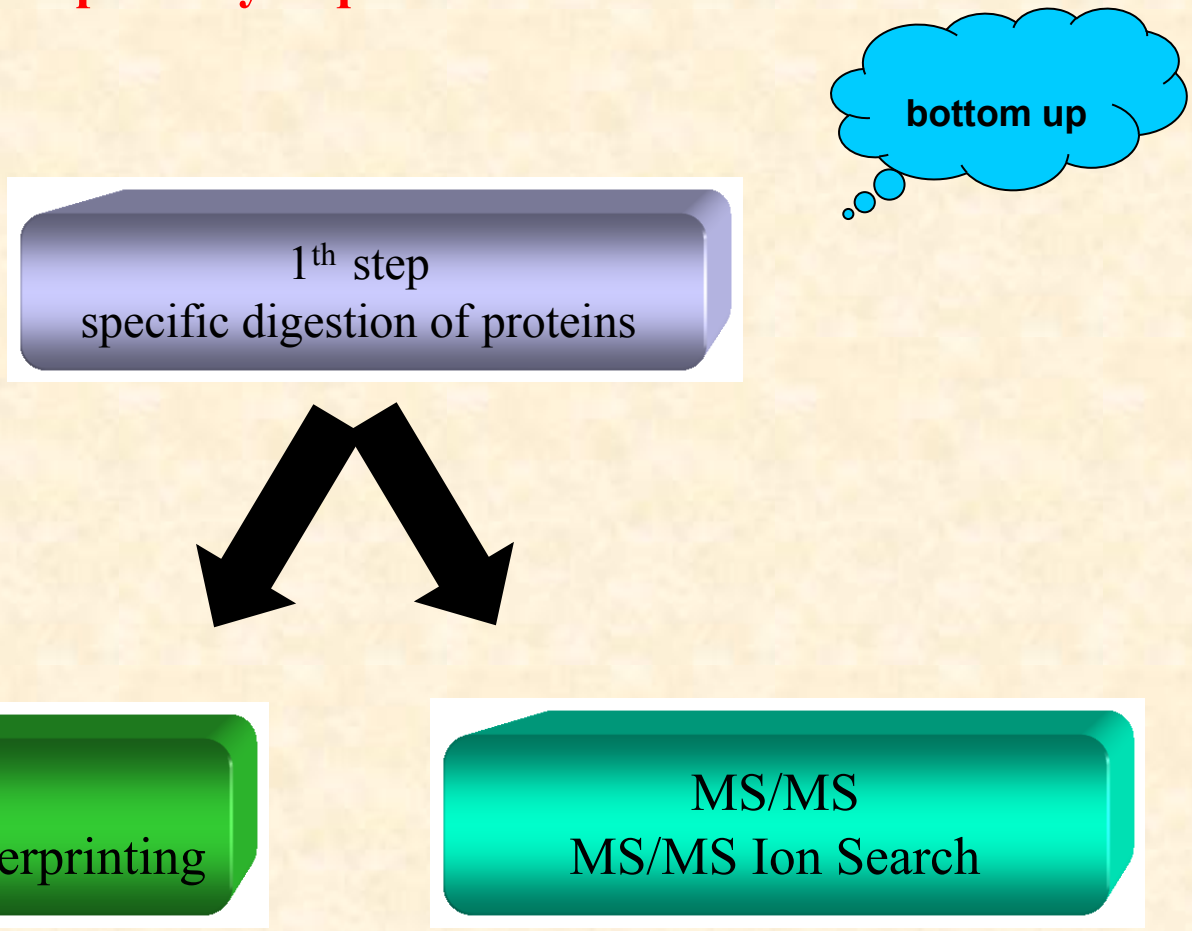
Functional Genomics and Proteomics
National Centre for Biomolecular Research
Faculty of Science Masaryk University

Protein identification

Protein identification using mass spectrometric data



Identification of known proteins is mainly performed at peptide level
- primary sequence is in database



Specific protein digestion

- Enzymatic digestion at selected aminoacids
- **examples of specific proteases**

trypsin	K-X, R-X	except P
Glu-C	E-X	except P
Asp-N	X-D	

- **„low-specifity“ proteases** (pepsin, thermolysin)

- **Chemical digestion**

CNBr(FA)	X-M

http://www.expasy.ch/tools/peptidecutter/

Tryptic digestion

QNGVQMLSPSEIPQRDWFPSDFTFGAATSAYQIEGAWNEDGKGESNWDHFCHN
HPERILDGSNSDIGANSYHMYKTDVRKLLKPMGMDAYRFSISWPRILPKGTKE
GGINPDGIKYYRNLINLLLENGIEP

digests always after **lysine (K)** or **arginine (R)**, if the next aminoacid is not **proline**

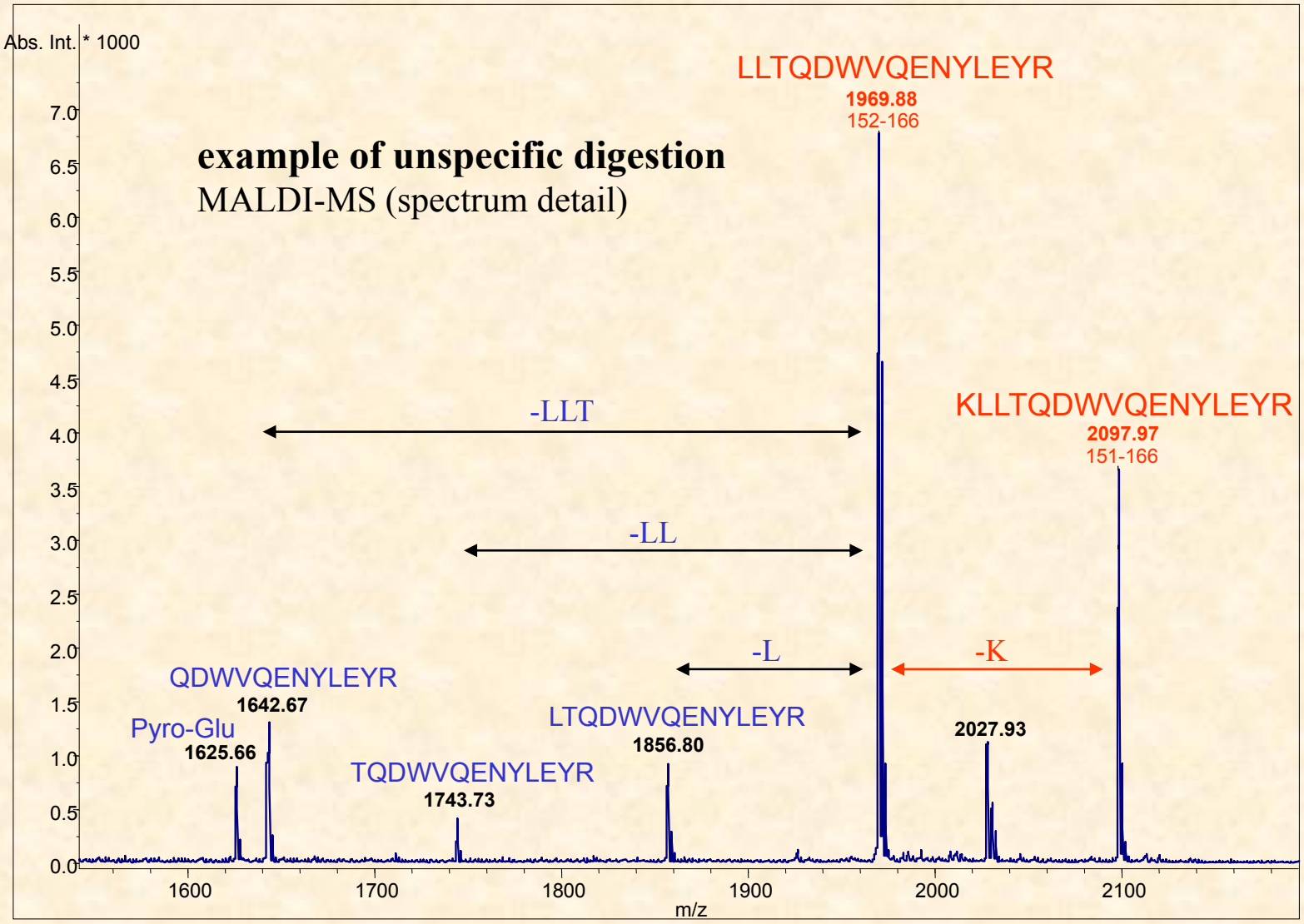
Peptide map

QNGVQMLSPSEIPQR	1-15	1683.848 Da
DWFPSDFTFGAATSAYQIEGAWNEDGK	16-42	3010.317 Da
GESNWDHFCHNHPER	43-57	1864.757 Da
ILDGSNSDIGANSYHMYK	58-75	1984.907 Da
TDVR	76-79	490.262 Da
...		

Specifity of digestion – one of crucial prerequisites of identification reliability

Set of masses of these formed peptides (i.e. peptide map) is characteristic for given protein similarly as fingerprint for human individual.

Tryptic digestion



Sequence of unspecifically digested peptides verified by MALDI-MS/MS


OM-CJF

MS
Peptide Mass Fingerprinting

principle:

- ***MS analysis*** *of specifically digested peptides*

  obtained set of ***"peptide masses" (peptide map)***

  = specific information suitable for protein identification

- *database search*

- method applicable for pro individual (separated) proteins

  → separation of proteins necessary

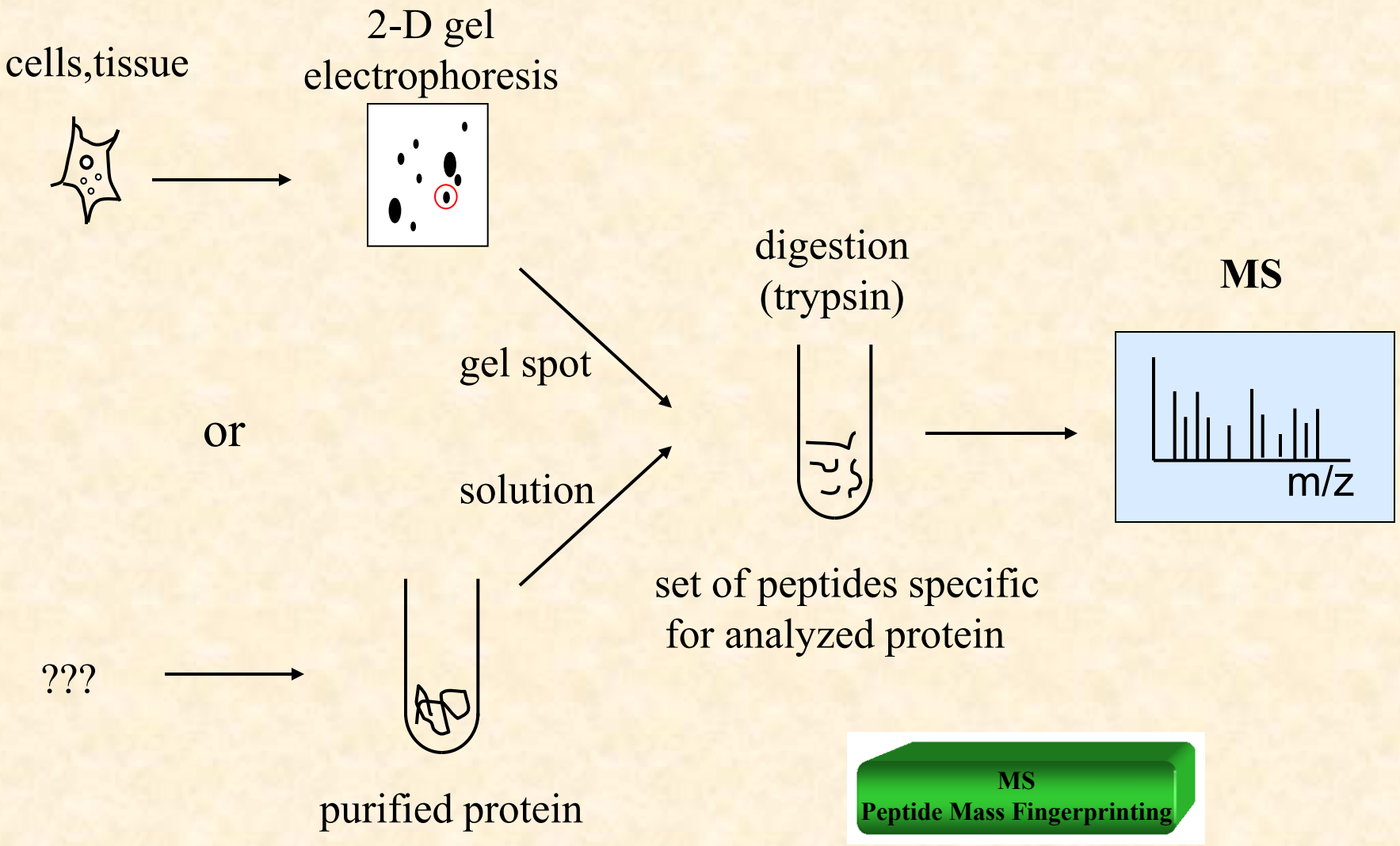
MS
Peptide Mass Fingerprinting

Measured peptide map is searched against database of protein sequences using database search engines.

