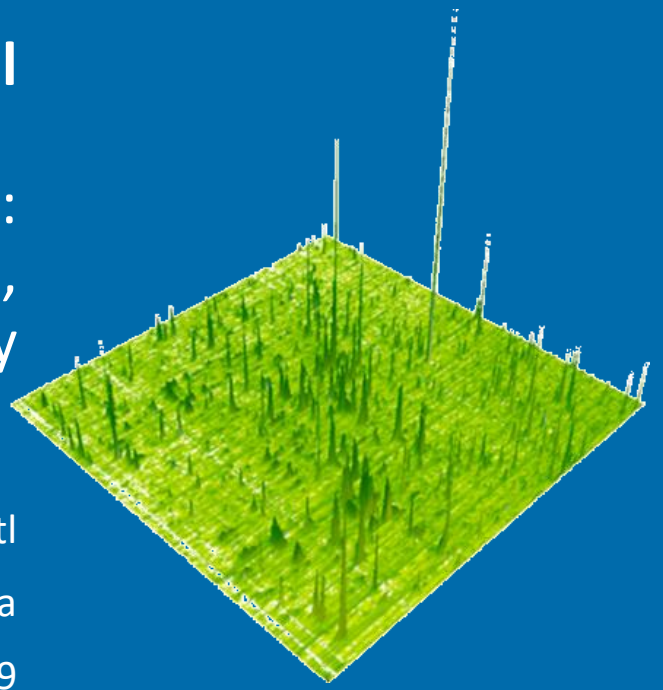
MAX F. PERUTZ LABORATORIES

universität wien

MEDIZINISCHE UNIVERSITÄT WIEN

NPG-net Winter School

A primer on structural proteomics: MS-basics, cross-linking, HDX, ion-mobility



Markus Hartl

MFPL Mass Spectrometry Facility, Vienna, Austria

Talk @ CEITEC, Brno, January 10, 2019

The purpose of this lecture

Explore the basics of mass spectrometry and get an idea of what it can do for you in terms of:

- Proteoform identification
- Structural information
- Protein dynamics

…and be honest about what the difficulties are.

What mass spectrometry is all about.

Analytical balances:

0.001 g to 1 g ± 0.0001 g

Mass spectrometers:

1E-24 g to 1E-19 g ± 1E-26 g

or

1 Da to 100.000 Da ± 0.01 Da

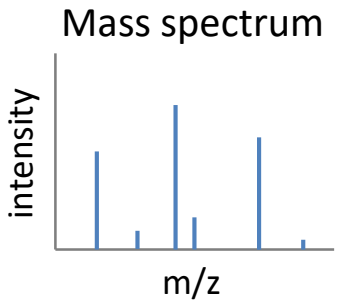
Why is that interesting?

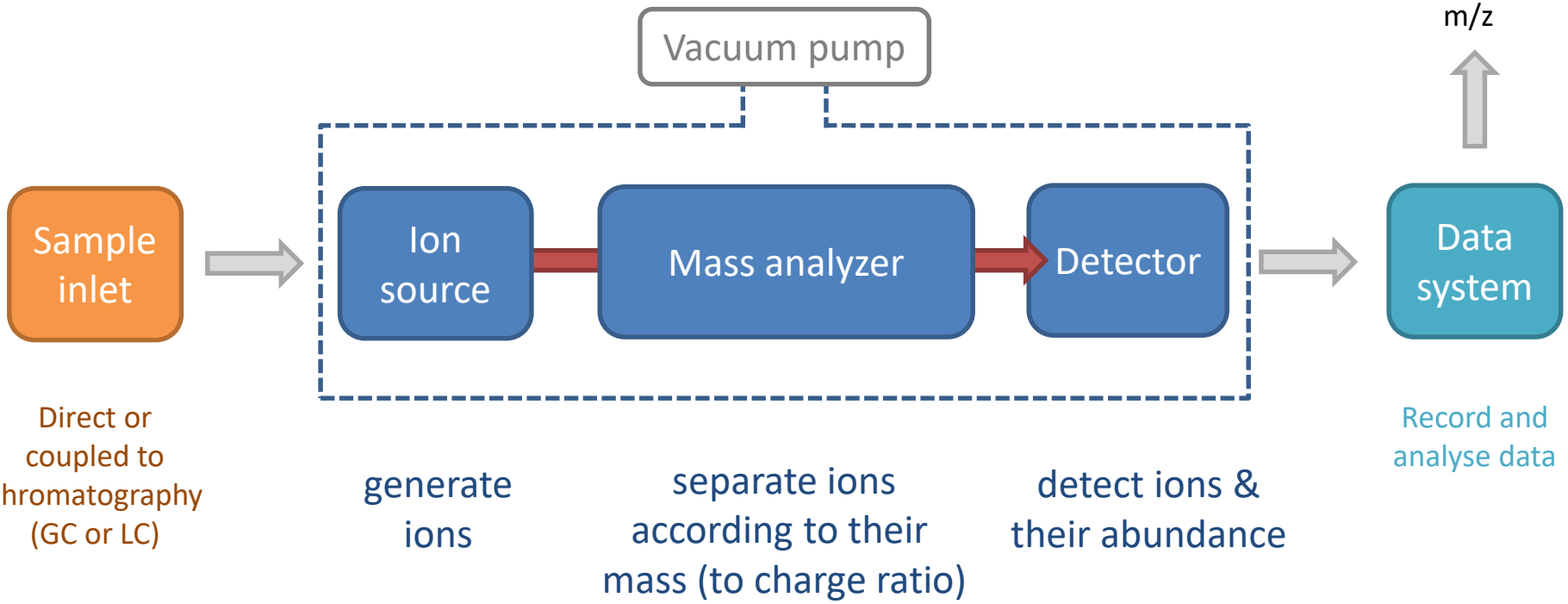
- Identify or verify compounds by their discrete atomic mass.
- Quantify compounds
- Gain information on structure/sequence

- Target analytes: biomolecules (metabolites, oligonucleotides, peptides, proteins), synthetic chemicals, polymers, drugs, etc.

How does it work?

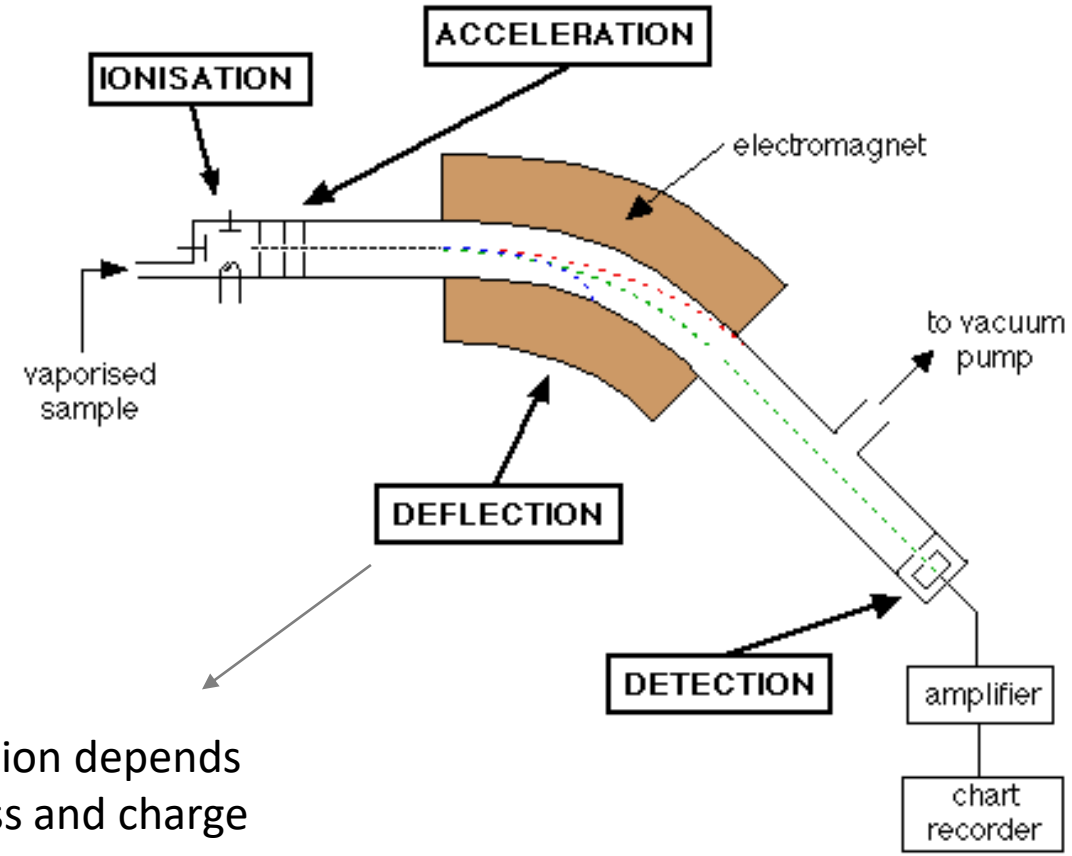
A mass spectrometer:





How does it work?

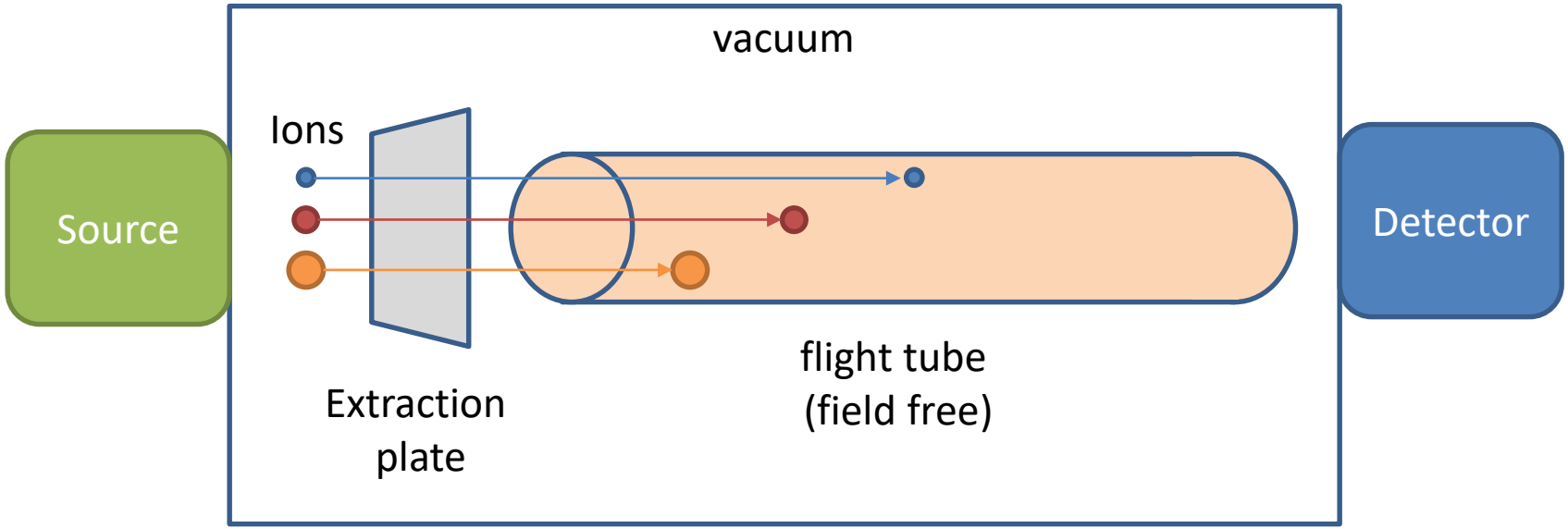
(a sector-field instrument as an example)



Deflection depends on mass and charge

http://www.chemguide.co.uk/analysis/masspec/howitworks.html

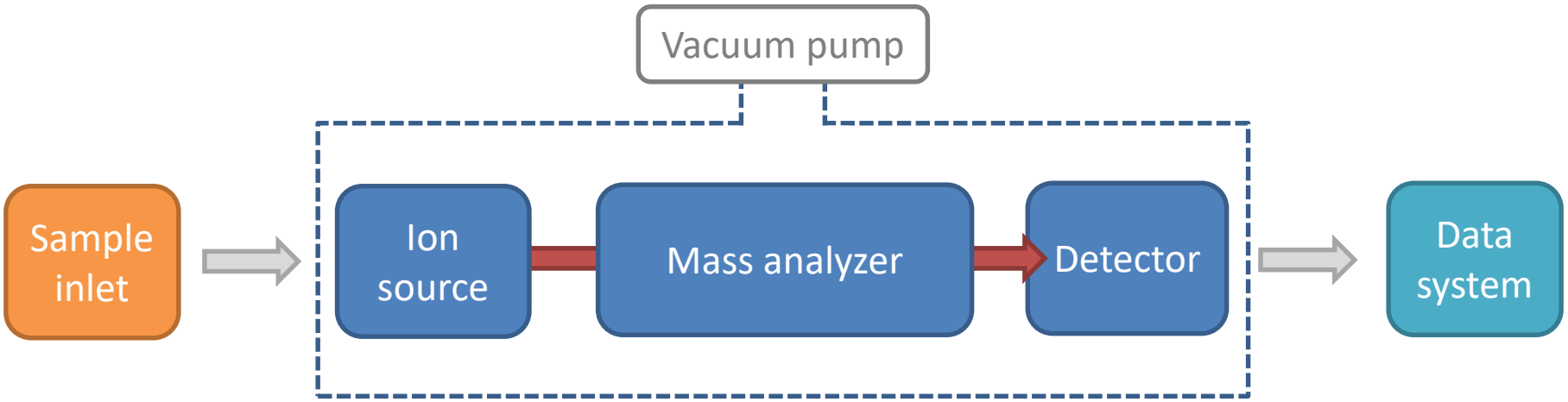
Time-of-flight analyzer



$$E_p = q \cdot U \qquad E_k = \frac{m \cdot v^2}{2} \leftrightarrow v = \sqrt{\frac{2 \cdot Ek}{m}} \leftrightarrow \frac{d}{t} = \sqrt{\frac{2 \cdot Ek}{m}} \rightarrow \frac{m}{q} = \frac{2 \cdot U \cdot t^2}{d^2}$$

Adapted from Lennart Marten's Youtube video: https://www.youtube.com/watch?v=vXsotPtOdRY&t=2s

How does it work?



generate ions

For example:
Electron impact (EI) — **hard**
Chemical ionis. (CI) — **hard**
Electrospray ionis. (ESI) — **soft**
MALDI — **soft**

separate ions according to their mass (to charge ratio)

For example:
Quadrupol — **Low res**
Iontrap — **Low res**
Time-of-Flight (TOF) — **High res**
Orbitrap — **High res**

detect ions & their abundance

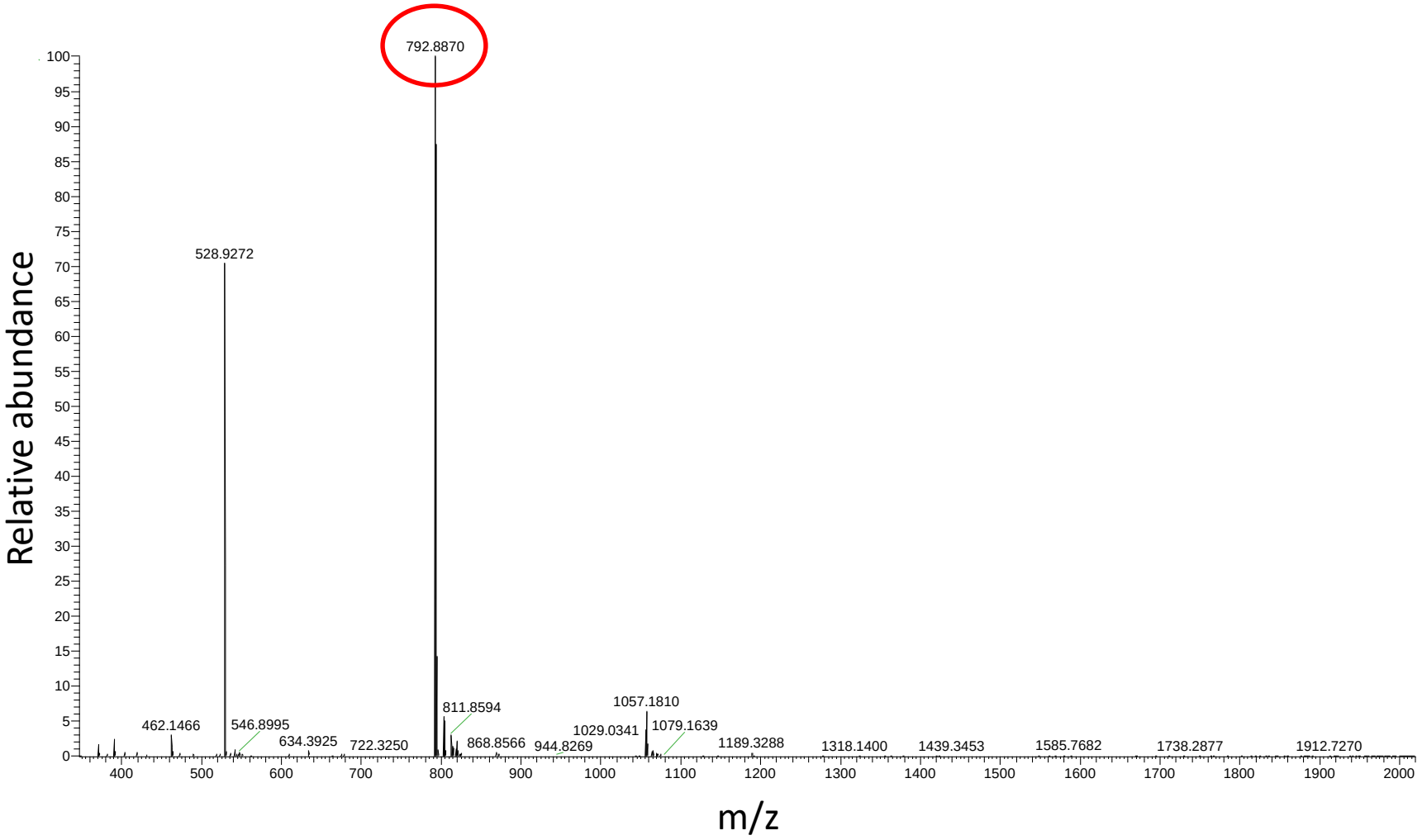
Instrument dependent:
e.g. Electron multiplier

Many combinations many acronyms

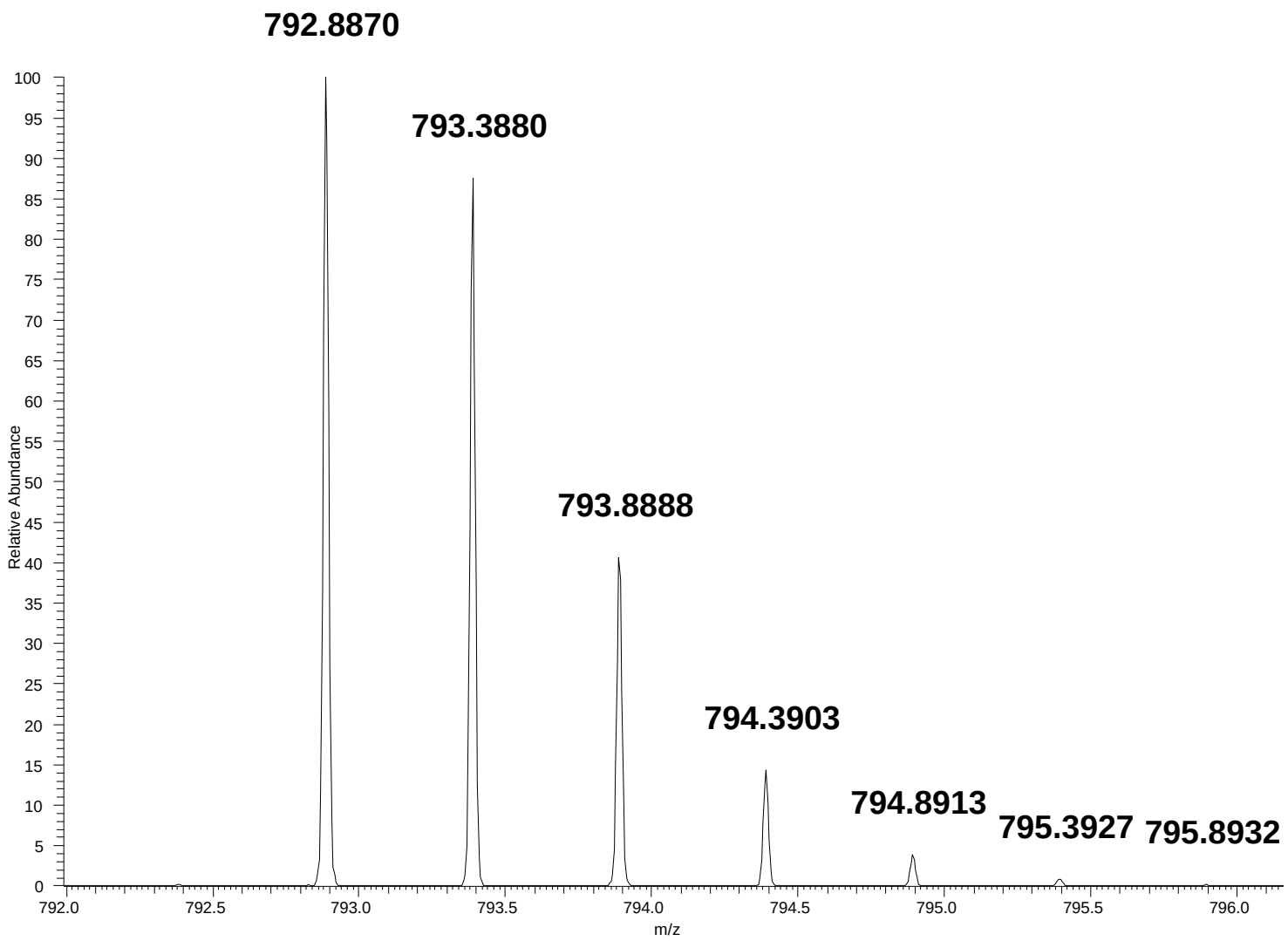
Separation	Ionisation	Mass Spectrometer type	
LC	ESI	TOF	LC-ESI-TOF
GC	(EI)	Quadrupole (Q)	GC-Quad
-	MALDI	TOF	MALDI-TOF
LC	(ESI)	Triple-Quadrupole (QQQ)	LC-TripleQ
LC	(ESI)	Quadrupole-Orbitrap	LC-Q Exactive
....			

Different instrumental setups for different questions.

The output: mass spectra



Zoom on signal at 792.8870 m/z



Atomic Weights and Isotopic Compositions for All Elements

Isotope			Relative Atomic Mass	Isotopic Composition
1	H	1	1.007 825 032 07(10)	0.999 885(70)
	D	2	2.014 101 777 8(4)	0.000 115(70)
	T	3	3.016 049 2777(25)	
6	C	12	12.000 000 0(0)	0.9893(8)
		13	13.003 354 8378(10)	0.0107(8)
		14	14.003 241 989(4)	
7	N	14	14.003 074 004 8(6)	0.996 36(20)
		15	15.000 108 898 2(7)	0.003 64(20)
8	O	16	15.994 914 619 56(16)	0.997 57(16)
		17	16.999 131 70(12)	0.000 38(1)
		18	17.999 161 0(7)	0.002 05(14)
15	P	31	30.973 761 998 42(70)	1
16	S	32	31.972 071 1744(14)	0.9499(26)
		33	32.971 458 9098(15)	0.0075(2)
		34	33.967 867 004(47)	0.0425(24)
		36	35.967 080 71(20)	0.0001(1)

Source: NIST (National Institute of Standards and Technology), http://physics.nist.gov/cgi-bin/Compositions/stand_alone.pl

Defining mass

Defining mass (according to IUPAC, https://goldbook.iupac.org/):

Unified atomic mass unit:

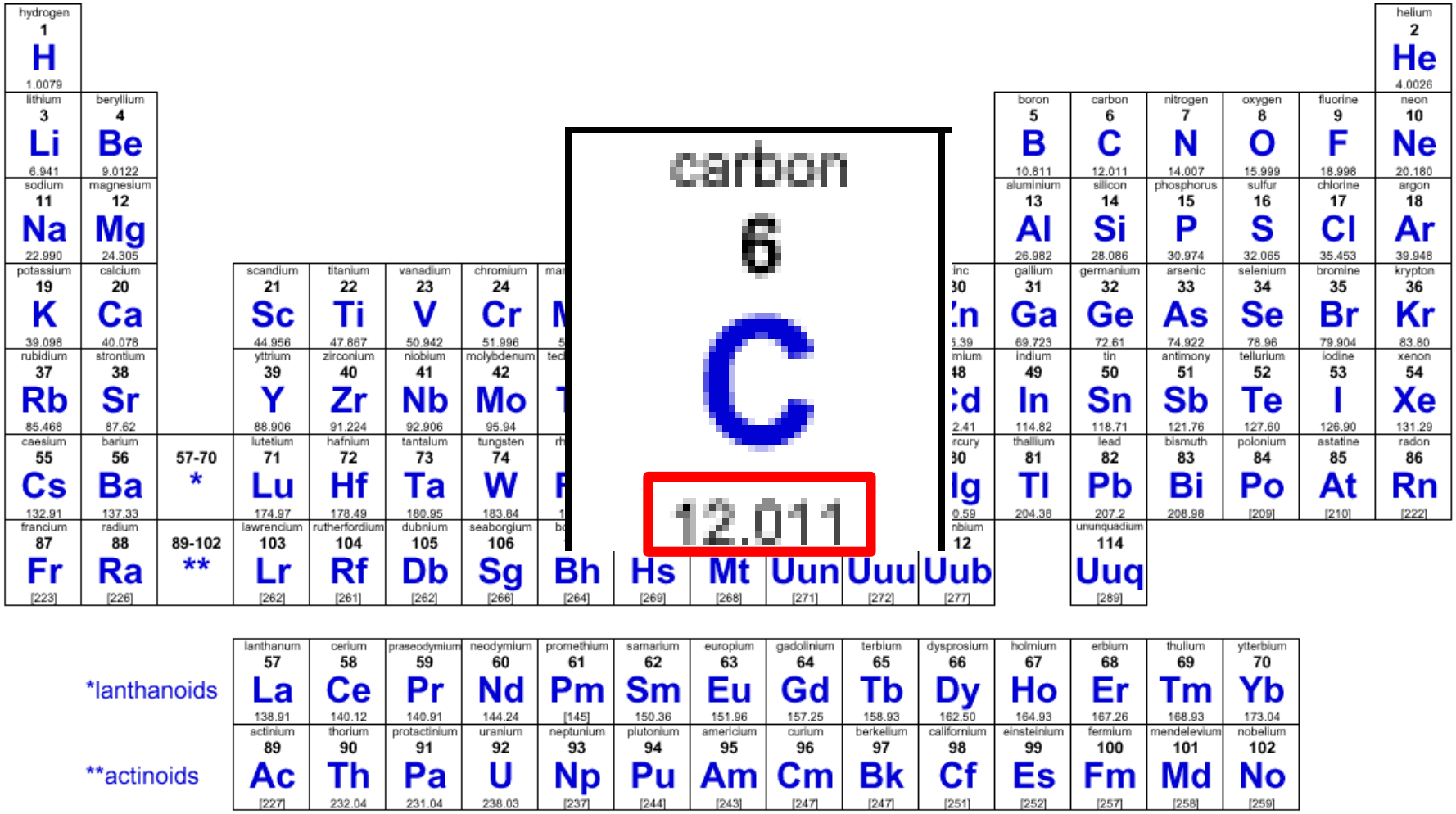
- $^{1}/_{12}$ of the mass of a carbon-12 atom in its ground state
- 1 u ≈ 1.660 5402 10 × 10 −27 kg
- Symbols: u or Da (for its equivalent dalton).

Relative atomic mass: mass of a discrete atomic particle or molecule expressed in unified atomic mass units.

