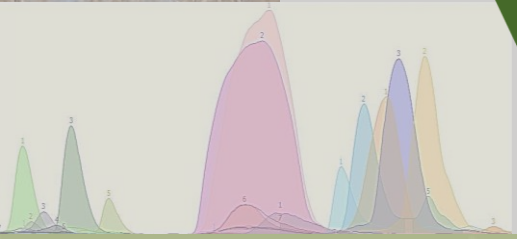
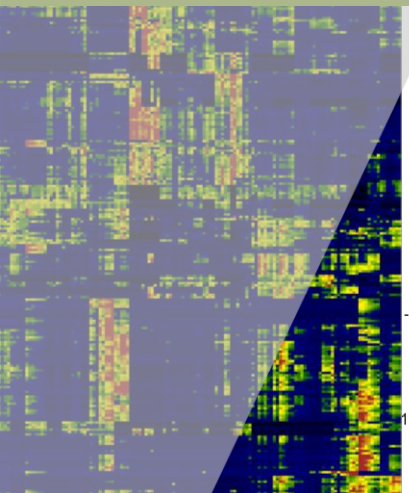
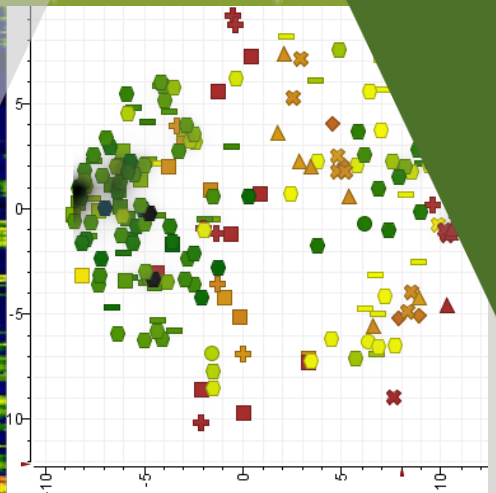






Analysis of peptides and proteins using MS-based proteomics.

Dr Delphine Vincent

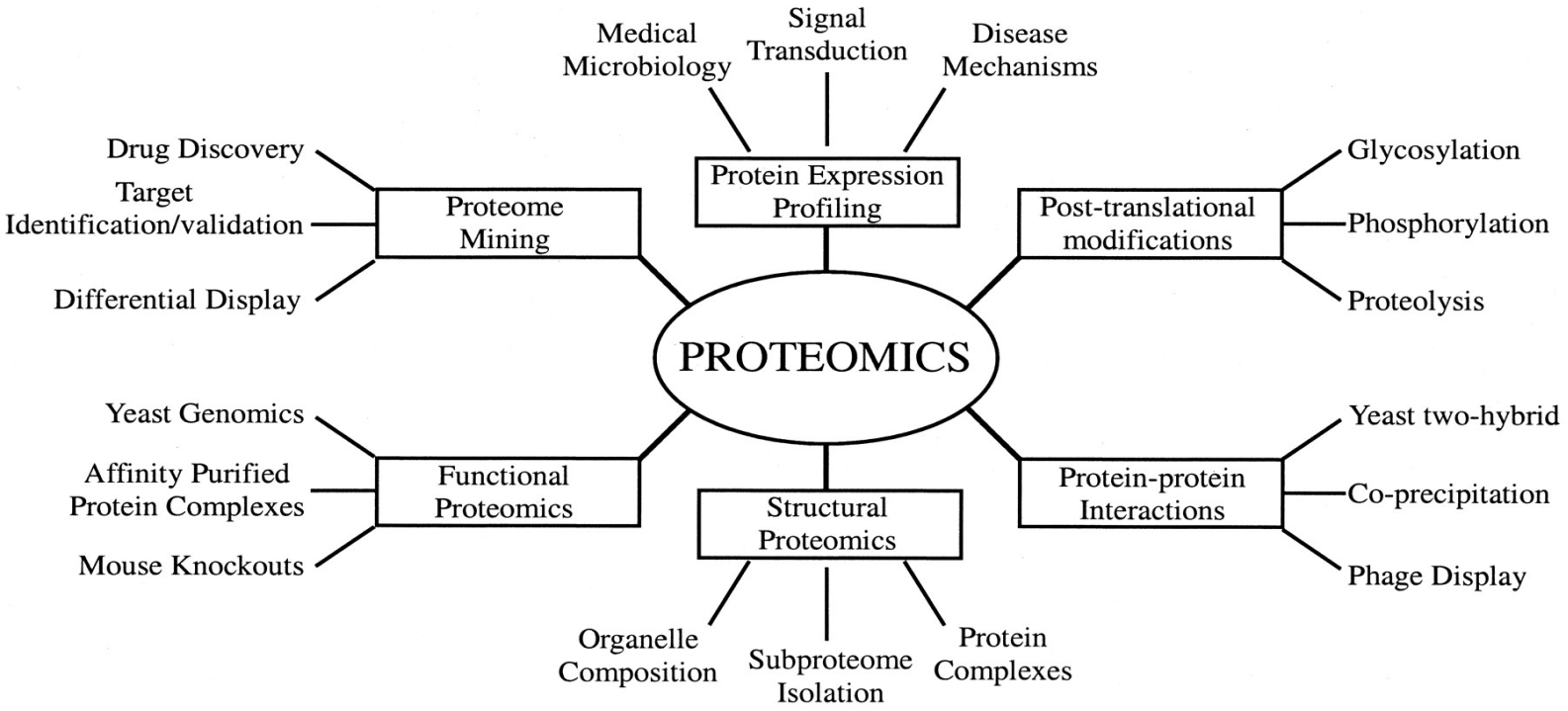
22/05/2019

INTRODUCTION

INTRODUCTION

Why do we do proteomics?

"Proteins form the structural fabric of cells and underpin all metabolic processes and regulatory mechanisms. …Therefore, understanding protein structure–function relationships in cell biology not only requires the identification of proteins but also the detailed analysis of the protein properties that constitute the dimensions of the proteome." (Larance & Lamond, Nature 2015)



Graves & Haystead, Microbiol. Mol. Biol. Rev. 2002;66:39-63

INTRODUCTION
Definitions

Proteomics: **large-scale study** of the functions, structures, and interactions of **proteins**.

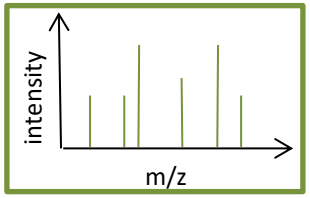
Proteome: the **entire complement of proteins** that is or can be expressed by a cell, tissue, or organism.

Proteoform: **all of the different molecular forms** in which the protein product of a single gene can be found, encompassing all forms of genetic variation, alternative splicing of RNA transcripts, and post-translational modifications (PTMs).

Post-translational modifications (PTMs): modifications that occur following the synthesis of the protein. Very diverse and account for much of the proteome complexity. **Phosphorylation, glycosylation, *etc*…**

Mass spectrometry (MS): identify the amount and type of chemicals present in a sample by measuring the mass-to-charge ratio and abundance of gas-phase ions.
Mass = weight, charge= number of proton

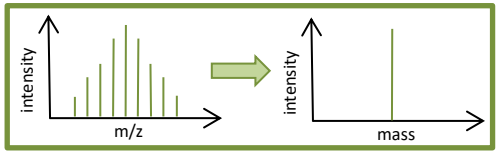
Spectra →



Bottom-up (BU) proteomics: identify proteins and characterize their amino acid sequences and post-translational modifications by proteolytic digestion of proteins prior to analysis by MS. **Peptides.**

Top-down (TD) sequencing: protein identification that uses an ion trapping mass spectrometer to store an isolated protein ion for mass measurement and tandem mass spectrometry analysis. **Intact proteins.**

Deconvolution: transformation of a charge state series into a molecular mass.
Protein mass.

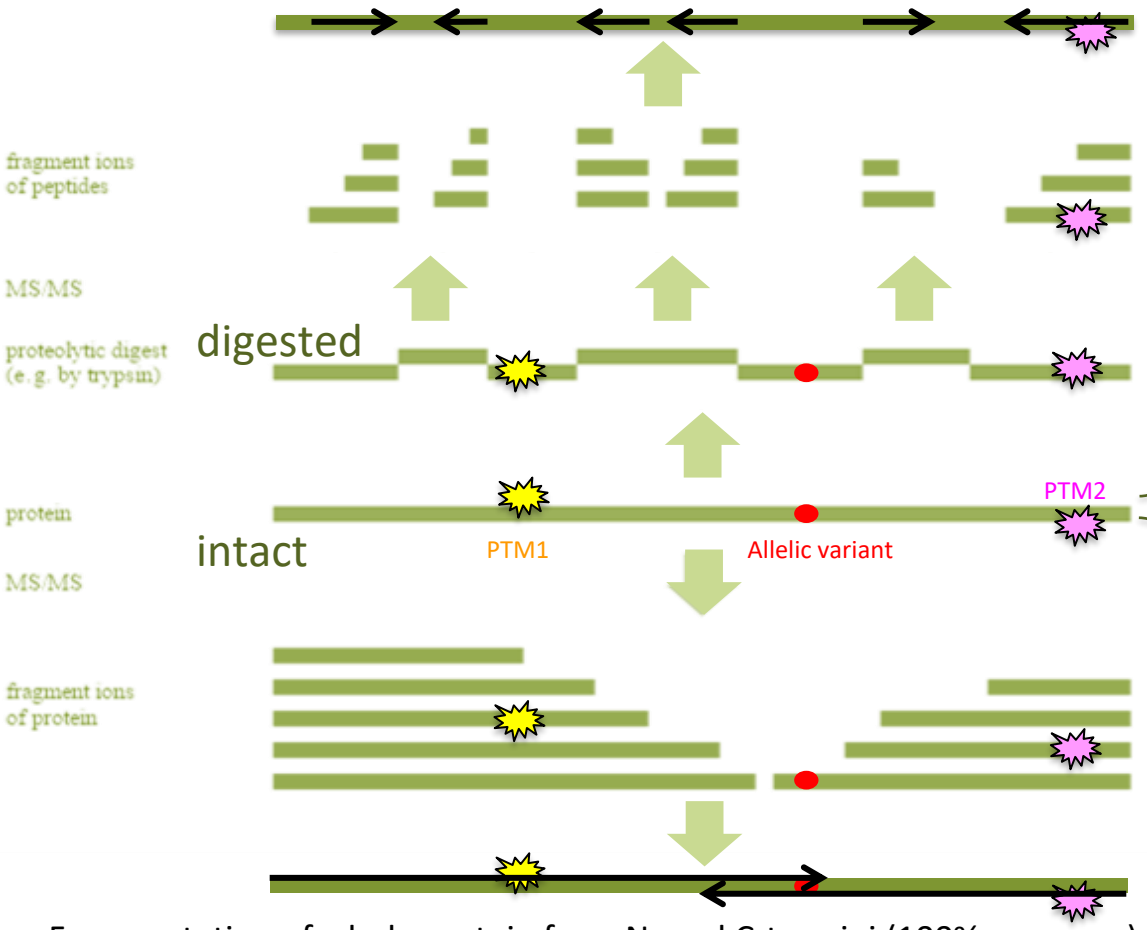


INTRODUCTION

Bottom-up vs. top-down

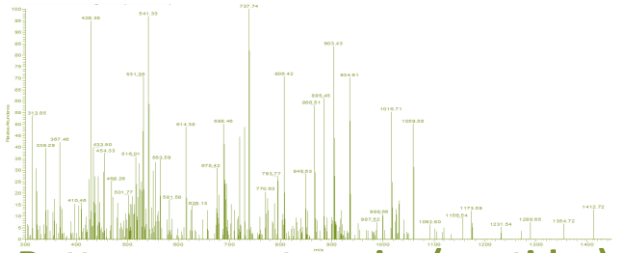
Highly complementary

Peptide fragments along the sequence (never 100% coverage, <10%)



Fragmentation of whole protein from N- and C-termini (100% coverage)

Top-down sequencing (TDS)



Bottom-up proteomics (peptides)

- \+ all MWs
- \- Dynamic range
- \+ Most mature
- \+ Widely used, sufficient tools
- \+ Can handle complex samples
- \- PTM analysis
- \- Protein processing
- \- Allelic variation
- \- Protein quantitation

up

bottom

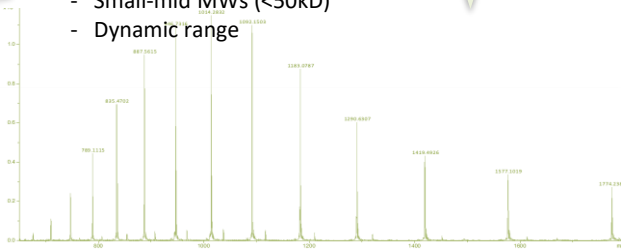
Top-down proteomics (proteins)

- \+ Protein quantitation
- \+ Allelic variation
- \+ PTM analysis on protein level
- \+ Protein processing
- \- Relatively new approach
- \- Complexity of the sample
- \- Small-mid MWs (<50kD)
- \- Dynamic range

top

proteoforms

down

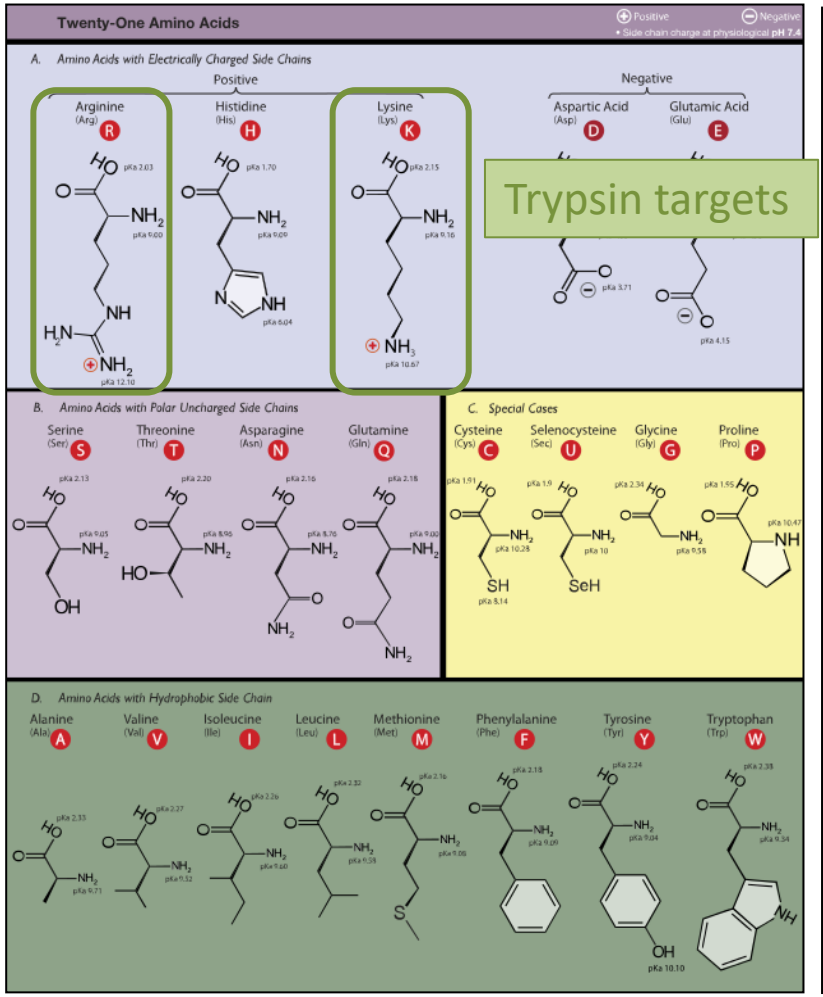


INTRODUCTION

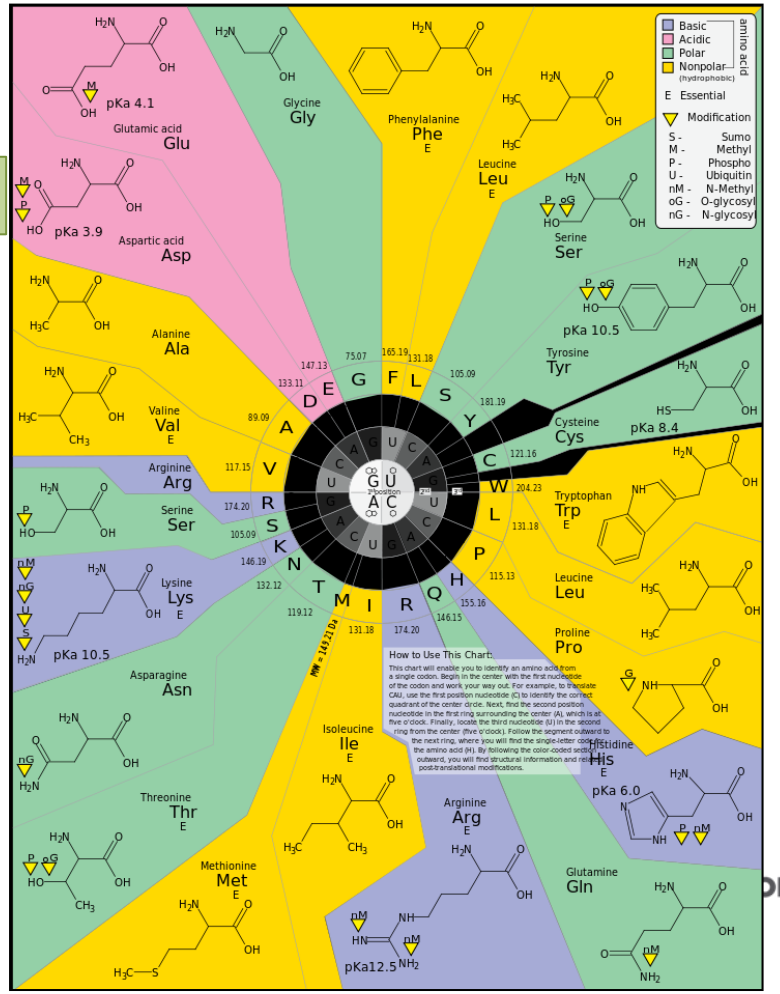
Bottom-up vs. top-down

Highly complementary

Bottom-up (AA sequence=protein ID)