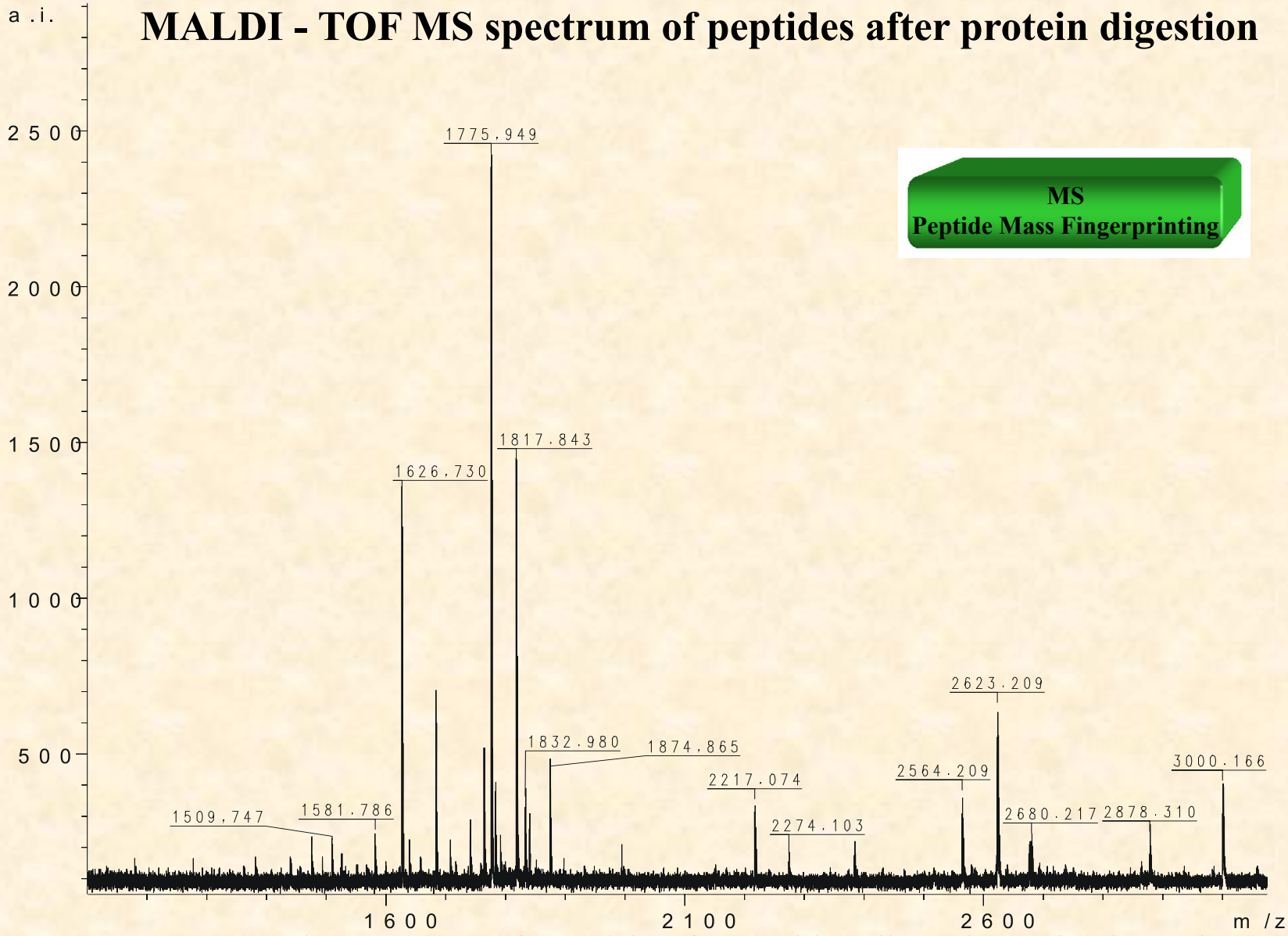
Database search engine calculates theoretical peptide map for each protein sequence in database (applying cleavage rules for selected protease) and stepwise compares experimentally obtained peptide map of our analysed protein with *in-silico* calculated peptide maps.

The search results in a list of proteins with most similar *in-silico* peptide maps. Similarity extent is given by score, all protein candidates with score value higher than the limit significant value (calculated by software) are considered as identified by search engine.

Experimental design



MALDI - TOF MS spectrum of peptides after protein digestion



MS
Peptide Mass Fingerprinting

MS spectrum contains masses of peptides formed by digestion of selected protein

C7250

MASCOT Peptide Mass Fingerprint

Your name: zz Email: zdrahal@sci.muni.cz

Search title:

Database(s): SwissProt, NCBInr, contaminants, cRAP

Enzyme: Trypsin

Allow up to: 1 missed cleavages

MS
Peptide Mass Fingerprinting

Taxonomy: All entries

Fixed modifications: --- none selected ---

Variable modifications: --- none selected ---

Acetyl (K)
Acetyl (N-term)
Acetyl (Protein N-term)
Amidated (C-term)
Amidated (Protein C-term)
Ammonia-loss (N-term C)
Biotin (K)
Biotin (N-term)
Carbamidomethyl (C)
Carbamyl (K)
Carbamyl (N-term)

Display all modifications

Protein mass: kDa

Peptide tol. ±: 1.2 Da

Mass values: MH^+ M_r $M\text{-}H^-$ Monoisotopic Average

Data file: Procházet... Soubor nevybrán.

Query

Data input:

http://www.matrixscience.com/cgi/search_form.pl?FORMVER=2&SEARCH=PMF

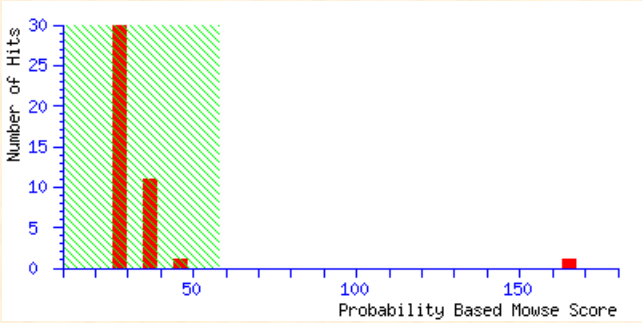
Decoy: Report top: AUTO hits

Start Search ... Reset Form

Mascot Search Results

MS
Peptide Mass Fingerprinting

C7250



Database : MSDB 20021127 (1019653 sequences)
Timestamp : 26 Jan 2003 at 10:36:50 GMT
Top Score : 165 for S18600, glutamate-ammonia ligase ....

1. S18600 **Mass:** 47780 **Total score:** 165 **Peptides matched:** 12
glutamate-ammonia ligase (EC 6.3.1.2) precursor, chloroplast (clone lambdaAtgsl1) - Arabidopsis thaliana
2. S32228 **Mass:** 47714 **Total score:** 76 **Peptides matched:** 7
glutamate-ammonia ligase (EC 6.3.1.2) precursor - rape - Brassica napus

Sequence Coverage: **44%**

```
1   MAQILAASPT CQMRVPKHSS VIASSSKLWS SVVLKQKKQS NNKVRGFRVL
51  ALQSDNSTVN RVETLLNLDT KPYSDRIIAE YIWIGGSGID LRSKSRTIEK
101 PVEDPSELPK WNYDGSSTGQ APGEDSEVIL YPQAIFRDPF RGGNNILVIC
151 DTWTPAGEPI PTNKRAKAAE IFSNKKVSGE VPWFGIEQEY TLLQQNVKWP
201 LGWPVGAFPG PQGPYYCGVG ADKIWGRDIS DAHYKACLYA GINISGTNGE
251 VMPGQWEFQV GPSVGIDAGD HVWCARYLLE RITEQAGVVL TLDPKPIEGD
301 WNGAGCHTNY STKSMREEGG FEVIKKAILN LSLRHKEHIS AYGEGNERRL
351 TGKHETASID QFSWGVANRG CSIRVGRDTE AKGKGYLEDR RPASNMDPYI
401 VTSLLAETTL LWEPTLEAEA LAAQKLSLNV
```

www.matrixscience.com

MS
Peptide Mass Fingerprinting

- ✓ fast identification technique
- ✓ usually MALDI-MS (storage of samples)

- ✗ only known proteins
- ✗ identification based only on mass of whole peptides
- ✗ detailed structural information is not possible
  (modification type – yes?, localization – no)
- ✗ protein separation always necessary

MS/MS
MS/MS Ion Search

principle:

- ***MS/MS analysis*** *of specifically digested peptides*

  obtained set of ***"masses" (m/z) of fragments of individual peptides***

  = specific information suitable for protein identification

- *database search*

- method suitable for protein mixtures

  → separation of intact proteins is not necessary

MS/MS
MS/MS Ion Search

Measured fragmentation maps (i.e. sets of masses (or m/z) of fragments formed during MS/MS of individual peptides) are searched against database of protein sequences by search engine.

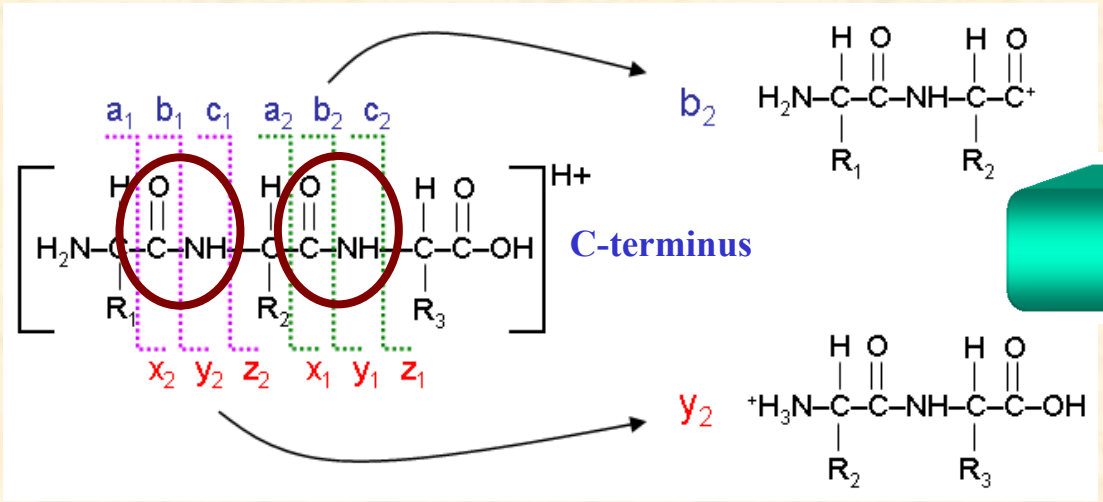
At first, database search engine prepares theoretical peptide map for a protein sequence in database, subsequently, it calculates theoretical fragmentation map for each peptide of the corresponding peptide map (according to given fragmentation rules) and then these *in-silico* prepared fragmentation maps are compared with our experimentally obtained fragmentation maps of analyzed peptides. The engine performs this operation for each protein sequence in database.

Software calculates individual score for each peptide, score value higher than limit peptide score determines significant similarity between theoretical and measured fragmentation map – significant peptide identification.
In final, search engine assort peptides to corresponding protein sequences (the more peptides with significant score per protein – the more reliable protein identification). The software also calculate protein score which is derived from individual peptide score as a tool for setting up results.