Standard atomic weight (or molecular weight or average mass): the weighted average of the masses of the naturally occurring isotopes. For example one carbon atom:

$$\frac{(98.8\% * 12.0 + 1.1\% * 13{,}003355)}{100\%} = 12.011$$

Standard atomic weight

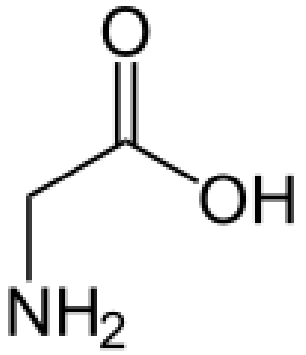


Example 1:

Atomic Weights and Isotopic Compositions for All Elements

Isotope		Relative Atomic Mass	Isotopic Composition	Standard Atomic Weight	Notes
1 H	1	1.007 825 032 07(10)	0.999 885(70)	[1.007 84, 1.008 11]	m
D	2	2.014 101 777 8(4)	0.000 115(70)		
T	3	3.016 049 2777(25)			
2 He	3	3.016 029 3191(26)	0.000 001 34(3)	4.002 602(2)	g,r
	4	4.002 603 254 15(6)	0.999 998 66(3)		
3 Li	6	6.015 122 795(16)	0.0759(4)	[6.938, 6.997]	m
	7	7.016 004 55(8)	0.9241(4)		
4 Be	9	9.012 182 2(4)	1.0000	9.012 1831(5)	
5 B	10	10.012 937 0(4)	0.199(7)	[10.806, 10.821]	m
	11	11.009 305 4(4)	0.801(7)		
6 C	12	12.000 000 0(0)	0.9893(8)	[12.0096, 12.0116]	
	13	13.003 354 8378(10)	0.0107(8)		
	14	14.003 241 989(4)			
7 N	14	14.003 074 004 8(6)	0.996 36(20)	[14.006 43, 14.007 28]	
	15	15.000 108 898 2(7)	0.003 64(20)		
8 O	16	15.994 914 619 56(16)	0.997 57(16)	[15.999 03, 15.999 77]	
	17	16.999 131 70(12)	0.000 38(1)		
	18	17.999 161 0(7)	0.002 05(14)		

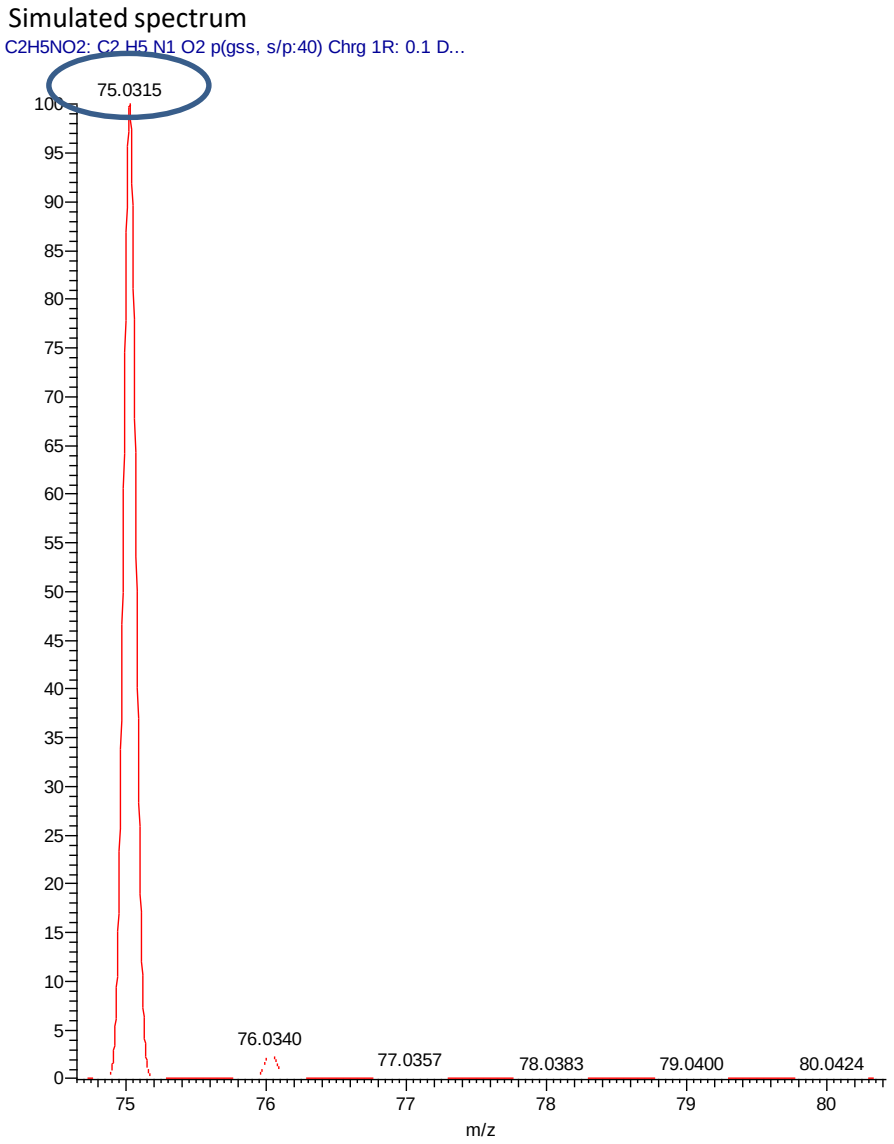
Source: National Institute of Standards and Technology (NIST), http://www.nist.gov/pml/data/comp.cfm

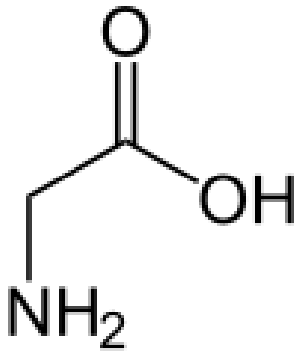


$C_2H_5NO_2$

Glycine:
Monoisotopic: 75.03203 u
Atomic weight: 75.06720 u
Nominal mass: 75 u

Example 1:





$C_2H_5NO_2$

Monoisotopic mass: 75.0315 u

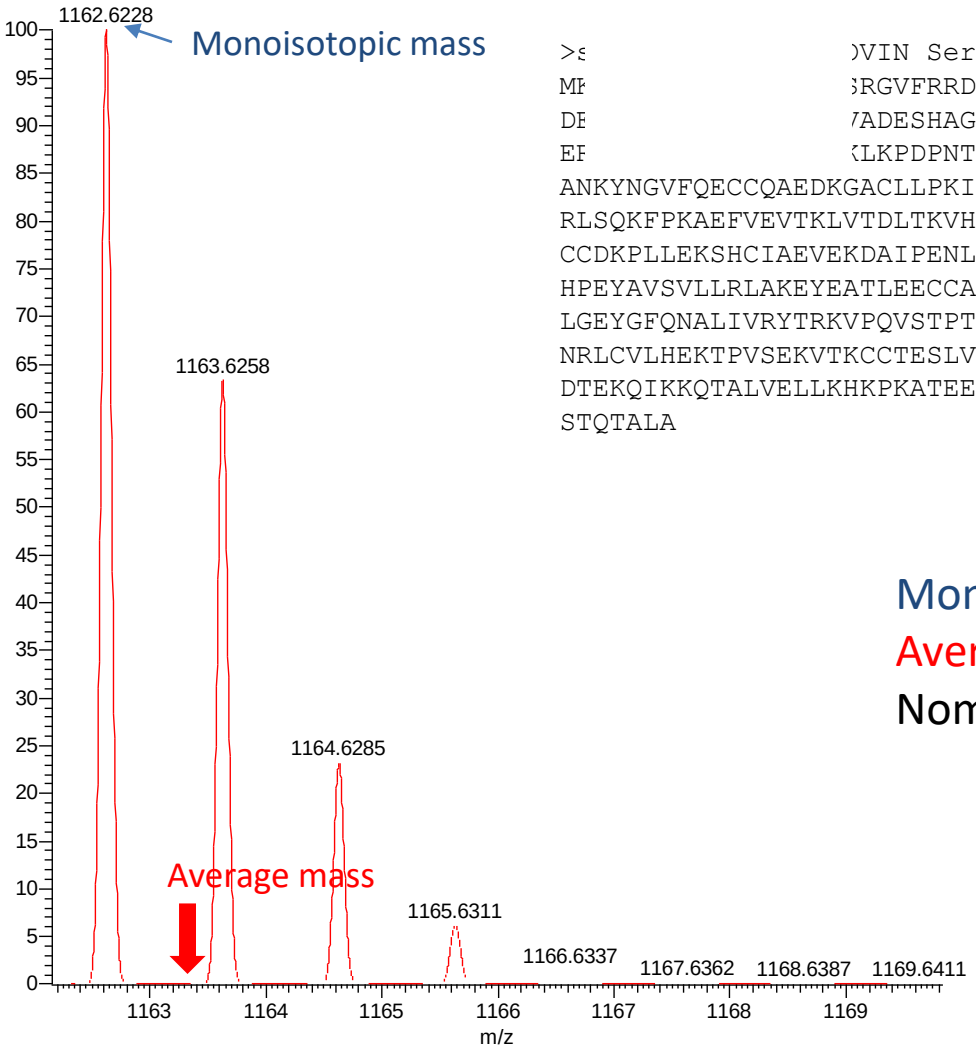
Average mass: 75.0672 u

Nominal mass: 75 u

Example 2: a peptide

Simulated spectrum

LVNELTEFAK +H2O: C53 H86 N12 O17 p(gss, s/p:40) Chrg...



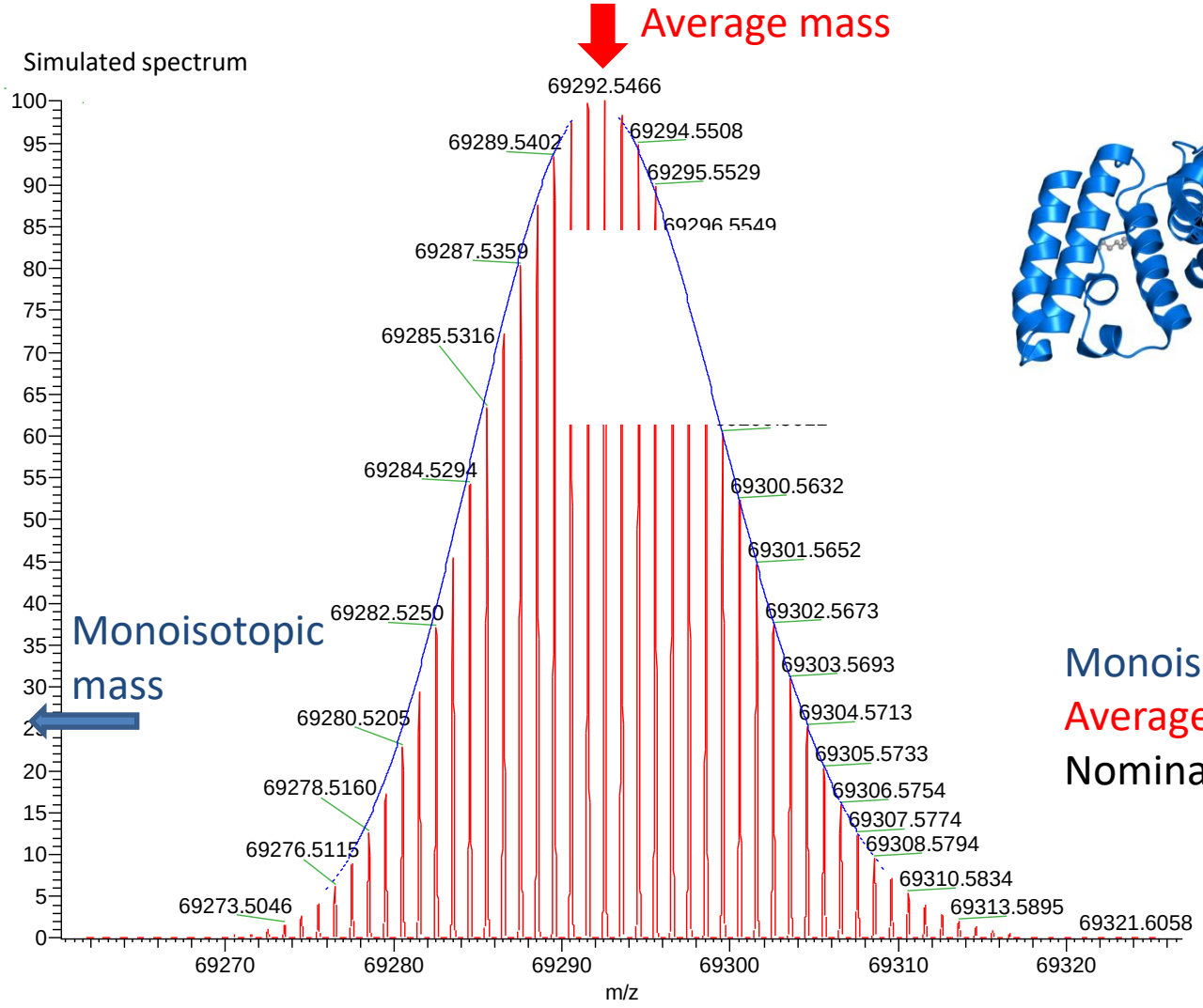
>s)VIN Serum albumin
MF ;RGVFRRDTHKSEIAHRFKDLGEEHFKGLVLIAFSQYLQQCPF
DE /ADESHAGCEKSLHTLFGDELCKVASLRETYGDMADCCEKQEP
EF {LKPDPNTLCDEFKADEKKFWGKYLYEIARRHPYFYAPELLYY
ANKYNGVFQECCQAEDKGACLLPKIETMREKVLASSARQRLRCASIQKFGERALKAWSVA
RLSQKFPKAEFVEVTKLVTDLTKVHKECCHGDLLECADDRADLAKYICDNQDTISSKLKE
CCDKPLLEKSHCIAEVEKDAIPENLPPLTADFAEDKDVCKNYQEAKDAFLGSFLYEYSRR
HPEYAVSVLLRLAKEYEATLEECCAKDDPHACYSTVFDKLKHLVDEPQNLIKQNCDQFEK
LGEYGFQNALIVRYTRKVPQVSTPTLVEVSRSLGKVGTRCCTKPESERMPCTEDYLSLIL
NRLCVLHEKTPVSEKVTKCCTESLVNRRPCFSALTPDETYVPKAFDEKLFTFHADICTLP
DTEKQIKKQTALVELLKHKPKATEEQLKTVMENFVAFVDKCCAADDKEACFAVEGPKLVV
STQTALA

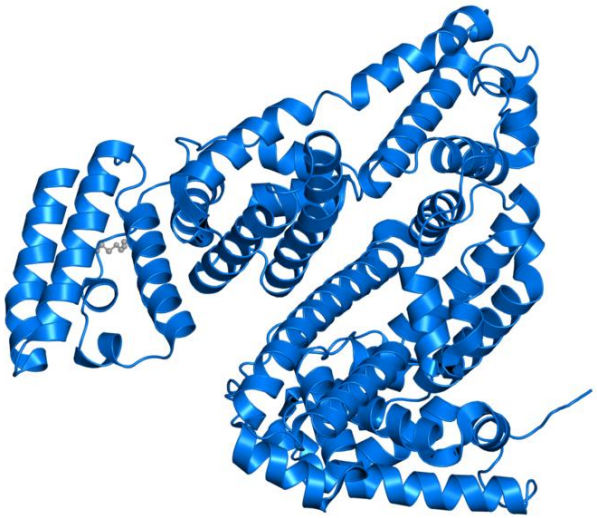
Monoisotopic mass: 1162.6228 u

Average mass: 1163.34 u

Nominal mass: 1163 u

Example 3: a protein





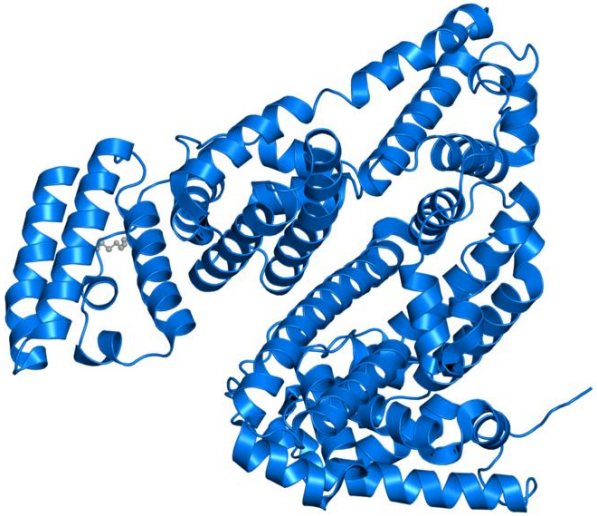
BSA

Monoisotopic mass: 69248.44 u

Average mass: 69293.41 u

Nominal mass: 69293 u

The basic idea



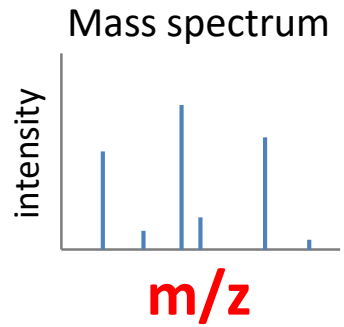
Bovine serum albumin
69293 Da

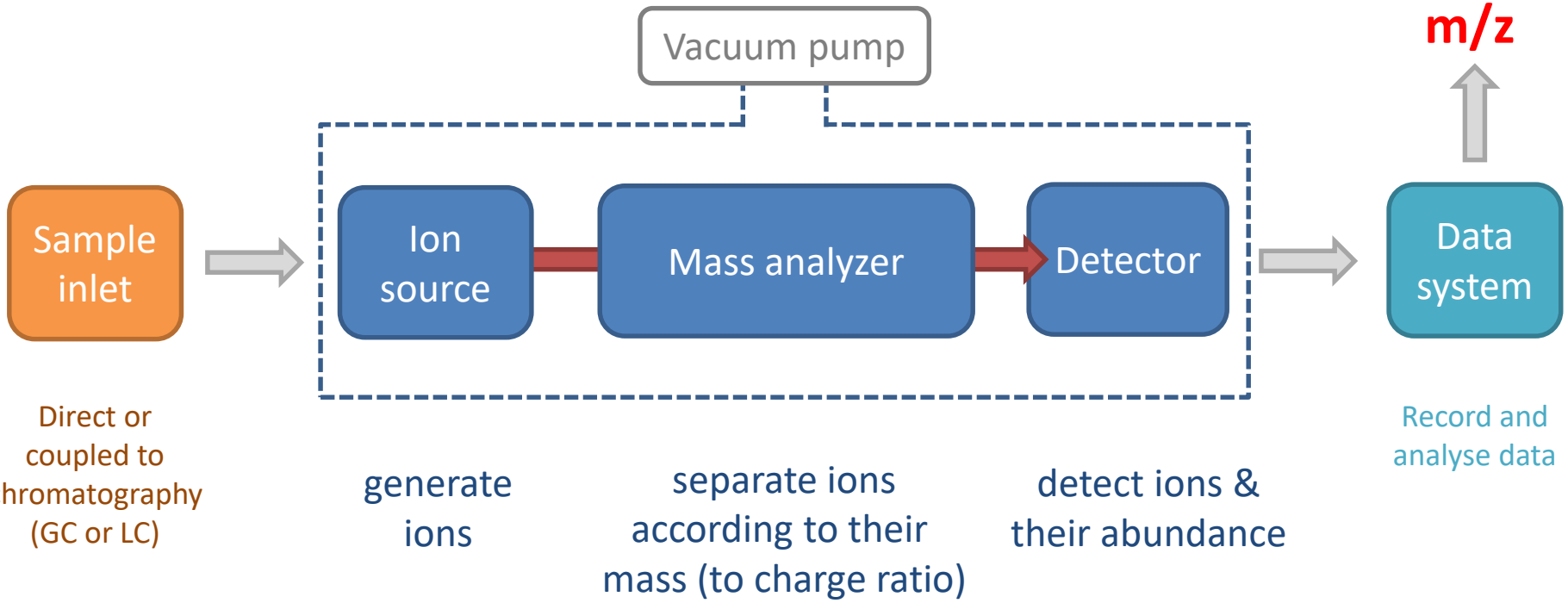
69293 Da measured

Which mass do you measure?
Isotopic, average?
How accurately?

MS measures mass / charge (m/z)

A mass spectrometer:

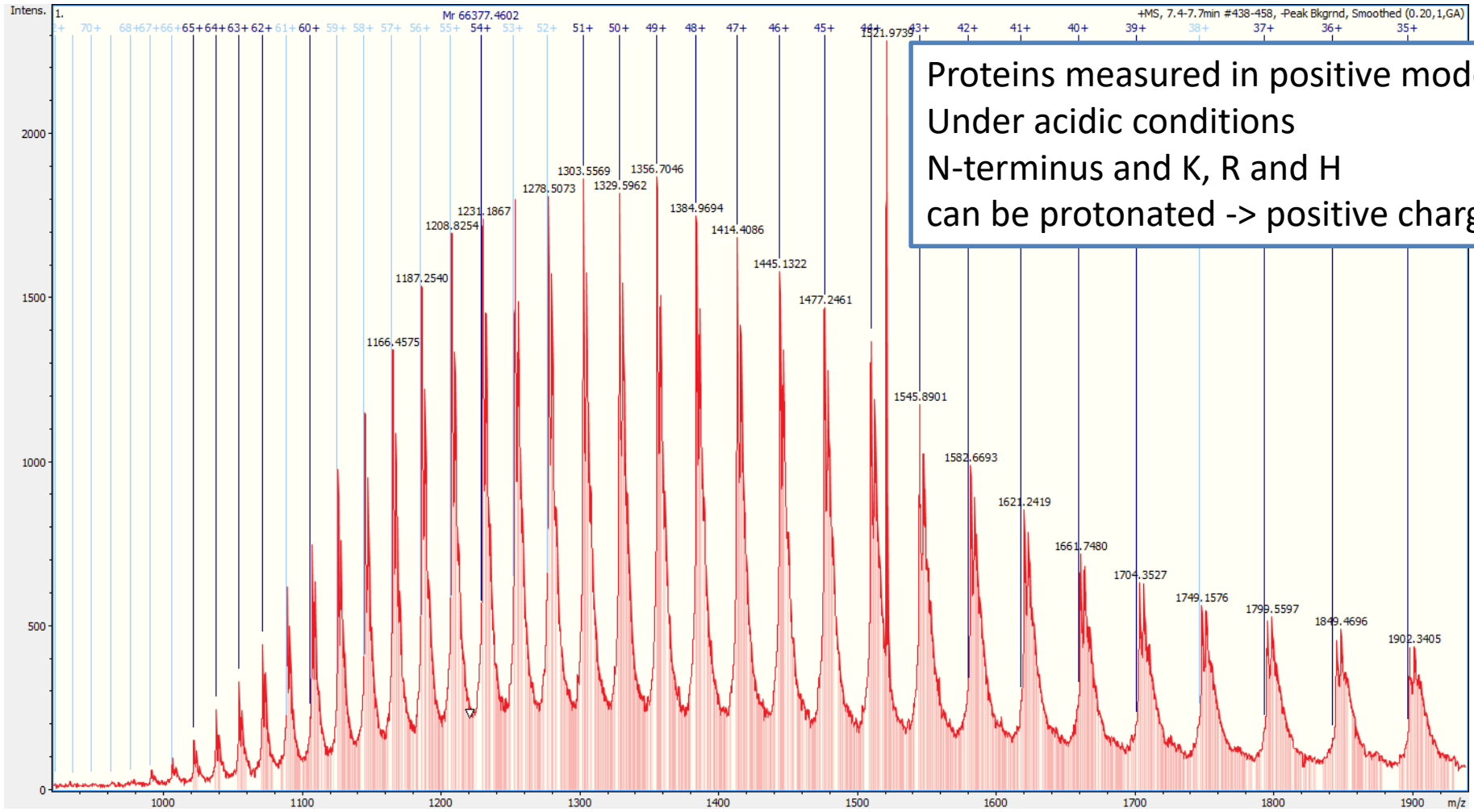




What is the charge of a protein in electrospray ionisation (ESI)?