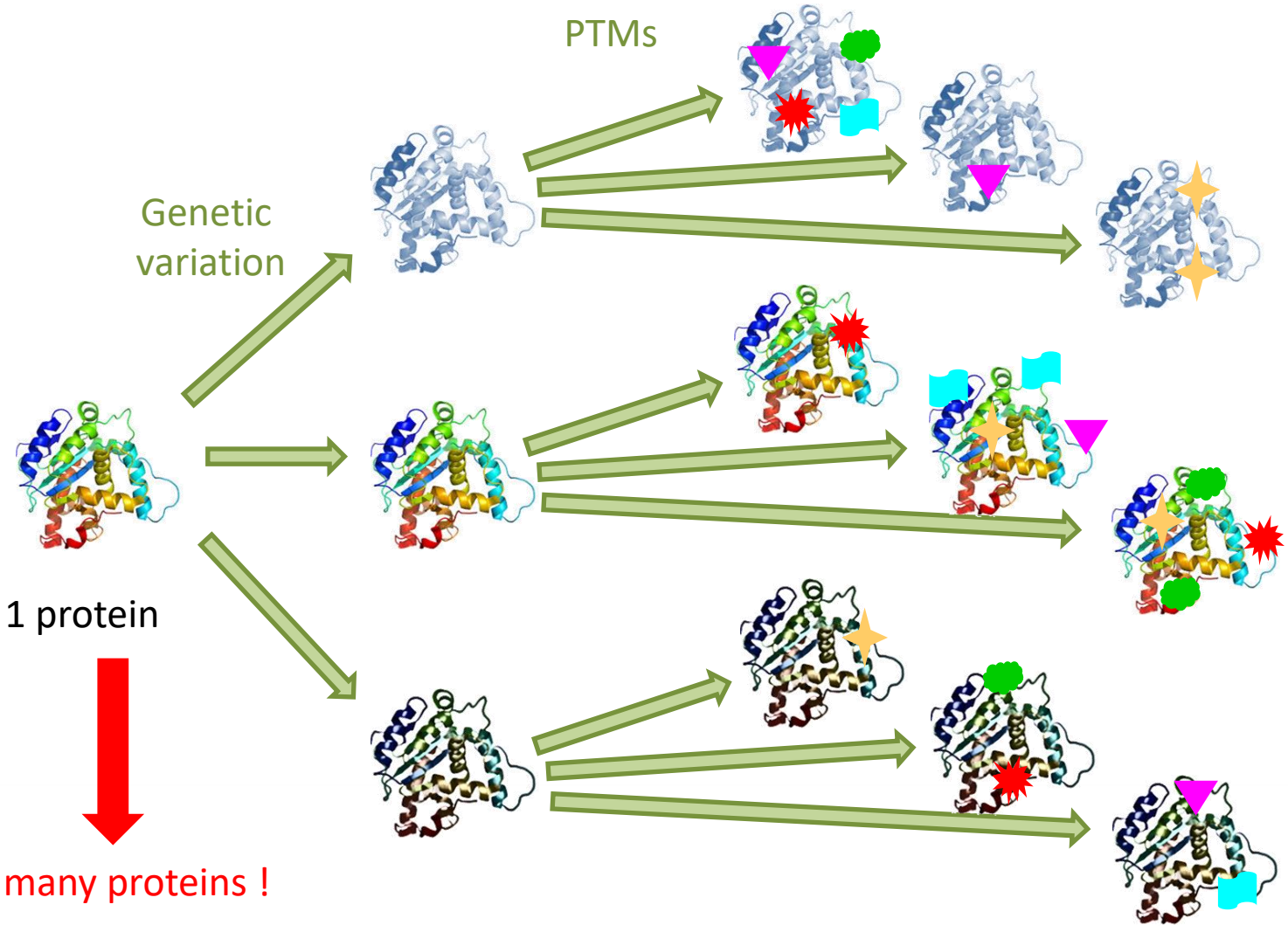


Top-down (variants & PTMs)



INTRODUCTION

Proteoforms



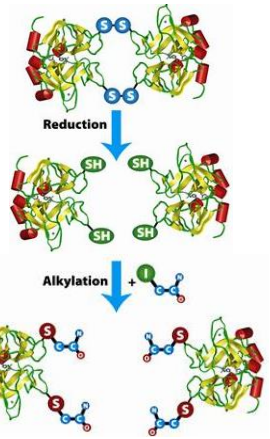
Many proteoforms in a given sample (e. g. milk).

Cells are crowded!

TECHNICALITIES

TECHNICALITIES
Extraction

Many protocols available (e.g. TCA/acetone precipitation, Phenol/ammonium acetate partition); most apply the same principles : denaturation and reduction of proteins (not native and not protein complexes) → subunits.



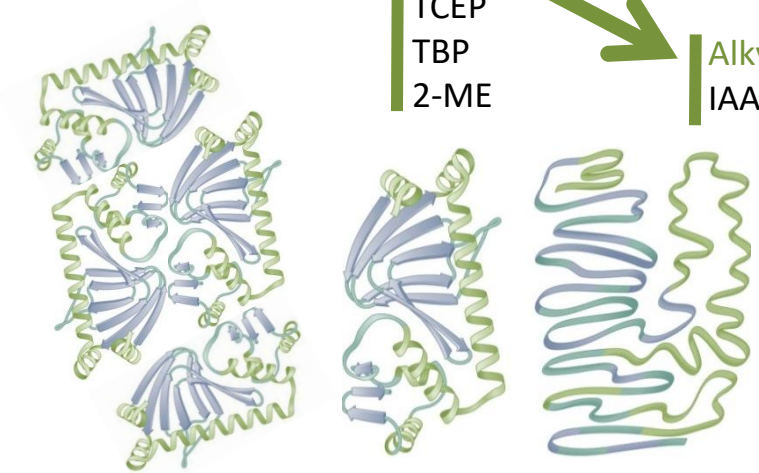
Denaturation using chaotropes (IV and III structures disrupted)
- Urea
- Thiourea
- Guanidine-HCl

→

Reduction of disulfide bonds (II structure disrupted)
- DTT
- TCEP
- TBP
- 2-ME

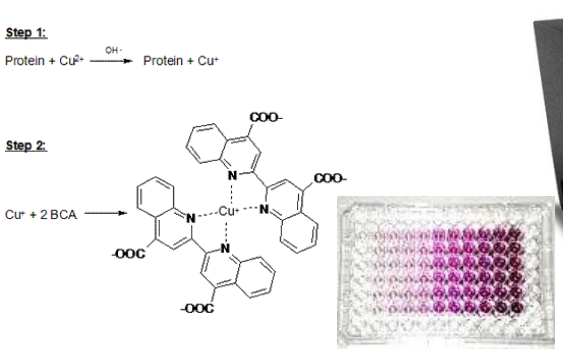
→

Alkylation of reduced bonds (I structure maintained)
- IAA



Complex (IV, III) → subunit (II) → denatured (I)

Protein content measured using a BCA assay



→ quality control, normalisation, trait

TECHNICALITIES

Digestion (bottom-up)

A protease (also called peptidase or proteinase) is any enzyme that performs proteolysis, by hydrolysis of the peptide bonds that link amino acids together in a polypeptide chain.
Trypsin (EC 3.4.21.4) is a serine protease, found in the digestive system of many vertebrates, where it hydrolyses proteins.

Proteases most widely used in proteomic analysis:

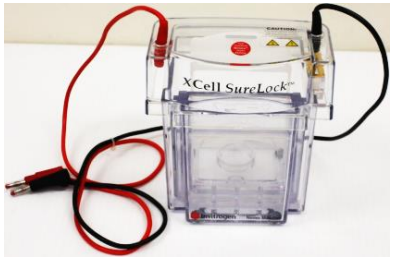
1.Trypsin (R, K) by far the most widely used protease in proteomic analysis.

2.Other proteases and cleavage reagents
- Glu-C (E)
- Lys-C (K)
- Chymotrypsin (F, W, Y)
- Asp-N (N)

3.Non specific proteases
- Subtilysin
- Pepsin (F, L)
- Proteinase K
- Pronase (mixture of endo- and exo-proteinases)
- Elastase (A, G, S, V)
- Thermolysine (I, M, F, W, Y, V)

TECHNICALITIES

Separation



SDS-PAGE

Gel-based proteomics not covered here.

2-DE

HPLC

Nano → capillary → normal → analytical flow

400nL/min 3uL/min 0.2mL/min 1mL/min

2-D-LC
(fraction collection)

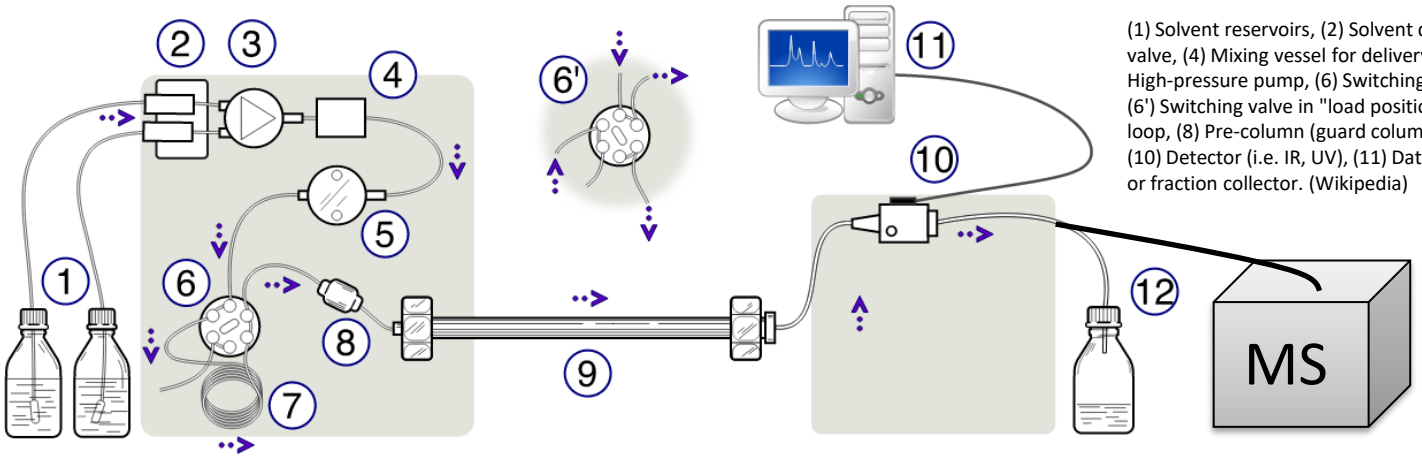
HPLC separation columns:
RPC (C18, C8, C4), IEC, SEC

high-throughput analyses.

TECHNICALITIES

HPLC

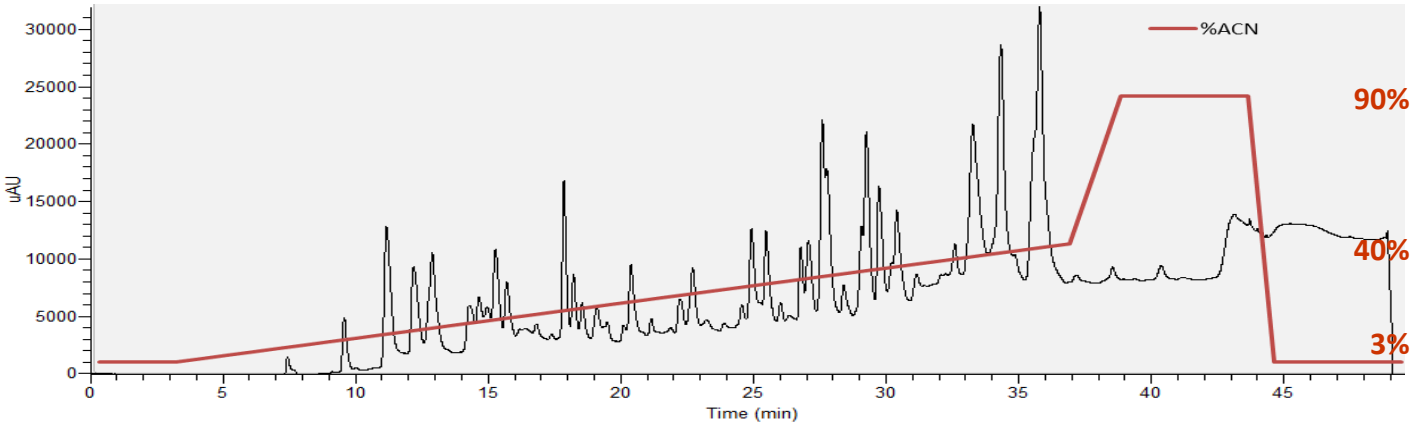
High-performance liquid chromatography (HPLC) separates the components in a mixture, to identify and quantify each component. The components bind to the separation column (stationary phase) and are eluted by increasing concentration of solvent (mobile phase). A component is then defined by its retention time (min) and abundance. Great diversity of methods (RP, NP, IEC, SEC…).