MS/MS fragmentation of peptides

- peptides consist of individual aminoacids which are connected by peptide bond
- during fragmentation (e.g. CID), peptide is fragmented preferentially at peptide bond and thus:
  - all peptide bonds might be fragmented (in each precursor molecule different ones) forming set of fragments with various number of aminoacids
  - differences in *m/z* (or mass) of „neighbouring fragments“ determines type of terminal aminoacid in the longer fragment
- serie of fragment ions are formed (b – y, a – x, c – z) which can be used for *de novo* primary structure elucidation; moreover they are predictable and they can be used for database search based protein identification even if they are not complete

Outline of tripeptide fragmentation

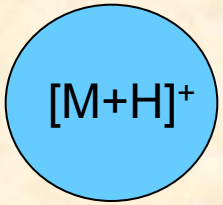


Roepstorff P. and Fohlman, J., Biomed. Mass Spectrom., 11 (11), 601 (1984)

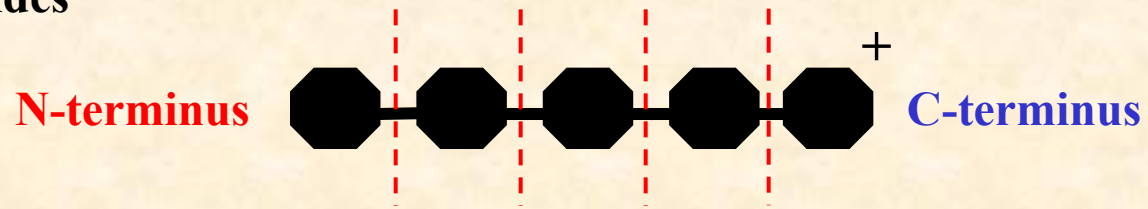
Peptide fragmentation - CID

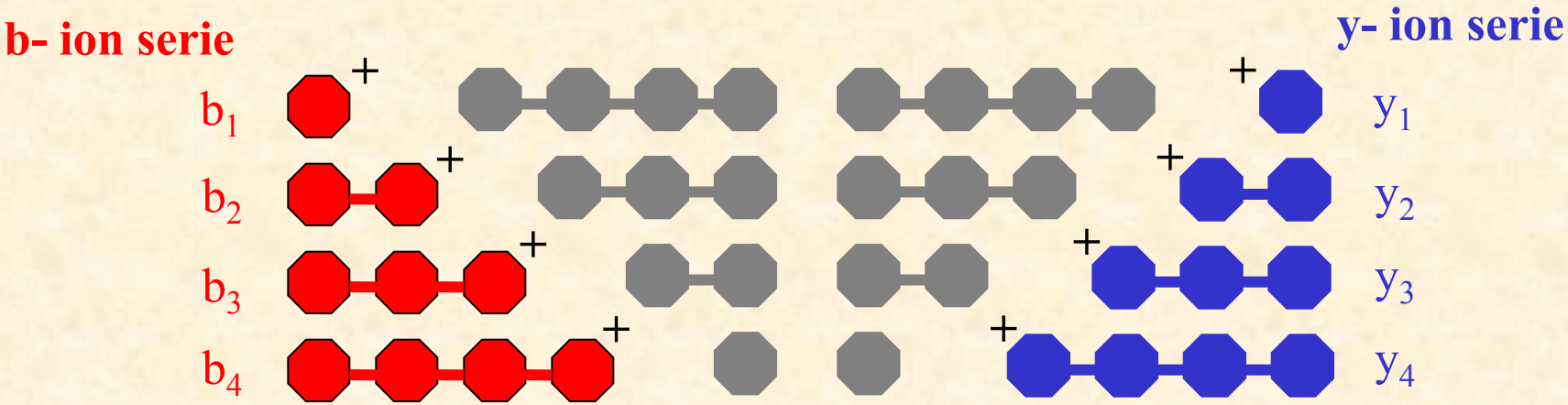
C7250

MS of peptides



MS/MS of peptides

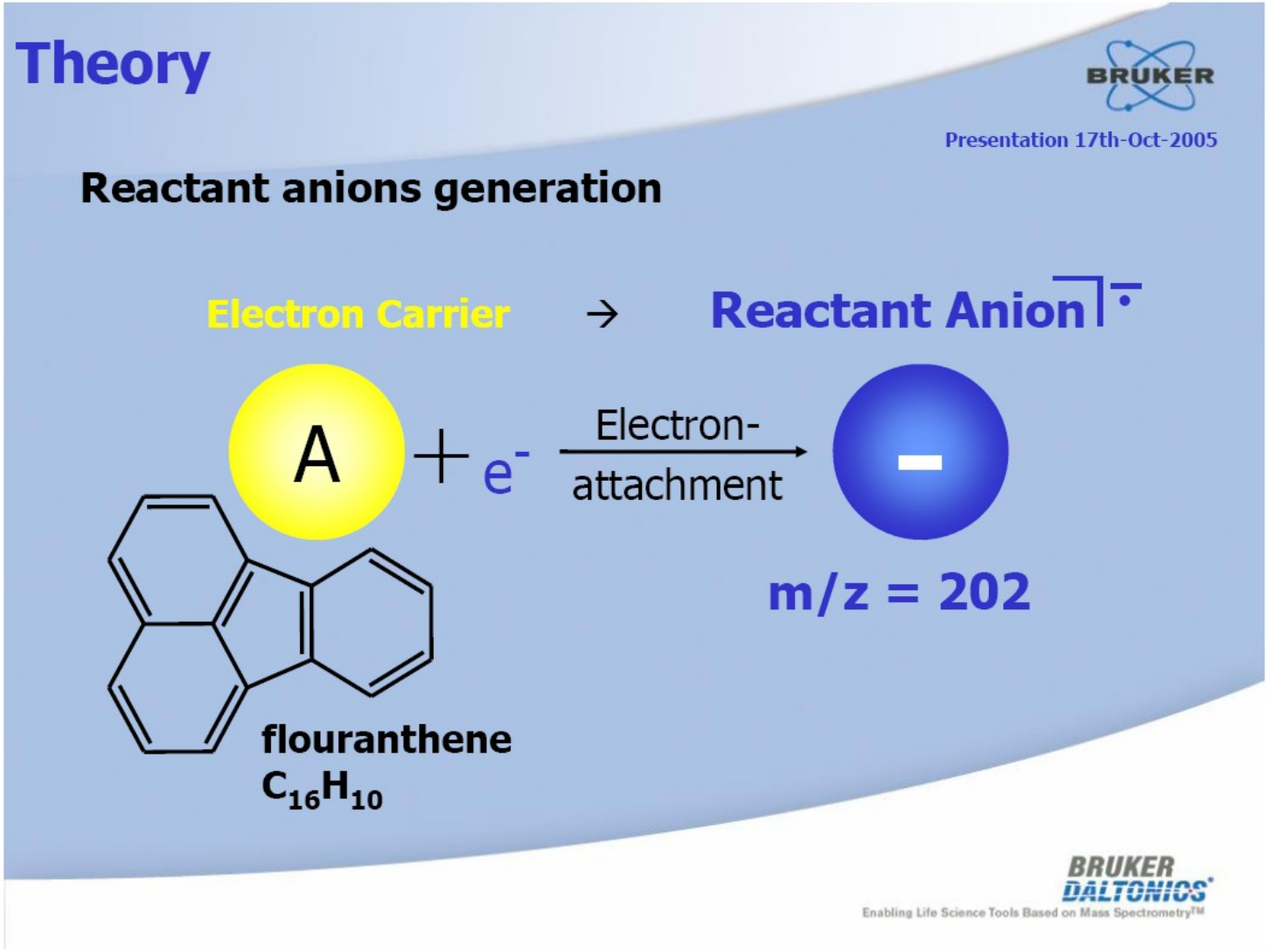




fragmentatation maps for individual peptides

Peptide fragmentation - ETD

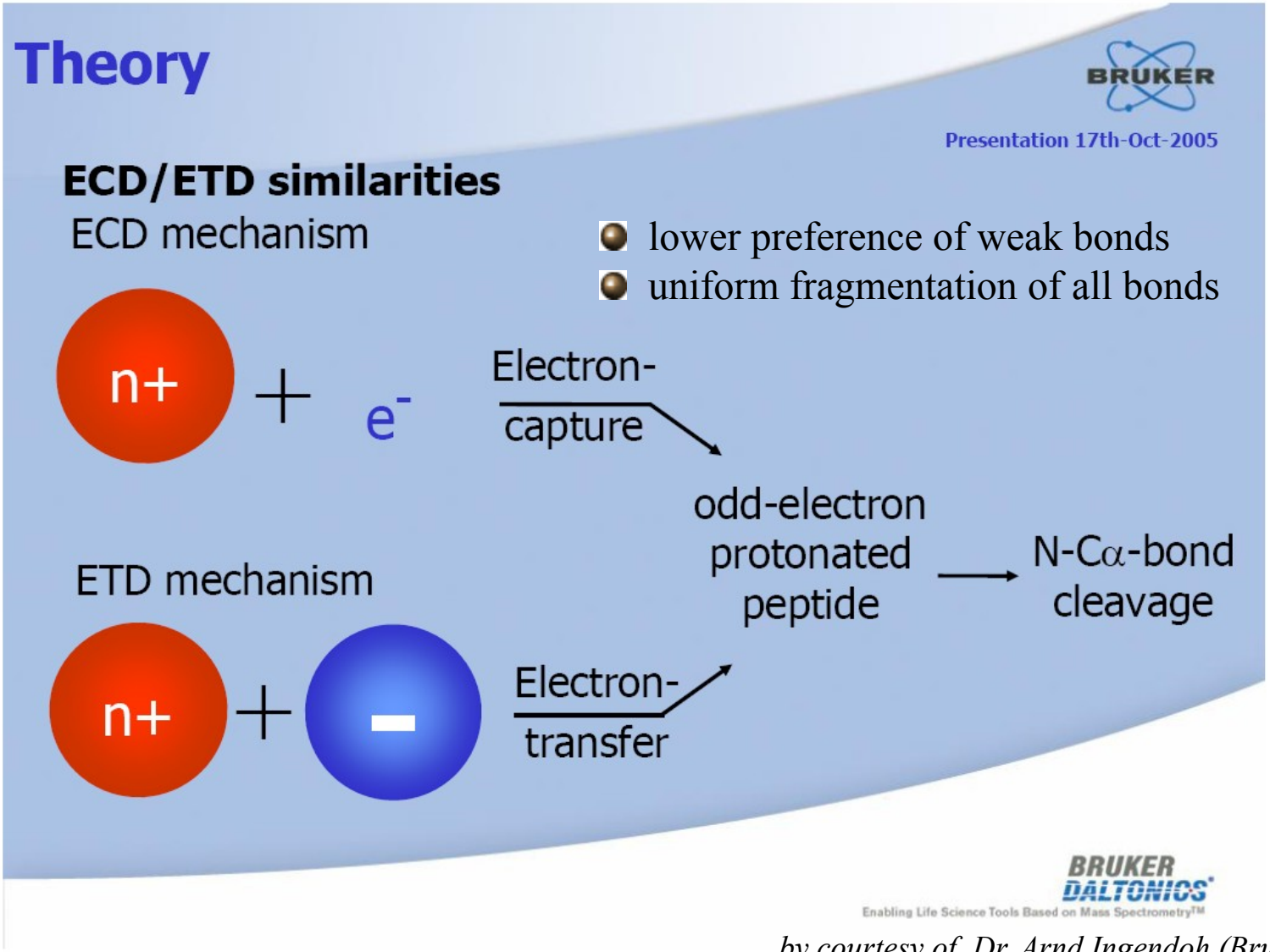
C7250



by courtesy of Dr. Arnd Ingendoh (Bruker)