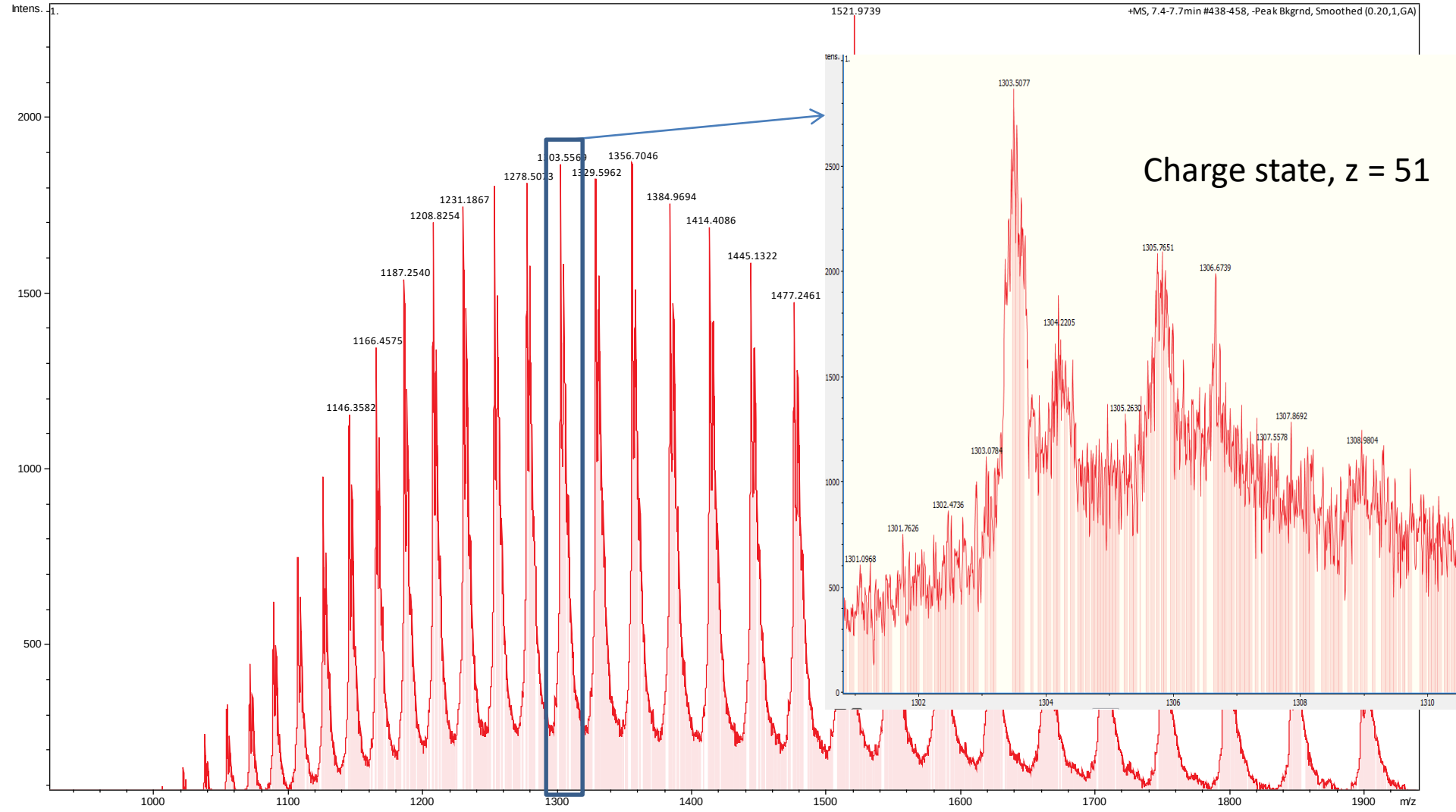
Charge envelopes in ESI: measuring m/z

55+ 50+ 45+

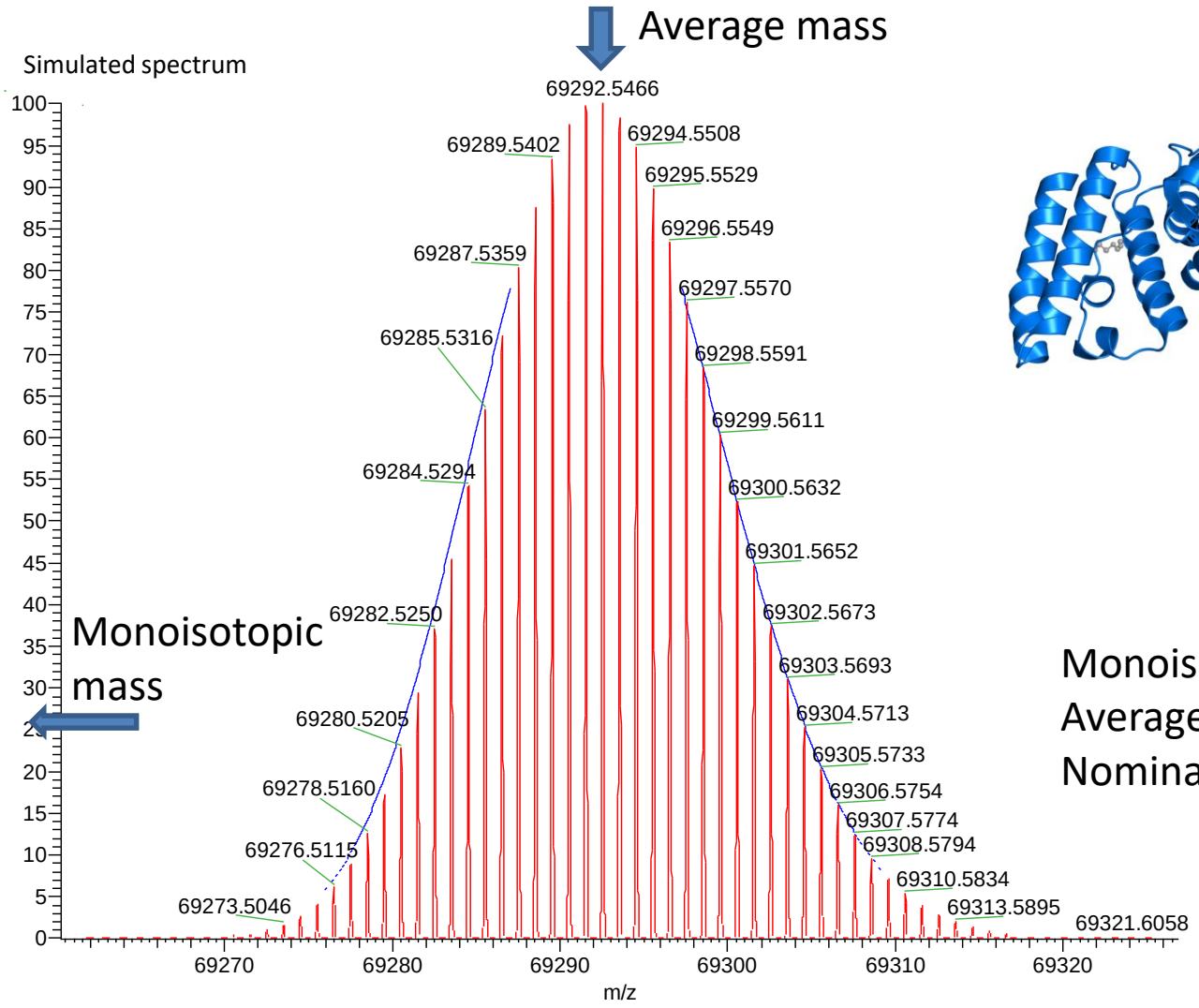


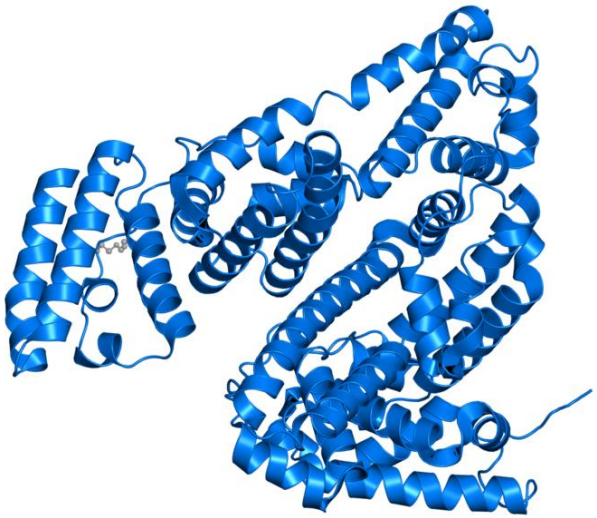
Proteins measured in positive mode:
Under acidic conditions
N-terminus and K, R and H
can be protonated -> positive charge

ESI-TOF spectrum of intact BSA



Theoretical uncharged spectrum

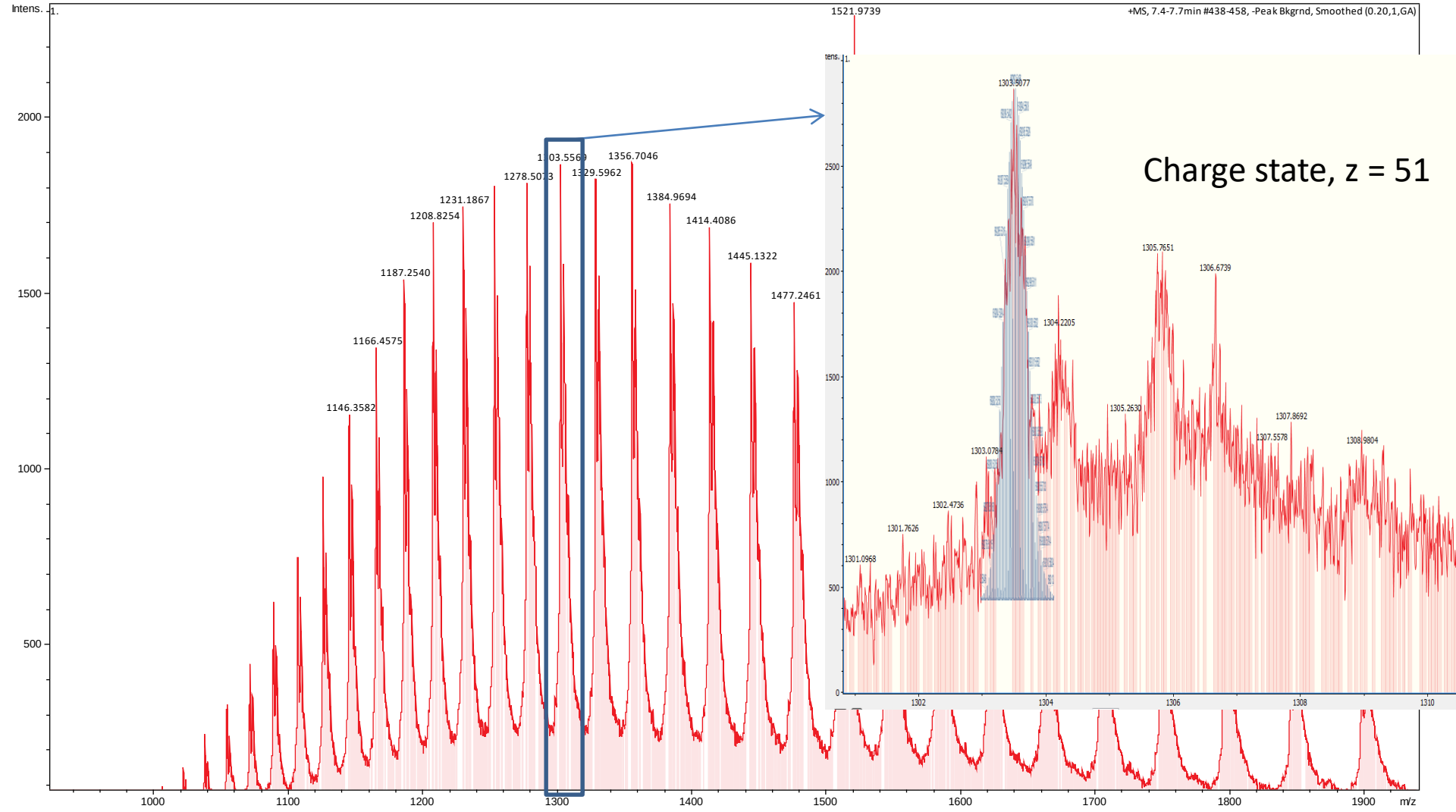




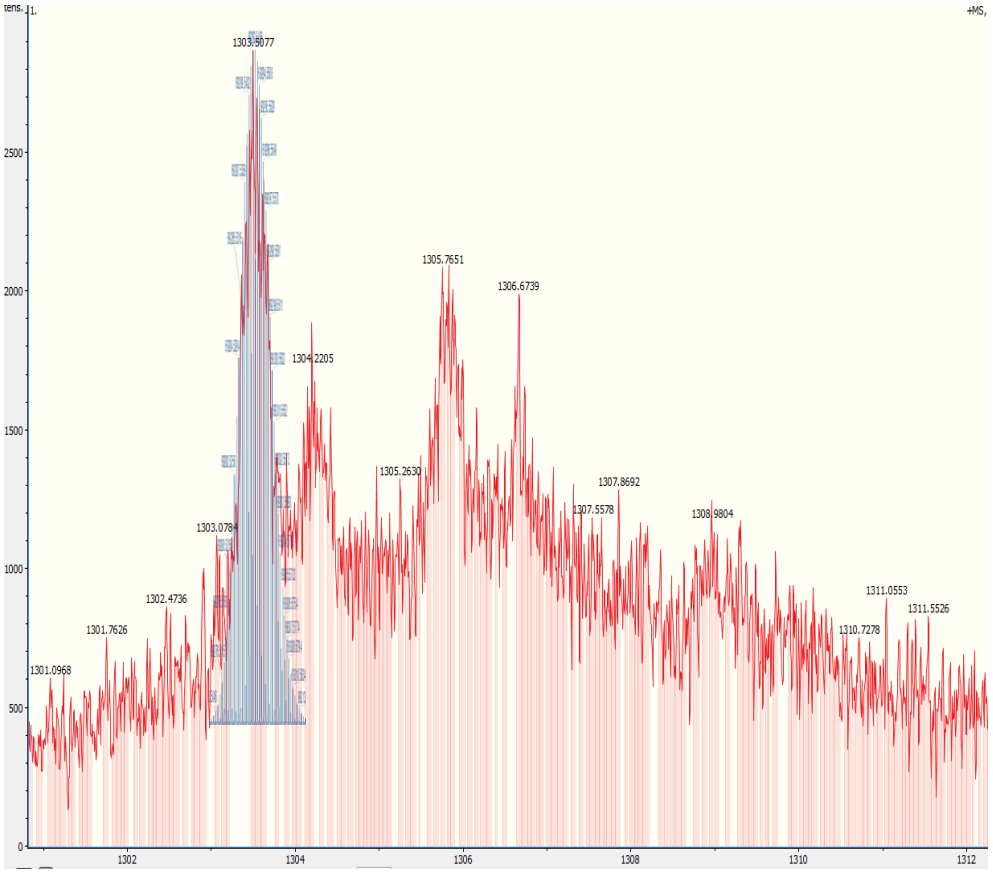
BSA

Monoisotopic mass: 69248.44 u
Average mass: 69293.41 u
Nominal mass: 69293 u

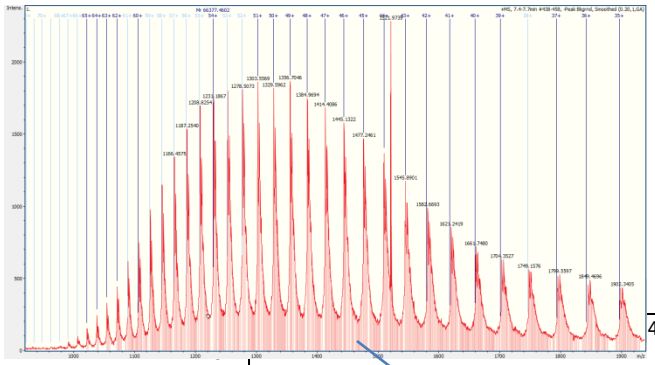
ESI-TOF spectrum of intact BSA

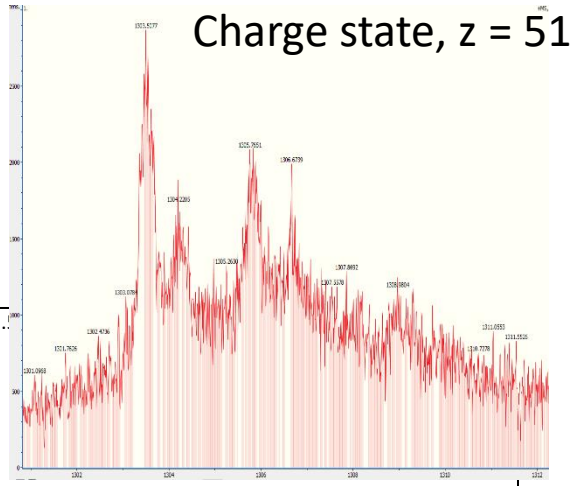


Intact proteins >30 kDa usually determined as average mass

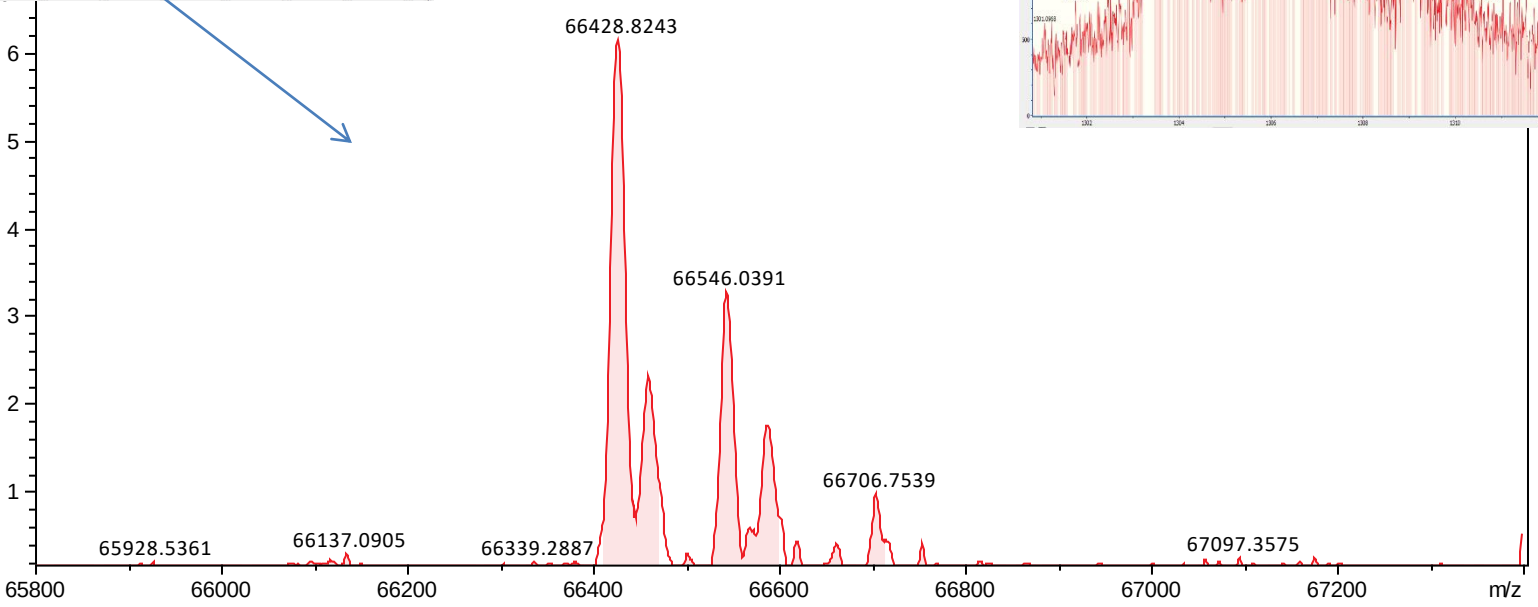


Using deconvolution algorithms





4-7.7min, -Peak Bkgrnd, Deconvoluted (MaxEnt, 1157.73-1499..

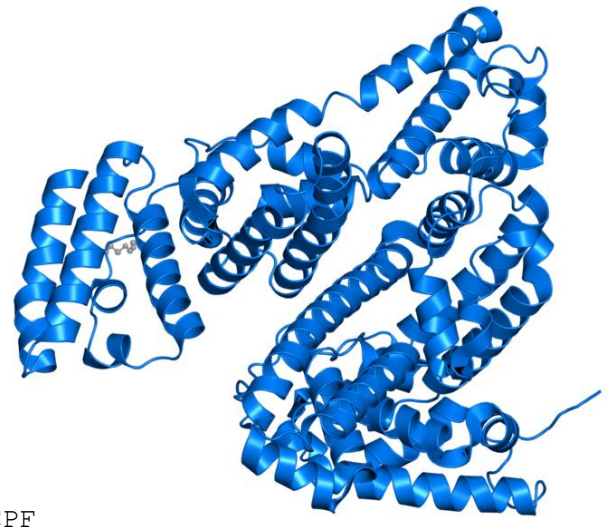


Take-home message

- Mass spectra depict m/z
- Proteins are usually multiply charged and display several charge states in ESI
- Proteins >30 kDa are usually determined as average mass
- Spectrum displaying several charge states needs to be deconvoluted to yield the non-charged average mass of the protein.

When the results do not match…

- ESI-TOF: 66428.8 Da
- Calc.: 69293.4 Da

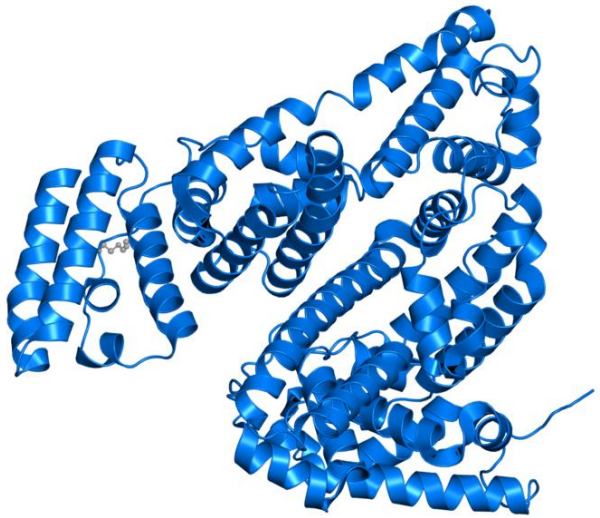


```
>sp|P02769|ALBU_BOVIN Serum albumin
MKWVTFISLLLLFSSAYSRGVFRRDTHKSEIAHRFKDLGEEHFKGLVLIAFSQYLQQCPF
DEHVKLVNELTEFAKTCVADESHAGCEKSLHTLFGDELCKVASLRETYGDMADCCEKQEP
ERNECFLSHKDDSPDLPKLKPDPNTLCDEFKADEKKFWGKYLYEIARRHPYFYAPELLYY
ANKYNGVFQECCQAEDKGACLLPKIETMREKVLASSARQRLRCASIQKFGERALKAWSVA
RLSQKFPKAEFVEVTKLVTDLTKVHKECCHGDLLECADDRADLAKYICDNQDTISSKLKE
CCDKPLLEKSHCIAEVEKDAIPENLPPLTADFAEDKDVCKNYQEAKDAFLGSFLYEYSRR
HPEYAVSVLLRLAKEYEATLEECCAKDDPHACYSTVFDKLKHLVDEPQNLIKQNCDQFEK
LGEYGFQNALIVRYTRKVPQVSTPTLVEVSRSLGKVGTRCCTKPESERMPCTEDYLSLIL
NRLCVLHEKTPVSEKVTKCCTESLVNRRPCFSALTPDETYVPKAFDEKLFTFHADICTLP
DTEKQIKKQTALVELLKHKPKATEEQLKTVMENFVAFVDKCCAADDKEACFAVEGPKLVV
STQTALA
```

What is wrong?

Protein processing

>sp|P02769|ALBU_BOVIN Serum albumin
MKWVTFISLLLLFSSAYSRGVFRRDTHKSEIAHRFKDLGEEHFKGLVLIAFSQYLQQCPF
DEHVKLVNELTEFAKTCVADESHAGCEKSLHTLFGDELCKVASLRETYGDMADCCEKQEP
ERNECFLSHKDDSPDLPKLKPDPNTLCDEFKADEKKFWGKYLYEIARRHPYFYAPELLYY
ANKYNGVFQECCQAEDKGACLLPKIETMREKVLASSARQRLRCASIQKFGERALKAWSVA
RLSQKFPKAEFVEVTKLVTDLTKVHKECCHGDLLECADDRADLAKYICDNQDTISSKLKE
CCDKPLLEKSHCIAEVEKDAIPENLPPLTADFAEDKDVCKNYQEAKDAFLGSFLYEYSRR
HPEYAVSVLLRLAKEYEATLEECCAKDDPHACYSTVFDKLKHLVDEPQNLIKQNCDQFEK
LGEYGFQNALIVRYTRKVPQVSTPTLVEVSRSLGKVGTRCCTKPESERMPCTEDYLSLIL
NRLCVLHEKTPVSEKVTKCCTESLVNRRPCFSALTPDETYVPKAFDEKLFTFHADICTLP
DTEKQIKKQTALVELLKHKPKATEEQLKTVMENFVAFVDKCCAADDKEACFAVEGPKLVV
STQTALA

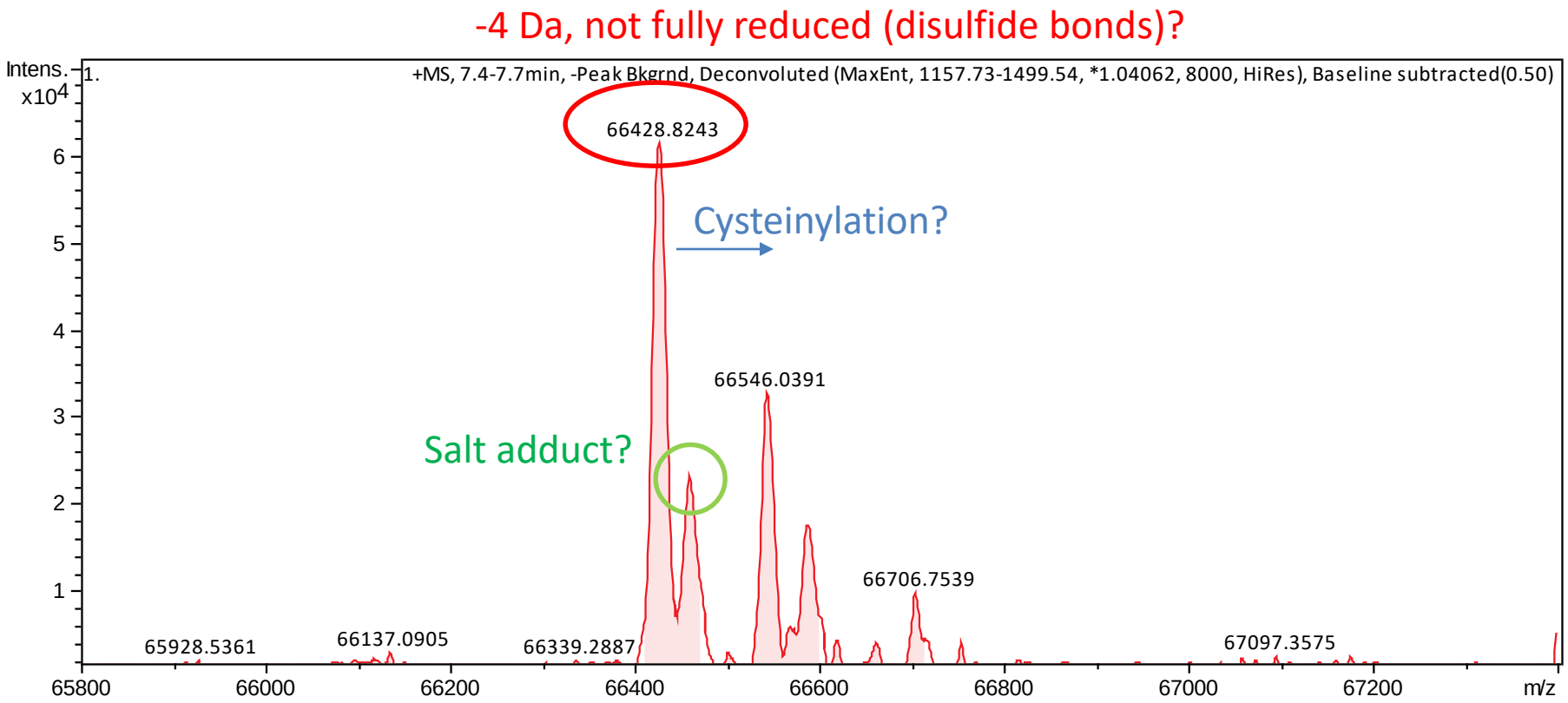


69293.41 Da full chain
66432.96 Da mature chain (reduced state)

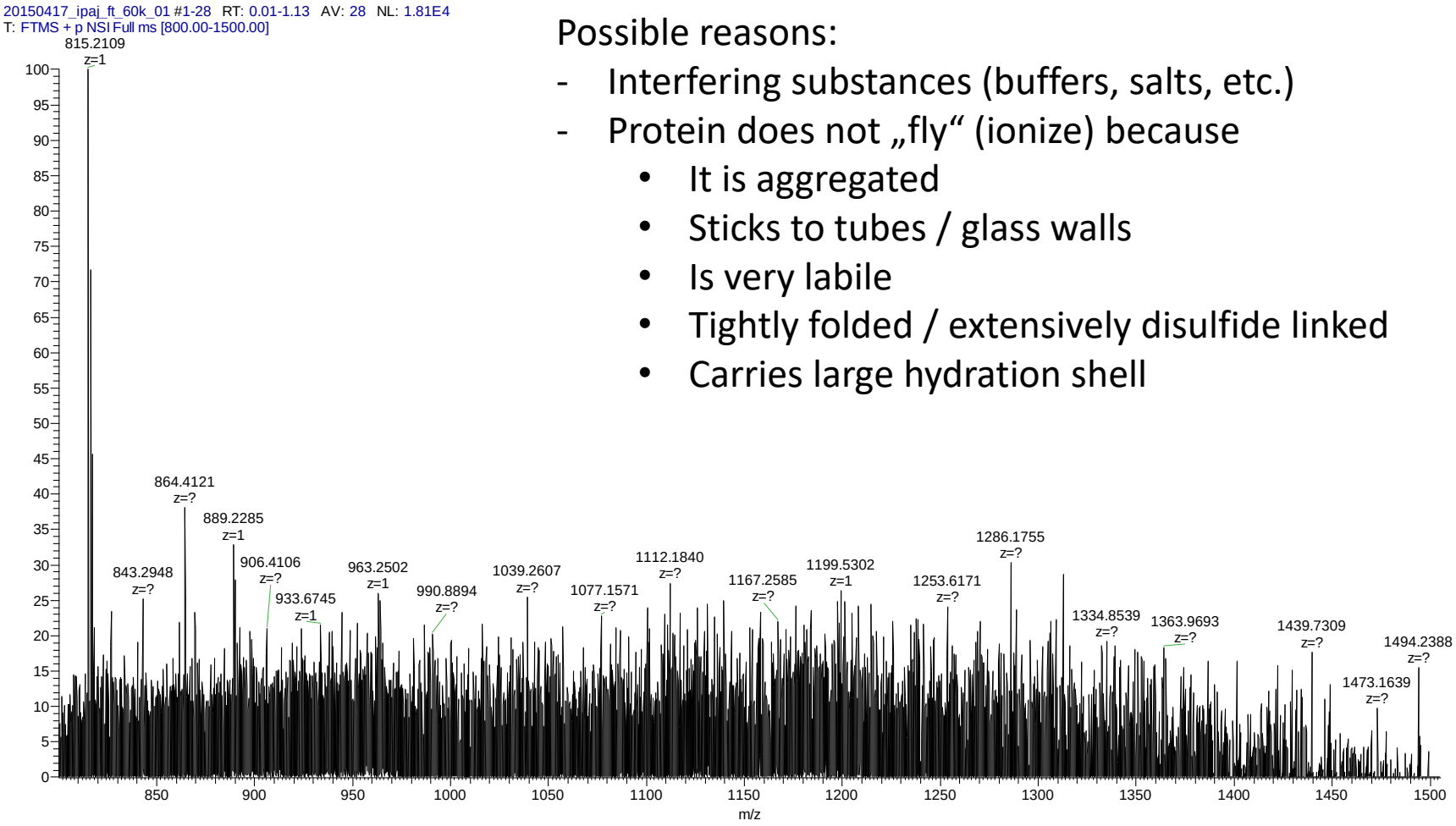
Plus:

- 17 disulfide bonds, 1 free –SH
- 12 other PTMs (Phospho, succinyl)
- Dimerizes
- Binds water, Ca, Na, K, Zn, fatty acids, hormones, drugs, etc.
- Chemical modifications

Interpreting the unknowns



Sometimes nothing works



Possible reasons:

- Interfering substances (buffers, salts, etc.)
- Protein does not „fly“ (ionize) because
  - It is aggregated
  - Sticks to tubes / glass walls
  - Is very labile
  - Tightly folded / extensively disulfide linked
  - Carries large hydration shell

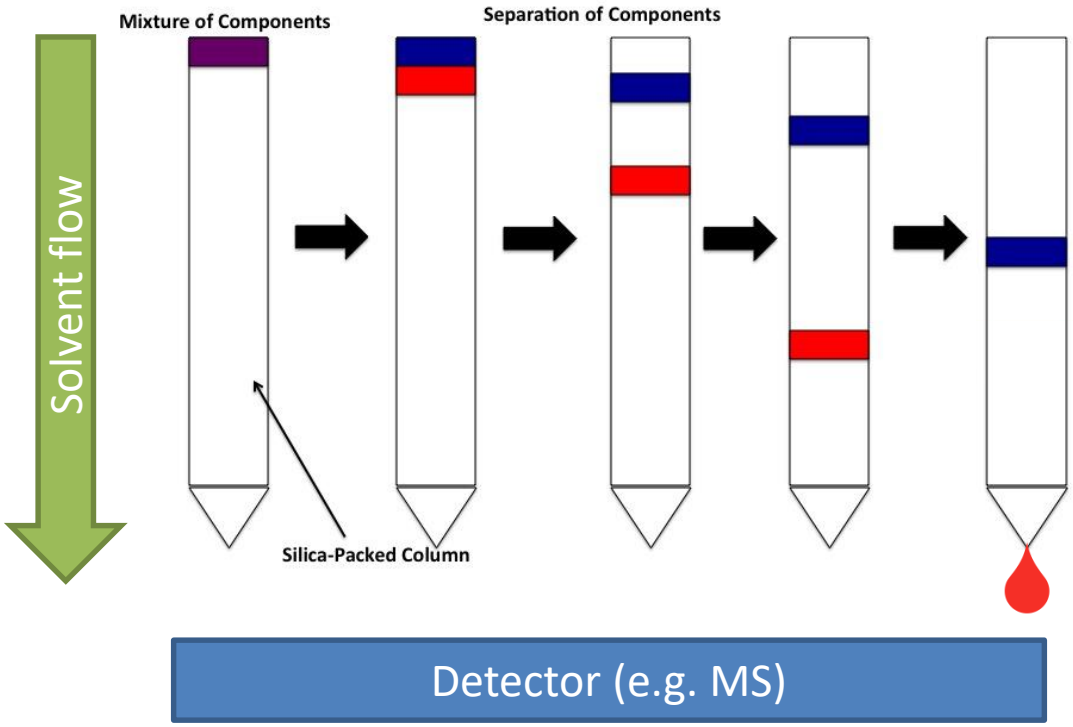
The limits of this idea?