(1) Solvent reservoirs, (2) Solvent degasser, (3) Gradient valve, (4) Mixing vessel for delivery of the mobile phase, (5) High-pressure pump, (6) Switching valve in "inject position", (6') Switching valve in "load position", (7) Sample injection loop, (8) Pre-column (guard column), (9) Analytical column, (10) Detector (i.e. IR, UV), (11) Data acquisition, (12) Waste or fraction collector. (Wikipedia)

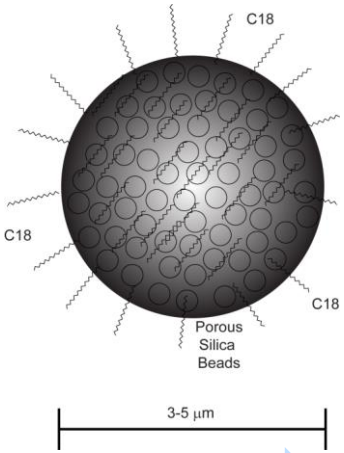
Parameters:

- Flow rate
- Mobile phases
- Temperature
- Stationary phase
- Length, Diameter
- Particle size
- Gradient
- Duration

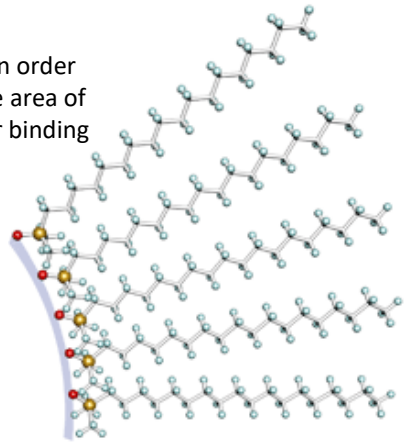


TECHNICALITIES
RP-HPLC

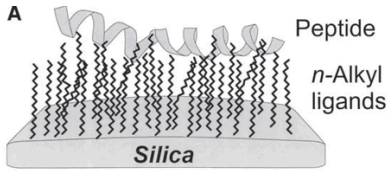
Reversed phase-HPLC (RP-HPLC) is typically used for peptides or proteins (non polar). One common stationary phase is a silica with a straight chain alkyl group such as C18, C8 or C4.



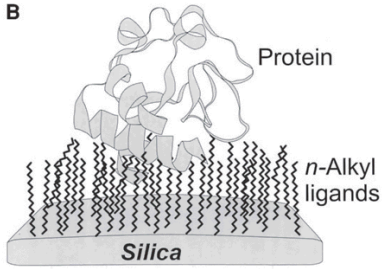
beads usually porous in order to increase the surface area of the beads available for binding



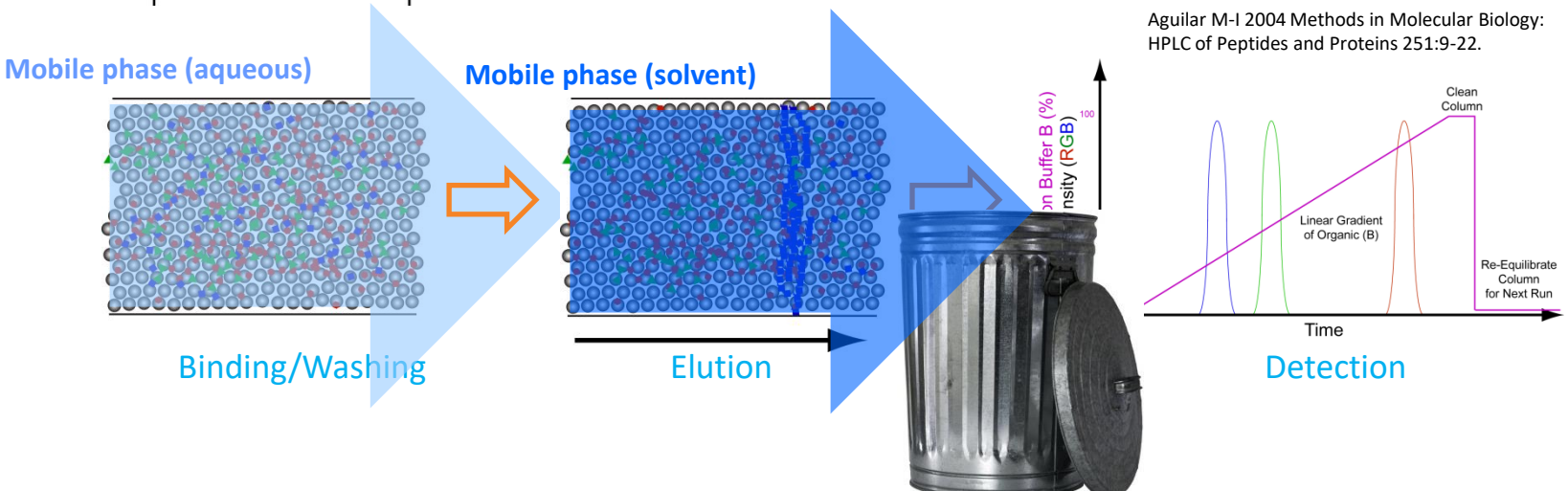
C18



C8
C4

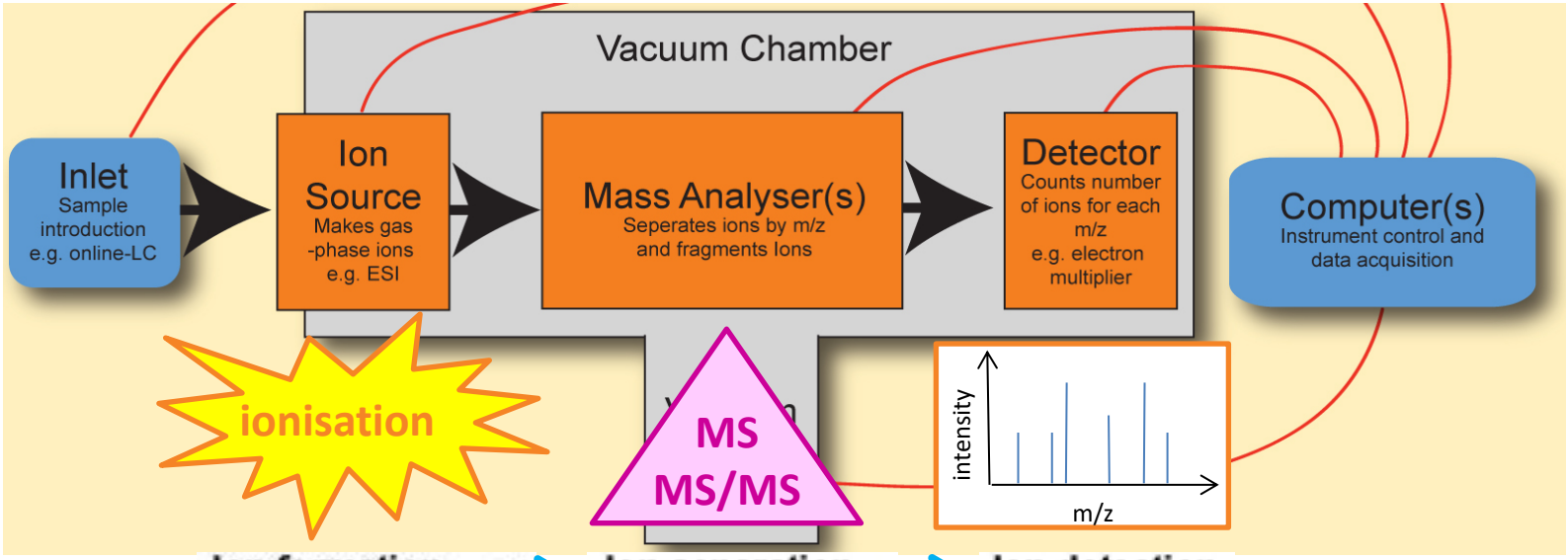


Aguilar M-I 2004 Methods in Molecular Biology: HPLC of Peptides and Proteins 251:9-22.



TECHNICALITIES

Mass spectrometry (MS)

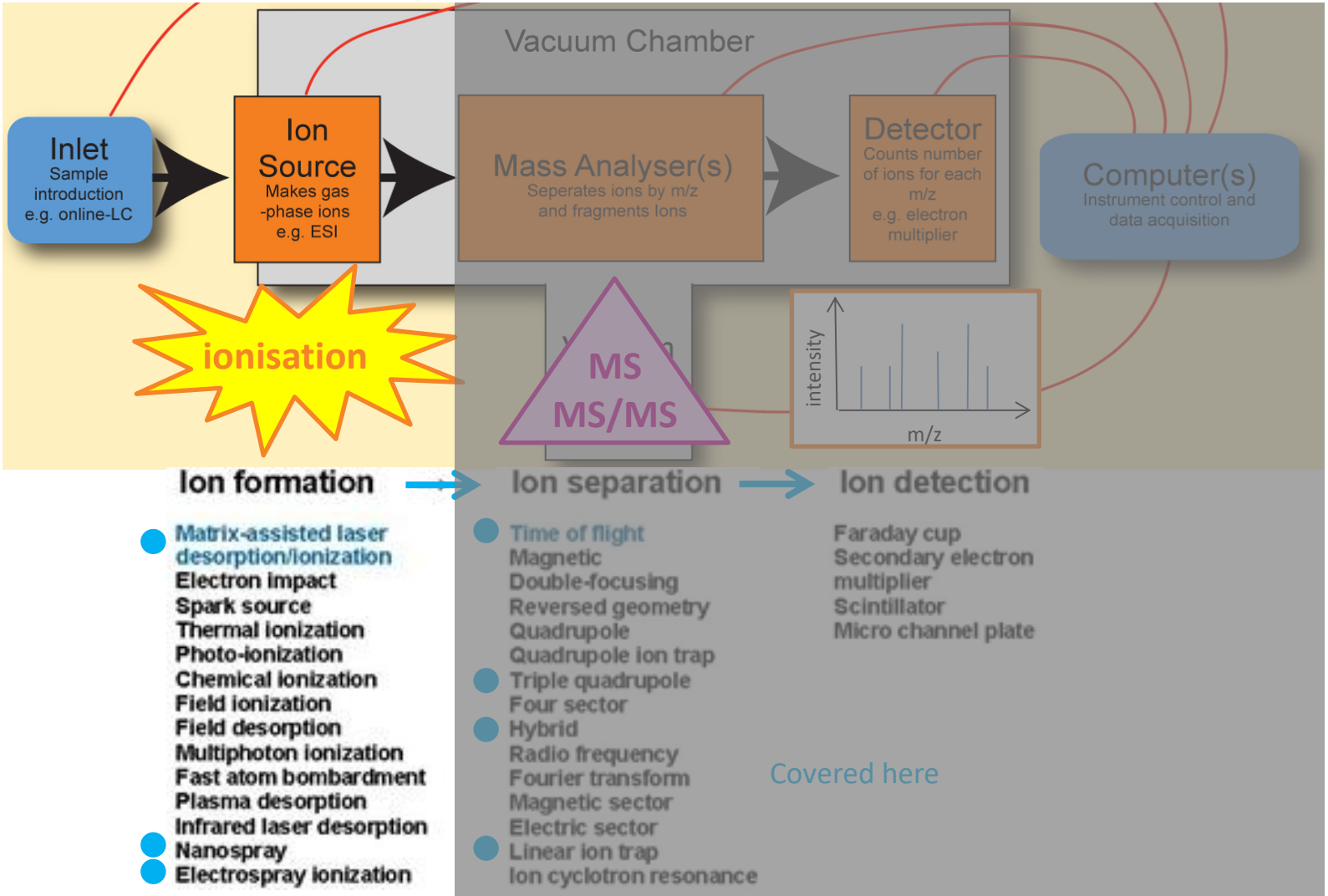


Ion formation	Ion separation	Ion detection
Matrix-assisted laser desorption/ionization	Time of flight	Faraday cup
Electron impact	Magnetic	Secondary electron multiplier
Spark source	Double-focusing	Scintillator
Thermal ionization	Reversed geometry	Micro channel plate
Photo-ionization	Quadrupole	
Chemical ionization	Quadrupole ion trap	
Field ionization	Triple quadrupole	
Field desorption	Four sector	
Multiphoton ionization	Hybrid	
Fast atom bombardment	Radio frequency	
Plasma desorption	Fourier transform	
Infrared laser desorption	Magnetic sector	
Nanospray	Electric sector	
Electrospray ionization	Linear ion trap	
	Ion cyclotron resonance	

Covered here

TECHNICALITIES

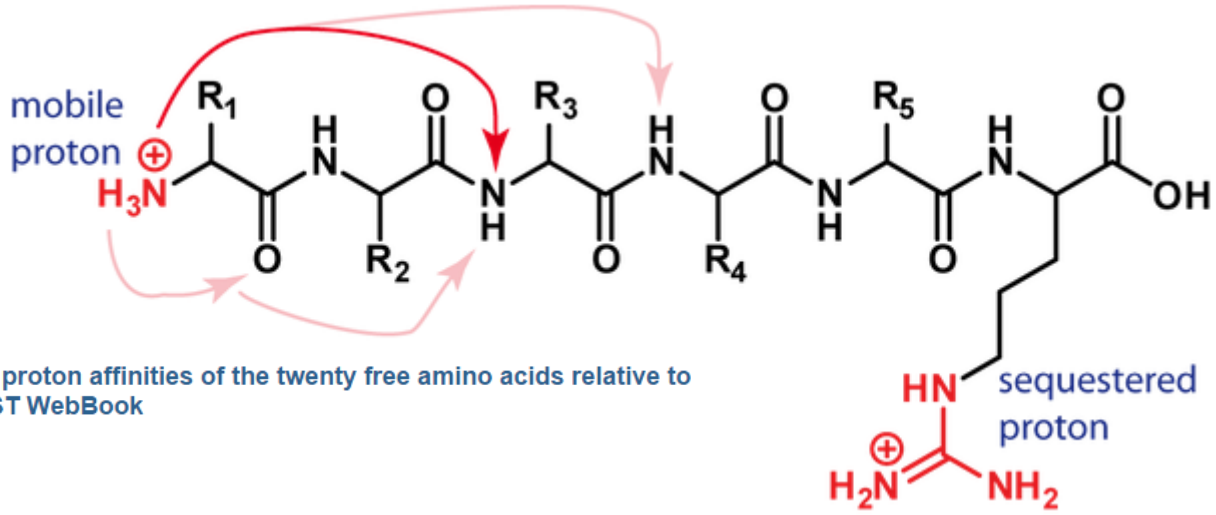
1/ ion formation



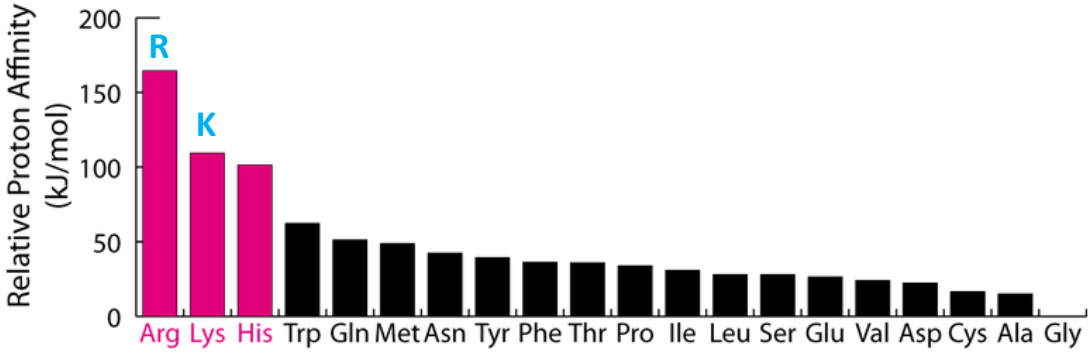
TECHNICALITIES
Ionisation

Ionisation is the process by which an atom or a molecule acquires a negative or positive charge by gaining or losing electrons to form ions. In mass spectrometry (MS), ionization refers to the production of gas phase ions. Ionization occurs in the instrument ion source.

Soft ionization refers to the processes which impart little residual energy onto the subject molecule and as such result in little fragmentation. In proteomics, electrospray ionization (ESI), or matrix-assisted laser desorption/ionization (MALDI) are commonly used.



‡ Experimental gas phase proton affinities of the twenty free amino acids relative to glycine, retrieved from NIST WebBook



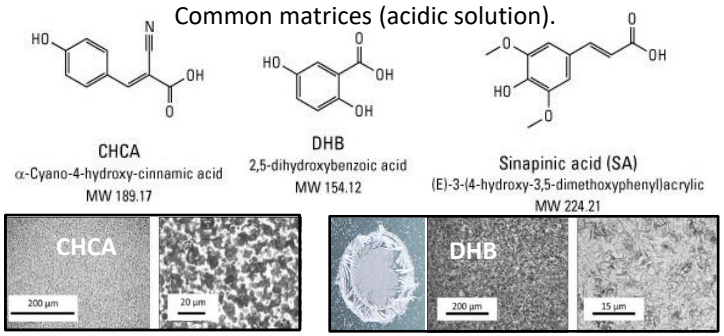
AGRICULTURE VICTORIA

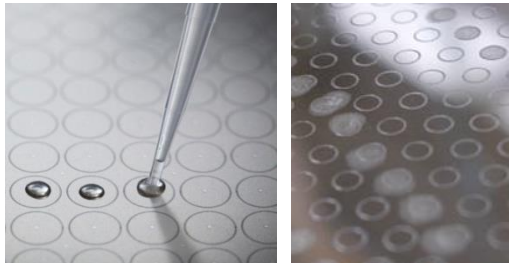
TECHNICALITIES
MALDI

Examples of applications:
1/ Microbial identification using Maldi Biotyper
2/ Protein identification following gel separation and trypsin digestion.