Peptide fragmentation - ETD

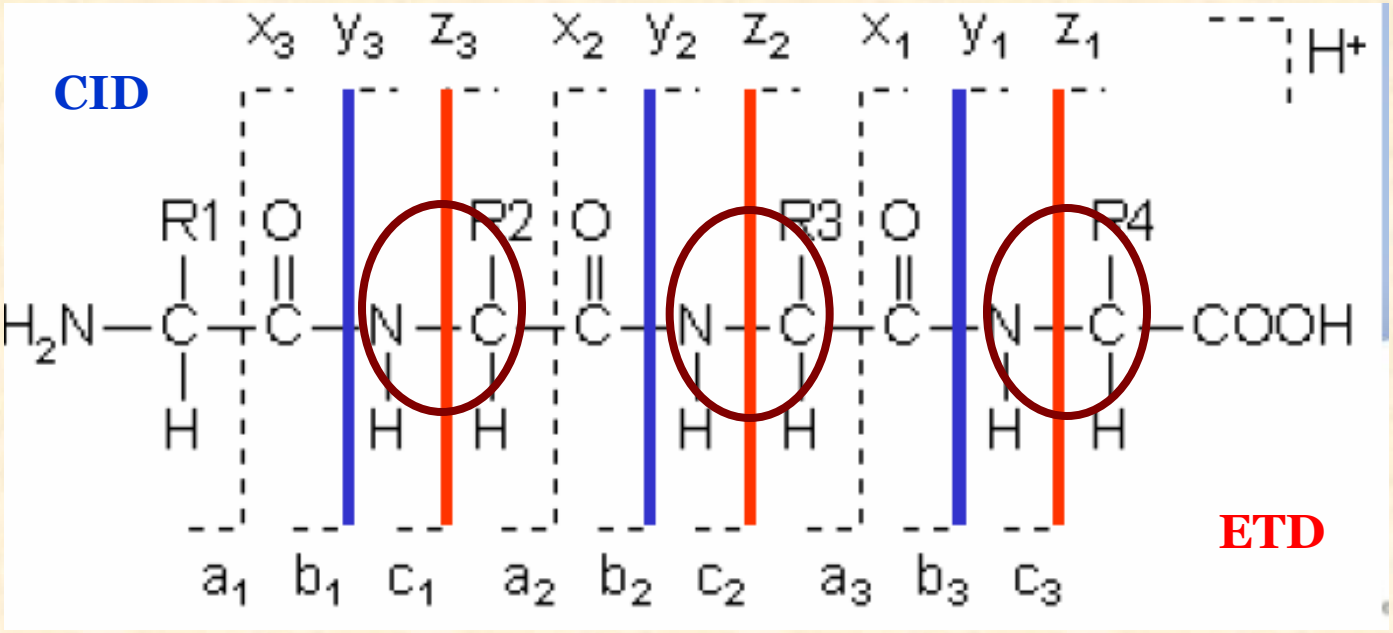
C7250



by courtesy of Dr. Arnd Ingendoh (Bruker)

Peptide fragmentation - ETD

C-terminus *(z- serie)*



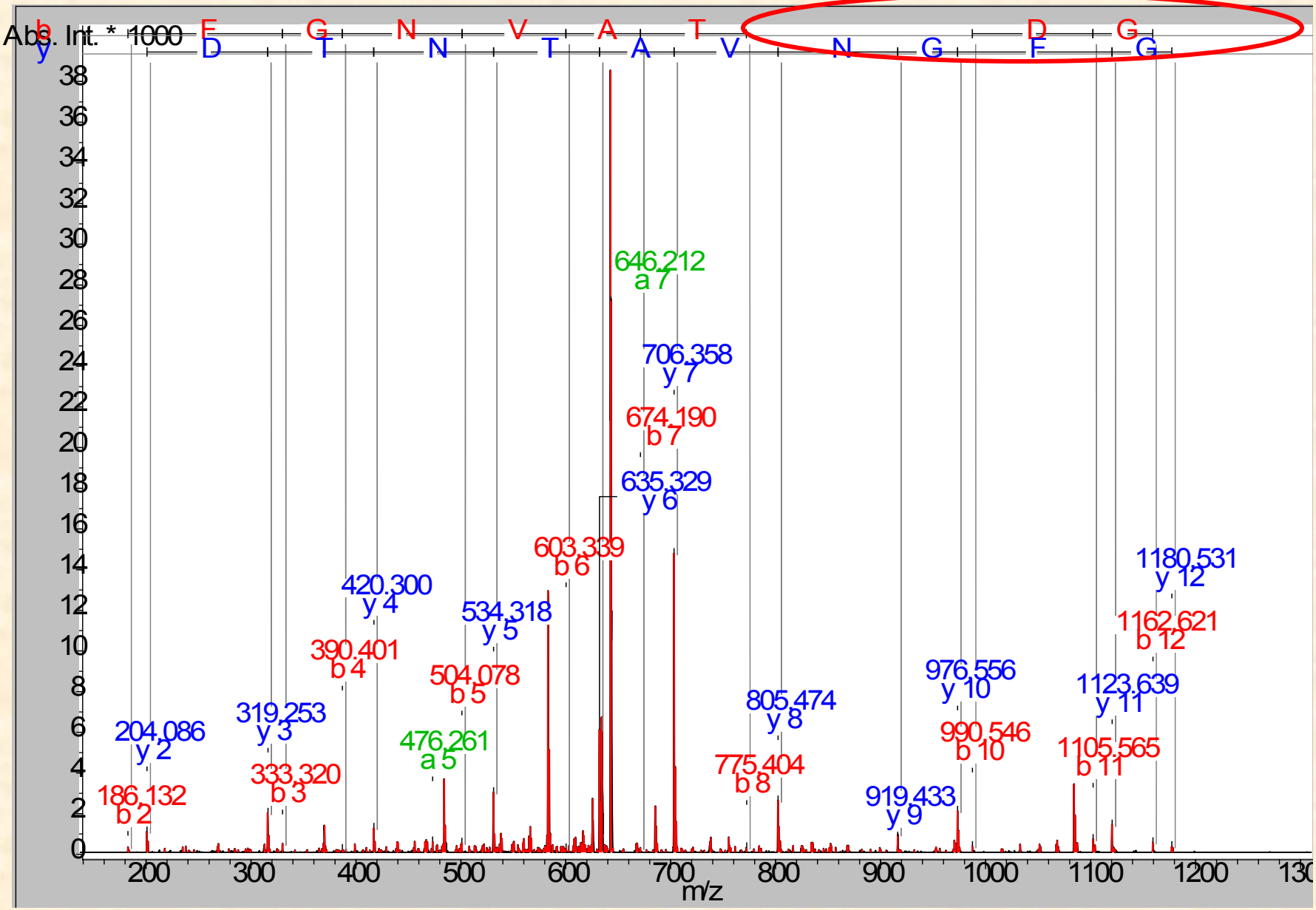
N-terminus *(c-serie)*

MS/MS
MS/MS Ion Search

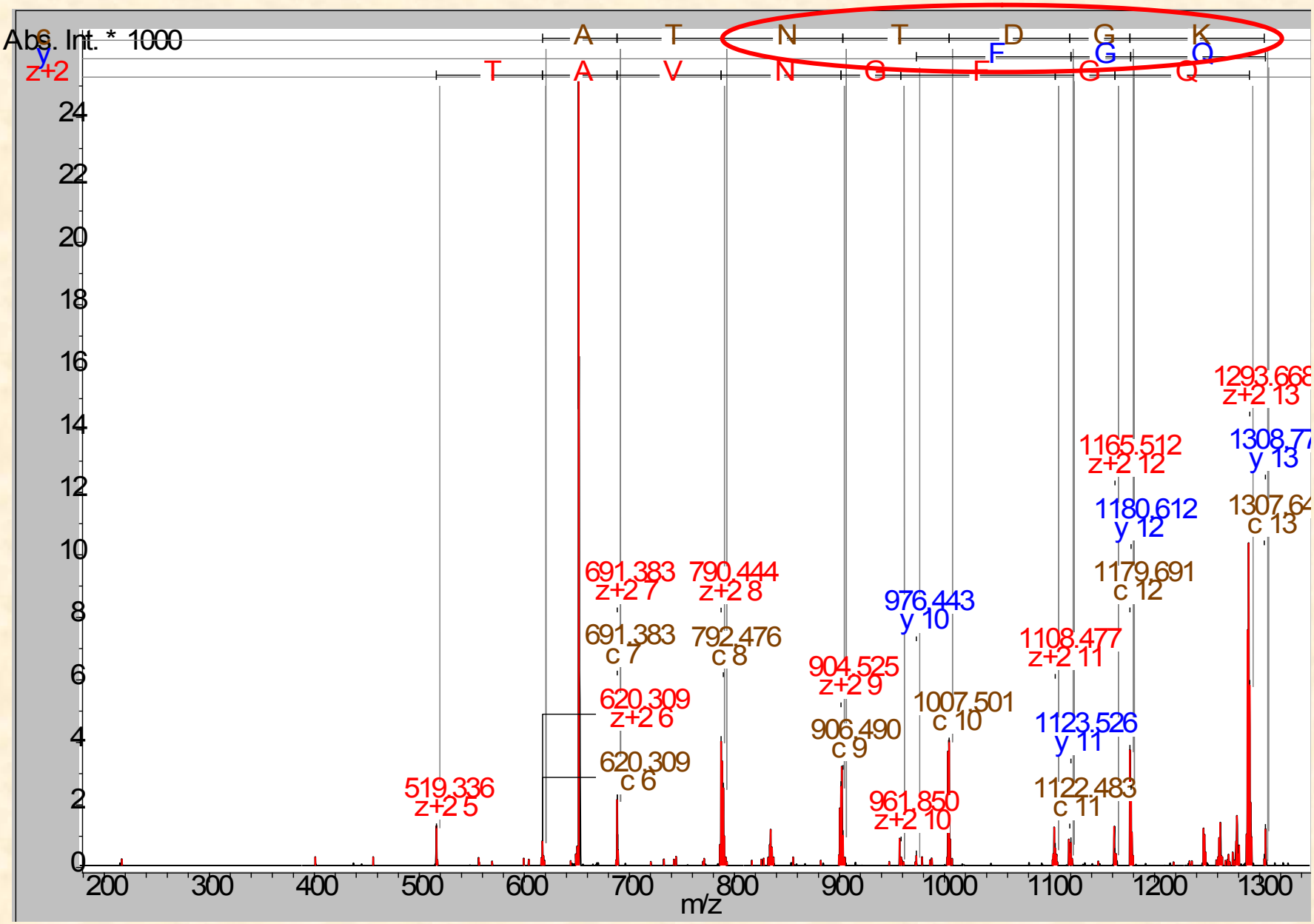
MS/MS fragmentation of peptide – *in silico*: ALELFR

a	b	c	Res:	x	y	z
44.050	72.045	89.071	1 **Ala** 6	-	-	-
157.134	185.129	202.156	2 **Leu** 5	701.363	675.384	658.358
286.177	314.172	331.198	3 **Glu** 4	588.279	562.300	545.273
399.261	427.256	444.282	4 **Leu** 3	459.237	433.257	416.231
546.329	574.324	591.351	5 **Phe** 2	346.153	320.173	303.147
-	-	-	6 **Arg** 1	199.084	173.105	156.078

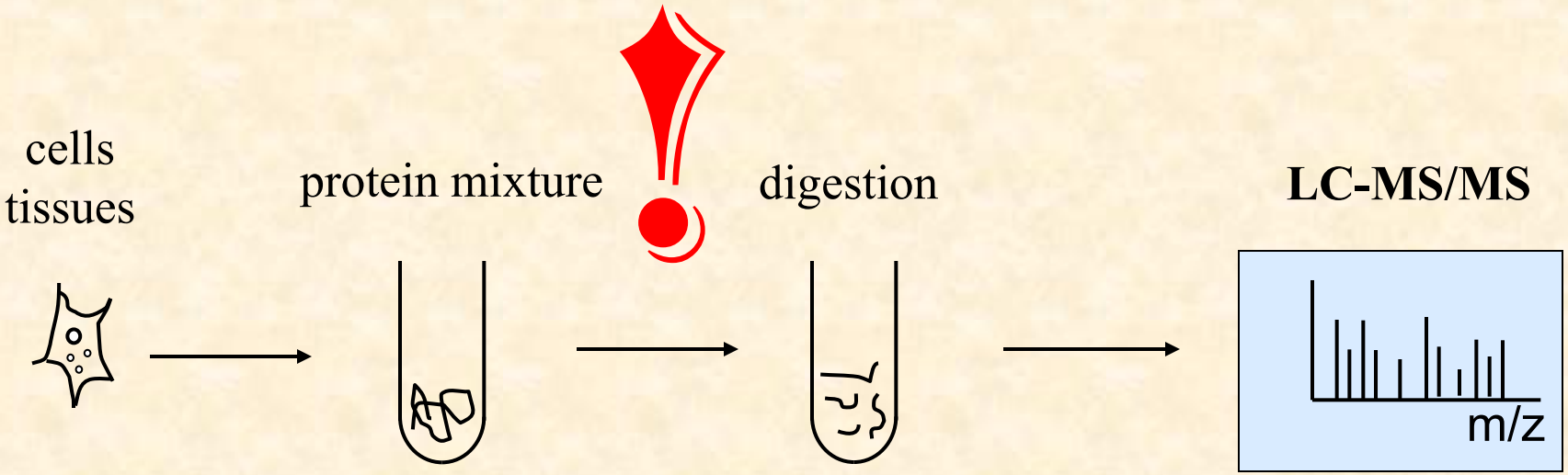
CID spectrum of peptide QGFGNVATNTDGK (b, y) C7250