Intact mass analysis limited by:

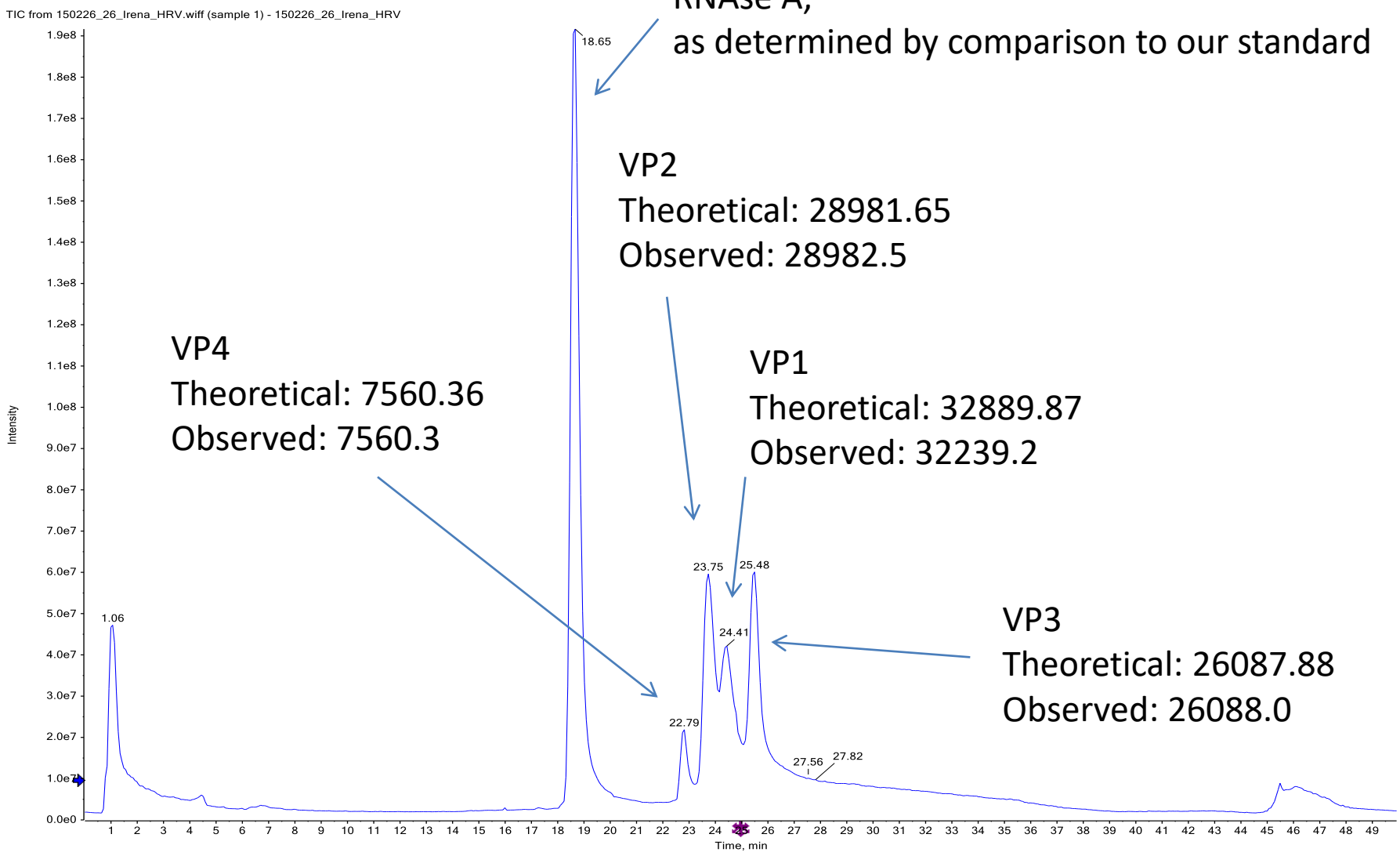
- Complex mixtures of unknown composition
- Analytes with isobaric mass
- Unknown modifications or processing of analytes
- Technical issues with biomolecules

The „LC“ in LC-MS

LC = liquid chromatography



Human rhinovirus



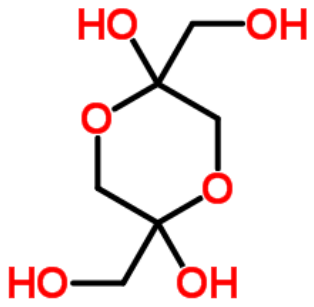
Intact mass analysis limited by:

- Complex mixtures of unknown composition
- Analytes with isobaric mass
- Unknown modifications or processing of analytes
- Technical issues with biomolecules

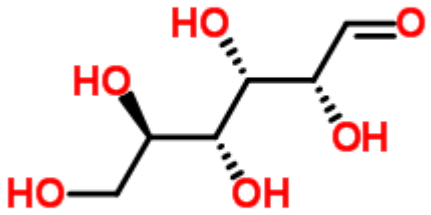
Sometimes one mass is not specific enough

$C_6H_{12}O_6$

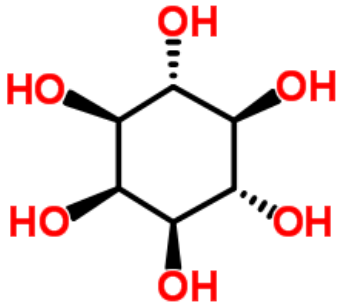
Monoisotopic mass: 180.0634 Da



1,3-Dihydroxyacetone Dimer



D-(+)-Glucose



Inositol

Sometimes one mass is not specific enough

ISGGDALQSCVDR	1320.4 Da
DVCSQLADGGSIR	1320.4 Da
IpSGGDALQSCVDR	1400.4 Da
ISGGDALQpSCVDR	1400.4 Da

No information about:

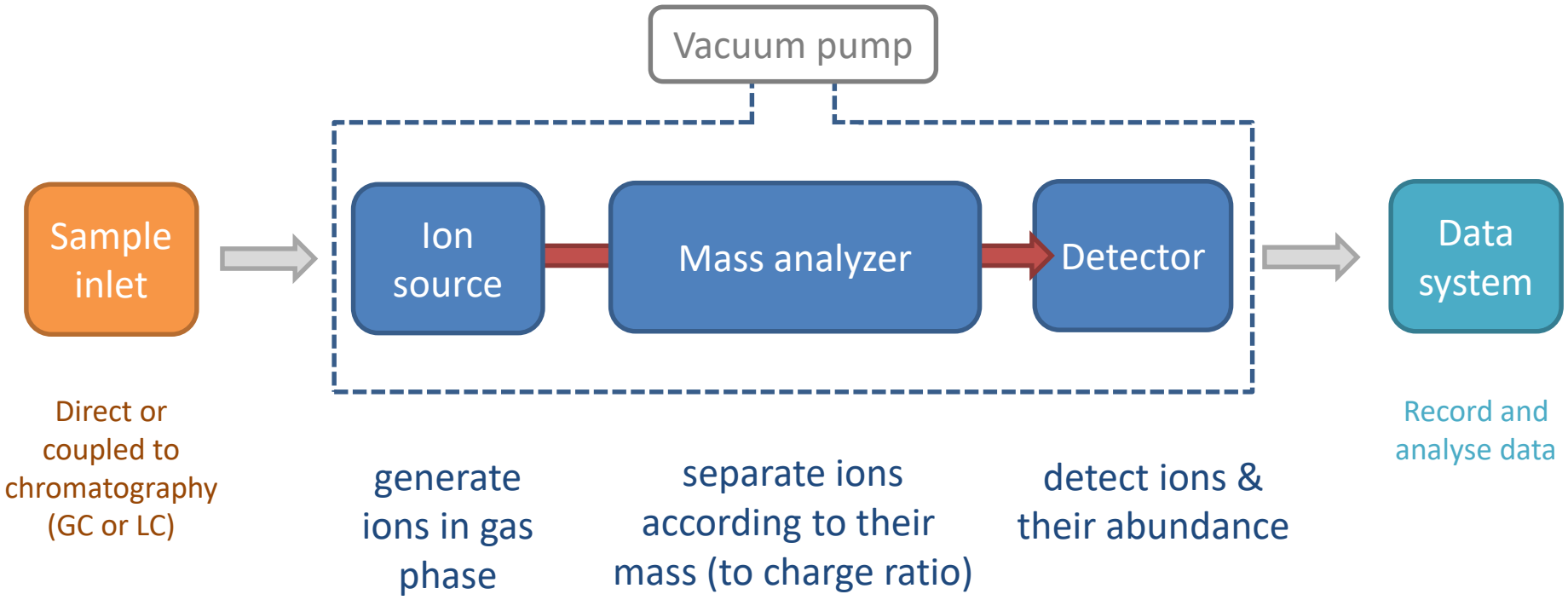
- positions of amino acids
- Position of post-translational modifications

Intact mass analysis limited by:

- Complex mixtures of unknown composition
- Analytes with isobaric mass
- Unknown modifications or processing of analytes
- Technical issues with biomolecules

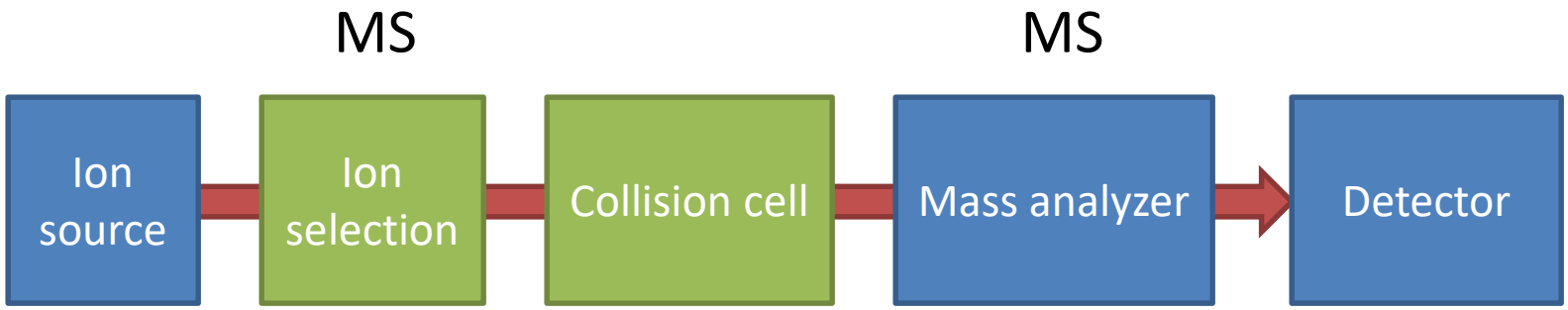
Generate more & specific information

A basic mass spectrometer

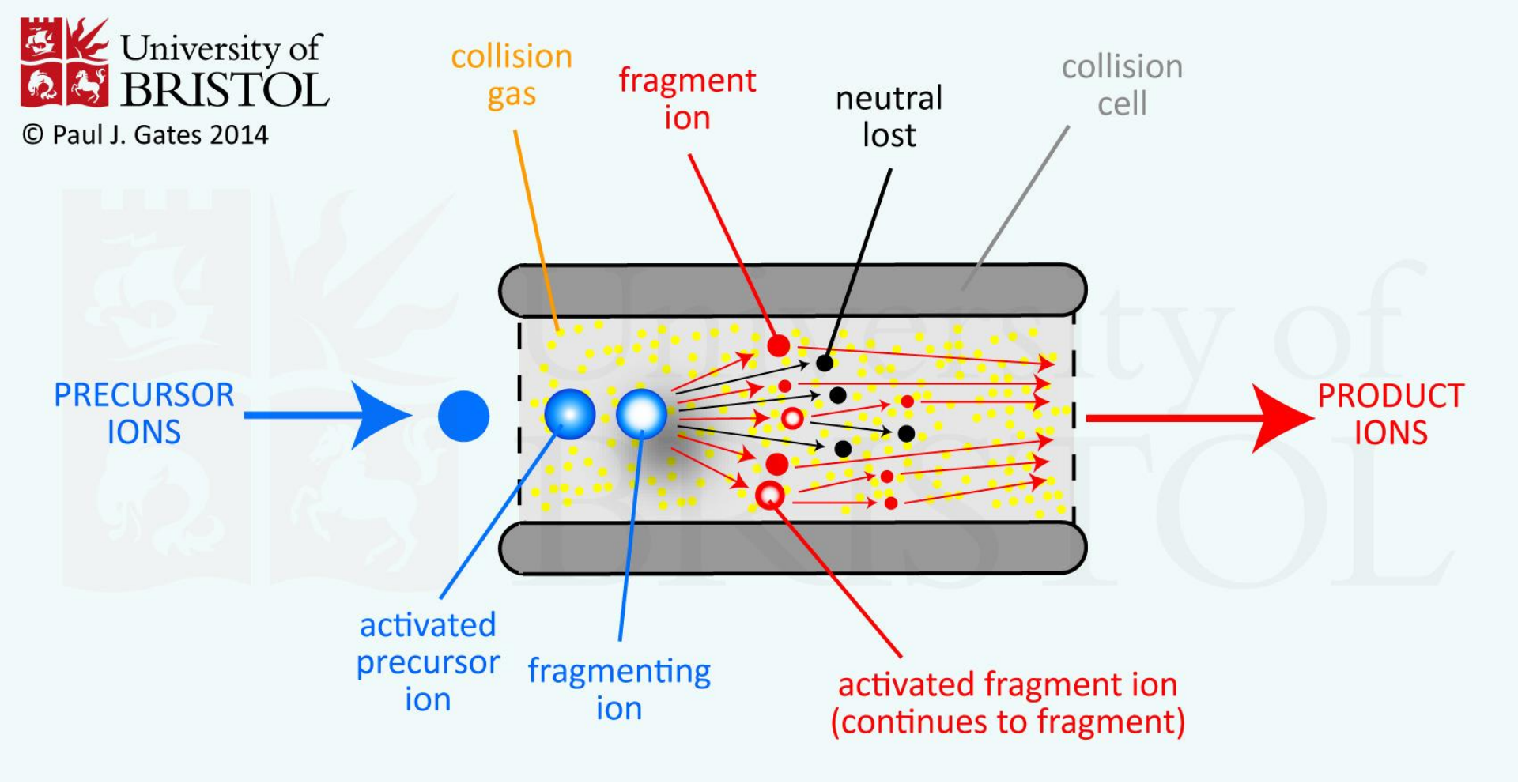


A basic mass spectrometer

Generating more & specific information: The MS/MS approach



A collision cell



Mapping fragments to a sequence

Query	Start – End	Observed	Mr(expt)	Mr(calc)	ppm	M	Score	Expect	Rank	U	Peptide
2	2 – 41	4631.2100	4630.2027	4630.1915	2.42	0	15	1.2e+002	1	U	M.SHHWGYGKHNGPEHWHKDFPIANGERQSPVDIDTKAVVQD.P + Acetyl (N-term)
61	180 – 260	9216.9194	9215.9122	9215.8929	2.09	0	280	4.9e-025	1	U	D.PGSLLPNVLDYWTYPGSLTTPPLLESVTWIVLKEPISVSSQQMLKFRTLNFNAEGEPELLMLANWRPAQPLKNRQVRGFPK.-
58	185 – 260	8749.6395	8748.6322	8748.6185	1.56	0	268	7.9e-024	1	U	L.PNVLDYWTYPGSLTTPPLLESVTWIVLKEPISVSSQQMLKFRTLNFNAEGEPELLMLANWRPAQPLKNRQVRGFPK.-
52	189 – 260	8326.3895	8325.3822	8325.3703	1.42	0	240	6.7e-021	1	U	L.DYWTYPGSLTTPPLLESVTWIVLKEPISVSSQQMLKFRTLNFNAEGEPELLMLANWRPAQPLKNRQVRGFPK.-
49	190 – 260	8211.3695	8210.3622	8210.3434	2.29	0	232	3.5e-020	1	U	D.YWTYPGSLTTPPLLESVTWIVLKEPISVSSQQMLKFRTLNFNAEGEPELLMLANWRPAQPLKNRQVRGFPK.-
46	191 – 260	8048.2995	8047.2922	8047.2801	1.51	0	228	8.6e-020	1	U	Y.WTYPGSLTTPPLLESVTWIVLKEPISVSSQQMLKFRTLNFNAEGEPELLMLANWRPAQPLKNRQVRGFPK.-
42	192 – 260	7862.2194	7861.2122	7861.2007	1.45	0	227	7.8e-020	1	U	W.TYPGSLTTPPLLESVTWIVLKEPISVSSQQMLKFRTLNFNAEGEPELLMLANWRPAQPLKNRQVRGFPK.-
37	193 – 260	7761.1794	7760.1722	7760.1531	2.46	0	222	2.4e-019	1	U	T.YPGSLTTPPLLESVTWIVLKEPISVSSQQMLKFRTLNFNAEGEPELLMLANWRPAQPLKNRQVRGFPK.-
29	194 – 260	7598.1094	7597.1022	7597.0897	1.64	0	222	2.1e-019	1	U	Y.PGSLTTPPLLESVTWIVLKEPISVSSQQMLKFRTLNFNAEGEPELLMLANWRPAQPLKNRQVRGFPK.-
25	195 – 260	7501.0594	7500.0522	7500.0370	2.03	0	204	1.6e-017	1	U	P.GSLTTPPLLESVTWIVLKEPISVSSQQMLKFRTLNFNAEGEPELLMLANWRPAQPLKNRQVRGFPK.-
22	196 – 260	7444.0295	7443.0222	7443.0155	0.90	0	189	4.3e-016	1	U	G.SLTTPPLLESVTWIVLKEPISVSSQQMLKFRTLNFNAEGEPELLMLANWRPAQPLKNRQVRGFPK.-
19	197 – 260	7356.9995	7355.9922	7355.9835	1.18	0	183	1.6e-015	1	U	S.LTTPPLLESVTWIVLKEPISVSSQQMLKFRTLNFNAEGEPELLMLANWRPAQPLKNRQVRGFPK.-
16	199 – 260	7142.8695	7141.8622	7141.8517	1.46	0	166	8.9e-014	1	U	T.TPPLLESVTWIVLKEPISVSSQQMLKFRTLNFNAEGEPELLMLANWRPAQPLKNRQVRGFPK.-
13	200 – 260	7041.8295	7040.8222	7040.8040	2.57	0	160	3.3e-013	1	U	T.PPLLESVTWIVLKEPISVSSQQMLKFRTLNFNAEGEPELLMLANWRPAQPLKNRQVRGFPK.-
10	210 – 260	5906.1894	5905.1822	5905.1763	0.99	0	87	8.3e-006	1	U	I.VLKEPISVSSQQMLKFRTLNFNAEGEPELLMLANWRPAQPLKNRQVRGFPK.-
7	213 – 260	5565.9495	5564.9422	5564.9289	2.39	0	70	0.00053	1	U	K.EPISVSSQQMLKFRTLNFNAEGEPELLMLANWRPAQPLKNRQVRGFPK.-
4	214 – 260	5436.9095	5435.9022	5435.8863	2.92	0	61	0.0037	1	U	E.PISVSSQQMLKFRTLNFNAEGEPELLMLANWRPAQPLKNRQVRGFPK.-

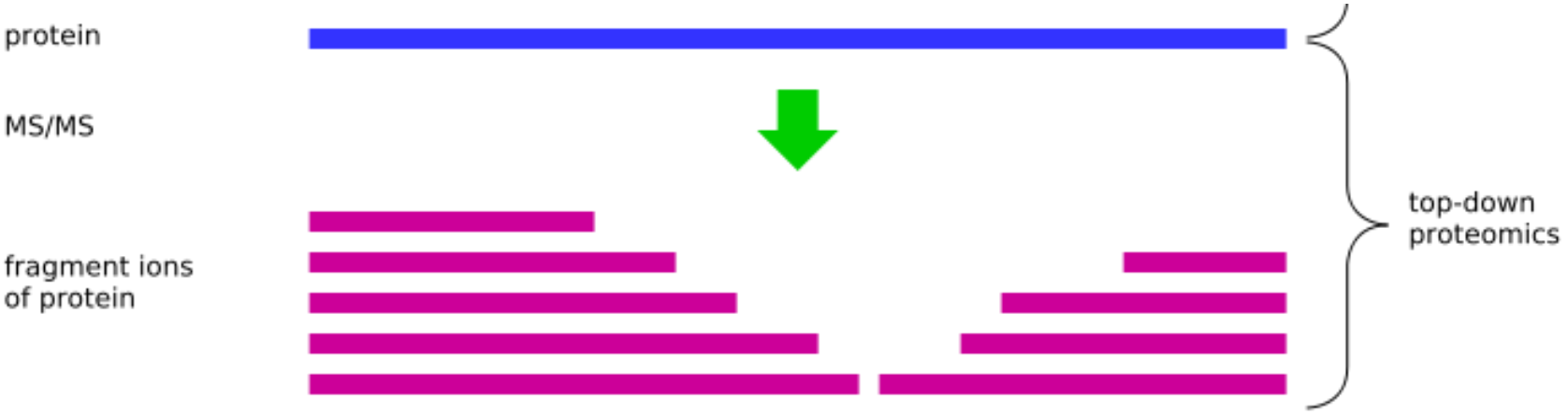
Protein sequence coverage: 46%

Matched peptides shown in ***bold red***.

```
   1 MSHHWGYGKH NGPEHWHKDF PIANGERQSP VDIDTKAVVQ DPALKPLALV
  51 YGEATSRRMV NNGHSFNVEY DDSQDKAVLK DGPLTGTYRL VQFHFHWGSS
 101 DDQGSEHTVD RKKYAAELHL VHWNTKYGDF GTAAQQPDGL AVVGVFLKVG
 151 DANPALQKVL DALDSIKTKG KSTDFPNFDP GSLLPNVLDY WTYPGSLTTP
 201 PLLESVTWIV LKEPISVSSQ QMLKFRTLNF NAEGEPELLM LANWRPAQPL
 251 KNRQVRGFPK
```

Database dependent!

Top-down proteomics



Drawbacks of top-down proteomics:

- Chromatographic separation of proteins is difficult
- Complex samples are very challenging (charge envelopes, overlapping signals)
- Analysis limited to proteins of 100 kDa max. (for complex mixtures)
- Data analysis algorithms not mature

What‘s the alternative?