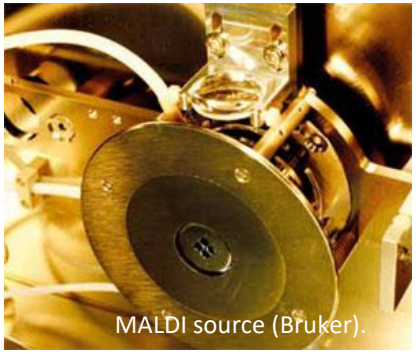
Matrix-assisted laser desorption/ionization (MALDI) is a soft ionization technique mostly producing singly-charged ions in the gas phase (mass readily obtained from the m/z spectra as z=1). The mass analyzer must be capable of measuring ions with very high m/z ratios.
Workflow: 1/ the sample is mixed with a suitable matrix (acidic) and applied to a metal plate, 2/ a pulsed laser irradiates the sample, triggering ablation and desorption of the sample and matrix, 3/ the analyte is ionized by being protonated or deprotonated in the hot plume of ablated gases, and can then be accelerated into a mass spectrometer.

Common matrices (acidic solution).

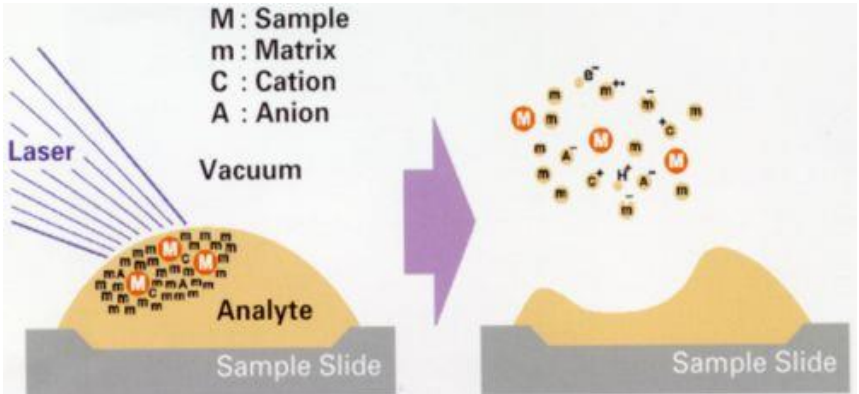


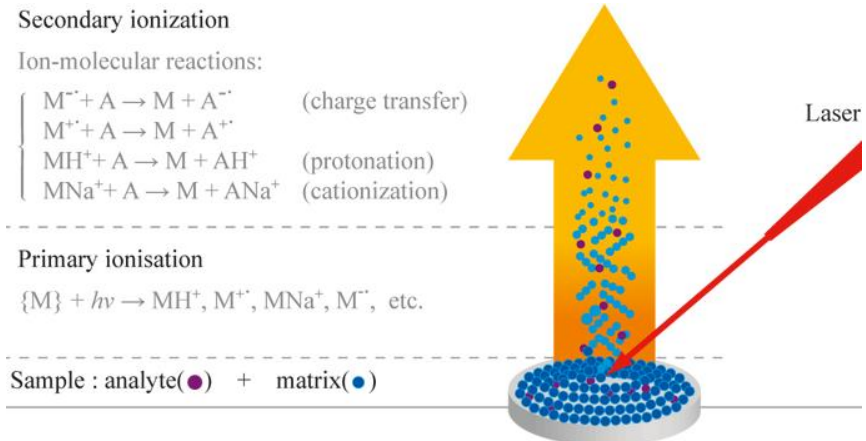


Spotting (1-2µL) on target plates. Dry spots.



MALDI source (Bruker).



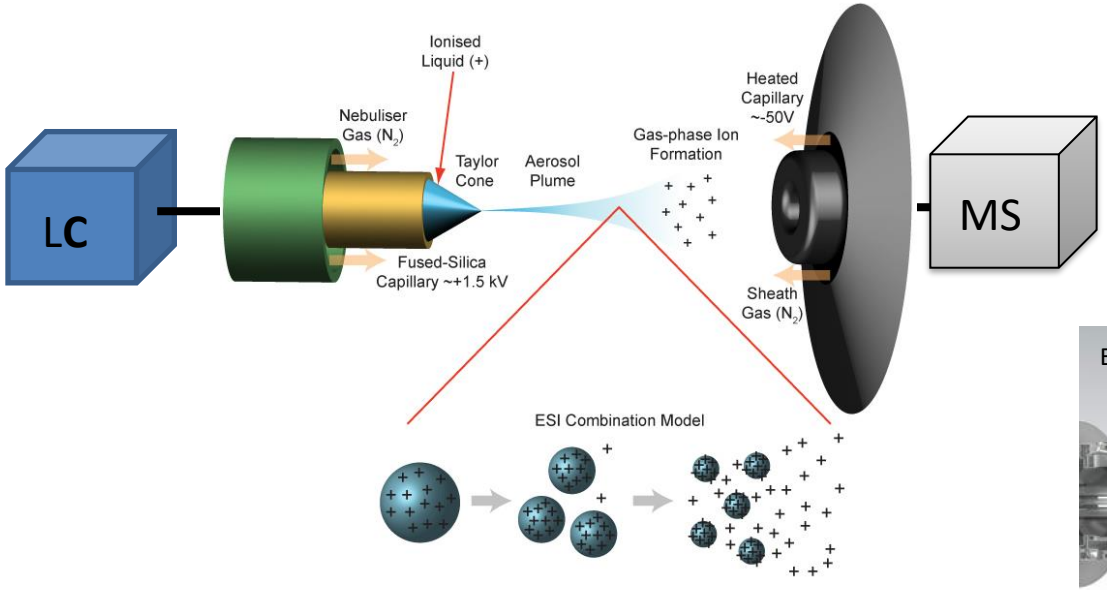


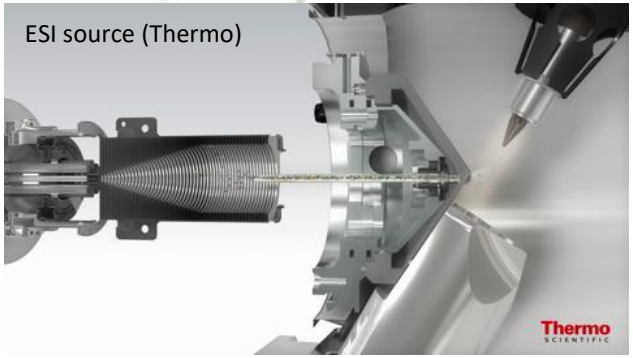
TECHNICALITIES

ESI

Electrospray ionization (ESI) produces ions using an electrospray in which a high voltage is applied to a liquid to create an aerosol. ESI produces multiply charged ions. The maximum number of charges is roughly dependent on the number of basic AA (K,R, H). This extends the mass range of the analyser to accommodate the kDa-MDa orders of magnitude observed in proteins and their associated polypeptide fragments. Another important advantage of ESI is that solution-phase information can be retained into the gas-phase.

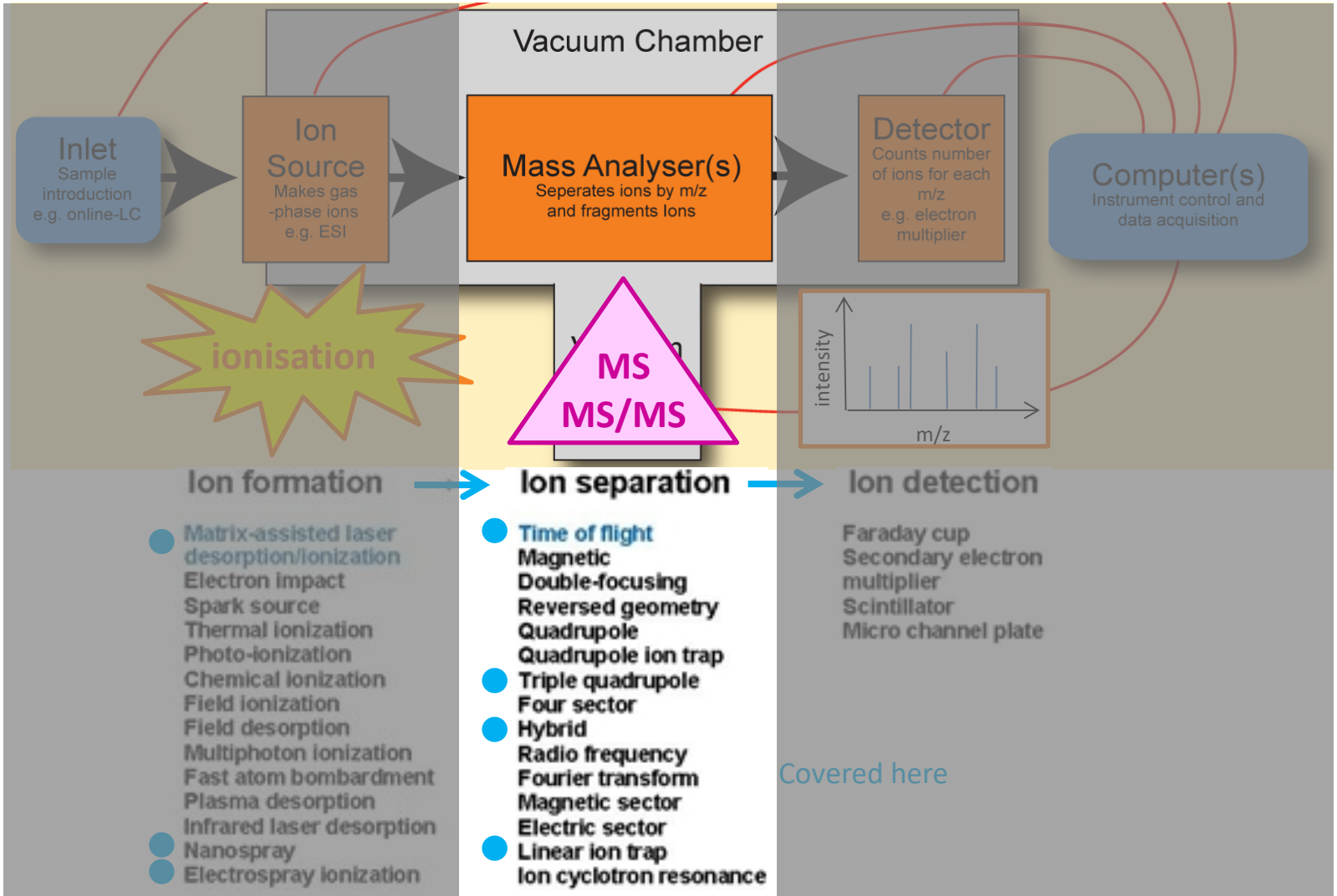
Highly compatible with HPLC which can be coupled online with the MS ESI source.





TECHNICALITIES

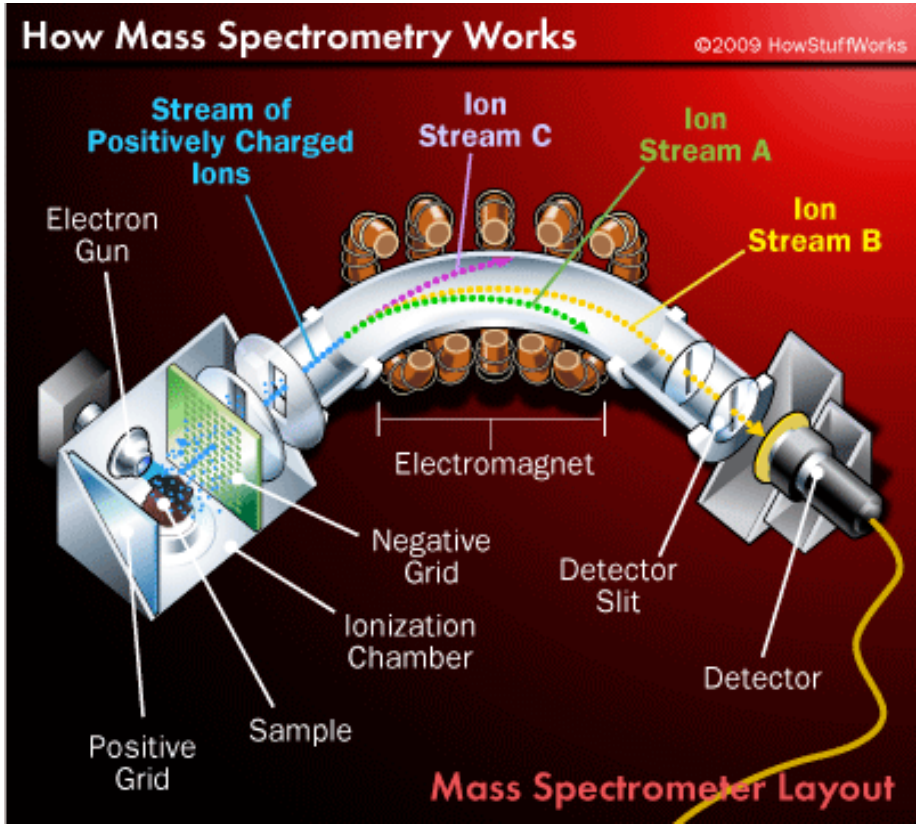
2/ ion separation

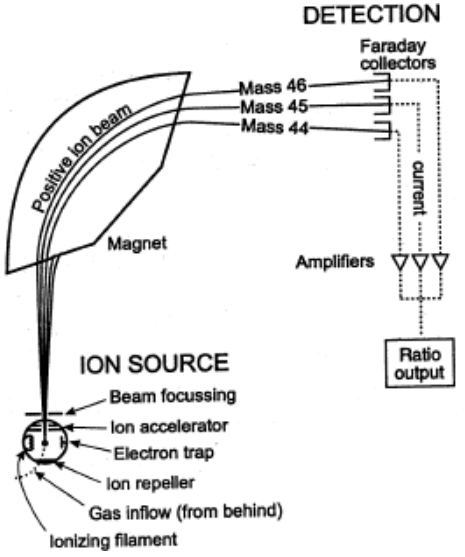


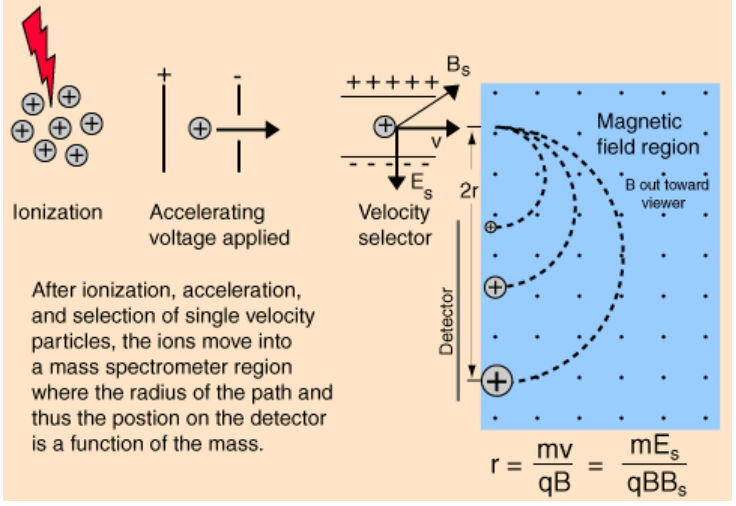
TECHNICALITIES

Mass analyser

The ions are accelerated so that they all have the same kinetic energy. The ions are then deflected by a magnetic field according to their masses. The lighter they are, the more they are deflected. The more the ion is charged, the more it gets deflected. The beam of ions passing through the machine is detected electrically.