ETD spectrum of peptide QGFGNVATNTDGK (c, y, z) C7250

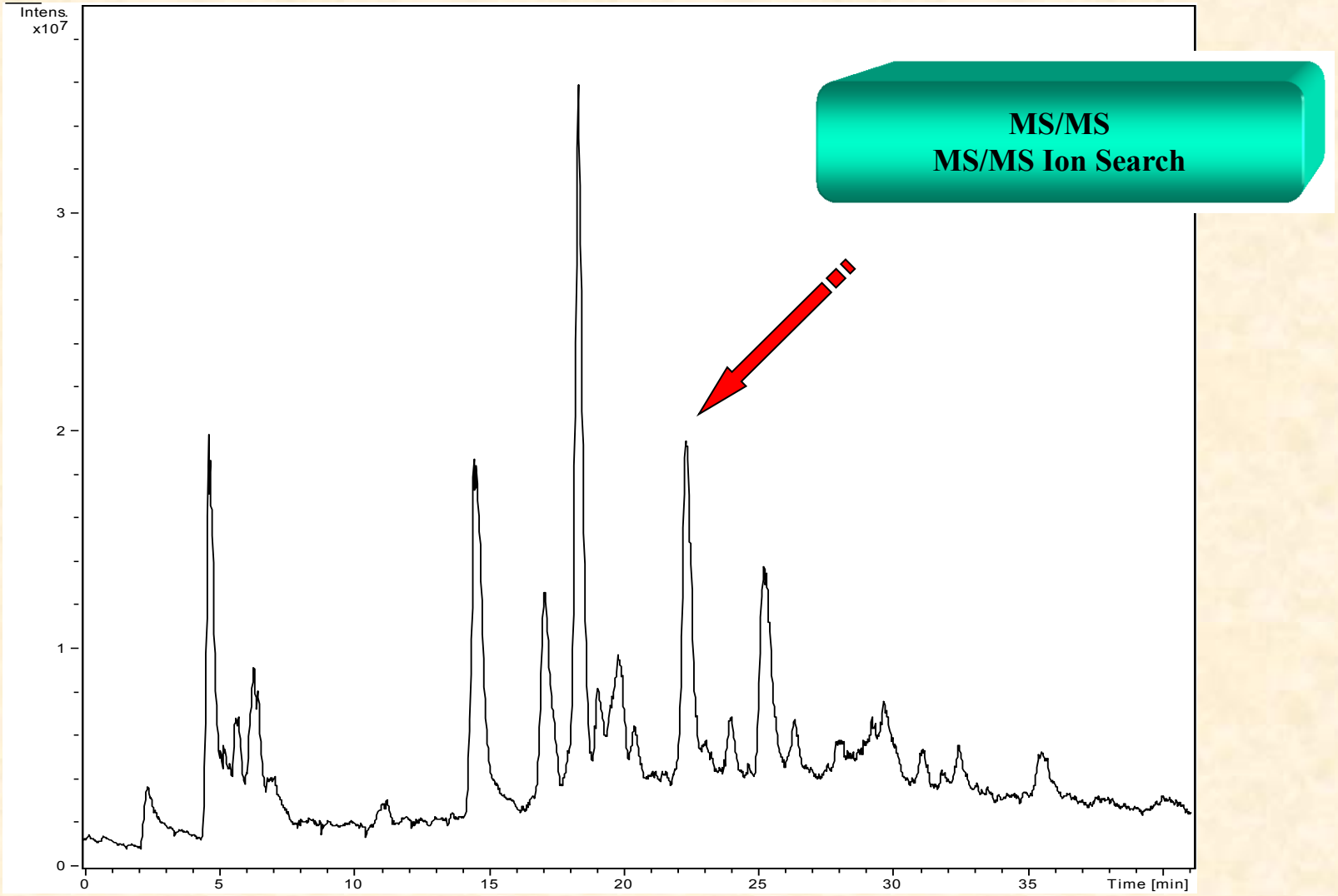


Experimental design



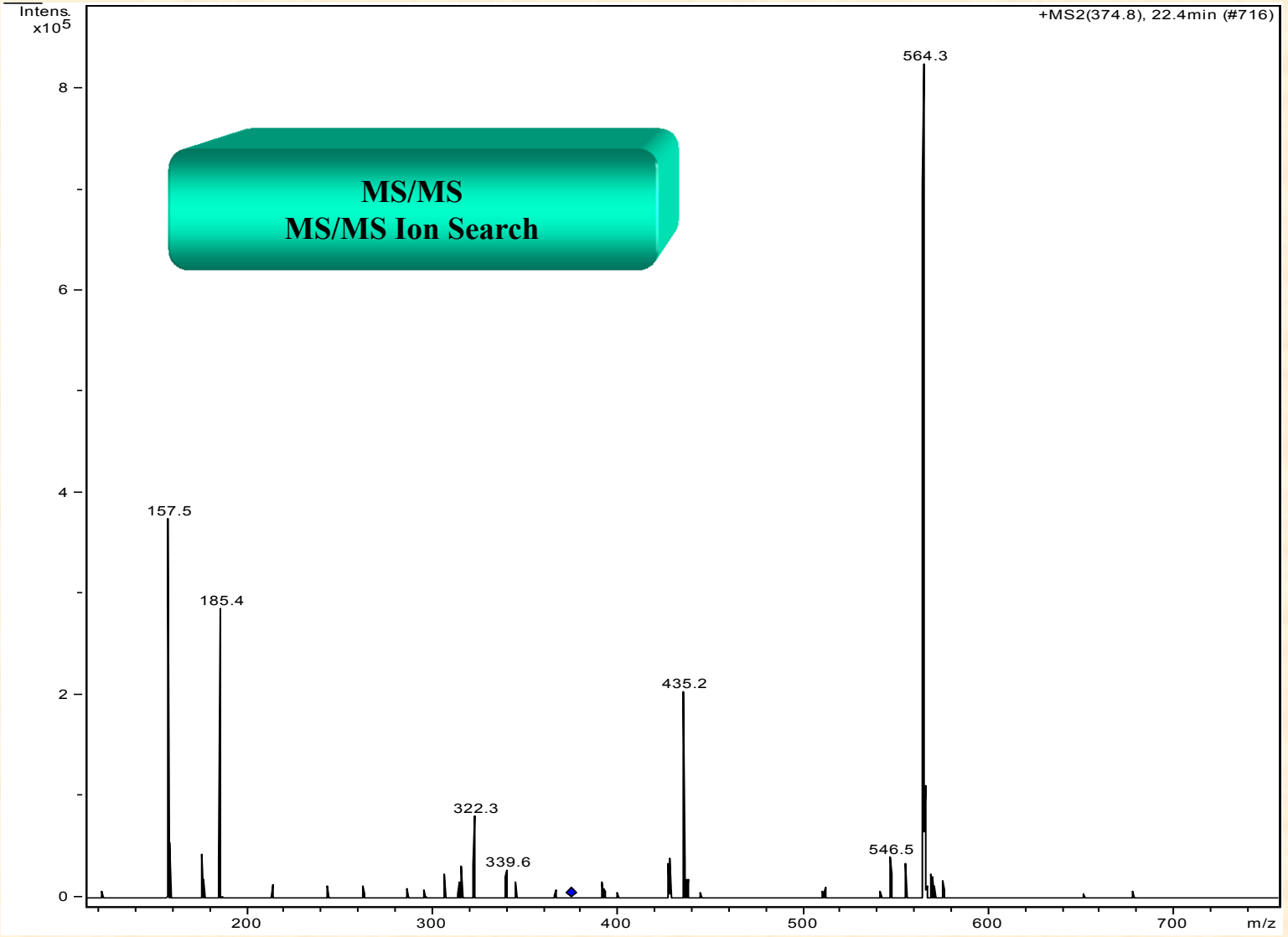
MS/MS
MS/MS Ion Search

Separation of myoglobin tryptic peptides

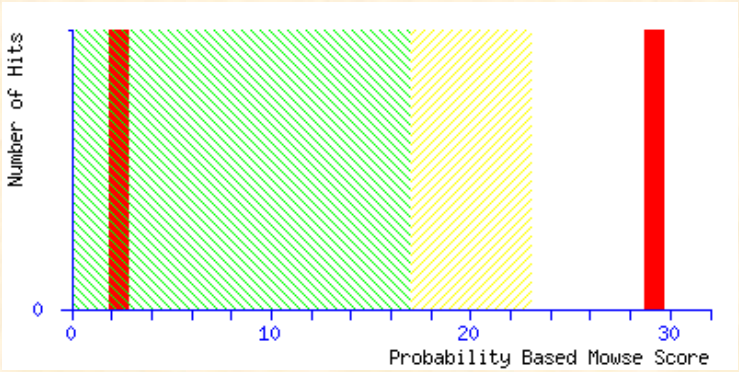


C7250

MS/MS spectrum of peptide, 374.8 Da, 2+



Mascot Search Results



Database : MSDB 20040329 (1457190 sequences)
Taxonomy : Other mammalia (30839 sequences)
Timestamp : 20 May 2004 at 06:55:04 GMT

MS/MS
MS/MS Ion Search

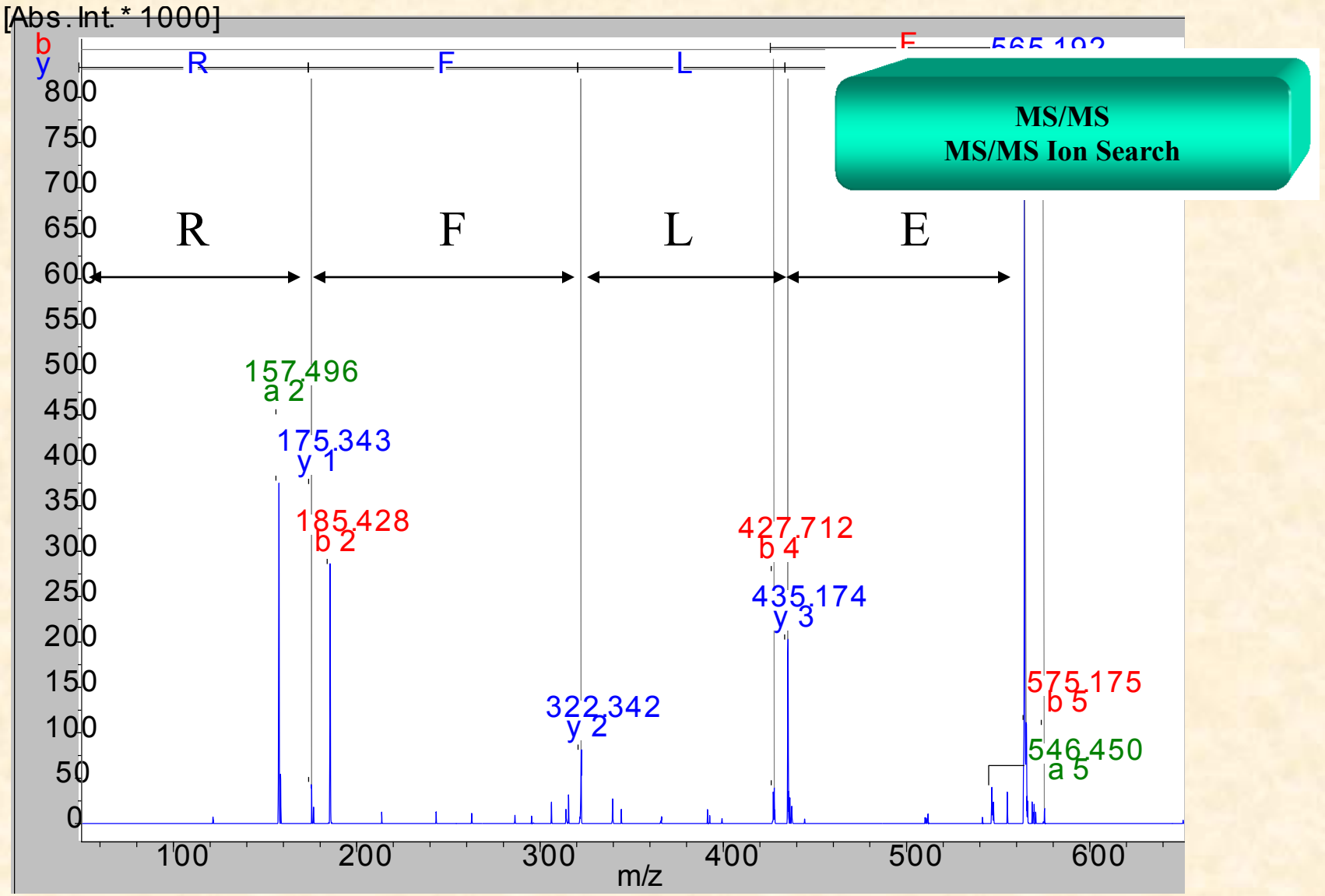
1. MYBD **Mass:** 16955 **Score:** 29 **Peptides matched:** 1
myoglobin - Eurasian badger (tentative sequence)