The alternative: break proteins into something more manageable before LC-MS-MS

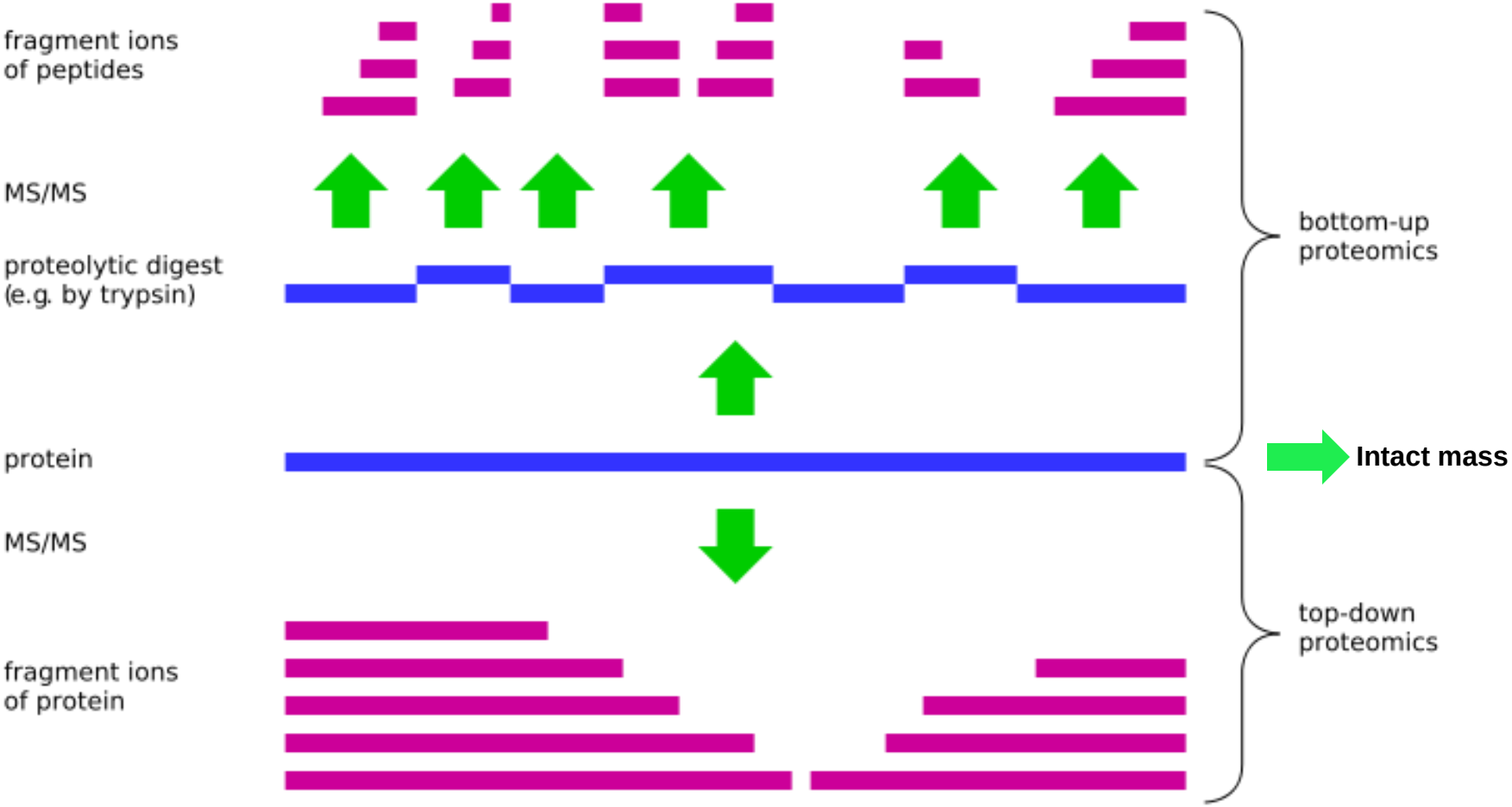
Digestion of BSA with trypsin: theoretically 75% coverage

```
        10         20         30         40         50         60
dthkseiahr fkDLGEEHFK GLVLIAFSQY LQQCPFDEHV KLVNELTEFA KTCVADESHA

        70         80         90        100        110        120
GCEKSLHTLF GDELCKvasl rETYGDMADC CEKqeperNE CFLSHKDDSP DLPKLKPDPN

       130        140        150        160        170        180
TLCDEFKade kkfwgkYLYE IARrHPYFYA PELLYYANKY NGVFQECCQA EDKgacllpk

       190        200        210        220        230        240
ietmrekvla ssarqrlrca siqkfgeral kawsvarlsq kfpkAEFVEV TKLVTDLTKv

       250        260        270        280        290        300
hkECCHGDLL ECADDRadla kYICDNQDTI SSKlkECCDK PLLEKSHCIA EVEKDAIPEN

       310        320        330        340        350        360
LPPLTADFAE DKdvckNYQE AKDAFLGSFL YEYSRrHPEY AVSVLLRlak EYEATLEECC

       370        380        390        400        410        420
AKDDPHACYS TVFDKlkHLV DEPQNLIKQN CDQFEKLGEY GFQNALIVRy trkVPQVSTP

       430        440        450        460        470        480
TLVEVSRslg kvgtrCCTKP ESERMPCTED YLSLILNRLC VLHEKtpvse kvtkCCTESL

       490        500        510        520        530        540
VNRRPCFSAL TPDETYVPKa fdekLFTFHA DICTLPDTEK qikkQTALVE LLKhkpkATE

       550        560        570        580
EQLKTVMENF VAFVDKccaa ddkEACFAVE GPKLVVSTQT ALA
```

Where does trypsin cut?

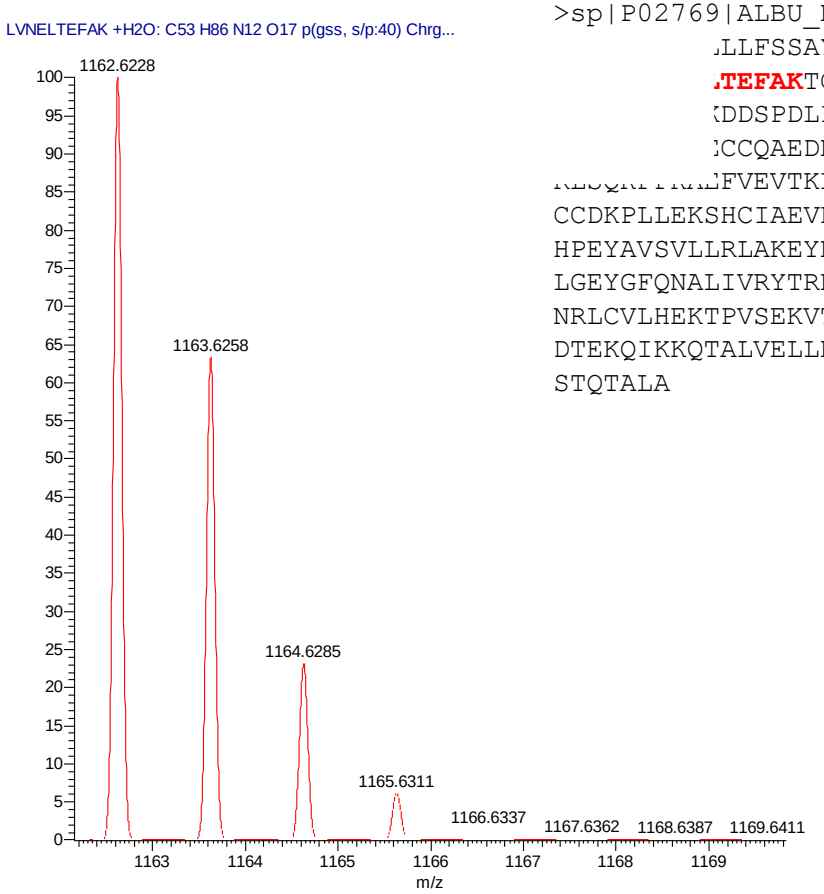
Why not 100% coverage?

Bottom-up vs. top-down proteomics



The advantage of bottom-up

- Easier chromatography and handling of peptides
- Simpler signals, with accurate monoisotopic masses:



>sp|P02769|ALBU_BOVIN Serum albumin
LLFSSAYSRGVFRRDTHKSEIAHRFKDLGEEHFKGLVLIAFSQYLQQCPF
TEFAKTCVADESHAGCEKSLHTLFGDELCKVASLRETYGDMADCCEKQEP
DDSPDLPKLKPDPNTLCDEFKADEKKFWGKYLYEIARRHPYFYAPELLYY
CCQAEDKGACLLPKIETMREKVLASSARQRLRCASIQKFGERALKAWSVA
FVEVTKLVTDLTKVHKECCHGDLLECADDRADLAKYICDNQDTISSKLKE
CCDKPLLEKSHCIAEVEKDAIPENLPPLTADFAEDKDVCKNYQEAKDAFLGSFLYEYSRR
HPEYAVSVLLRLAKEYEATLEECCAKDDPHACYSTVFDKLKHLVDEPQNLIKQNCDQFEK
LGEYGFQNALIVRYTRKVPQVSTPTLVEVSRSLGKVGTRCCTKPESERMPCTEDYLSLIL
NRLCVLHEKTPVSEKVTKCCTESLVNRRPCFSALTPDETYVPKAFDEKLFTFHADICTLP
DTEKQIKKQTALVELLKHKPKATEEQLKTVMENFVAFVDKCCAADDKEACFAVEGPKLVV
STQTALA

Monoisotopic mass: 1162.6228 u
Average mass: 1163.34 u
Nominal mass: 1163 u

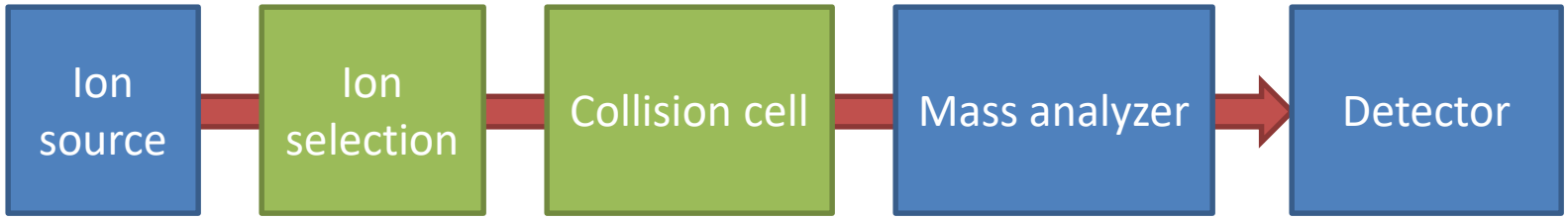
Peptide mass is not enough

ISGGDALQSCVDR	1320.4 Da
DVCSQLADGGSIR	1320.4 Da
IpSGGDALQSCVDR	1400.4 Da
ISGGDALQpSCVDR	1400.4 Da

No information about:

- positions of amino acids
- Position of post-translational modifications

More & specific info: the MS/MS approach



Peptide fragmentation (CID)

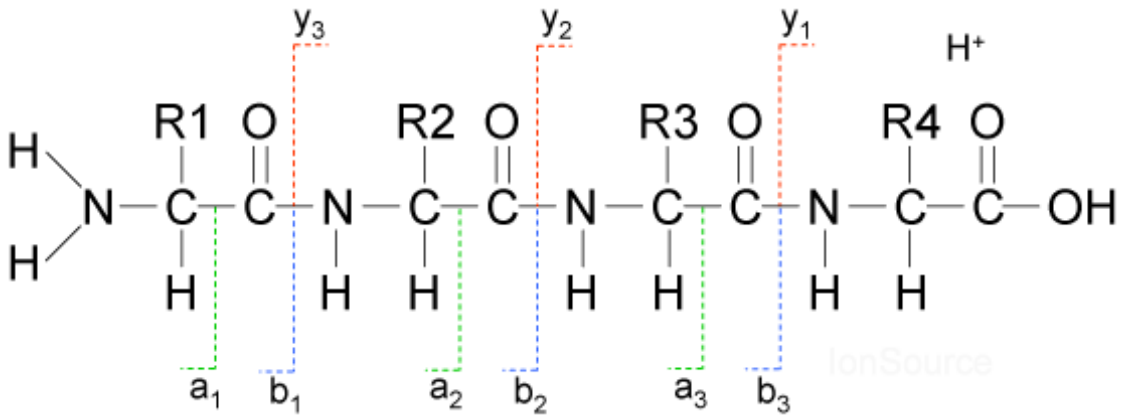
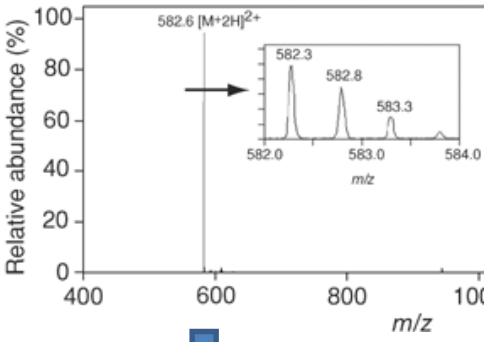


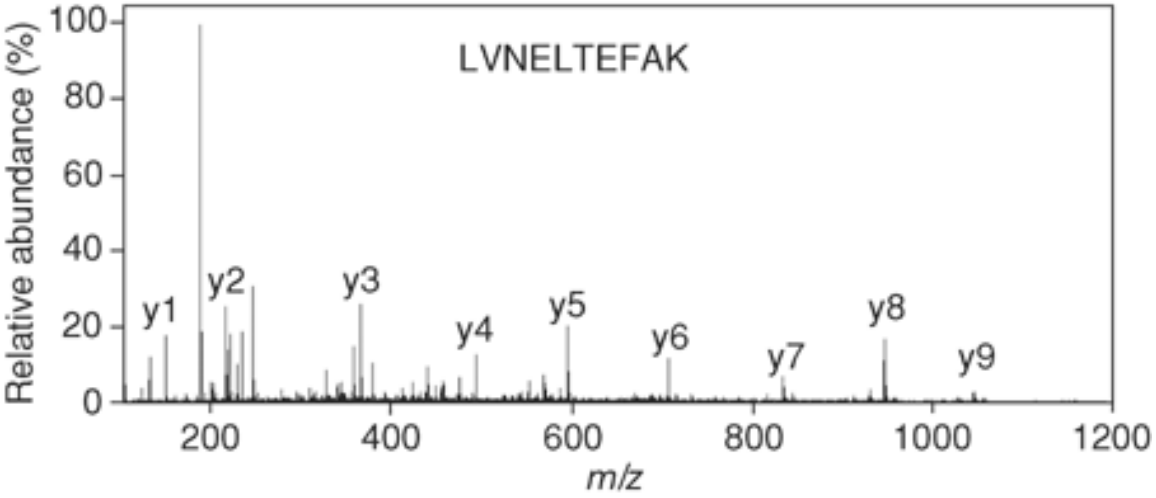
Image: www.ionsource.com

How to identify a peptide from spectrum

MS1 precursor mass

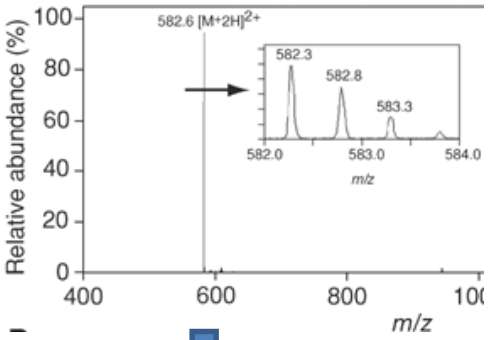


MS/MS fragment masses

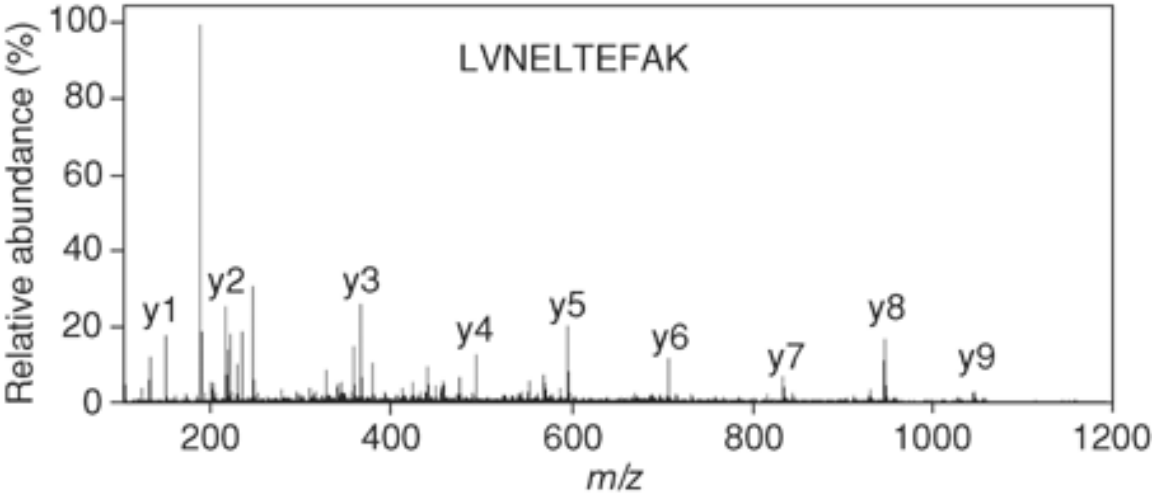


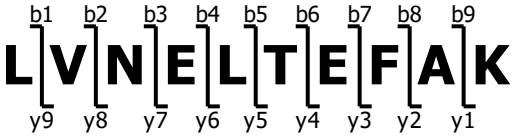
How to identify a peptide from MS/MS spectrum

MS1 precursor mass



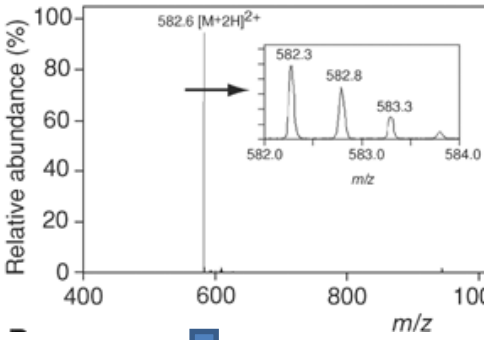
MS/MS fragment masses



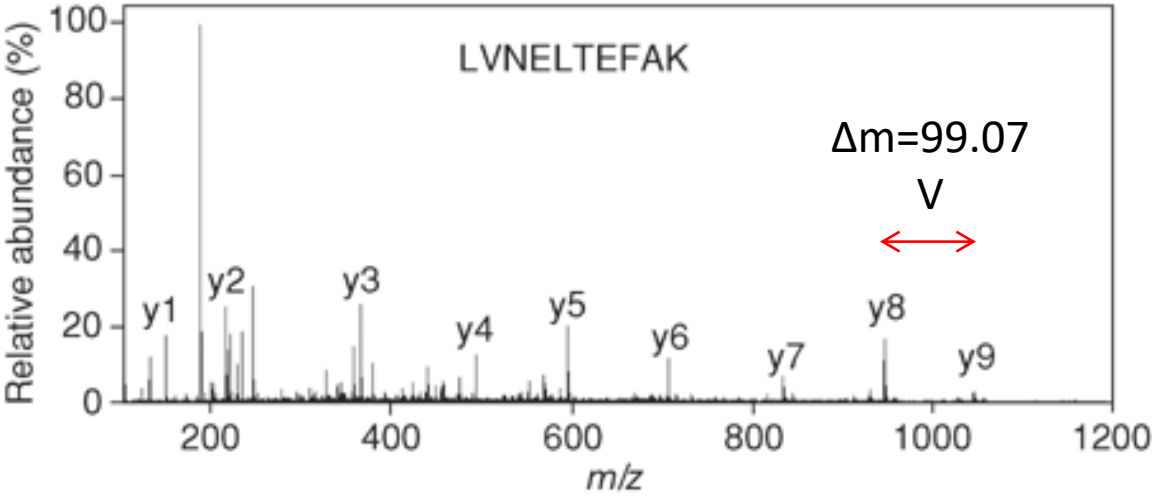


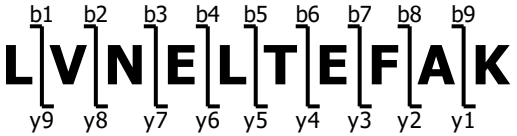
How to identify a peptide from MS/MS spectrum

MS1 precursor mass



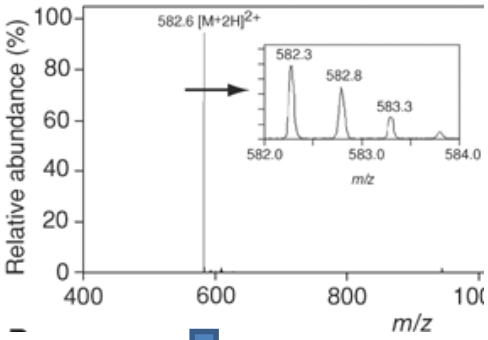
MS/MS fragment masses



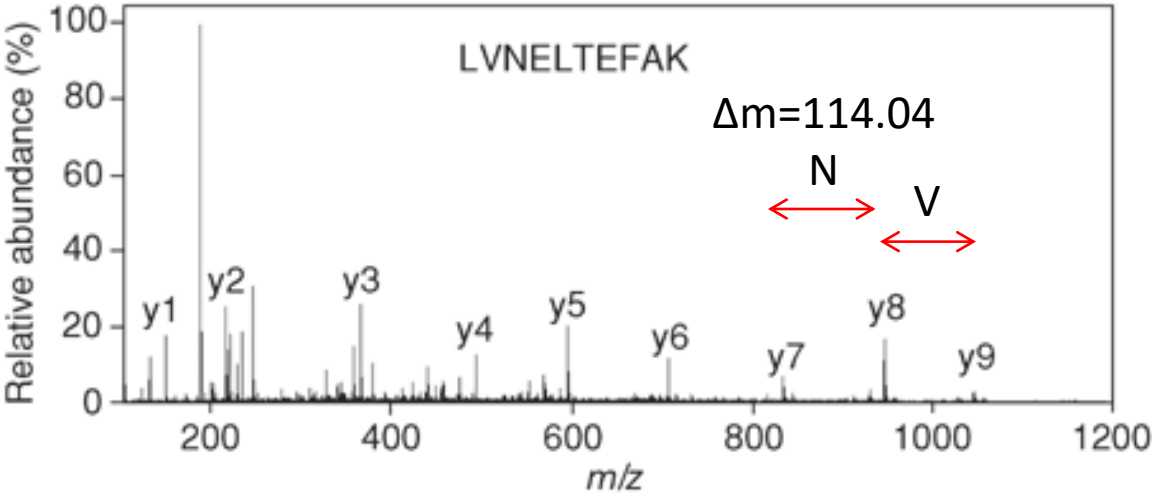


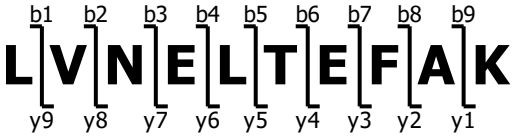
How to identify a peptide from MS/MS spectrum

MS1 precursor mass



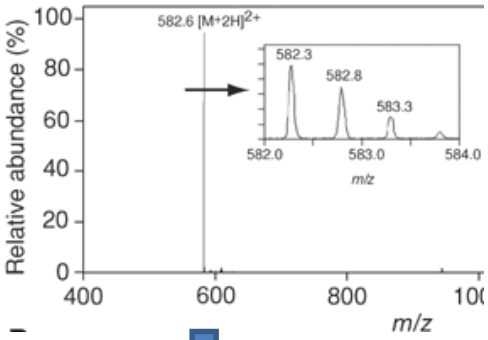
MS/MS fragment masses



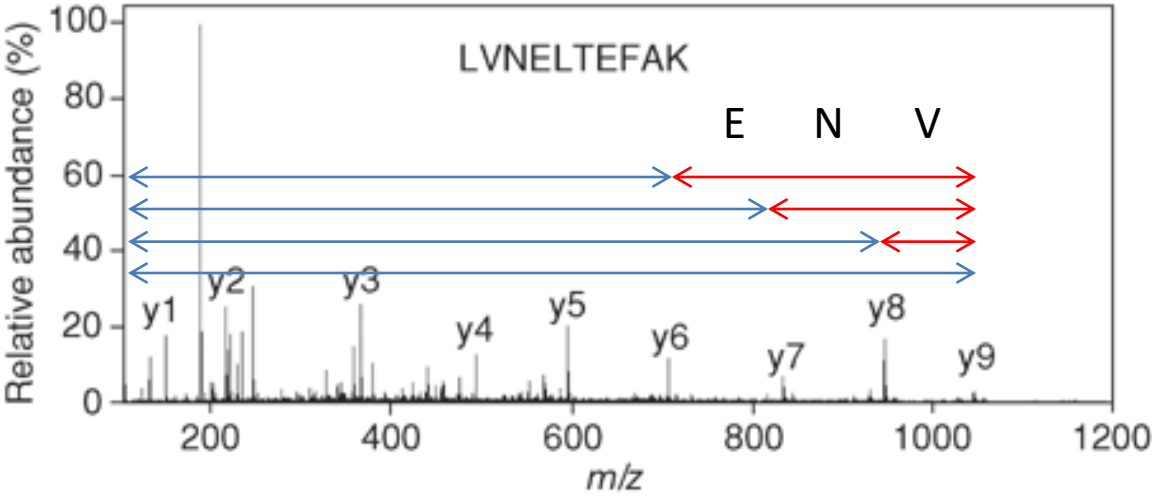


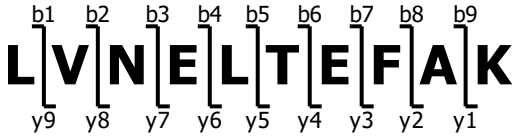
How to identify a peptide from MS/MS spectrum

MS1 precursor mass



MS/MS fragment masses





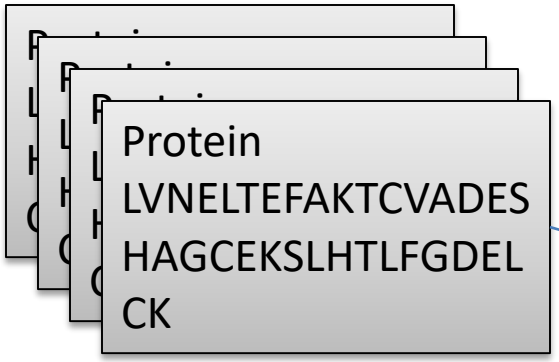
b	Res.	y
114.092	1 L 10	
213.160	2 V 9	1050.547
327.203	3 N 8	951.478
456.246	4 E 7	837.435
569.330	5 L 6	708.393
670.378	6 T 5	595.309
799.420	7 E 4	494.261
946.489	8 F 3	365.218
1017.526	9 A 2	218.150
	10 K 1	147.113

Information:
Precursor mass
Ion-series

De-novo sequencing or
comparison with database

The basic idea of a DB searches

Database of protein sequences:



Take a sequence database and calculate masses of all tryptic peptides and their fragment ions to generate a mass database.

Mass database

In-silico digest

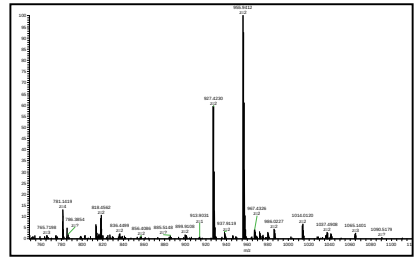
Peptide list	**Precursor masses**	**Fragment ion masses**
LVNELTEFAK	1163.6306	b1 (114.0919), b2, ..., y1 (147.1134), ...
TCVADESHAGCEK	1349.5460	b1 (102.0556), b2, ..., y1 (147.1134), ...
SLHTLFGDELCK	1362.6722	b1 (88.0399), b2, ..., y1 (147.1134), ...

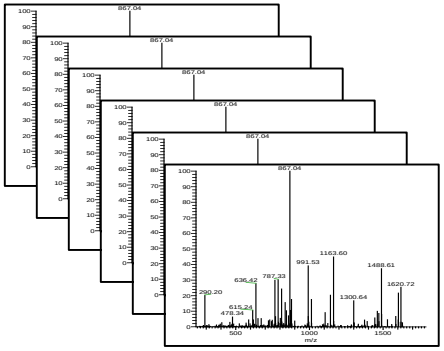
The basic idea of a DB searches

Acquired raw data

MS2: product ion masses

MS 1: precursor masses





Mass database

Match measured spectra to theoretical spectra

Coverage?

How many proteins can you detect?

In a 2 h run on the newest generation instrument:

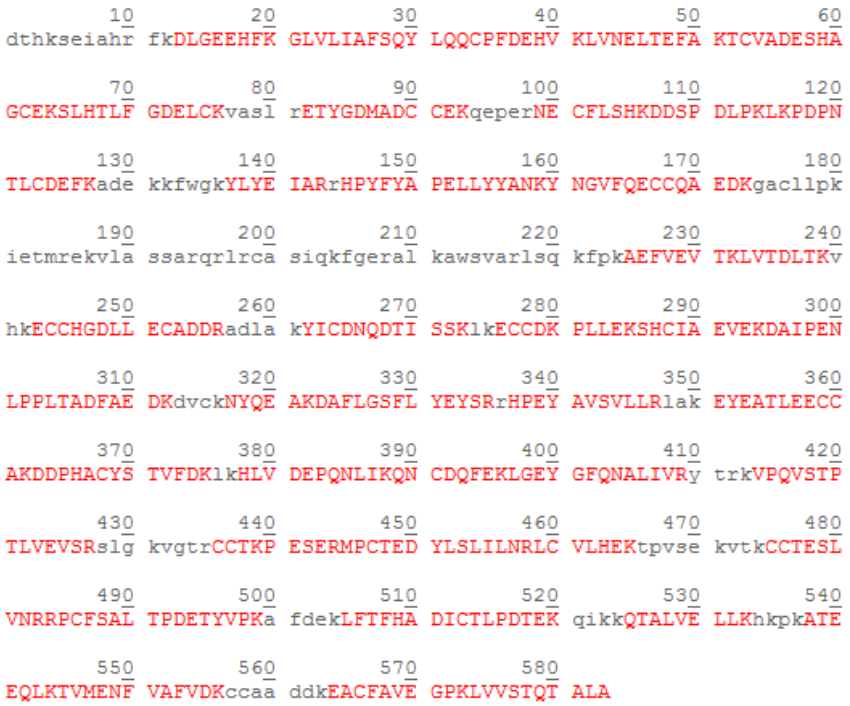
- 25.000 peptides (>50.000 spectra)
- >4.000 proteins (depending on organism)
- at 1% FDR!