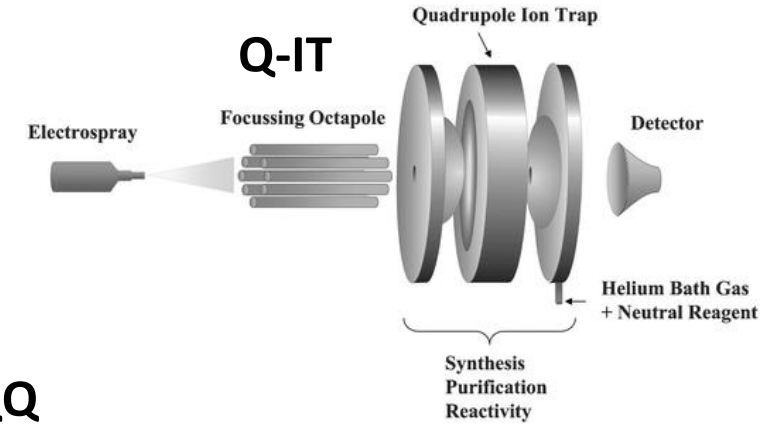




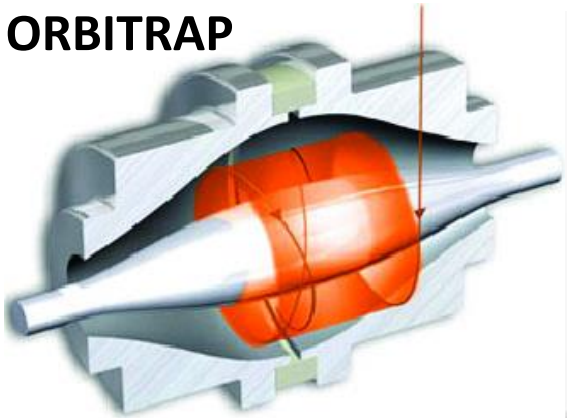
TECHNICALITIES

Types of mass analysers

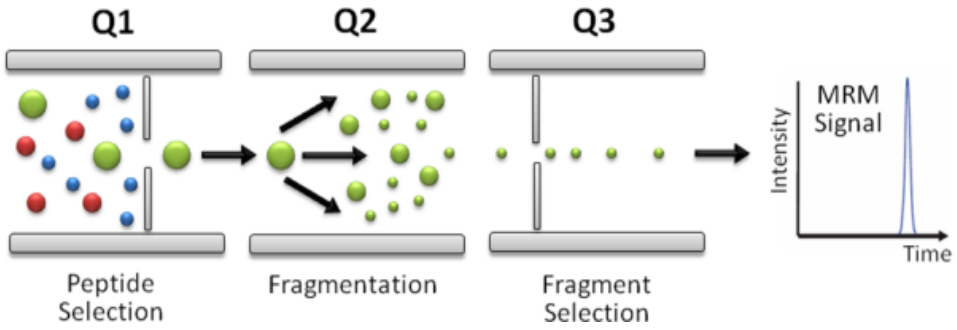
Q-IT



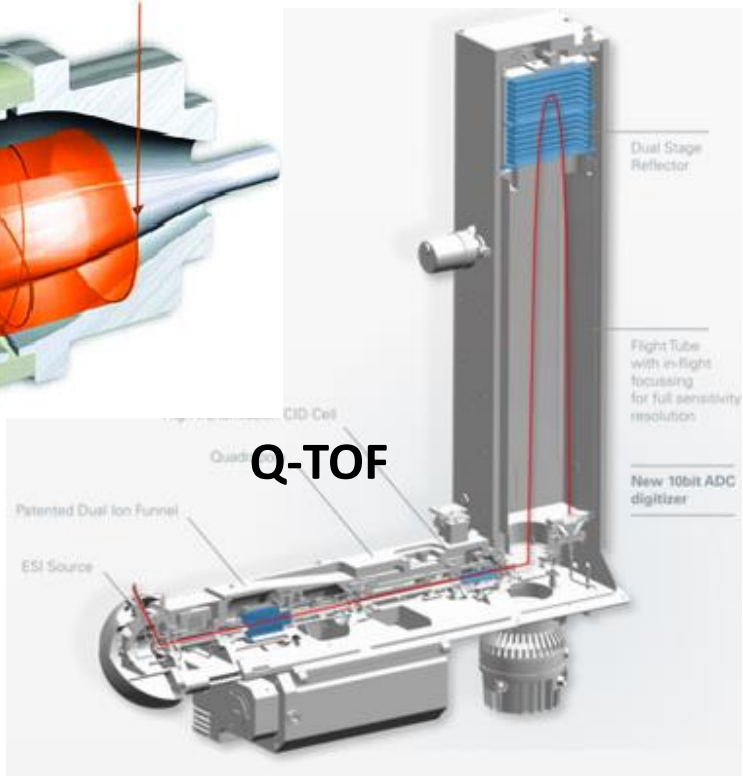
ORBITRAP



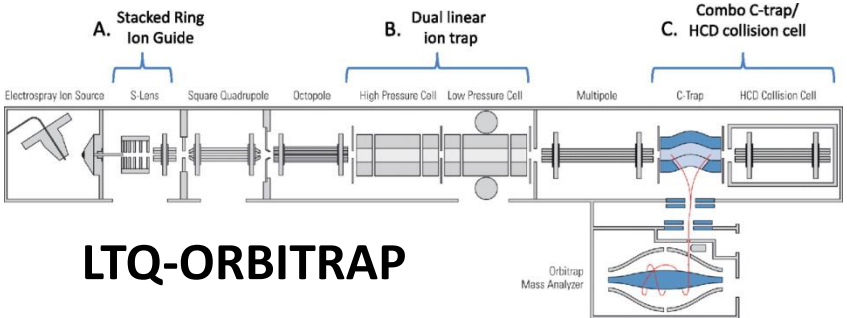
QQQ



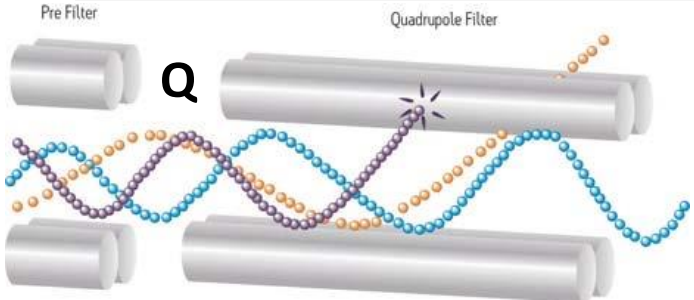
Q-TOF



LTQ-ORBITRAP

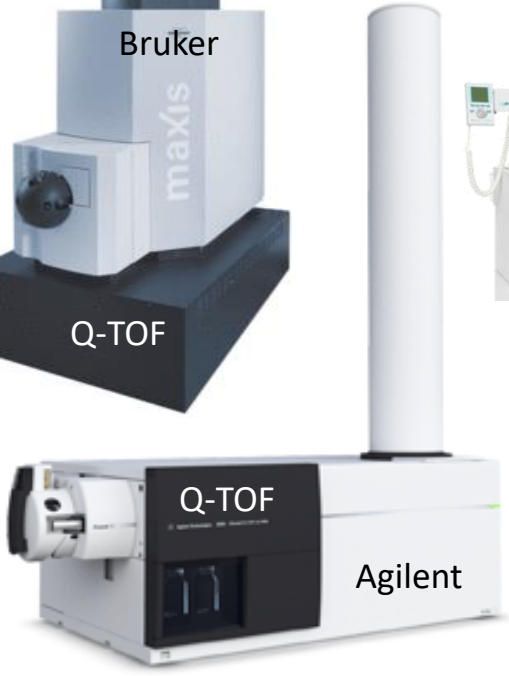


Q



TECHNICALITIES

Example of Mass Spectrometers

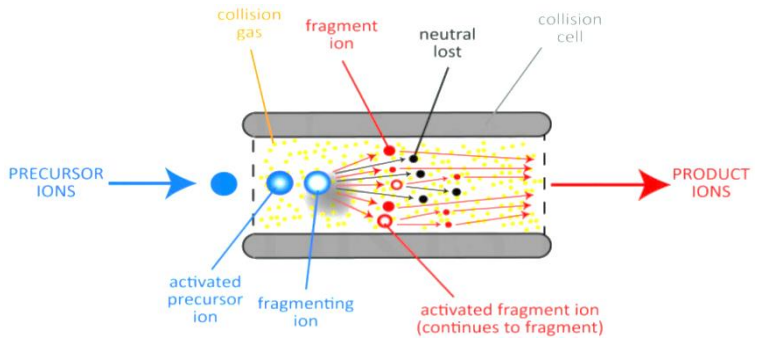


TECHNICALITIES

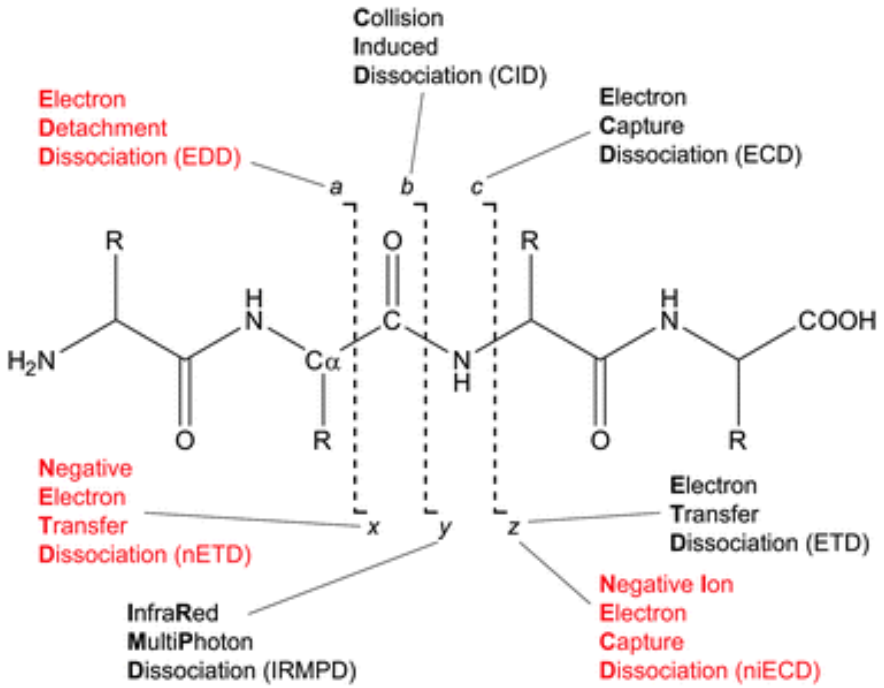
Fragmentation modes (MS/MS)

Fragmentation is a type of chemical dissociation used in MS to find the structural formula of a molecule through mass spectrum analysis, a process called structural elucidation.
It can occur in the ion source (in-source fragmentation, i.e. MALDI) where it is generally not a desired effect as it is less controlled.
Desired fragmentation is made in the collision zone (post-source fragmentation) of a tandem mass spectrometer. Different types of mass fragmentation:

-collision-induced dissociation (CID),
-electron-capture dissociation (ECD),
-electron-transfer dissociation (ETD),
-negative electron-transfer dissociation (NETD),
-electron-detachment dissociation (EDD),
-photodissociation,
-surface-induced dissociation (SID),
-Higher-energy C-trap dissociation (HCD),
-charge remote fragmentation.

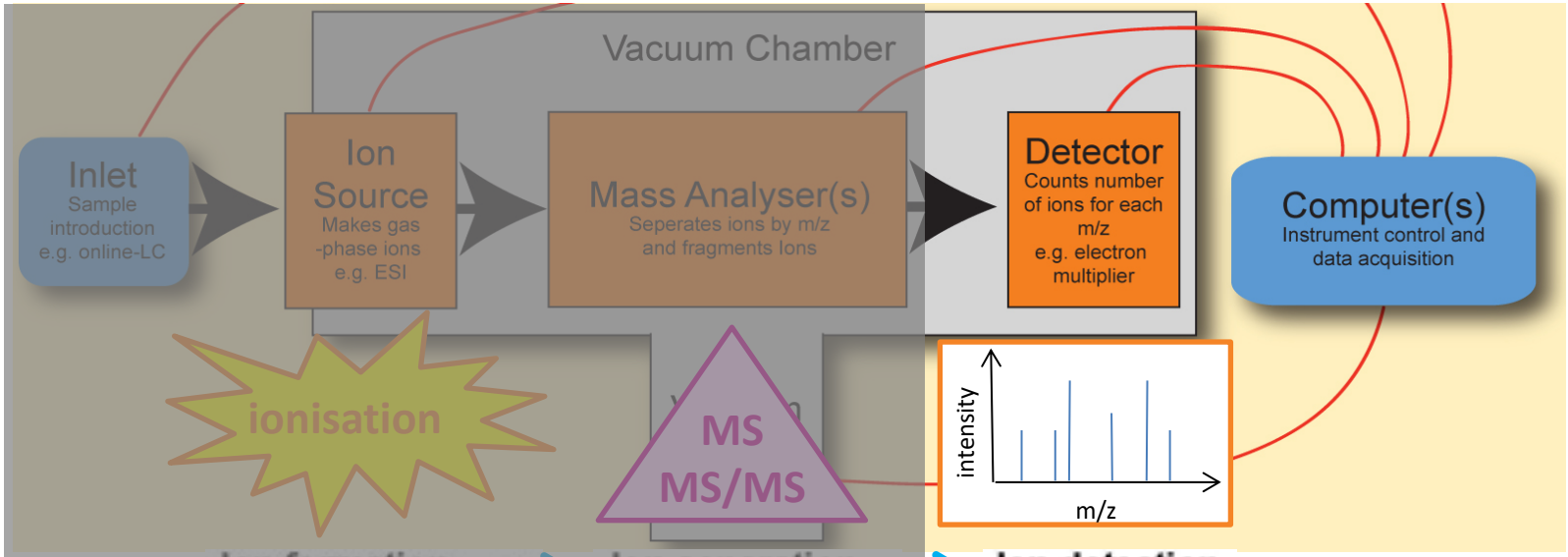


CID



TECHNICALITIES

3/ ion detection



Ion formation → **Ion separation** → **Ion detection**

- Matrix-assisted laser desorption/ionization
- Electron impact
- Spark source
- Thermal ionization
- Photo-ionization
- Chemical ionization
- Field ionization
- Field desorption
- Multiphoton ionization
- Fast atom bombardment
- Plasma desorption
- Infrared laser desorption
- Nanospray
- Electrospray ionization

- Time of flight
  - Magnetic
  - Double-focusing
  - Reversed geometry
  - Quadrupole
  - Quadrupole ion trap
- Triple quadrupole
  - Four sector
- Hybrid
  - Radio frequency
  - Fourier transform
  - Magnetic sector
  - Electric sector
- Linear ion trap
  - Ion cyclotron resonance