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
1.	374.81	747.60	747.43	0.18	0	29	0.014	1	ALELFR

	Accession	Mass	Score	Description
1.	Q865L4	3798	55	Myoglobin (Fragment).- Bos taurus (Bovine).
2.	1A6K	17004	50	myoglobin - sperm whale
3.	1MNJB	16734	50	myoglobin (met, ph 7.1) mutant with his 64 replaced by val …..
4.	1MNKA	16722	50	myoglobin (aquomet, ph 7.1) mutant with his 64 replaced by val…..
5.	1DTMA	17052	50	recombinant sperm whale myoglobin variant h93g mutant YES - sperm whale

C7250

MS/MS spectrum of peptide, 747 Da, ALELFR



C7250

Probability is mathematics only

PD, taxonomy: All entries Individual ions scores > **25** indicate identity or extensive homology

1. ct74_rgenePd06_2913 **Mass:** 49299 **Score:** 46 **Queries matched:** 1 **emPAI:** 0.07
ct74_rgenePd06_2913

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
272	711.4	1420.7	1420.698	0.10	0	46	0.00025	1	R.WFSLDEINELR.R

NCBInr, taxonomy: Homo Sapiens Individual ions scores > **39** indicate identity or extensive homology

6. gi|94982457 **Mass:** 105502 **Score:** 229 **Queries matched:** 7 **emPAI:** 0.20
actinin alpha 1 isoform b [Homo sapiens]

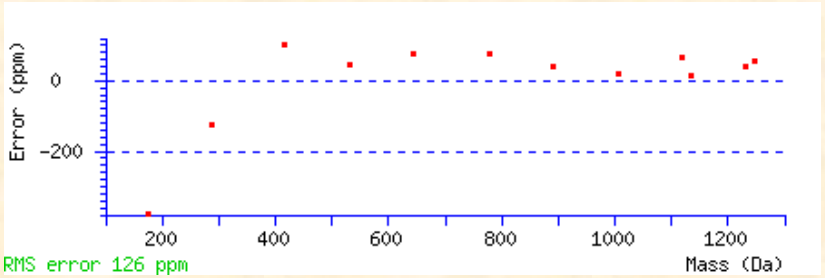
Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
157	600.3	1198.6	1198.623	0.04	0	40	0.057	1	R.DLLLDPAWEK.Q
219	663.4	1324.8	1324.648	0.20	1	15	16	7	R.RDQALTEEHAR.Q
241	686.9	1371.8	1371.779	0.11	0	72	3.4e-005	1	K.LMLLLEVISGER.L
248	693.9	1385.9	1385.766	0.14	0	87	1.1e-006	1	R.VGWEQLLTTIAR.T
272	711.4	1420.7	1420.698	0.10	0	60	0.00046	1	K.GYEEWLLNEIR.R
274	715.4	1428.8	1428.757	0.05	0	79	5.3e-006	1	R.TINEVENQILTR.D
334	780.4	2338.3	2338.180	0.16	0	66	0.0001	1	K.IDQLEGDHQLIQEALIFDNK.H

MS/MS
MS/MS Ion Search

C7250

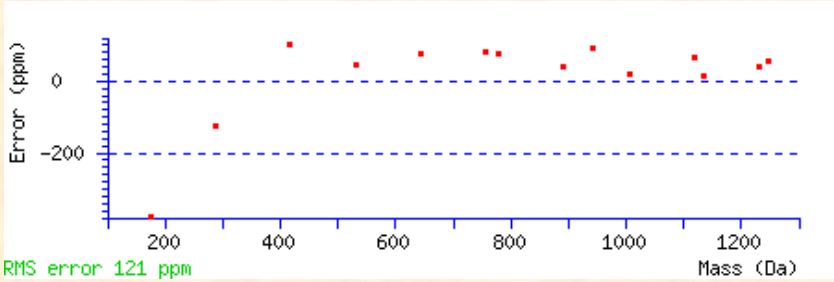
WFSLDEINELR

#	b	Seq.	y	#
1	187.0866	W		11
2	334.1550	F	1235.6266	10
3	421.1870	S	1088.5582	9
4	534.2711	L	1001.5262	8
5	649.2980	D	888.4421	7
6	778.3406	E	773.4152	6
7	891.4247	I	644.3726	5
8	1005.4676	N	531.2885	4
9	1134.5102	E	417.2456	3
10	1247.5943	L	288.2030	2
11		R	175.1190	1



GYEEWLLNEIR

#	b	Seq.	y	#
1	58.0287	G		11
2	221.0921	Y	1364.6845	10
3	350.1347	E	1201.6212	9
4	479.1773	E	1072.5786	8
5	665.2566	W	943.5360	7
6	778.3406	L	757.4567	6
7	891.4247	L	644.3726	5
8	1005.4676	N	531.2885	4
9	1134.5102	E	417.2456	3
10	1247.5943	I	288.2030	2
11		R	175.1190	



MS/MS
MS/MS Ion Search

- ✓ more reliable identification based on peptide fragmentation *(protein is identifiable based on MS/MS spectrum of one peptide)*
- ✓ MS/MS data allows sequence determination of unknown proteins (*de novo sequencing*)
- ✓ MS/MS techniques are suitable for detailed characterization of sequence and PTMs

- ✘ more technically (financially) demanding and more time-consuming than MS techniques

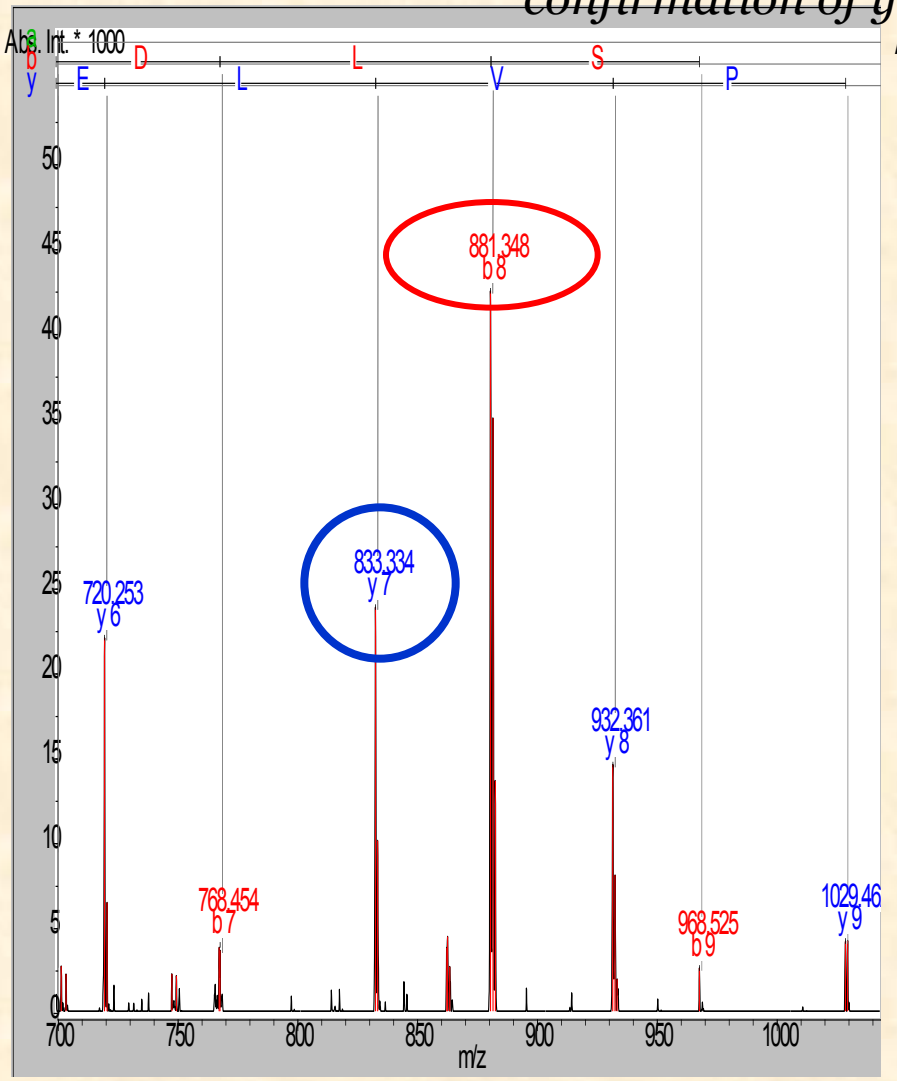
Determination of structure of unknown proteins

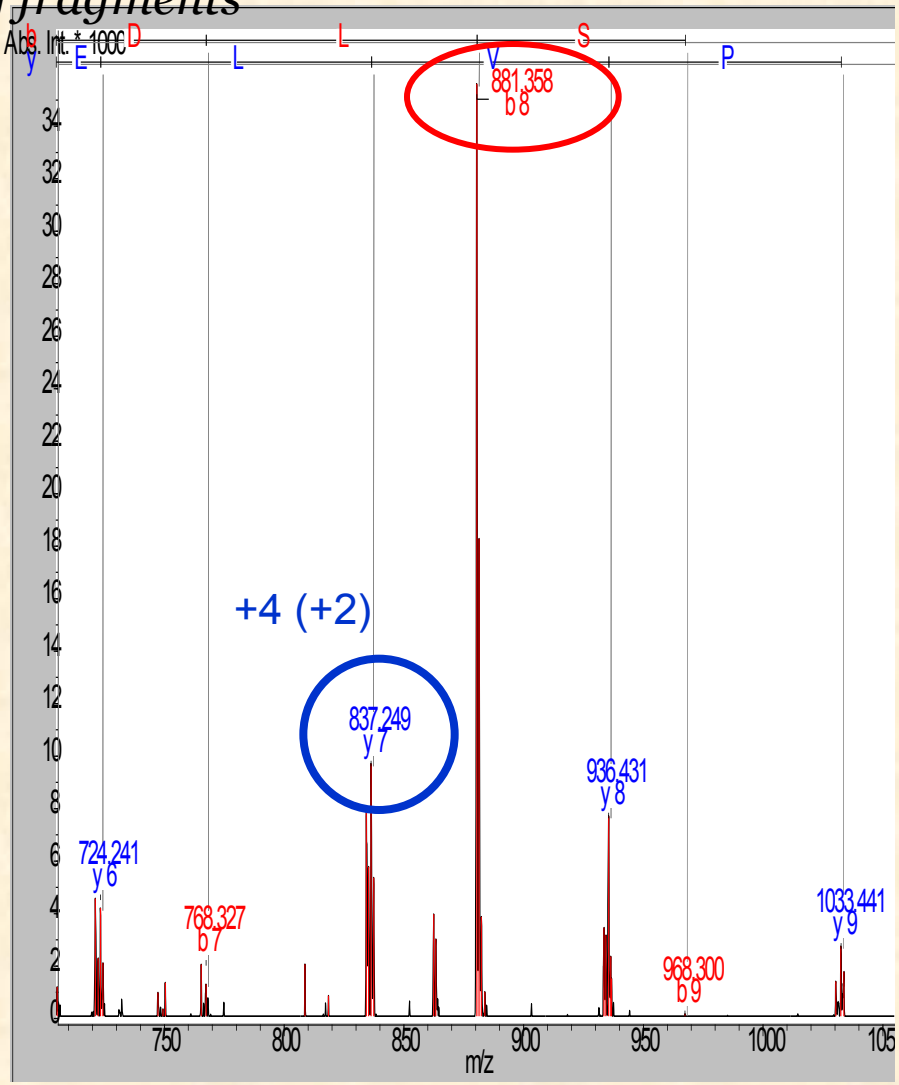
MS/MS
De novo sequencing

- different proteases ➡ overlap of peptides ➡ succession in sequence
- MS/MS of peptides (MALDI-MS/MS, LC-ESI-MS/MS)
- MS/MS spectra interpretation (manual, automatic - SWs)
- supporting information (BLAST…)