Sample requirement: 2 µg peptide sample

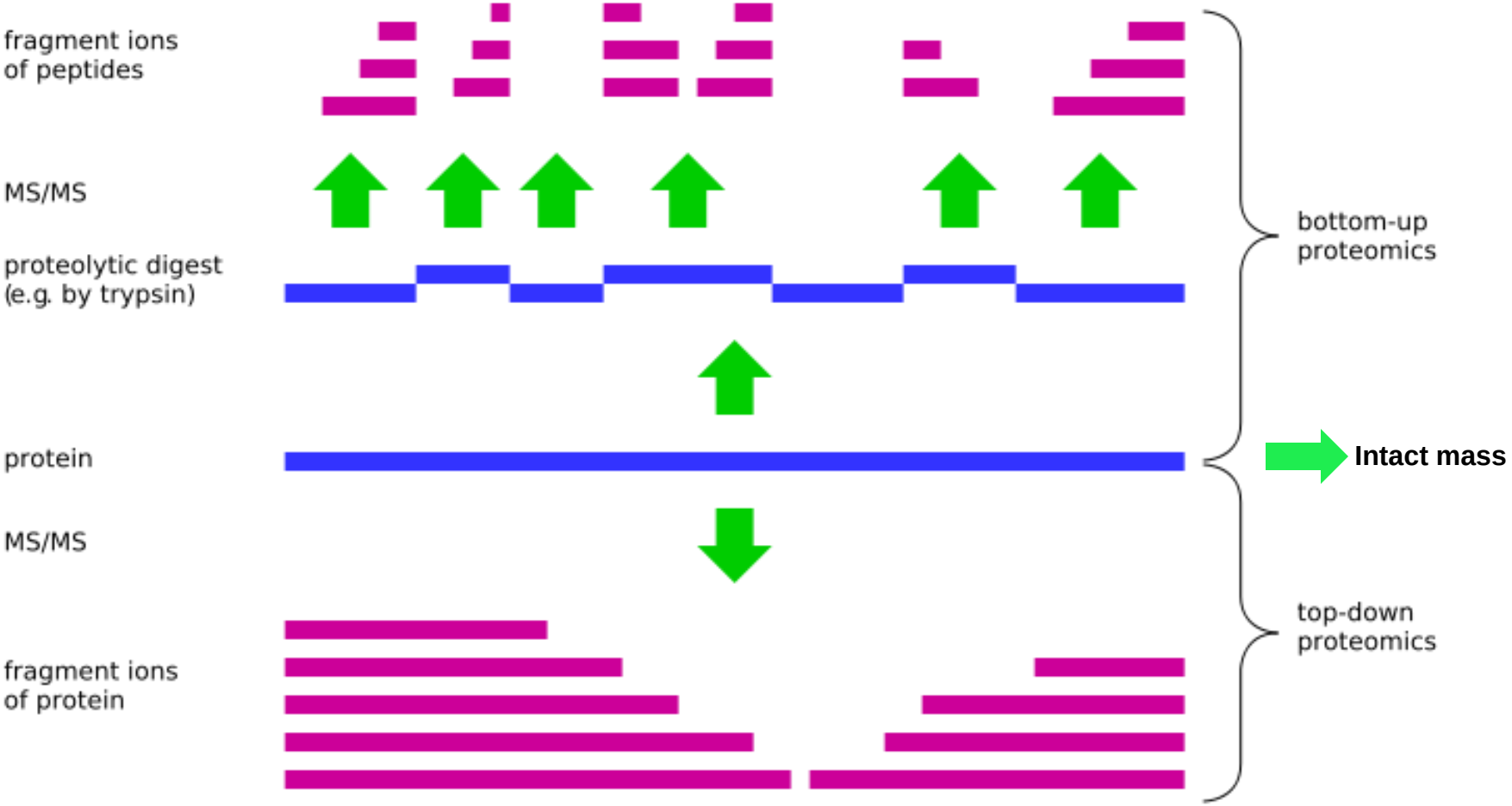
Sample preparation time: 6h plus digest time

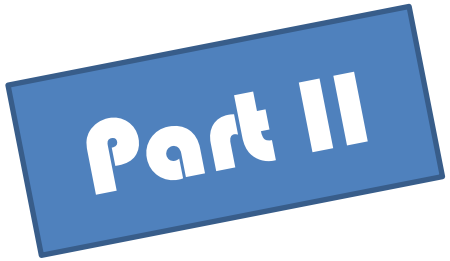
Coverage at protein level?

- Almost never 100%, typically 1-80%
- Modifications with low stoichiometry lost



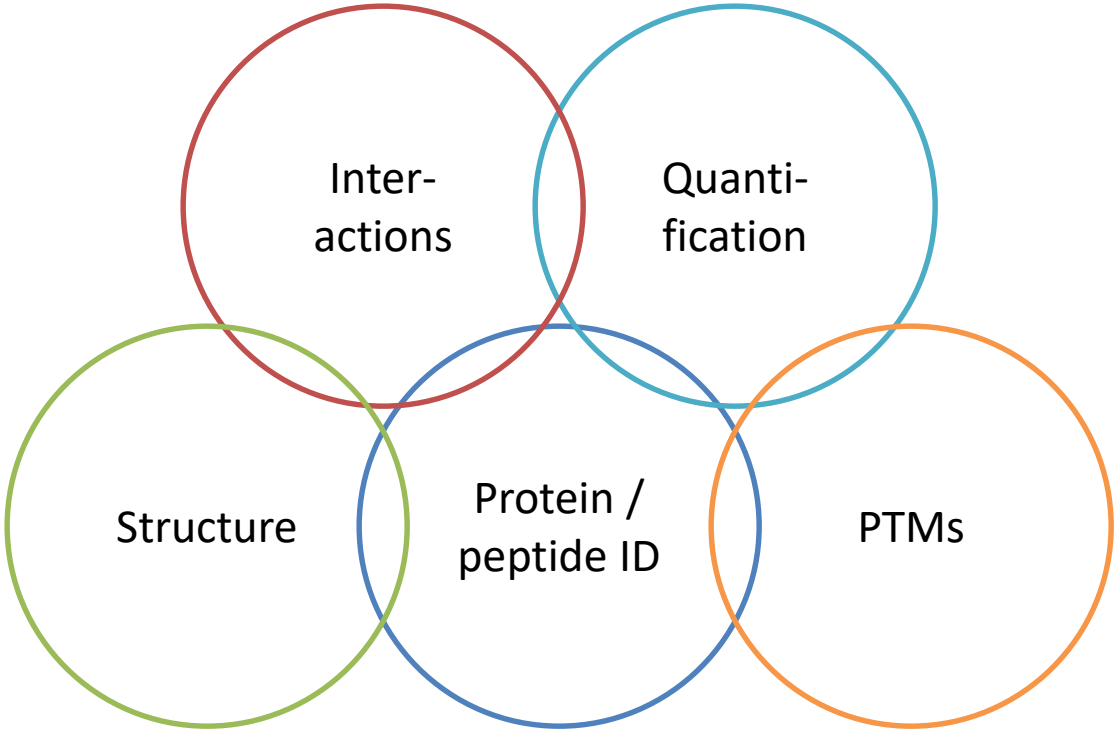
Bottom-up vs. top-down proteomics vs. intact mass analysis

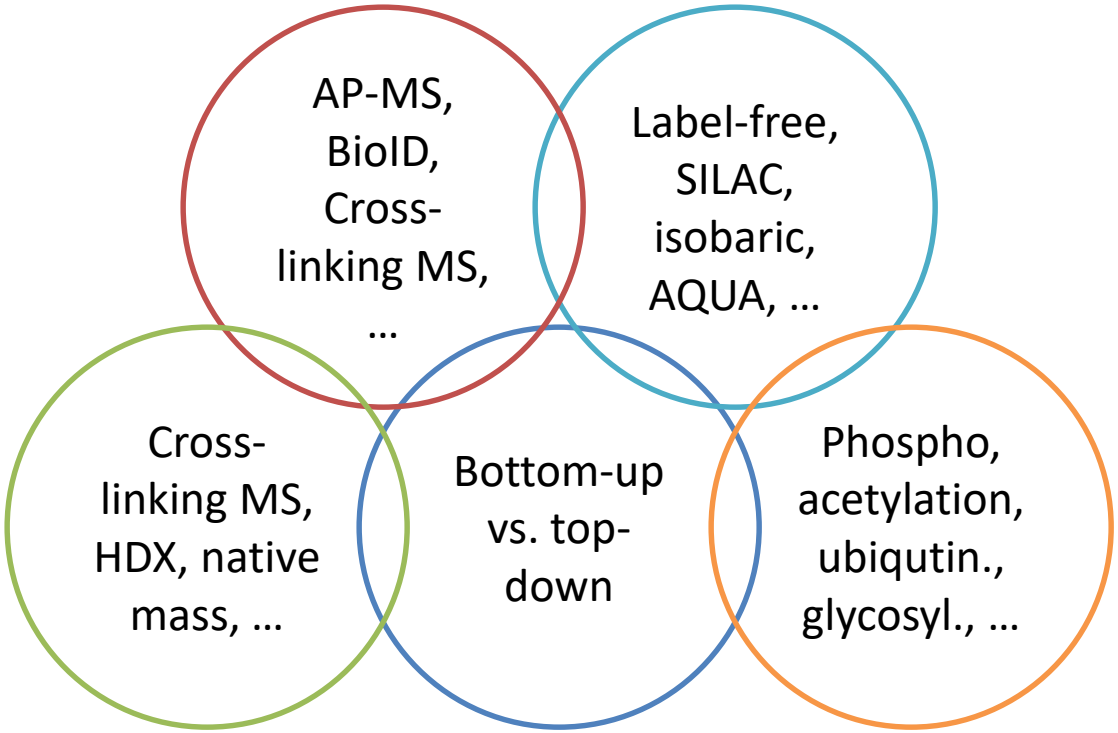




What is this all useful for?

Applications in proteomics




AP-MS, BioID, Cross-linking MS, …
Label-free, SILAC, isobaric, AQUA, …
Cross-linking MS, HDX, native mass, …
Bottom-up vs. top-down
Phospho, acetylation, ubiqutin., glycosyl., …

Characterizing structure & interactions with MS

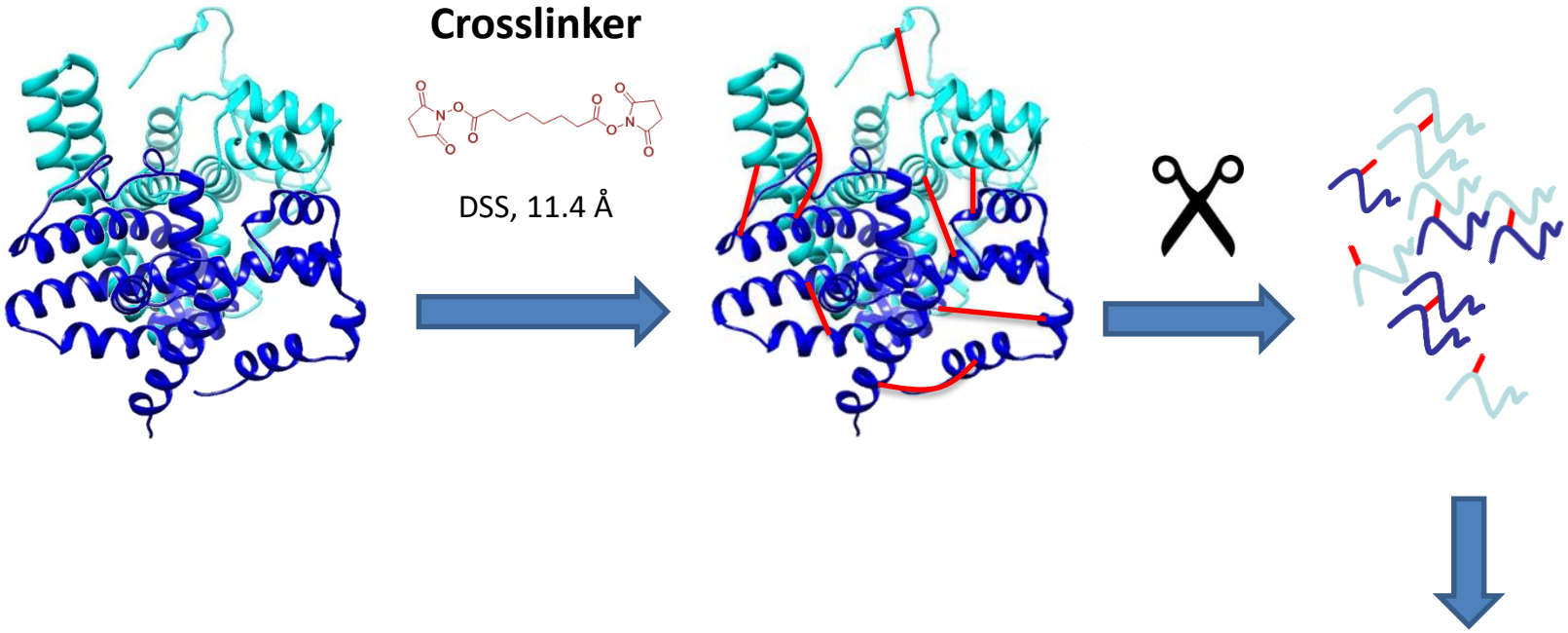
- Identify interaction partners (e.g. Affinity purification MS, BioID, etc.)
- Crosslinking MS
- Hydrogen-deuterium exchange (HDX)
- Native mass spectrometry

Why crosslinking?

- High-resolution structural tools not always applicable, especially for higher order protein complexes or flexible regions (e.g. IDPs)
- XL-MS is a complementary low-resolution tool
- XL-MS monitors proteins in solution

- What you can gain:
  - Distance constraints
  - Identification of interaction partners
  - Protein complex and network analysis

The principle of XL-MS



Data processing & evaluation

Identify cross-linked peptides ⇨ Distance constraints for structural modeling

LC-MS/MS analysis

Illustration (modified): Evelyn Rampler, IMP

The principle of XL-MS

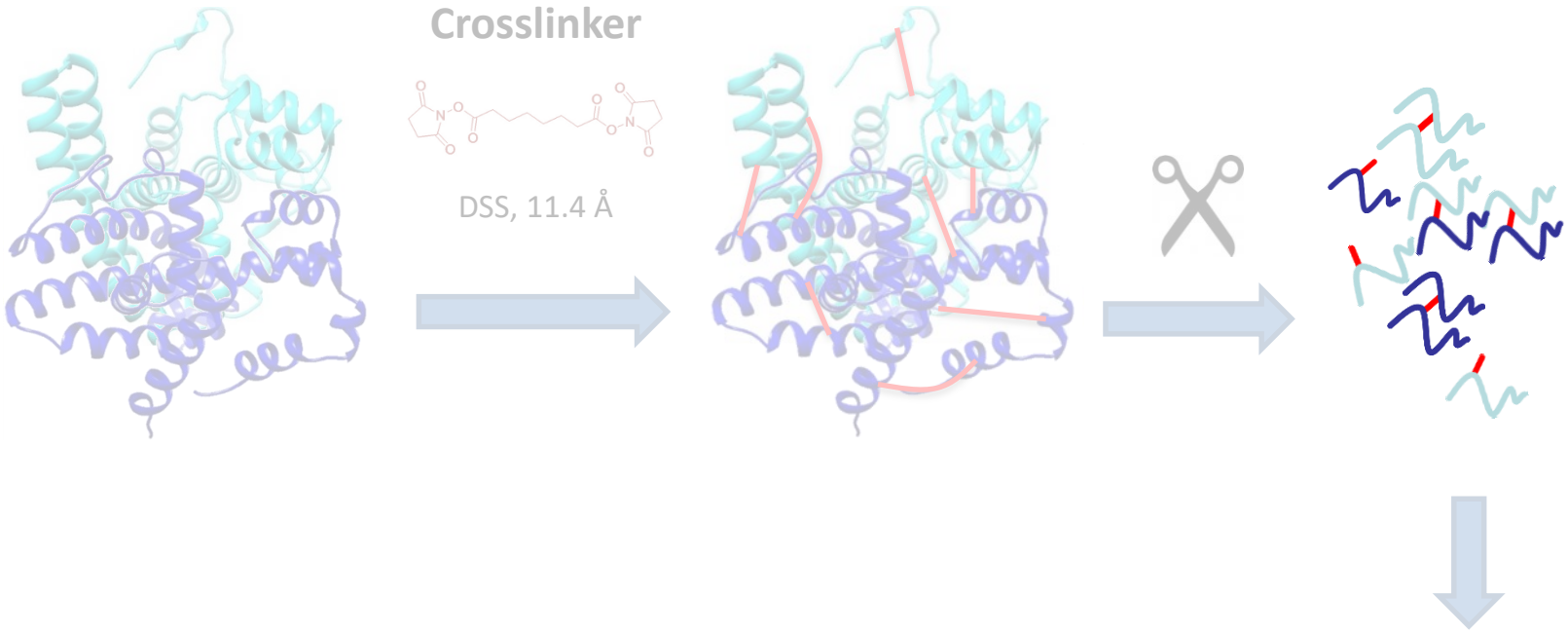
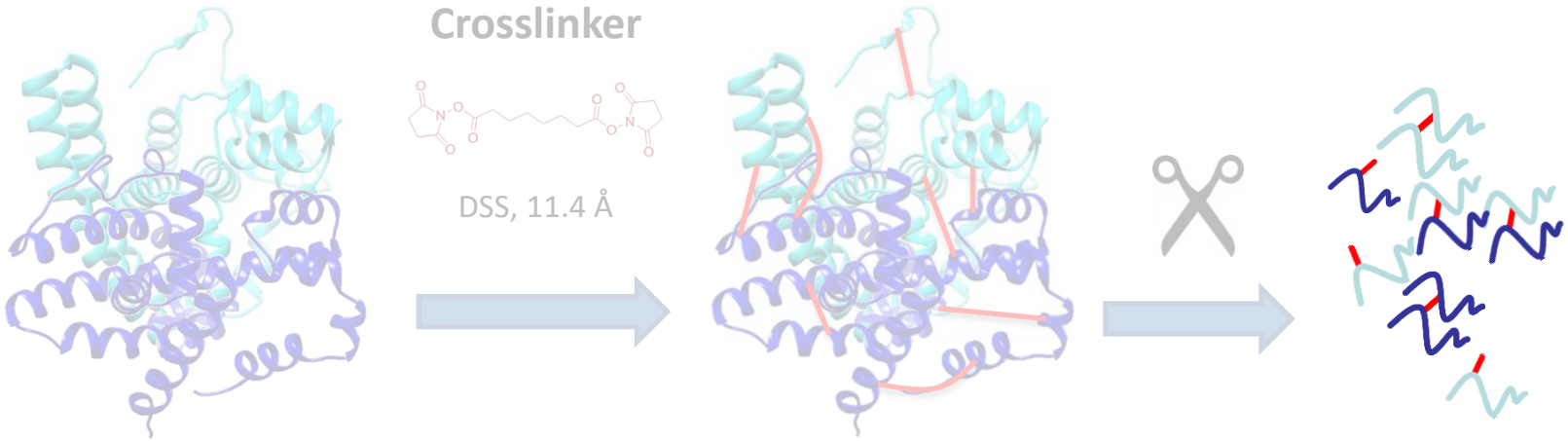


Illustration (modified): Evelyn Rampler, IMP

The principle of XL-MS



1. excess of uncross-linked peptides
2. low abundant cross-linked peptides

➔ low yield of cross-link information

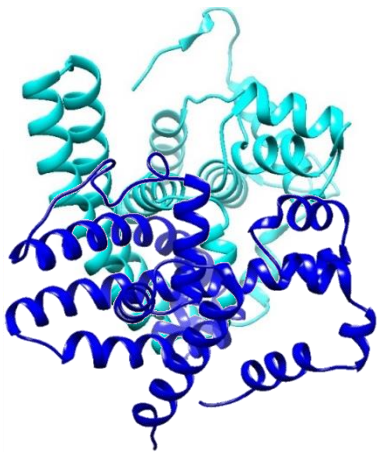
LC-MS/MS analysis

Data processing & evaluation

Identify cross-linked peptides
⇨ Distance constraints for structural modeling

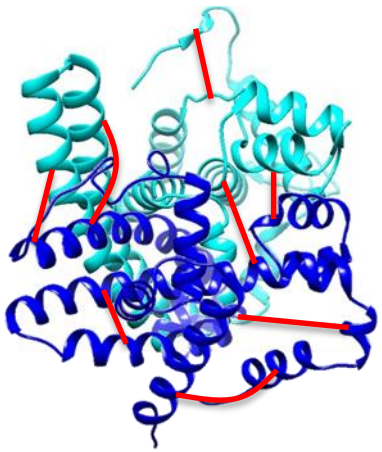
Illustration (modified): Evelyn Rampler, IMP

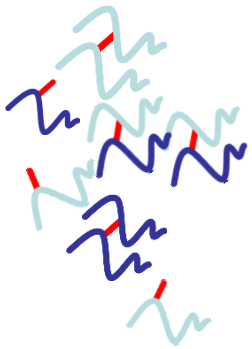
The principle of XL-MS



Crosslinker

DSS, 11.4 Å





enrichment of cross-linked peptides: e.g. SEC, SCX

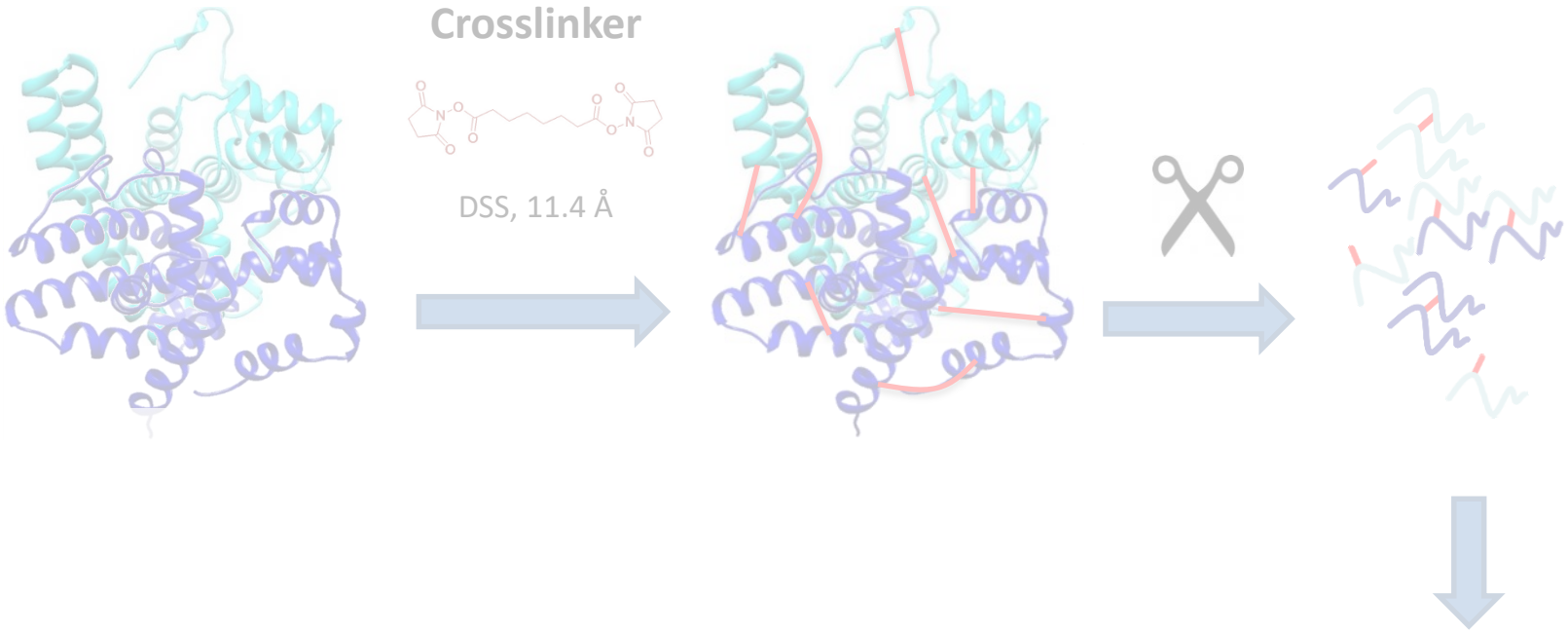
LC-MS/MS analysis

Data processing & evaluation

Identify cross-linked peptides ⇨ Distance constraints for structural modeling

Illustration (modified): Evelyn Rampler, IMP

The principle of XL-MS



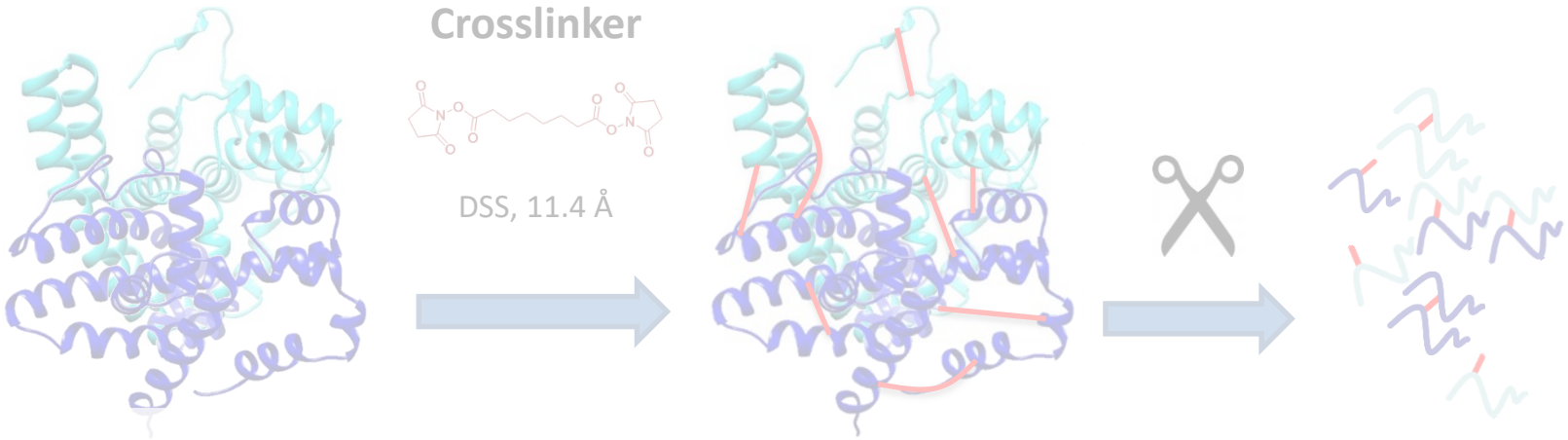
Data processing & evaluation

Identify cross-linked peptides
⇨ Distance constraints for structural modeling

LC-MS/MS analysis

Illustration (modified): Evelyn Rampler, IMP

The principle of XL-MS



1. complex fragmentation spectra
2. search space problem (n^2)
3. estimation of false positives

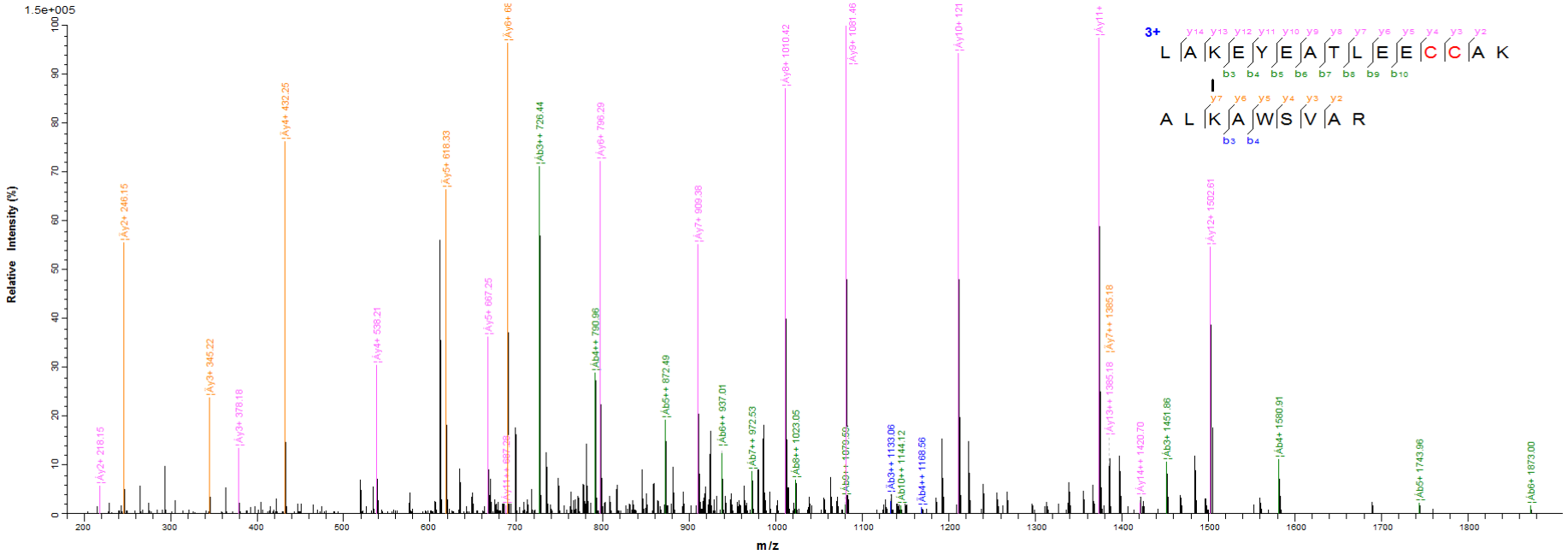
Data processing & evaluation

LC-MS/MS analysis

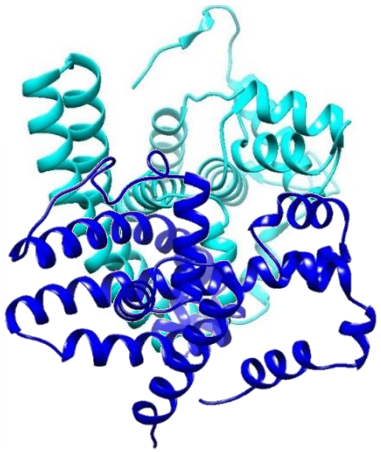
Identify cross-linked peptides
⇨ Distance constraints for structural modeling

Illustration (modified): Evelyn Rampler, IMP

Cross-linked peptide spectra are more complex

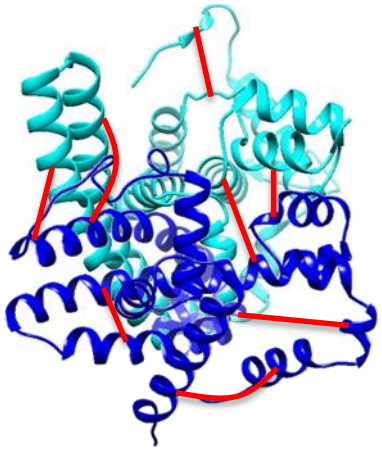


Choose your weapons wisely & protein-specific

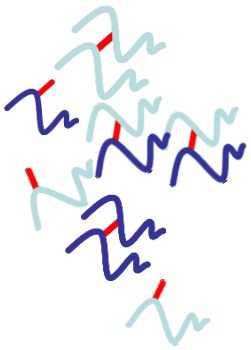


Crosslinker

DSS, 11.4 Å



Enzyme



Enrichment?