Faraday cup
Secondary electron multiplier
Scintillator
Micro channel plate

Covered here

AGRICULTURE VICTORIA

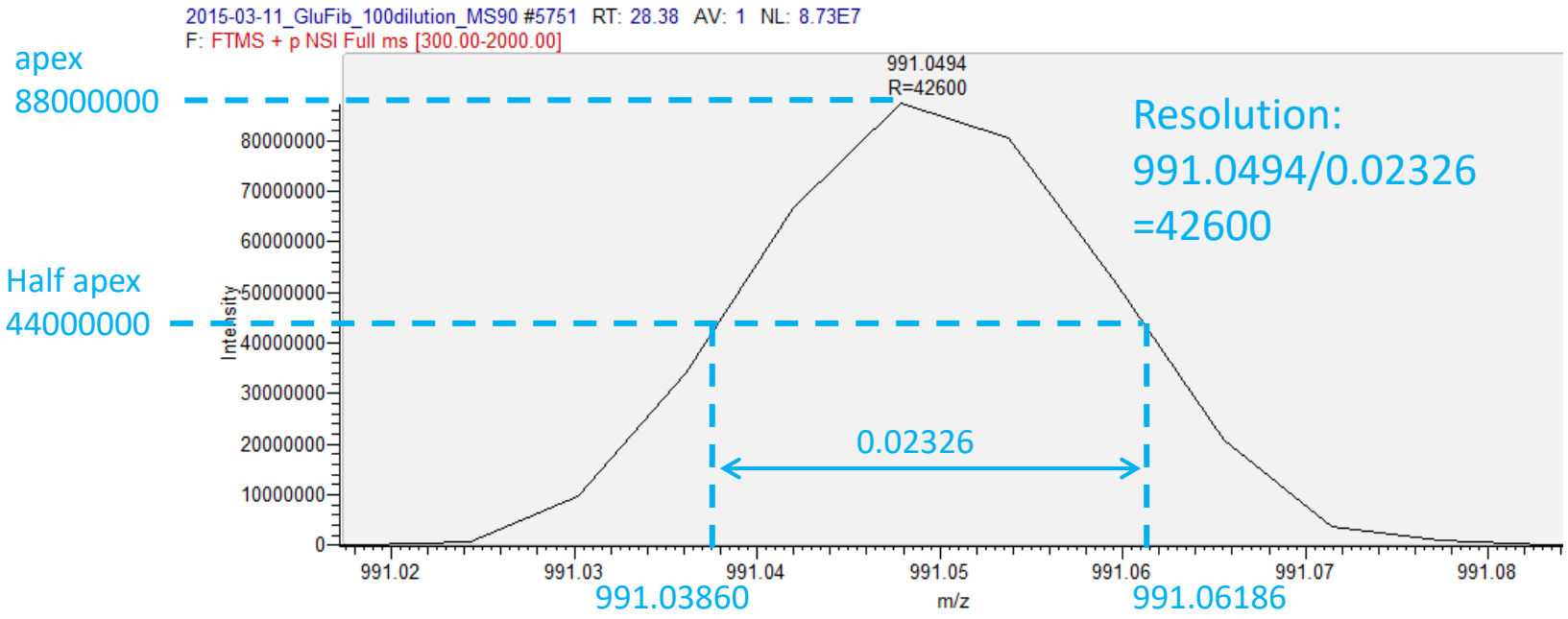
TECHNICALITIES

Mass resolution

High resolution → isotope pattern → monoisotopic mass → high mass accuracy → reliable ID
Low resolution → isotopes not resolved → average mass → low mass accuracy → less reliable

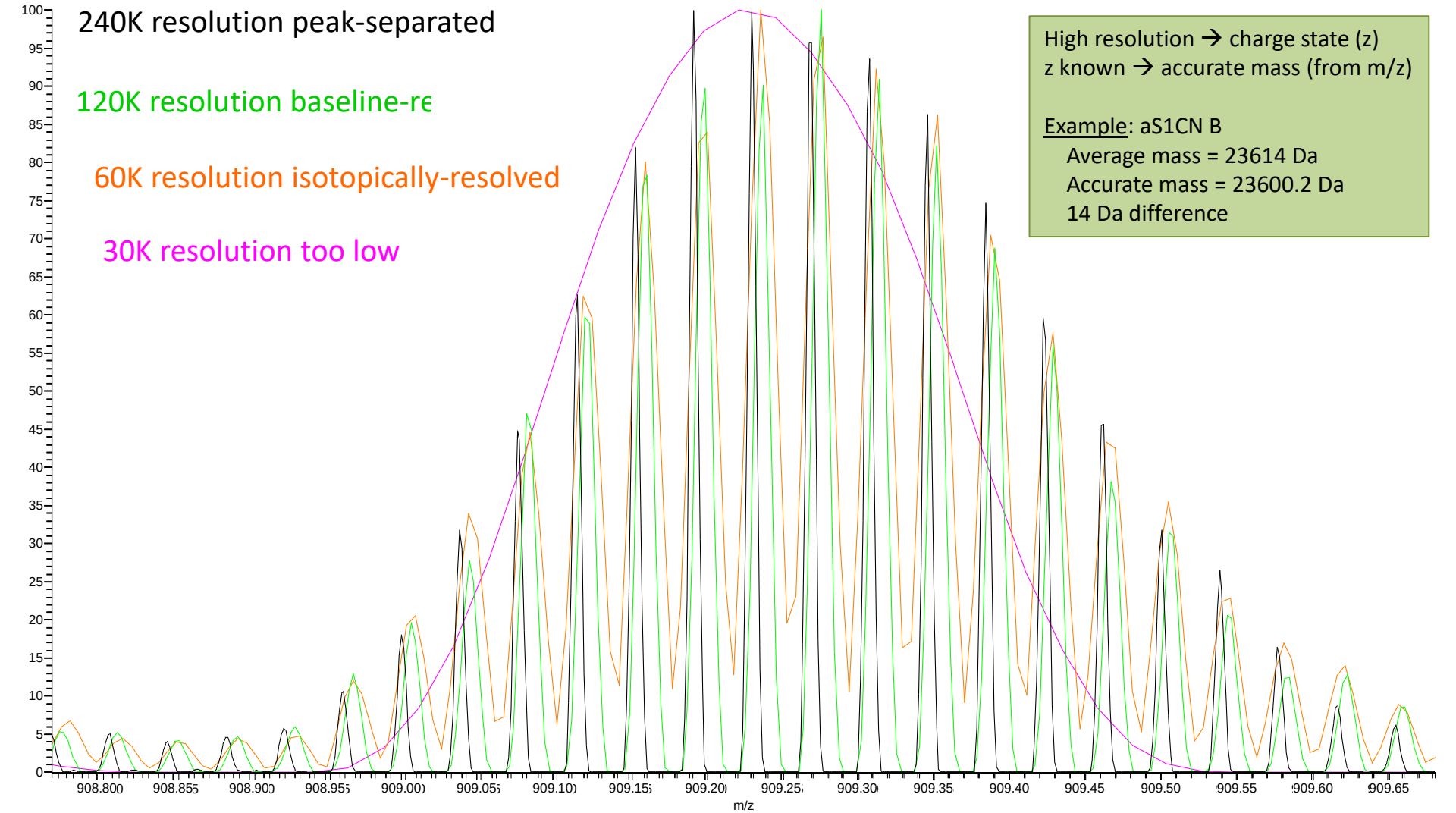
If the mass analyser used has sufficient resolution it will allow the separate detection of the different isotopic forms of an ion.

To determine the full width at half-maximum (FWHM) resolution of a peak in a mass spectrum divide the m/z of the peak by the width of that peak at 50% of its intensity. An value of this number that is greater than 10,000 will give excellent separation of isotopic peaks.



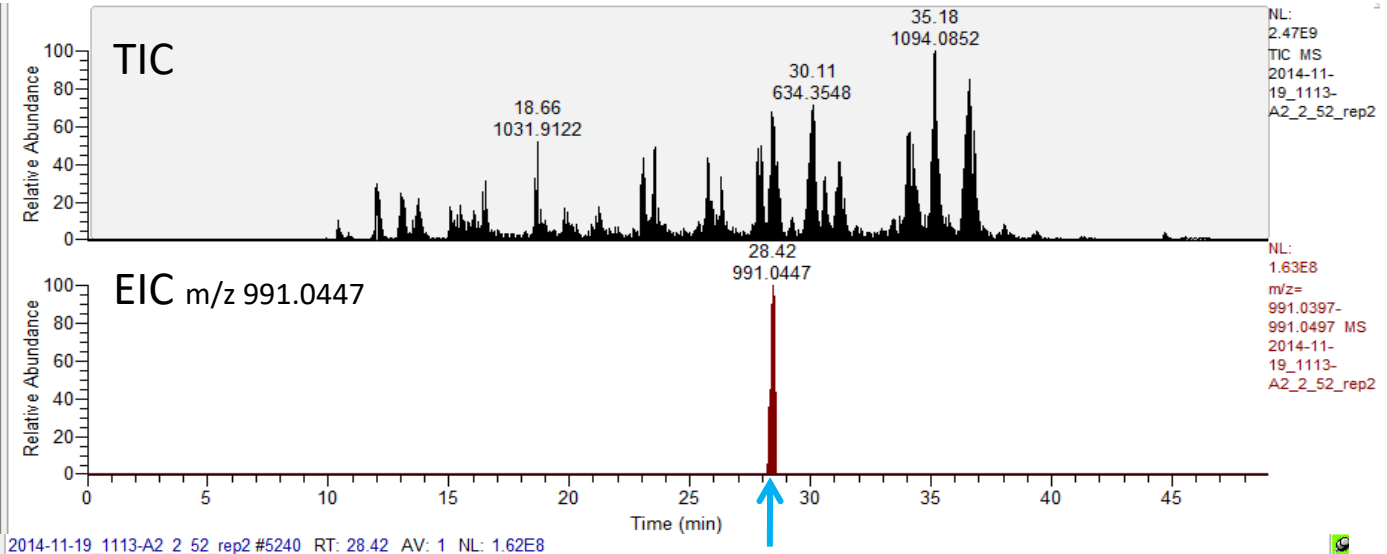
TECHNICALITIES

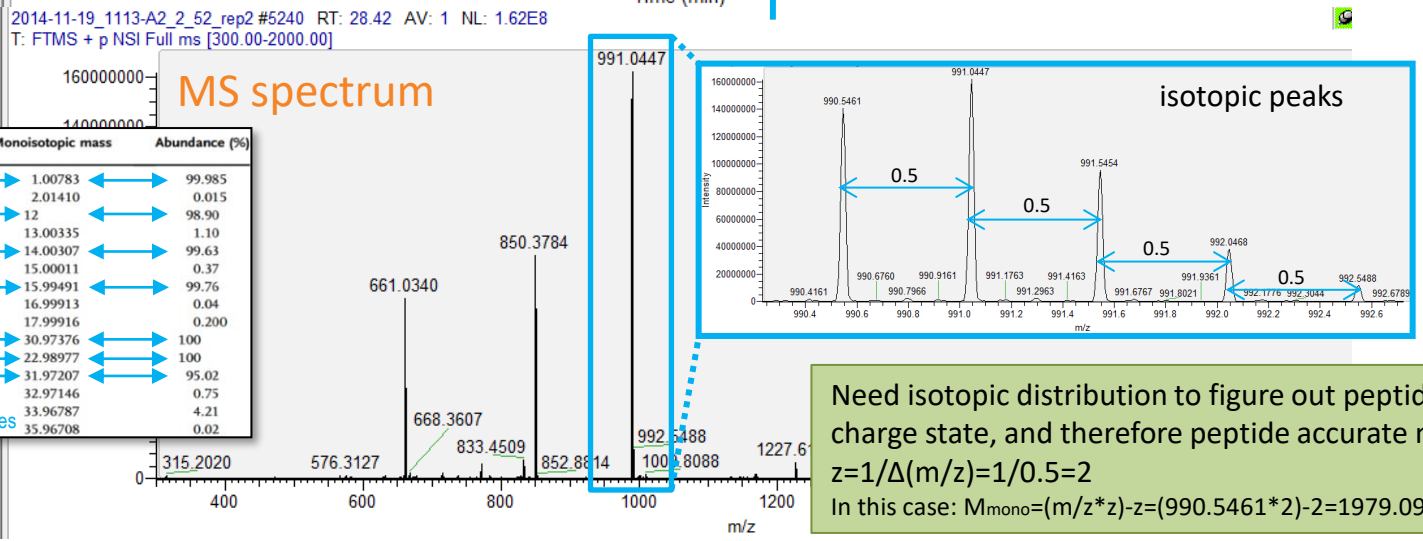
Why does resolution matter?



TECHNICALITIES

Charge state of peptide (MS1)





Element	Average mass	Isotope	Monoisotopic mass	Abundance (%)
Hydrogen	1.008	^{1}H	1.00783	99.985
		^{2}H	2.01410	0.015
Carbon	12.011	^{12}C	12	98.90
		^{13}C	13.00335	1.10
Nitrogen	14.007	^{14}N	14.00307	99.63
		^{15}N	15.00011	0.37
Oxygen	15.999	^{16}O	15.99491	99.76
		^{17}O	16.99913	0.04
		^{18}O	17.99916	0.200
Phosphorus	30.974	^{31}P	30.97376	100
Sodium	22.990	^{23}Na	22.98977	100
Sulfur	32.064	^{32}S	31.97207	95.02
		^{33}S	32.97146	0.75
		^{34}S	33.96787	4.21
		^{36}S	35.96708	0.02

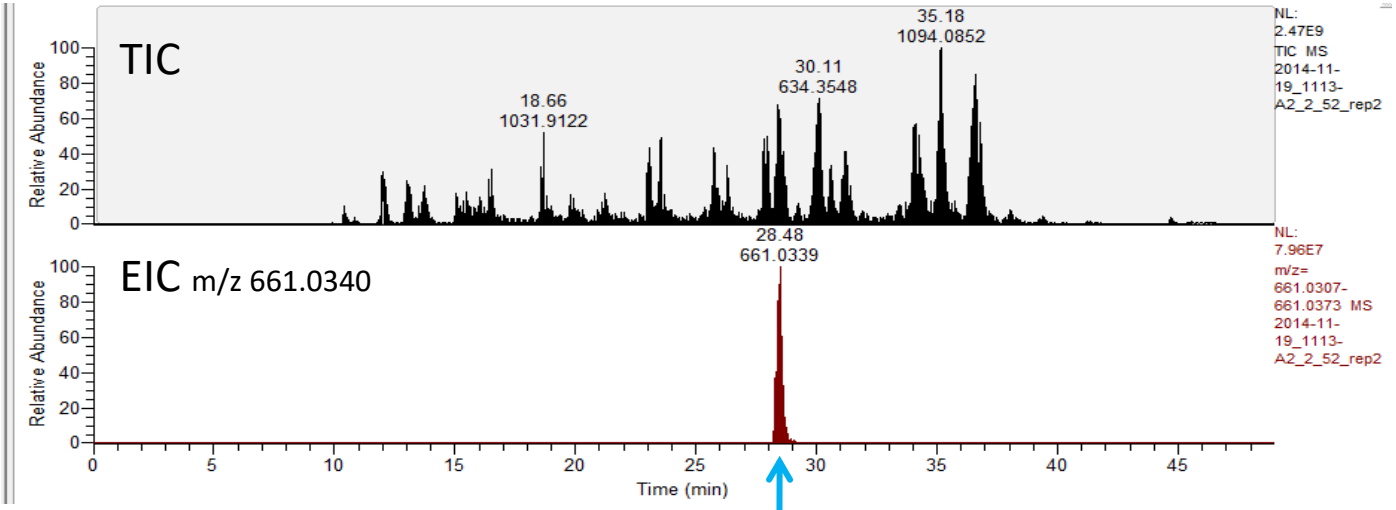
Monoisotopic mass=most abundant isotopes

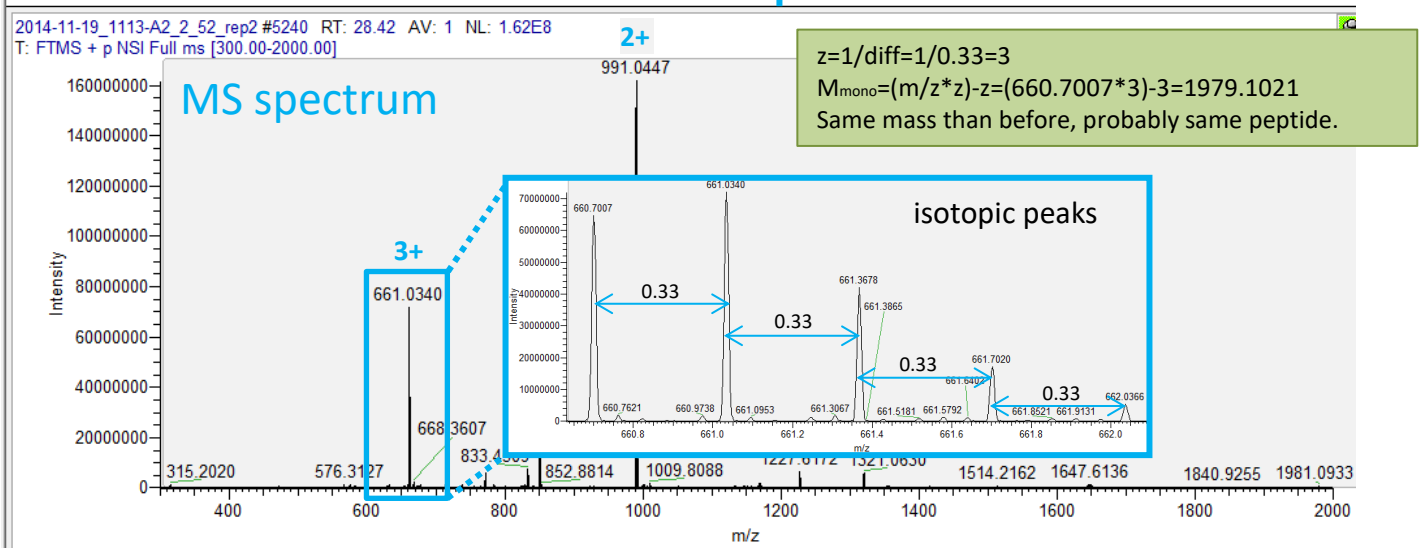
Need isotopic distribution to figure out peptide charge state, and therefore peptide accurate mass.
$z=1/\Delta(m/z)=1/0.5=2$
In this case: $M_{mono}=(m/z*z)-z=(990.5461*2)-2=1979.0922$

TECHNICALITIES

Charge state of peptide (MS1)

The greater the charge state (ESI only), the greater the chance to detect the ion along the m/z scanning window. Essential for large molecules such as proteins.