De novo sequencing of peptides – MS/MS + digestion in $H_2{}^{18}O$

confirmation of y fragments





Top Down

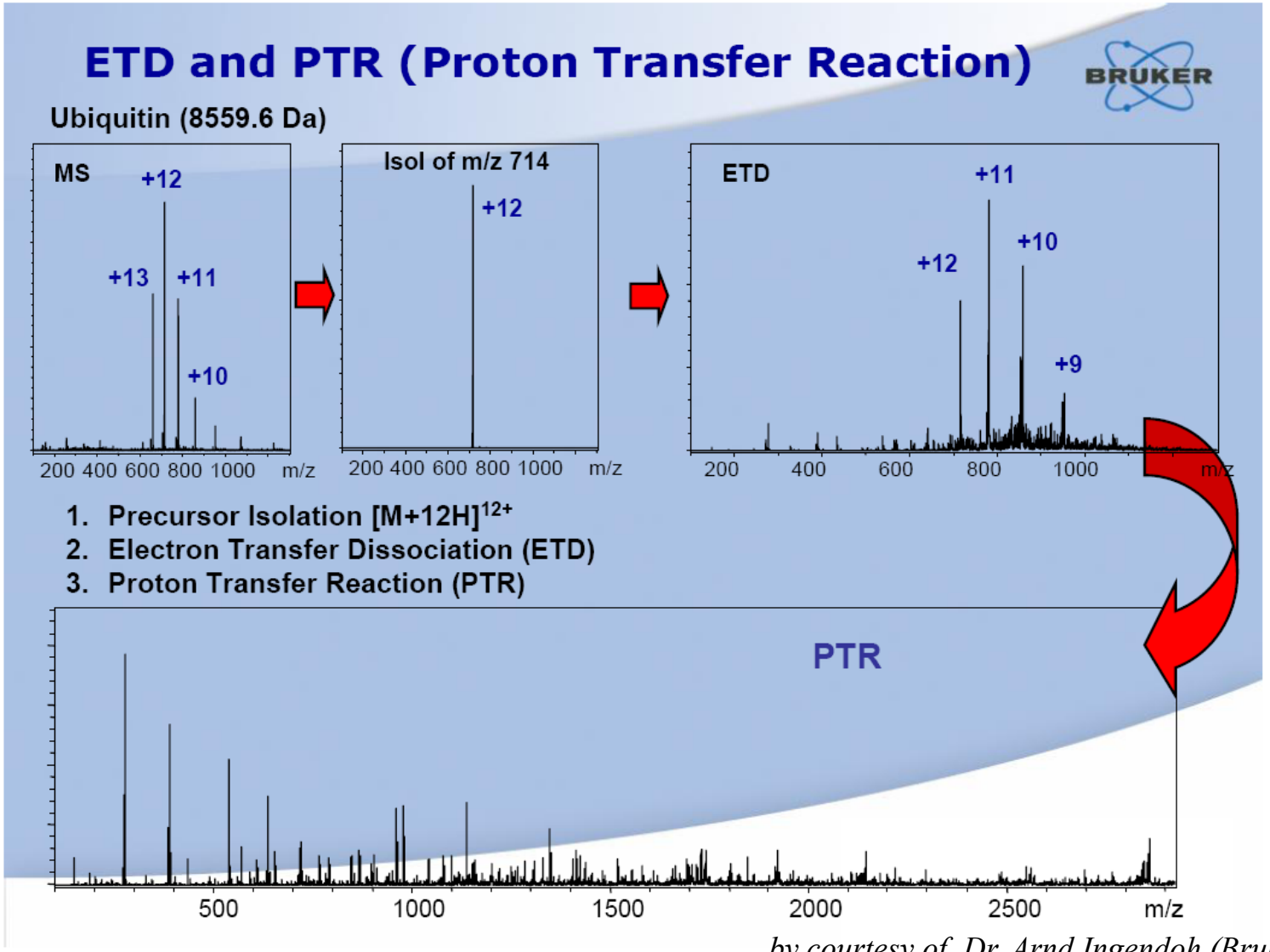
The ***top-down*** approach uses the mass of the intact proteins, individually or in mixtures, and then fragments the intact proteins inside the mass spectrometer without prior enzymatic digestion[3].

The advantages of top-down proteomics are the ability to measure the actual intact protein molecular weight, preserving both the entire protein sequence and the integrity of post-translational modifications. Currently, top-down proteomics are limited to FTICR instruments because of requirements for high resolving power, mass accuracy and complementary fragmentation methods.

Intact protein and fragment molecular weights can be searched against a corresponding database in a manner similar to that of the bottom up approach in order to provide protein identification[4-6].

3. Reid GE, McLuckey SA. *J Mass Spectrom.* 2002; **37**: 663.
4. Senko MW, Beu SC, McLafferty FW. *Anal Chem.* 1994; **66**: 415.
5. Mortz E, O'Connor PB, Roepstorff P, Kelleher NL, Wood TD,McLafferty FW, Mann M. *Proc Natl Acad Sci U S A* 1996; **93**: 8264.
6. Meng F, Cargile BJ, Patrie SM, Johnson JR, McLoughlin SM, Kelleher NL.*Anal Chem.* 2002; **74**: 2923.

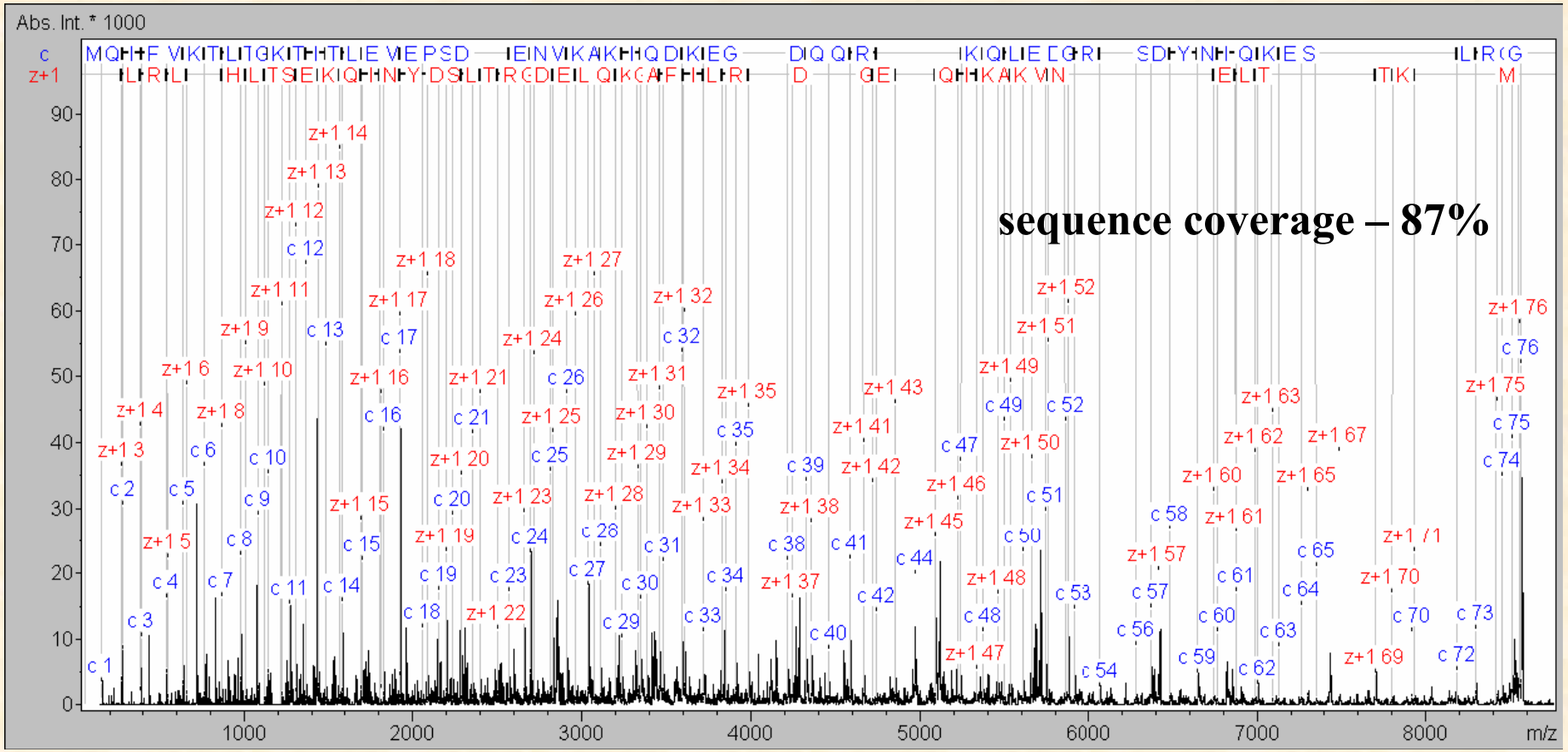
ETD/PTR (ESI-IT)



by courtesy of Dr. Arnd Ingendoh (Bruker)

ETD/PTR (ESI-IT)

Ubiquitin (8559.6 Da)

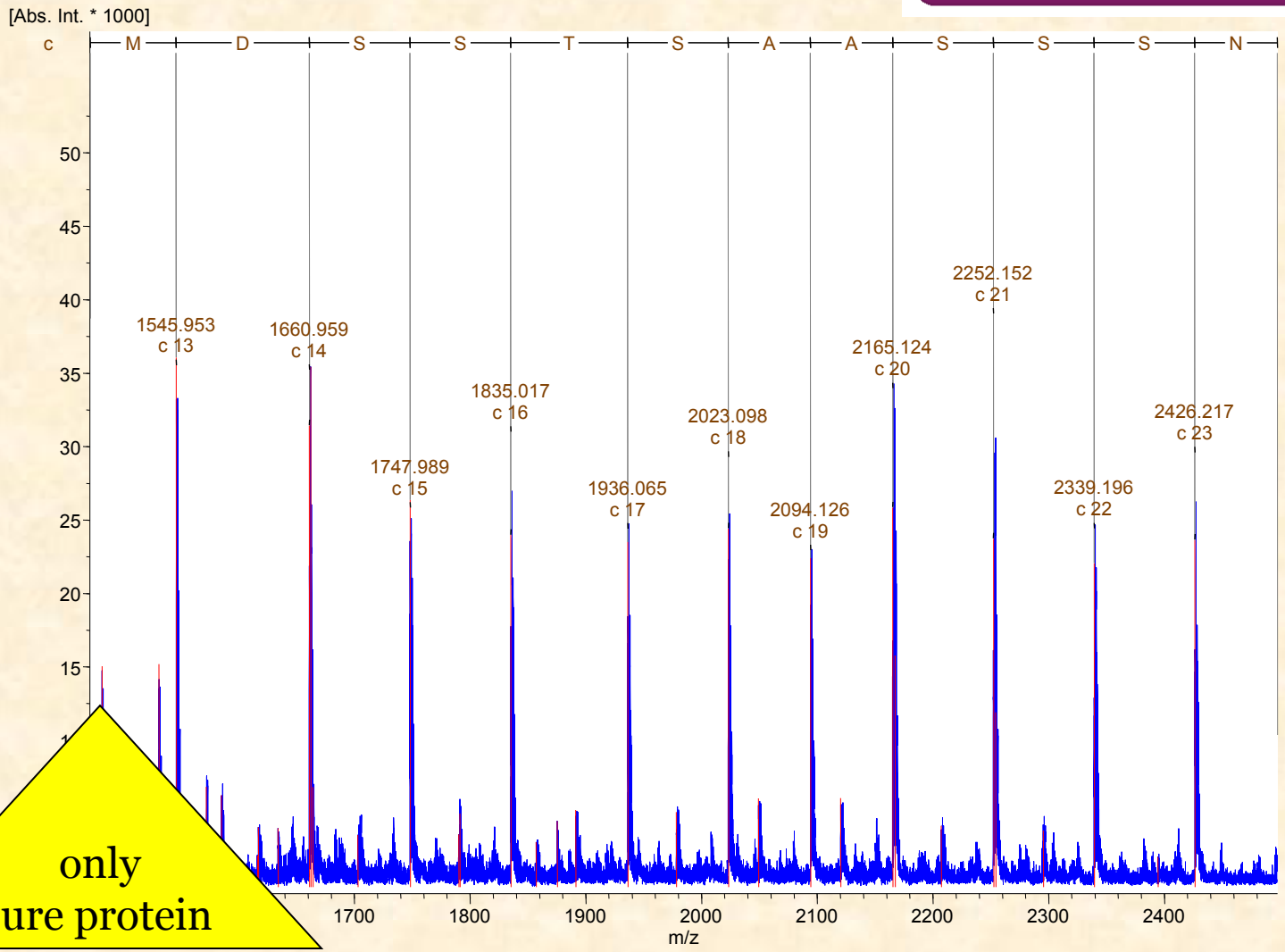


by courtesy of Dr. Arnd Ingendoh (Bruker)