LC-MS/MS analysis

Workflow?

Data processing & evaluation

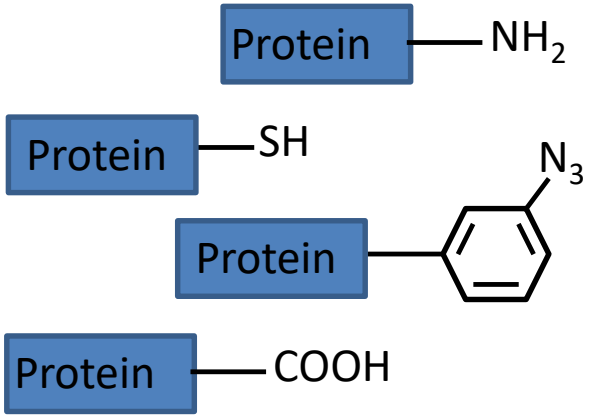
Identify cross-linked peptides ⇨ Distance constraints for structural modeling

Illustration (modified): Evelyn Rampler, IMP

Crosslinker chemistry

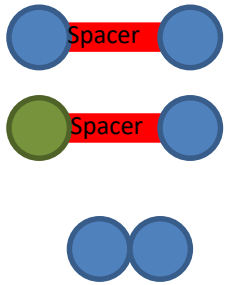
Types

- NH_2-reactive crosslinker
- Sulfhydryl-reactive crosslinker
- Photoreactive crosslinkers
- COOH-reactive crosslinker



Design:

- Homobifunctional
- Heterobifunctional
- Zero-length

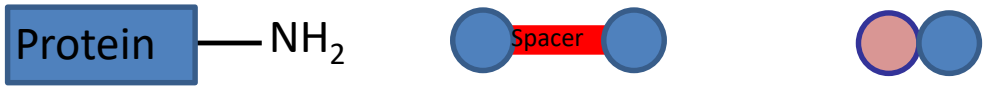


According to:

A. Sinz, Journal of Mass Spectrometry, *38:1225-1237*, 2010

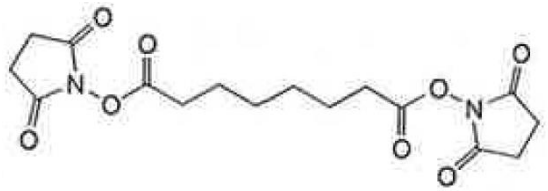
Slide kindly provided by Evelyn Rampler

Crosslinkers with different lengths



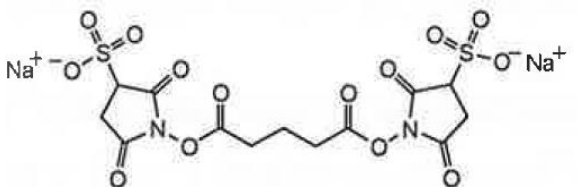
Crosslinker	Chemical name	spacer
DSS	*Di*(succinimidyl)suberate	11.4 Å
BS2G	*Bis*(sulfosuccinimidyl)glutarate	7.7 Å
$BS2G_{d0/d6}$	Di(sulfosuccinimidyl)glutarate	7.7 Å
EDC	1-Ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride	0 Å

DSS



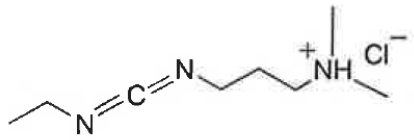
11.4 Å

BS2G



7.7 Å

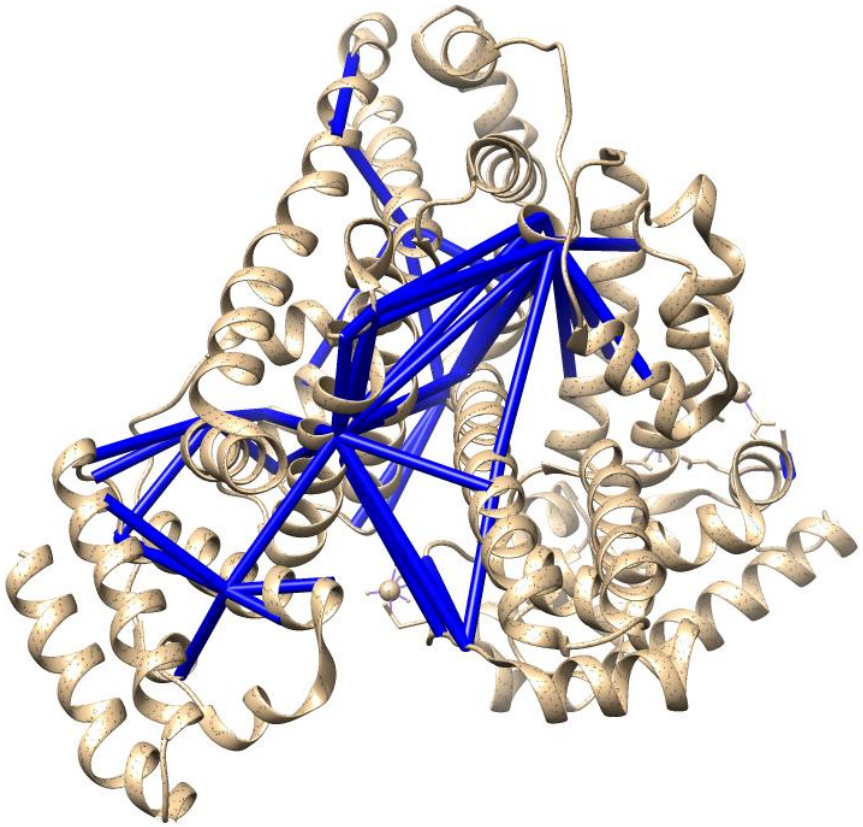
EDC



0 Å

Slide kindly provided by Evelyn Rampler

Example results: BSA



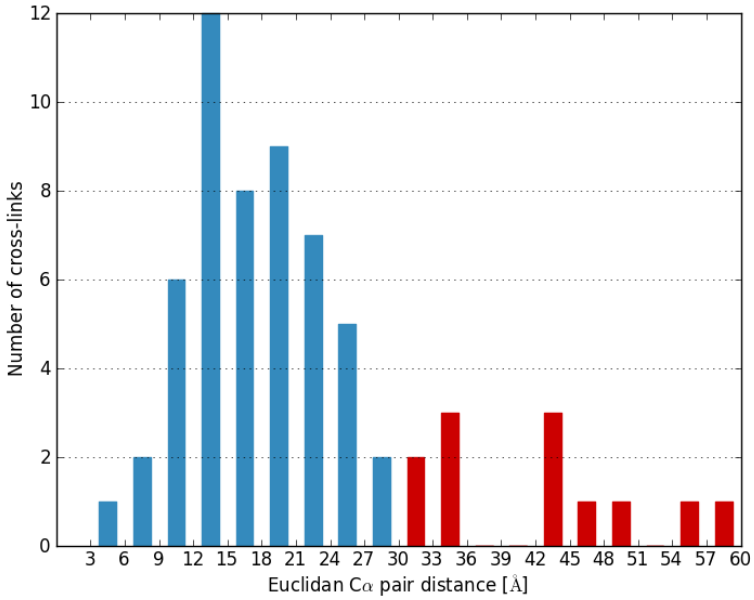
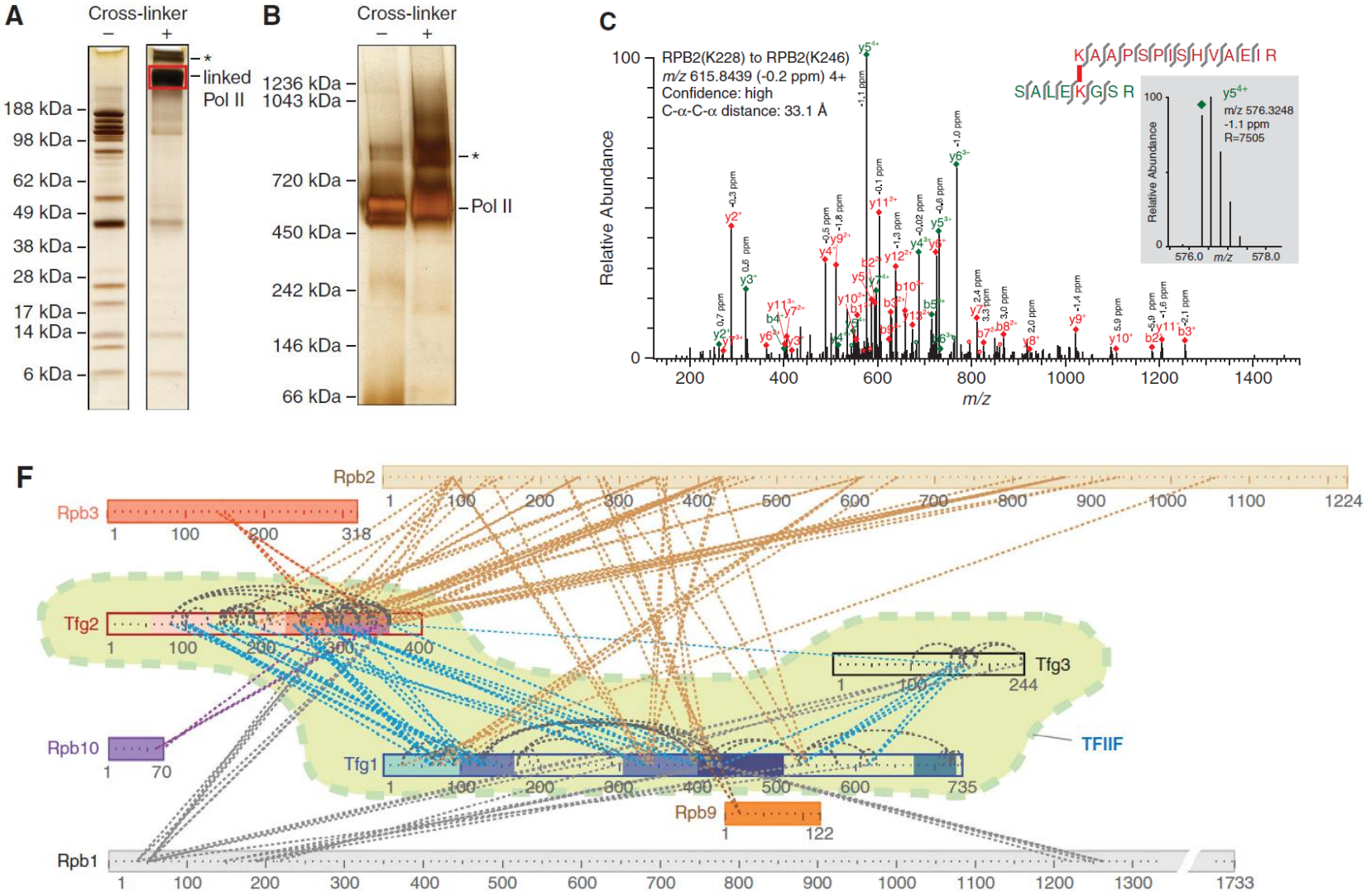


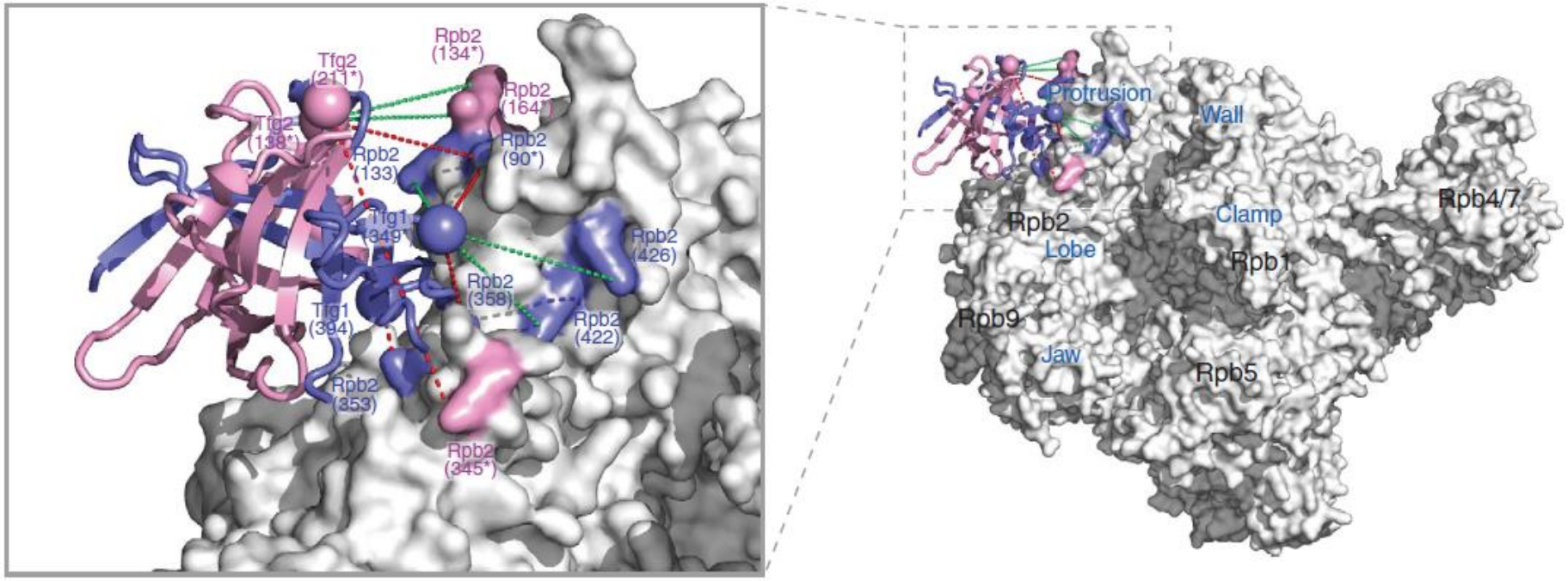
Illustration: Thomas Gossenreiter, MFPL

Example results: RNA Polymerase II and TFIIF



From: Chen, Z. A., Jawhari, A., Fischer, L., Buchen, C., Tahir, S., Kamenski, T., … Rappsilber, J. (2010). Architecture of the RNA polymerase II-TFIIF complex revealed by cross-linking and mass spectrometry. *The EMBO Journal*, *29*(4), 717–726. doi:10.1038/emboj.2009.401

Binding of TFIIF to RNA Polymerase II



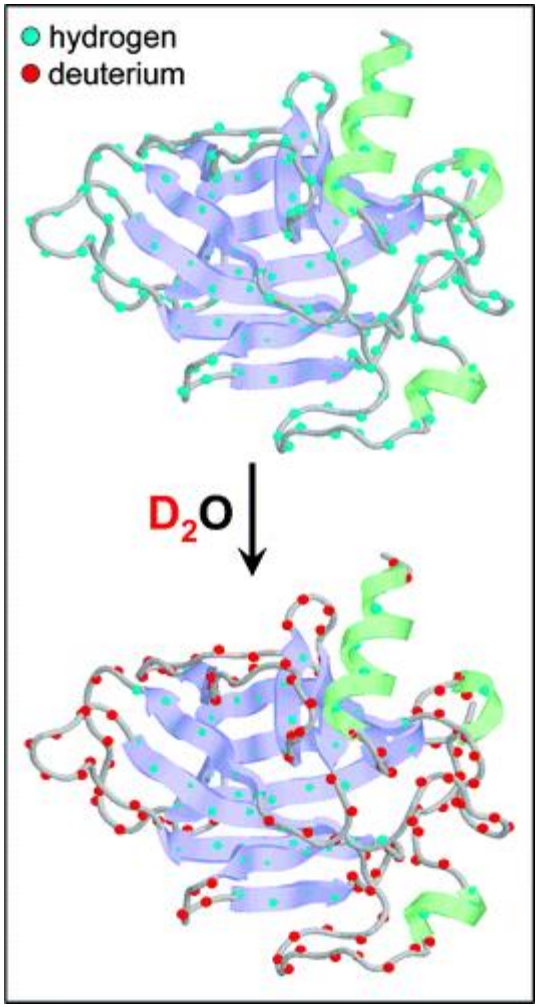
From: Chen, Z. A., Jawhari, A., Fischer, L., Buchen, C., Tahir, S., Kamenski, T., … Rappsilber, J. (2010). Architecture of the RNA polymerase II-TFIIF complex revealed by cross-linking and mass spectrometry. *The EMBO Journal*, *29*(4), 717–726. doi:10.1038/emboj.2009.401

Characterizing structure & interactions with MS

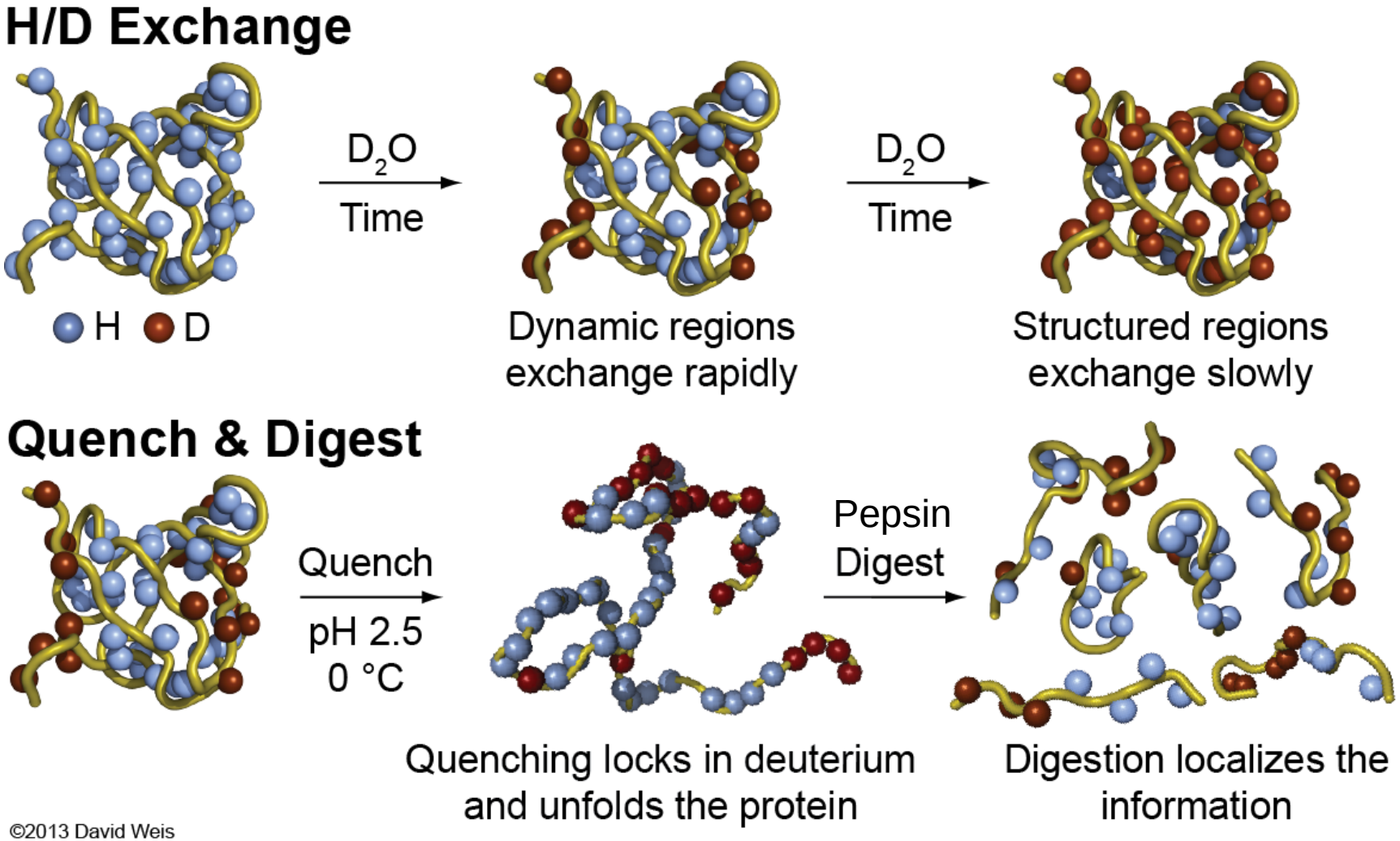
- Identify interaction partners (e.g. Affinity purification MS, BioID, etc.)
- Crosslinking MS
- Hydrogen-deuterium exchange (HDX)
- Native mass spectrometry

H/D Exchange

- In a solution of D_2O amide-hydrogens can be exchanged against D
- Depends on:
  - Hydrogen bonds
  - Accessibility of H -> structure!
  - pH of solution
  - temperature

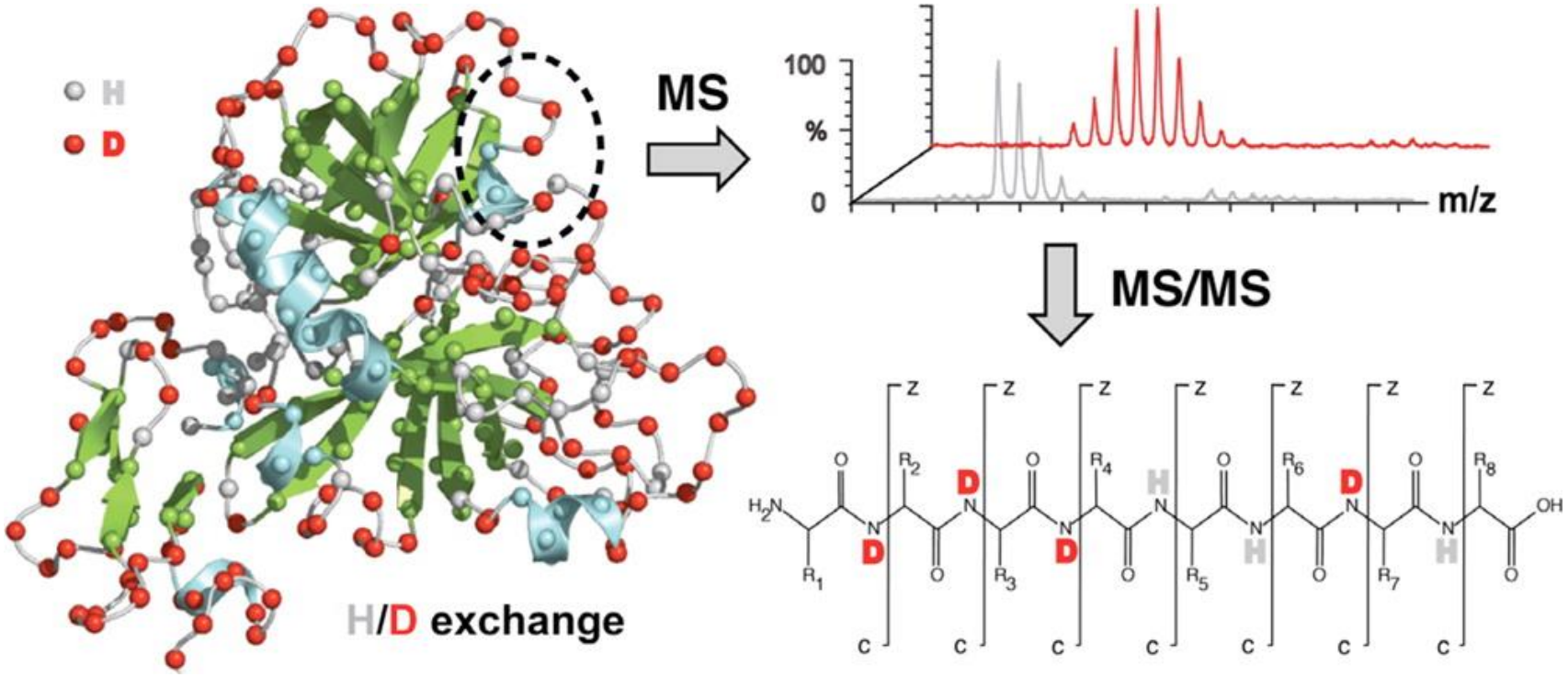


H/D Exchange MS



http://mvsc.ku.edu/content/hydrogen-deuterium-exchange-mass-spectrometry

Amino acid specific information



From: Measuring the Hydrogen/Deuterium Exchange of Proteins at High Spatial Resolution by Mass Spectrometry: Overcoming Gas-Phase Hydrogen/Deuterium Scrambling