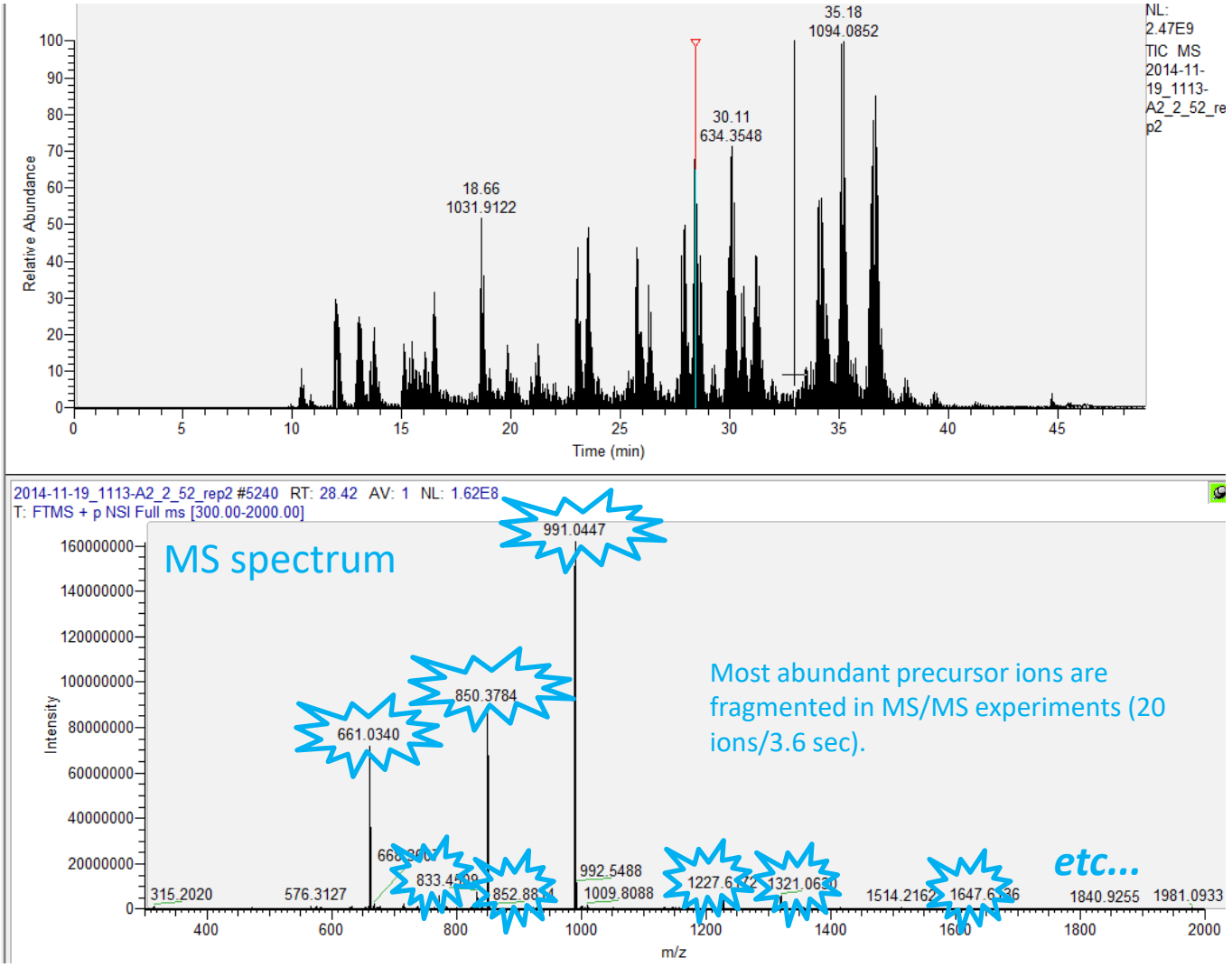




z=1/diff=1/0.33=3
M_{mono}=(m/z*z)-z=(660.7007*3)-3=1979.1021
Same mass than before, probably same peptide.

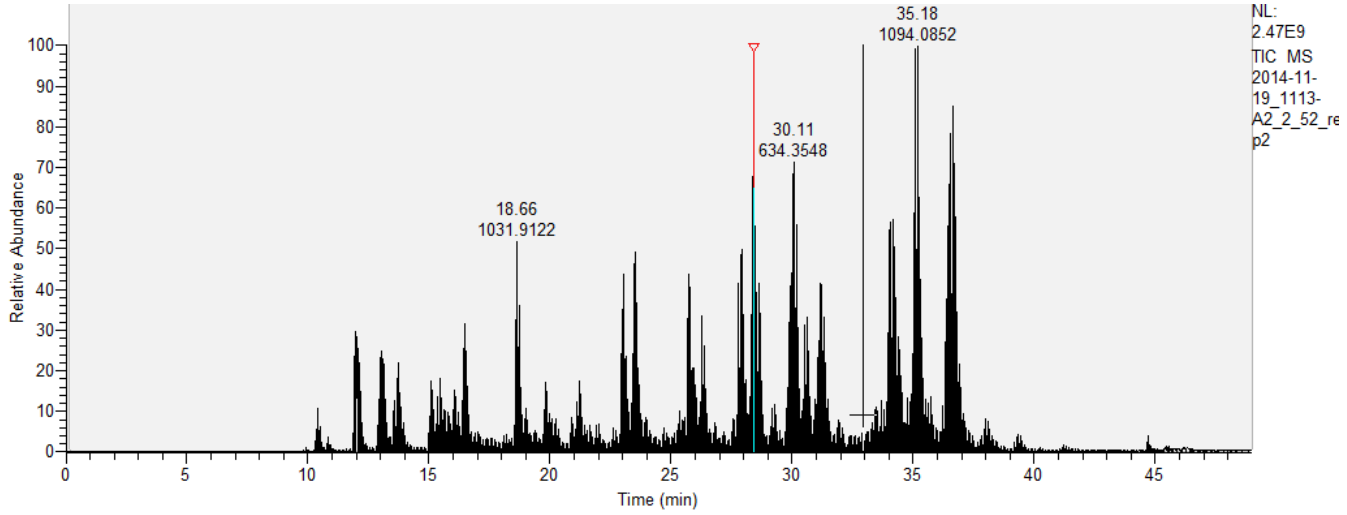
TECHNICALITIES

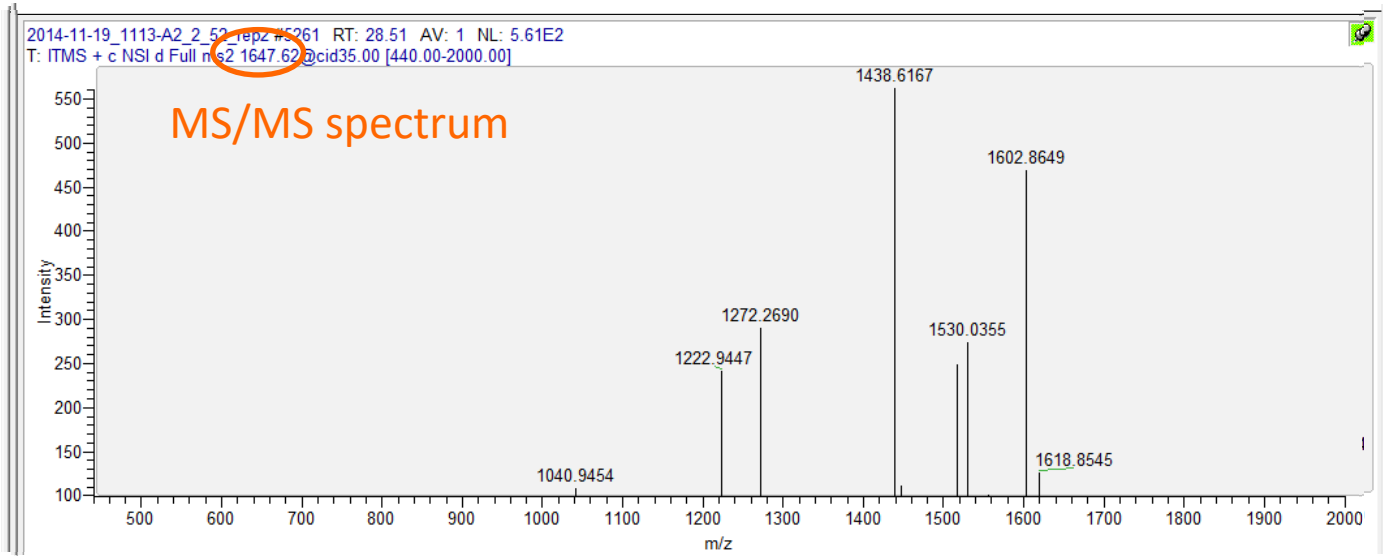
Fragmentation of peptides (MS2)



TECHNICALITIES

Fragmentation of peptides (MS2)



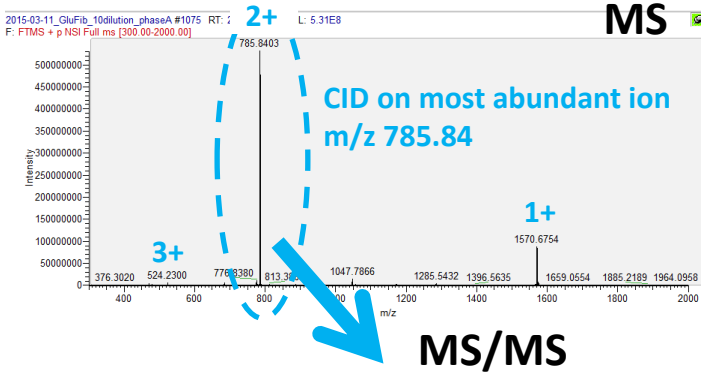


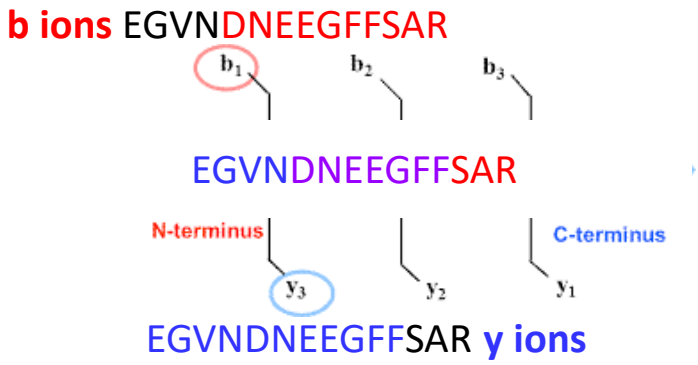
TECHNICALITIES

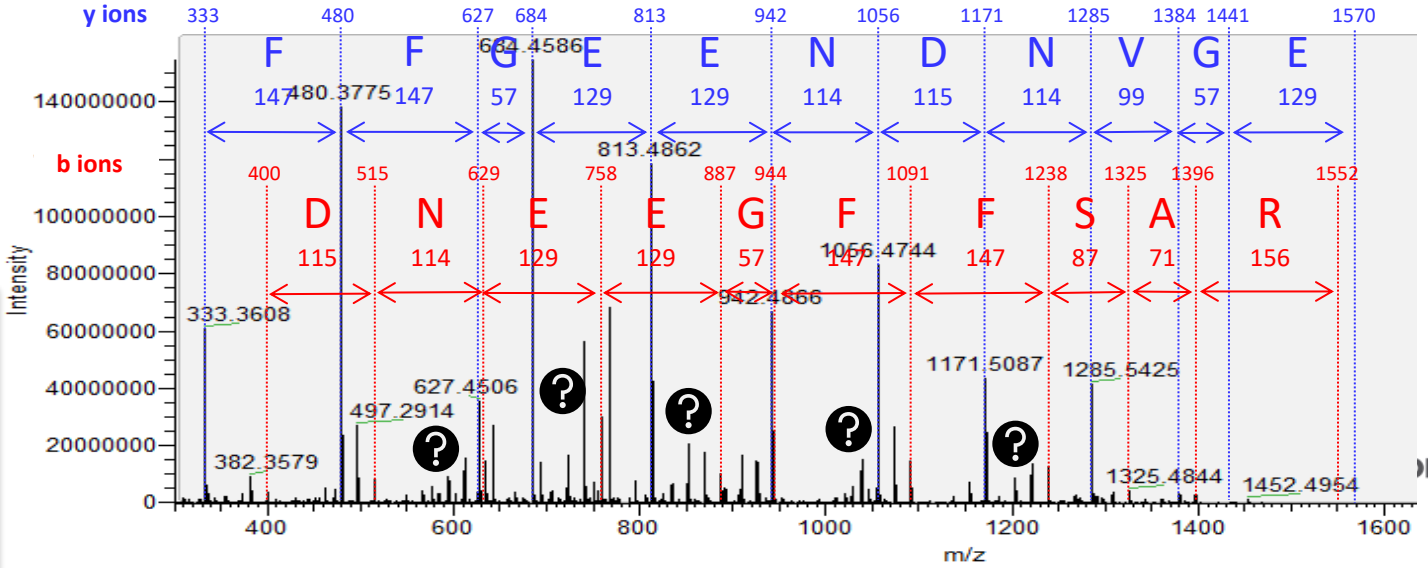
Peptide sequencing (MS2)

The diagram below illustrates how peptides are sequenced using MS/MS data.