C7250

In-source decay, ISD (Rnase B, 13,7 kDa)
MALDI- MS

Top-Down ISD

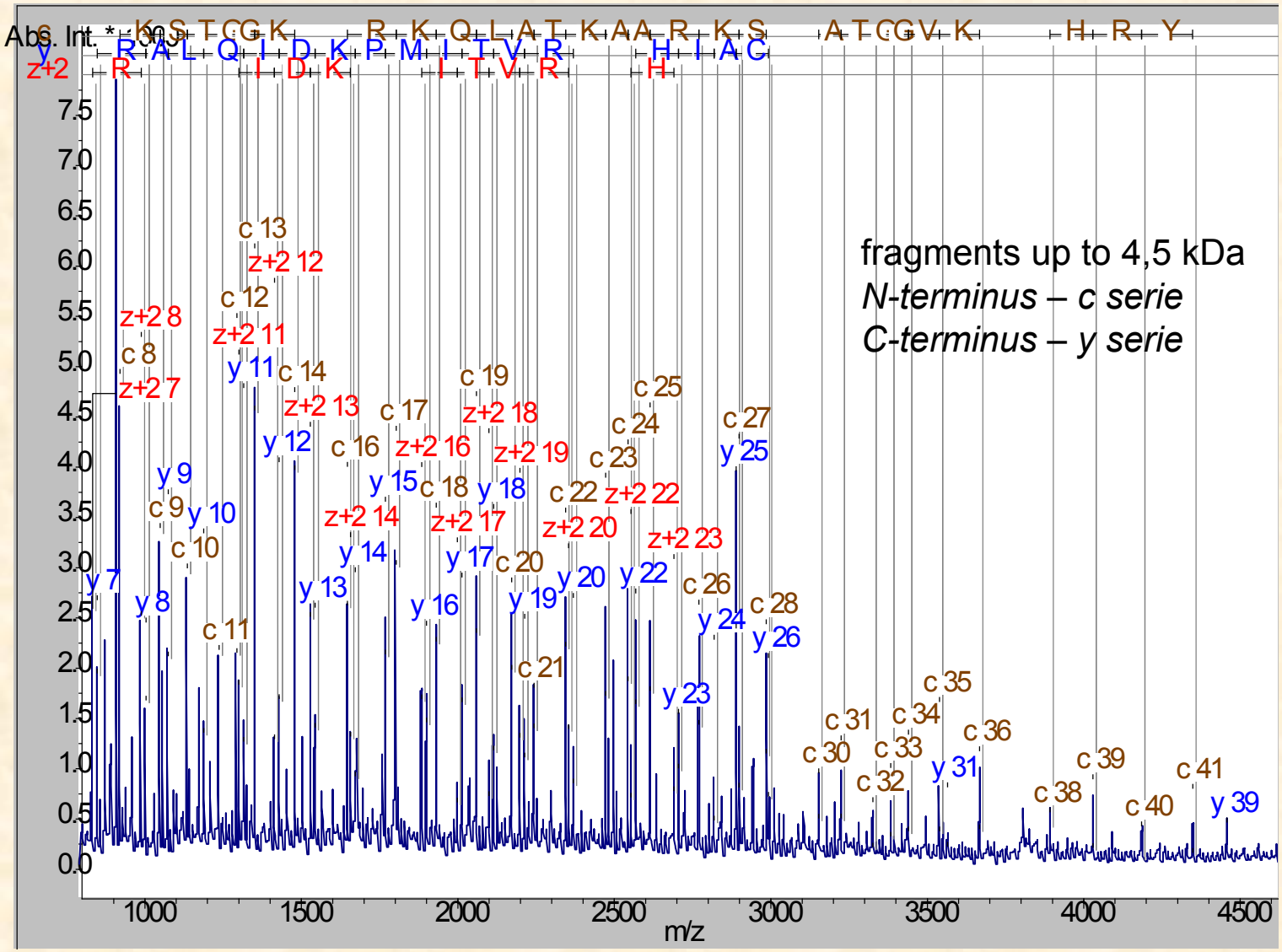


only
pure protein

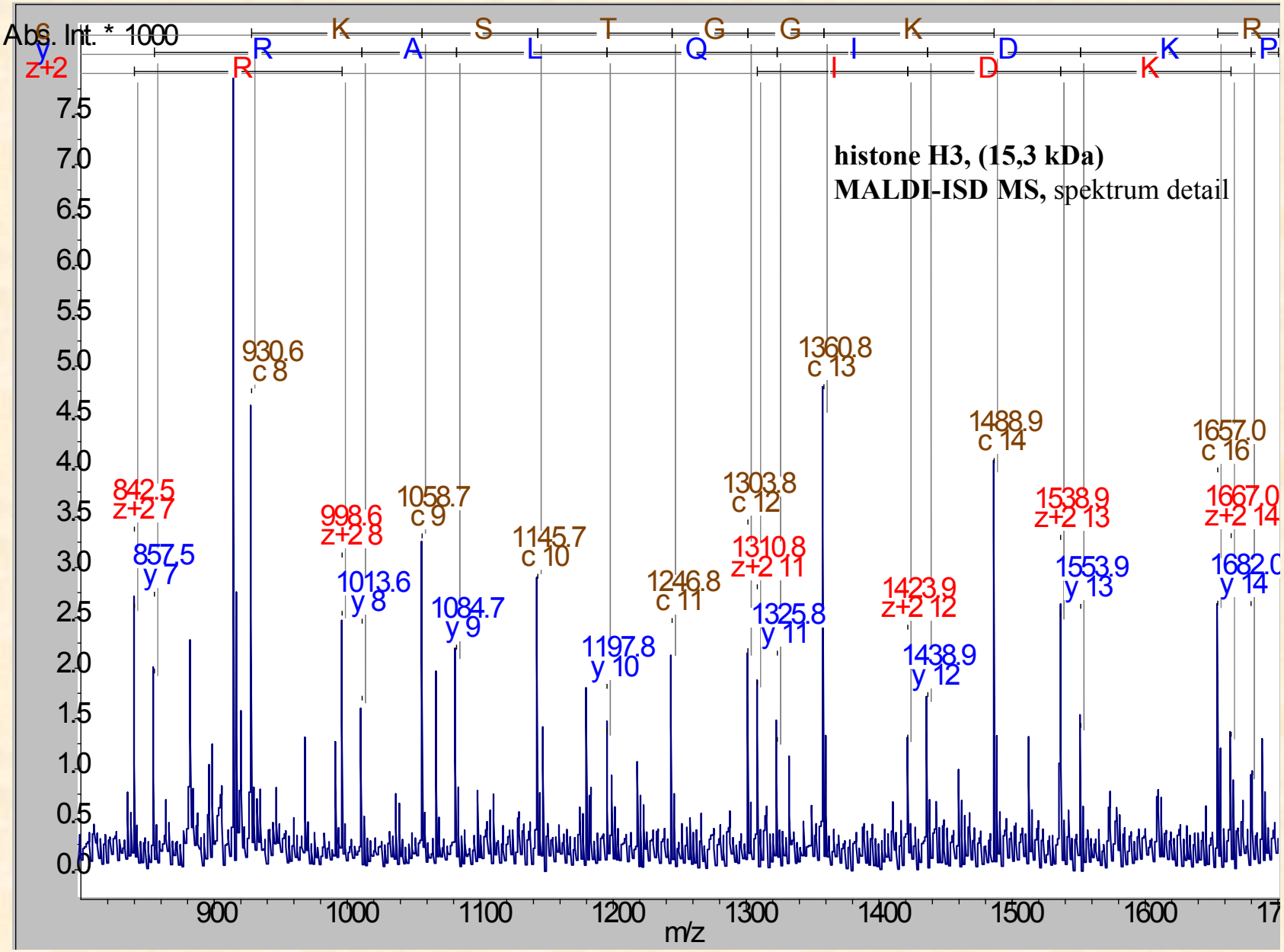
C7250

histone H3, (15,3 kDa)
MALDI-ISD MS

ISD



ARTKQTARKSTGGKAPR... .... KDIQLARRIRGERA



Characterization of therapeutic antibodies

C7250

Assessment of N- and C-terminal modification status.

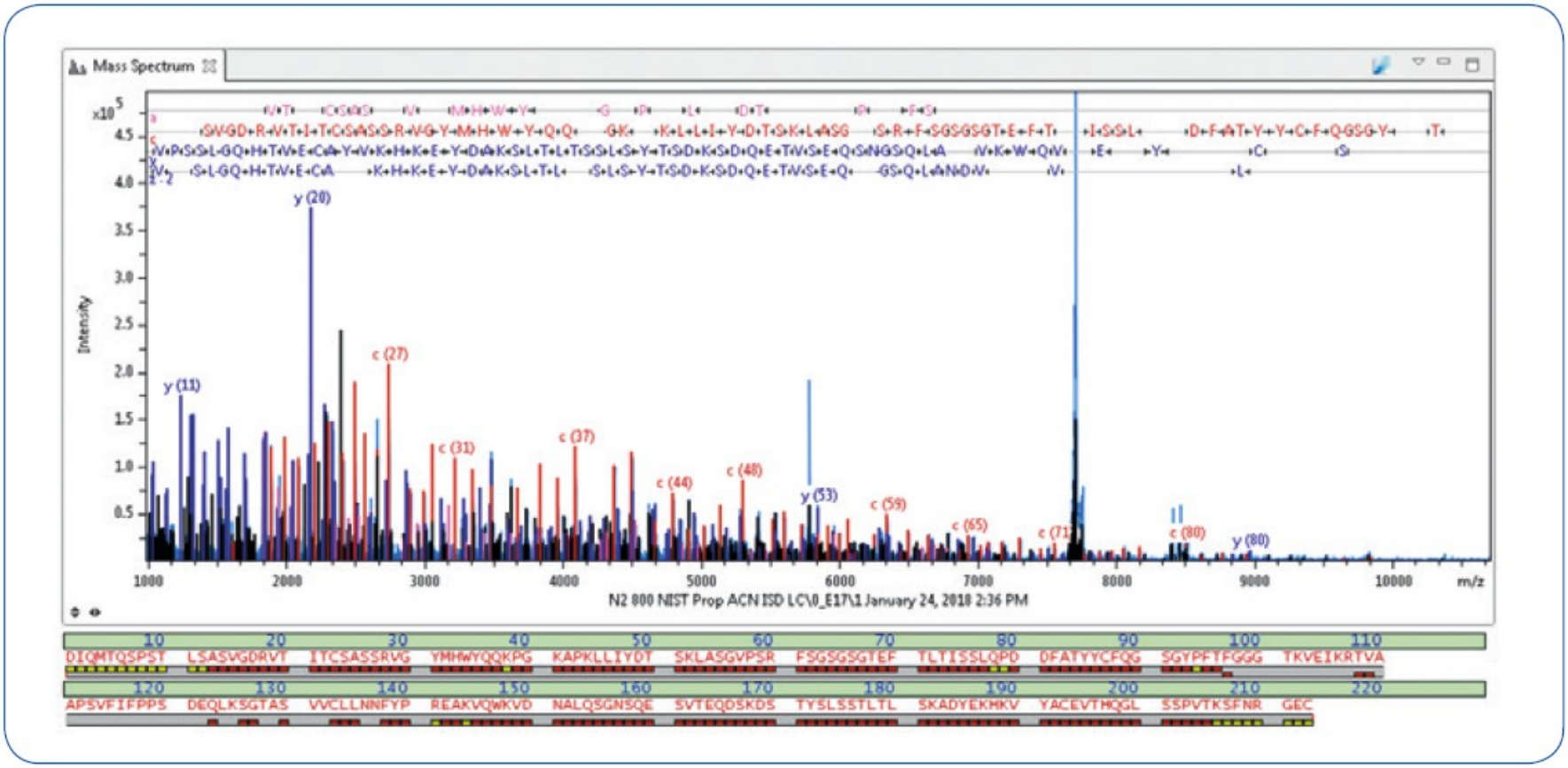


Figure 4: MALDI-TDS spectrum of the NISTmAb LC with annotion of N- and C-terminal ion series in red (c-, a.ions) and blue (y, z+2-ions), respectively. The residues covered by ISD fragments are shown in the protein sequence coverage view sa colored "bricks" (bottom). N-terminal 97 and C-terminal 77 residues were confirmed including proper terminal status

TDS – Top-Down-Sequencing


End