Kasper D. Rand, Martin Zehl, and Thomas J. D. Jørgensen

Accounts of Chemical Research **2014** *47* (10), 3018-3027

DOI: 10.1021/ar500194w

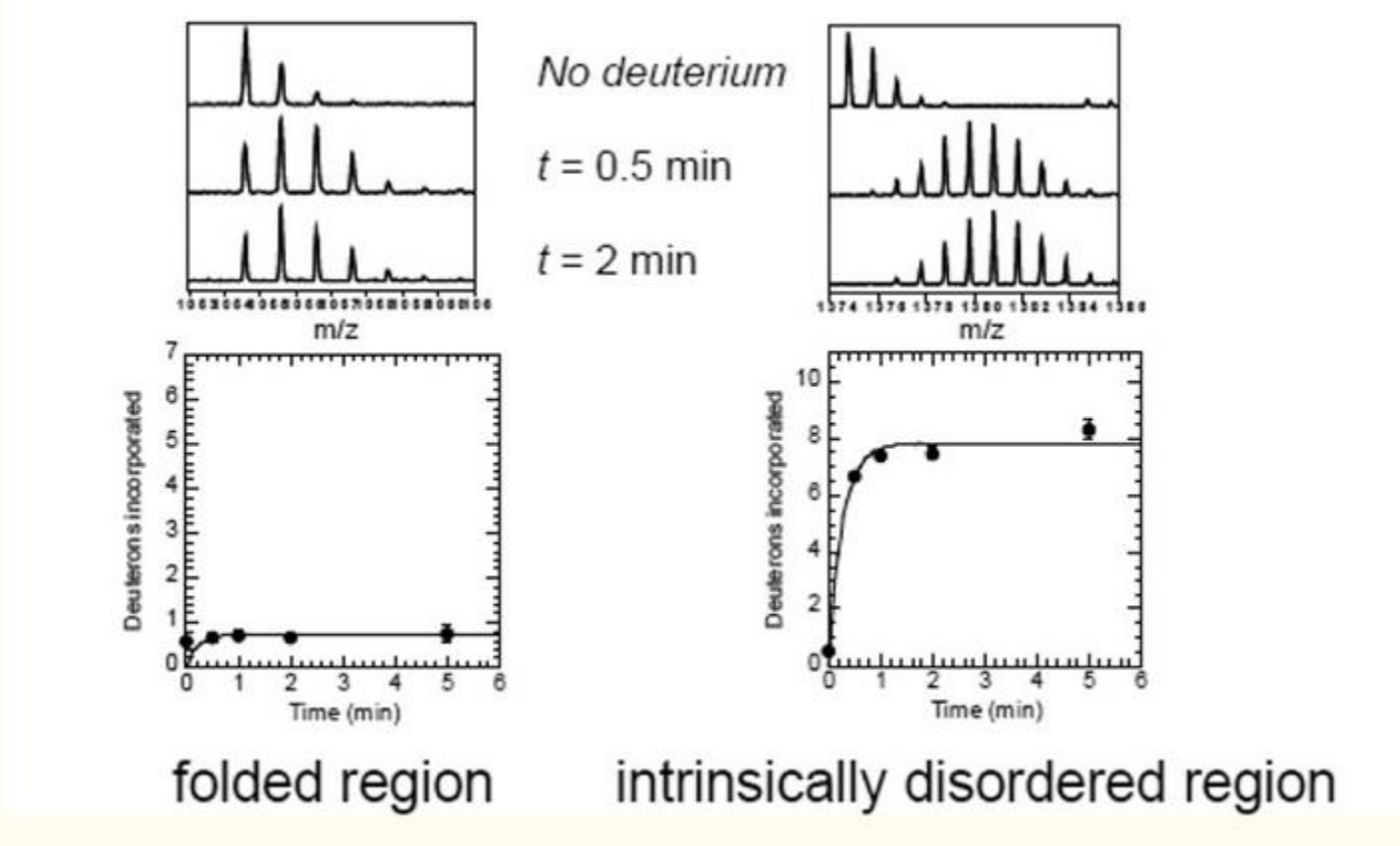
H/D Exchange applications

Provides structural information on:

- Large proteins
- Protein-ligand interaction
- Protein complexes
- Viral particles

Provides information on protein dynamics and conformational state

Differentiate folded and disordered regions

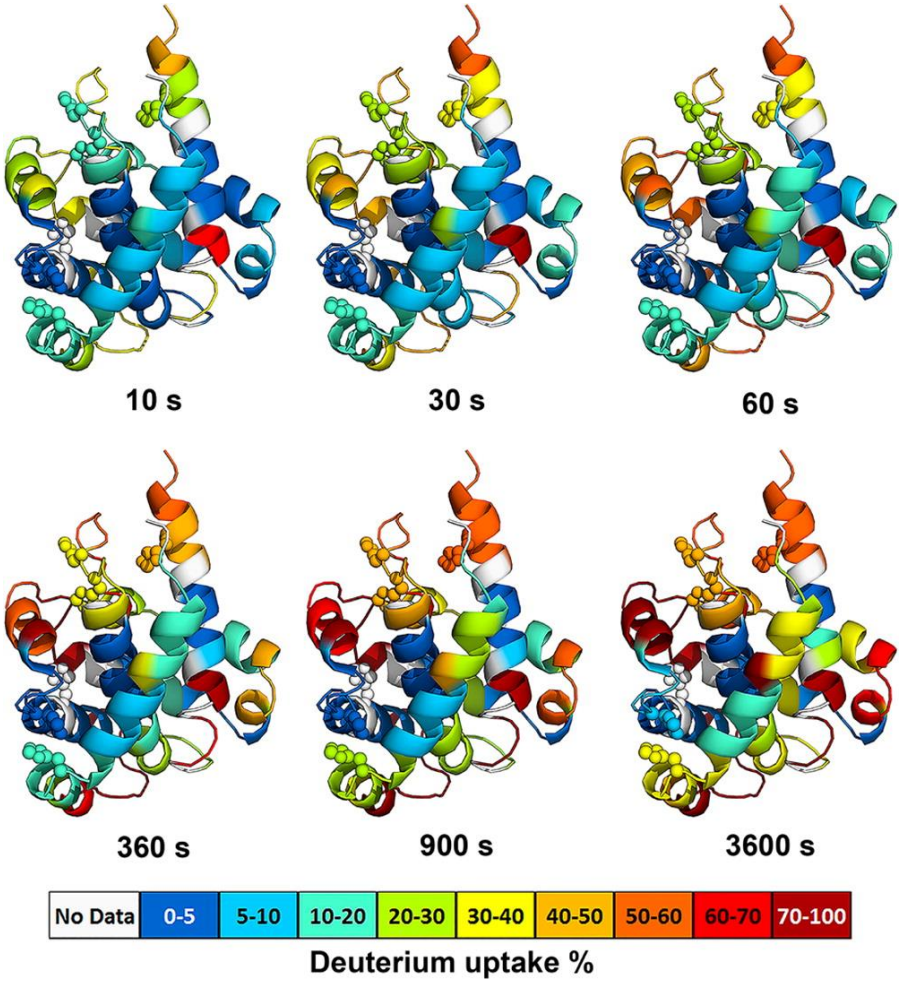


From: Balasubramaniam & Komives, Biochim Biophys Acta. 2013, 1834(6): 1202–1209.

H/D exchange time series

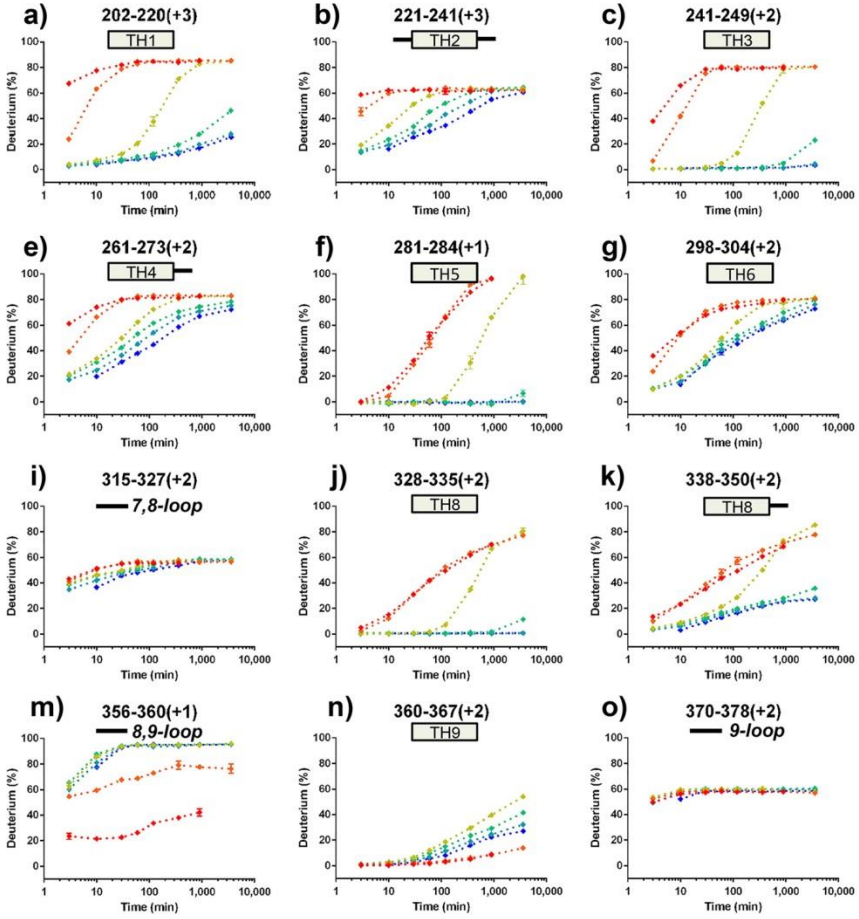
Example:

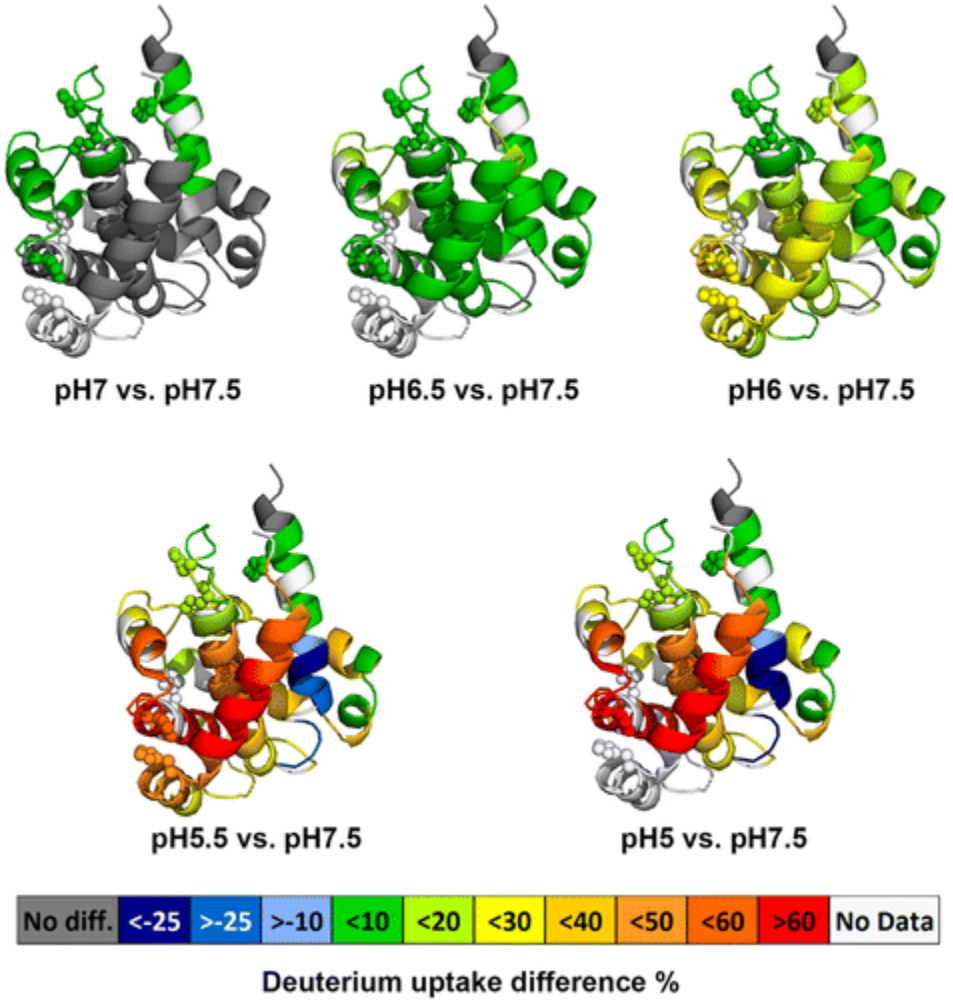
Diphteria toxin enters cells via endosomal pathway and undergoes a pH dependent conformational change



From: Hydrogen–Deuterium Exchange and Mass Spectrometry Reveal the pH-Dependent Conformational Changes of Diphtheria Toxin T Domain
Jing Li, Mykola V. Rodnin, Alexey S. Ladokhin, and Michael L. Gross
Biochemistry **2014** *53* (43), 6849-6856

pH dependent conformational change



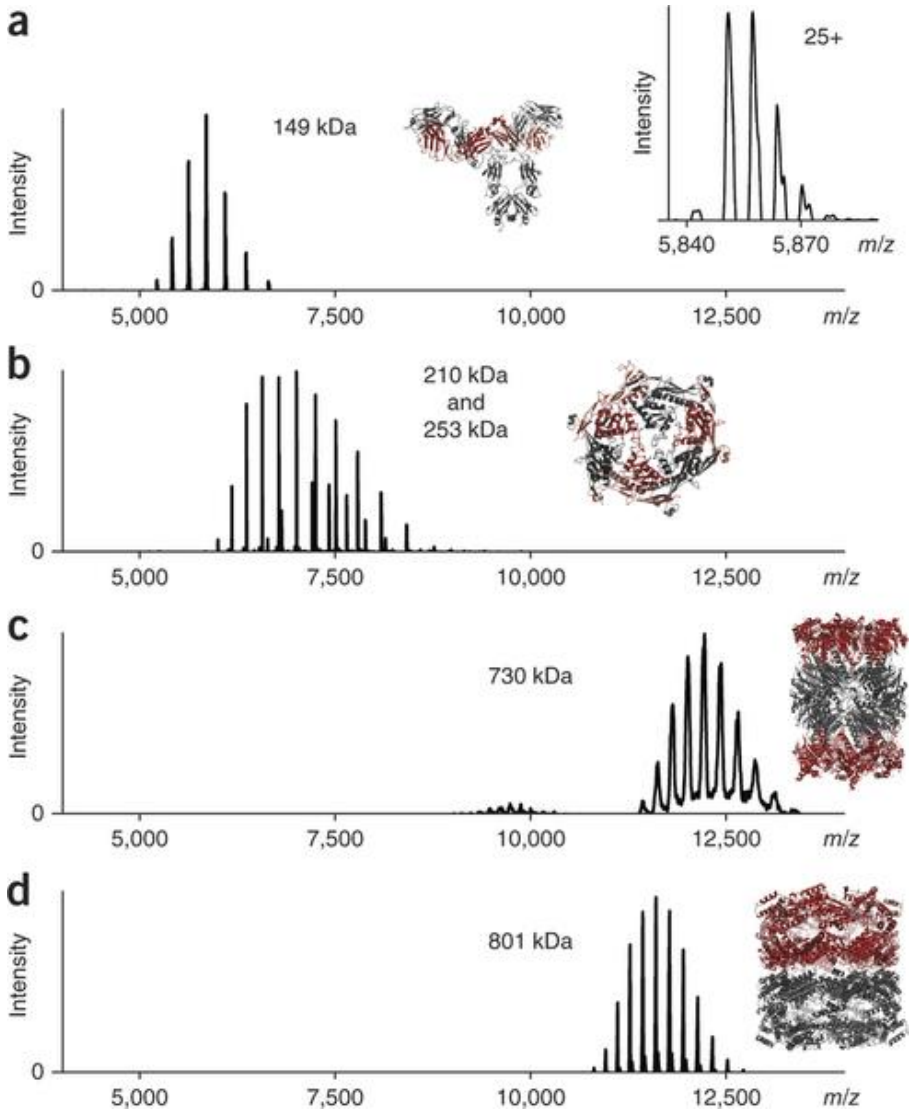


From: Hydrogen–Deuterium Exchange and Mass Spectrometry Reveal the pH-Dependent Conformational Changes of Diphtheria Toxin T Domain
Jing Li, Mykola V. Rodnin, Alexey S. Ladokhin, and Michael L. Gross
Biochemistry **2014** *53* (43), 6849-6856

Requirements for HDX

- Pure protein preparations,
- a special LC-MS system that allows digest and sepration at low temperatures
- specialised software
- and an experienced application specialist

Back to the start: intact native MS



MS of intact proteins under native conditions (e.g. in ammonium acetate at physiological pH)

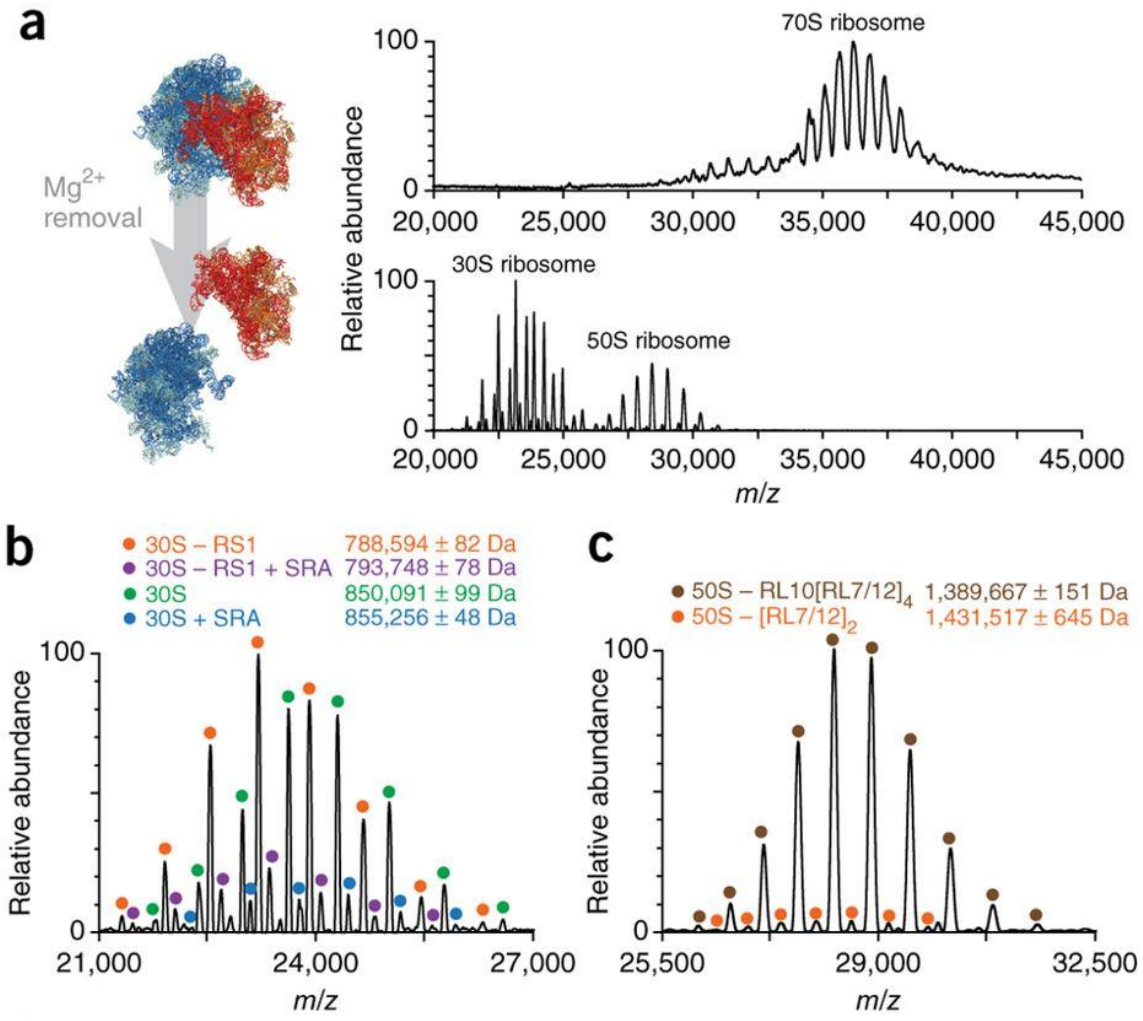
(**a**–**d**) Native mass spectra of IgG antibody (**a**), bacteriophage HK97 capsid pentamers and hexamers (**b**), yeast 20S proteasome (**c**) and *E. coli* GroEL (**d**).

Nat Methods. 2012 Nov;9(11):1084-6. doi: 10.1038/nmeth.2208. Epub 2012 Oct 14.
High-sensitivity Orbitrap mass analysis of intact macromolecular assemblies.
Rose RJ1, Damoc E, Denisov E, Makarov A, Heck AJ.

Native MS

- Buffers that retain native structure (?)
- Samples: min 20 microliters of a 1-5 mg/mL solution
- Mainly manual acquisition

Native MS to the extreme: intact ribosomes



Waterbeemd et al., Nature Methods **volume 14**, pages 283–286 (2017)

Native MS to the extreme: complexes ejected from native membranes

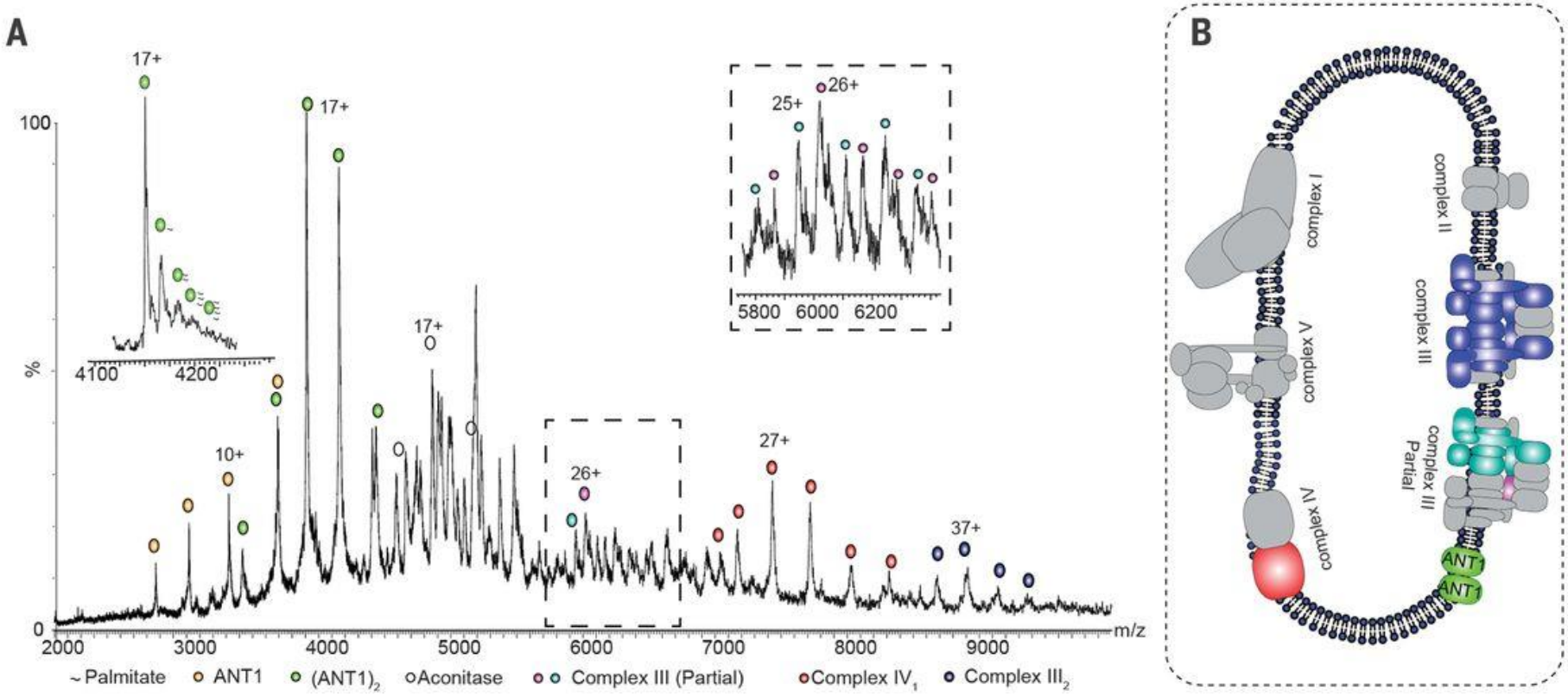
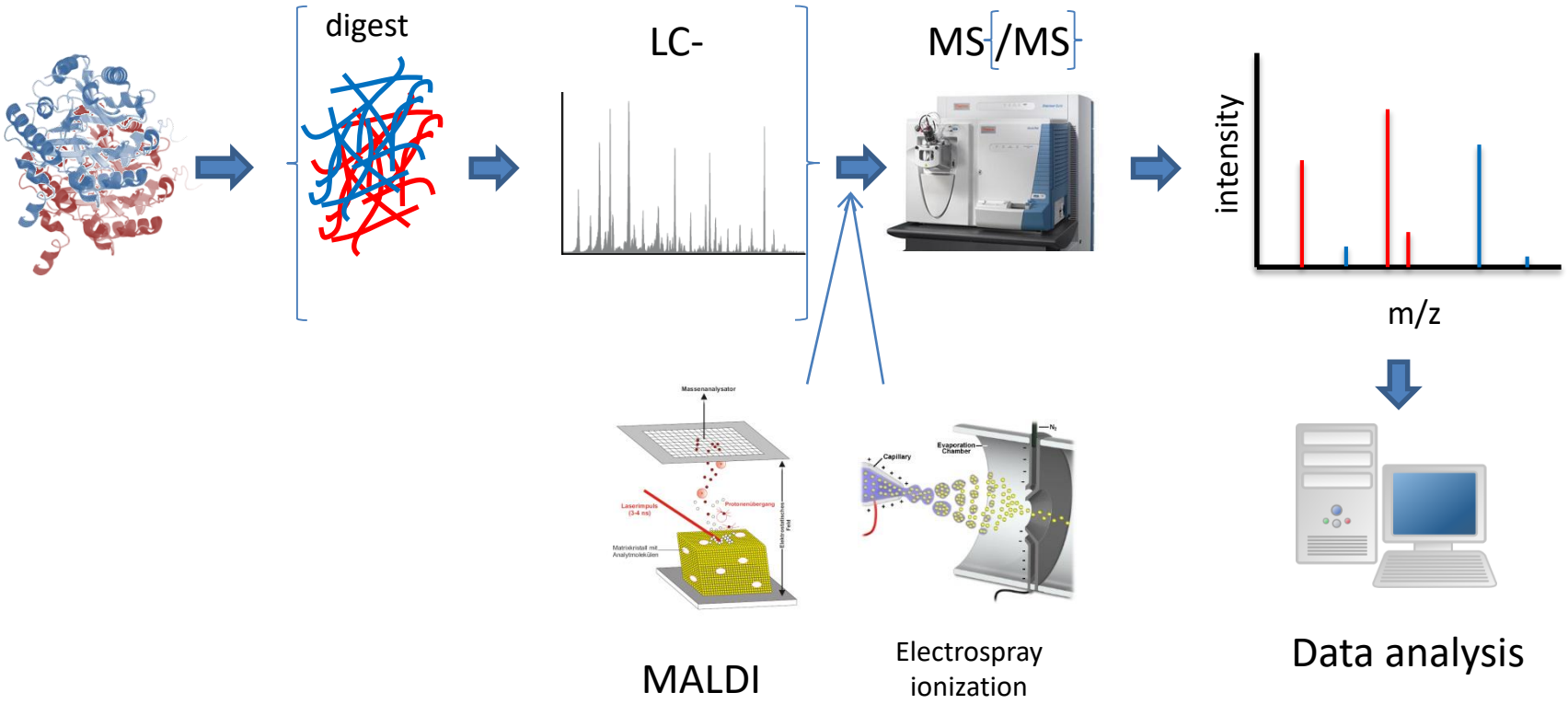


Fig. 3 Intact mitochondria and inner membranes yield complexes I, III, IV, and V, as well as ANT-1 (adenine nucleotide translocase 1) with palmitate transport through the dimer interface.

Dror S. Chorev et al. Science 2018;362:829-834

Published by AAAS

Summary: Protein mass spectrometry workflow



Summary:

- Basic principles of MS
- Intact vs. Top-down vs. Bottom-up
- MS/MS
- Sequencing & Database searching
- Applications: XLMS, HDX, native MS

Thanks!