1-letter code	Monoisotopic
A	71.03711
R	156.10111
N	114.04293
D	115.02694
C	103.00919
E	129.04259
Q	128.05858
G	57.02146
H	137.05891
I	113.08406
L	113.08406
K	128.09496
M	131.04049
F	147.06841
P	97.05276
S	87.03203
T	101.04768
W	186.07931
Y	163.06333
V	99.06841



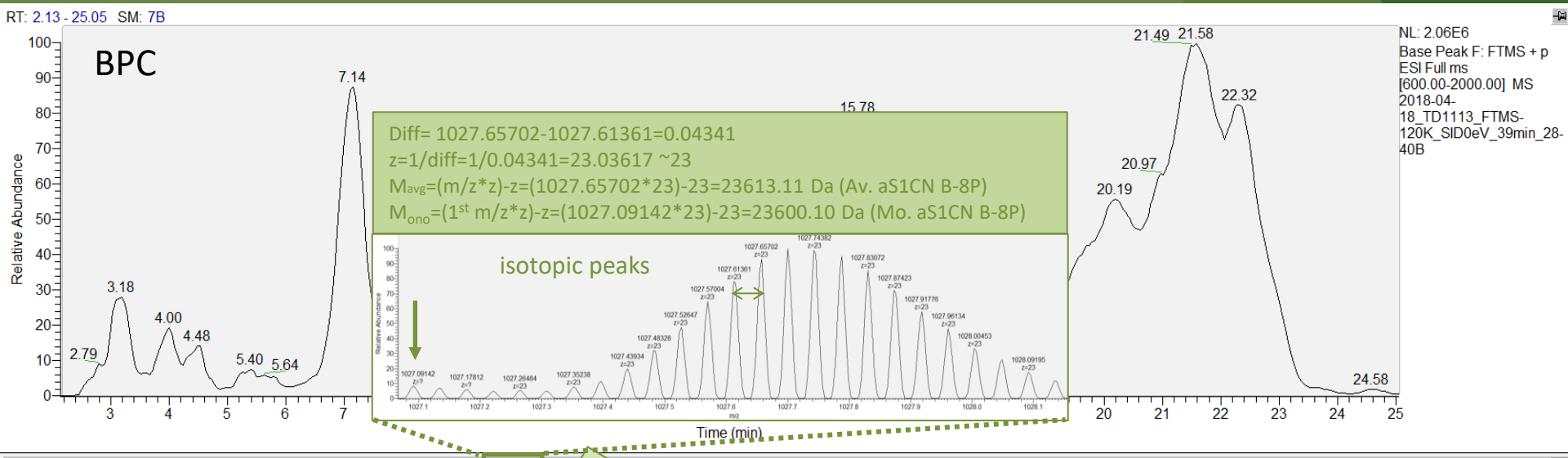


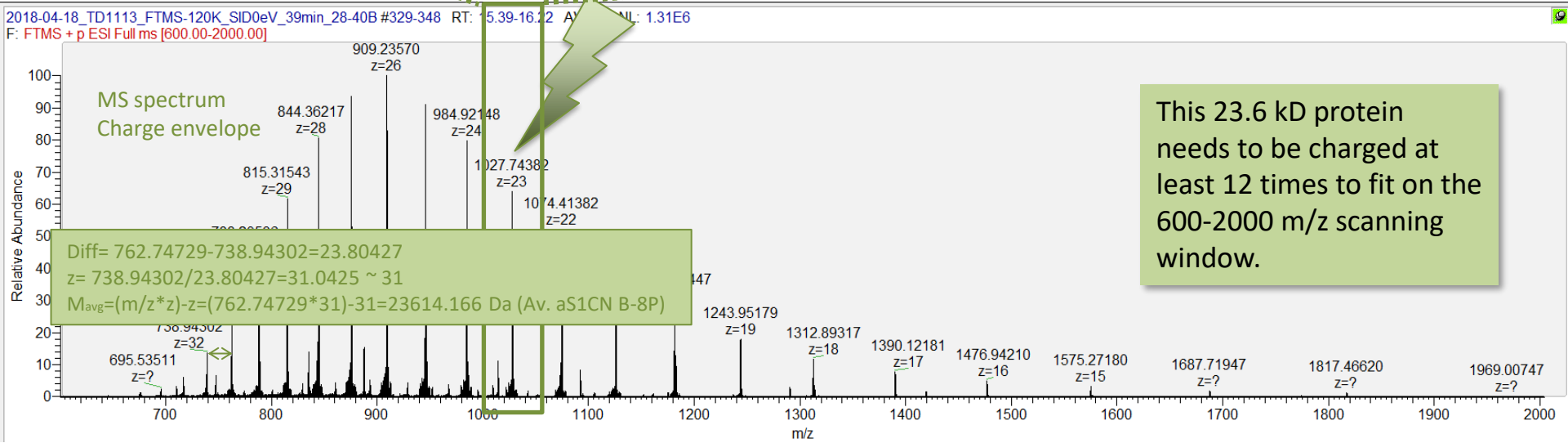


Not every spectral peak is annotated for this peptide, hence the need for powerful algorithm to figure this out!

TECHNICALITIES

Charge state of protein (MS1)



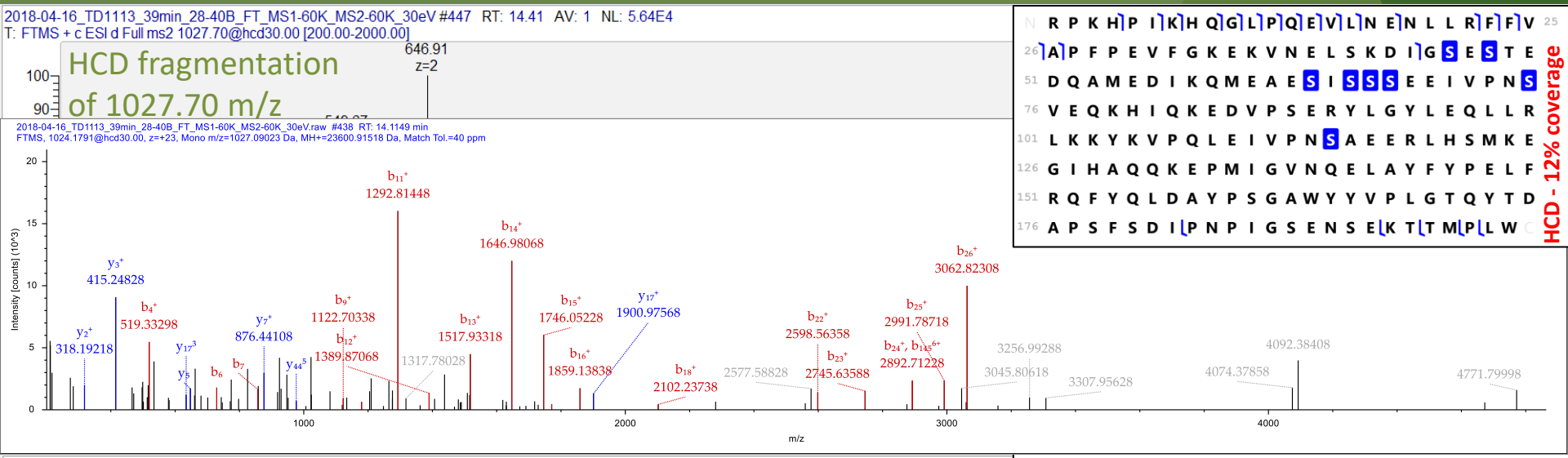


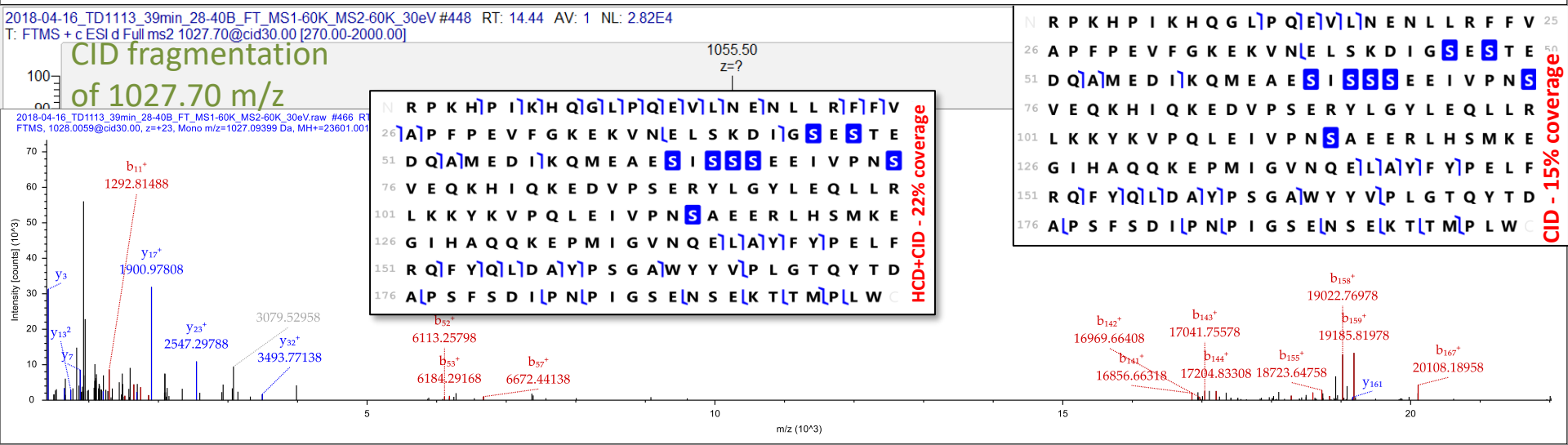
This 23.6 kD protein needs to be charged at least 12 times to fit on the 600-2000 m/z scanning window.

TECHNICALITIES

Fragmentation of a protein (MS2)

aS1CN B-8P 23600.1 Da

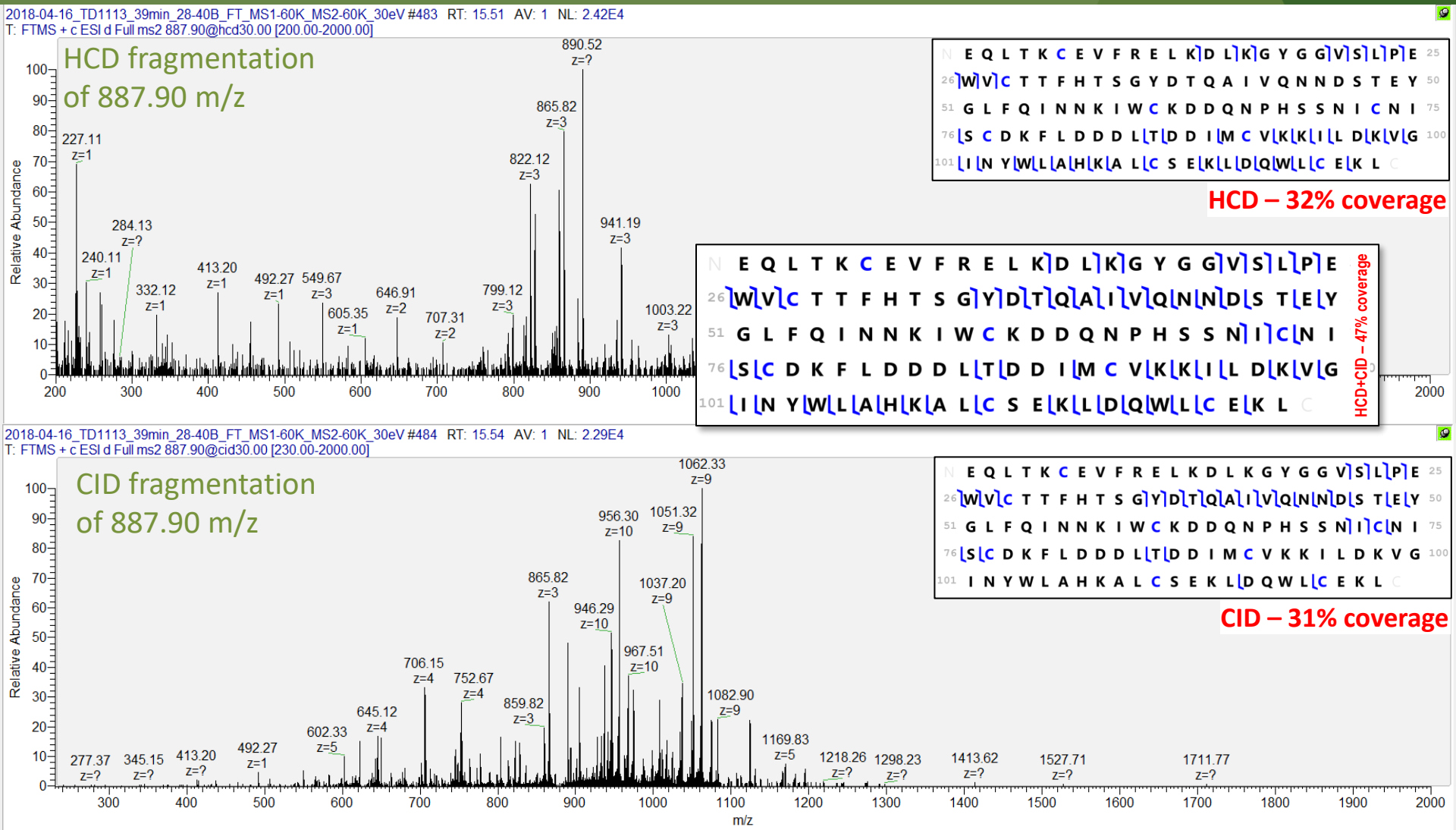




TECHNICALITIES

Fragmentation of a protein (MS2)

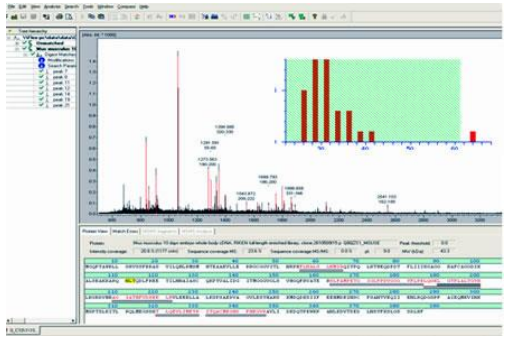
aLA B 14177.23 Da

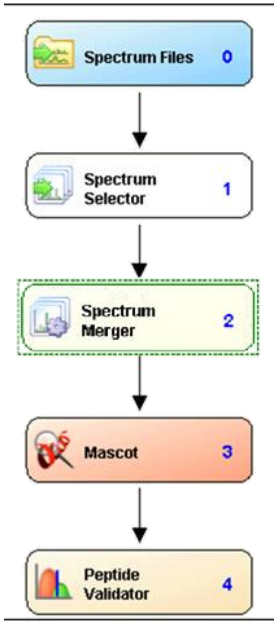


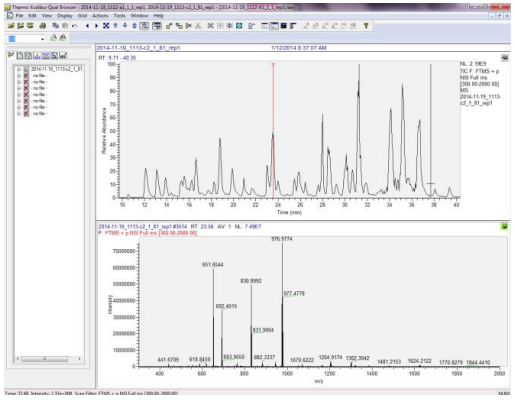
TECHNICALITIES

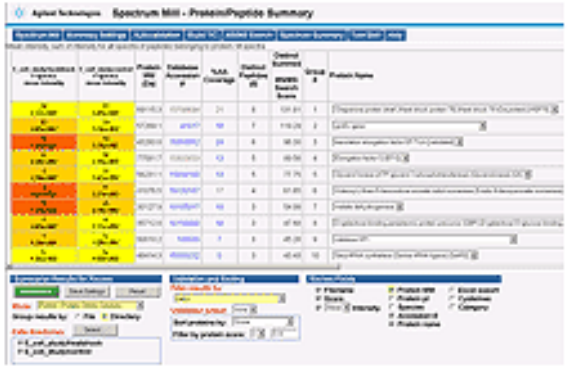
Data Analysis

Softwares:
Xcalibur, Proteome Discoverer, ProSightPC, Protein Deconvolution (Thermo)
Mass Hunter, Spectrum Mill (Agilent)
FlexAnalysis, DataAnalysis, Biotools, SequenceEditor (Bruker)
Genedata expressionist





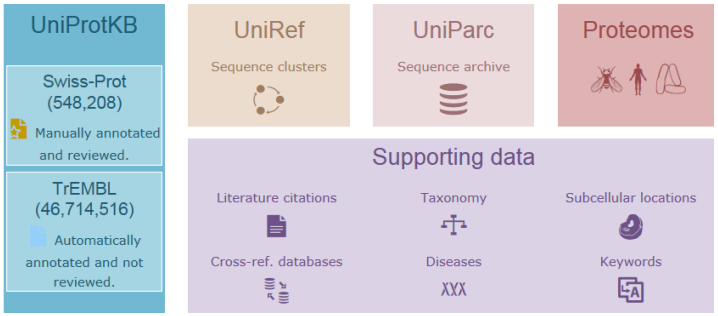




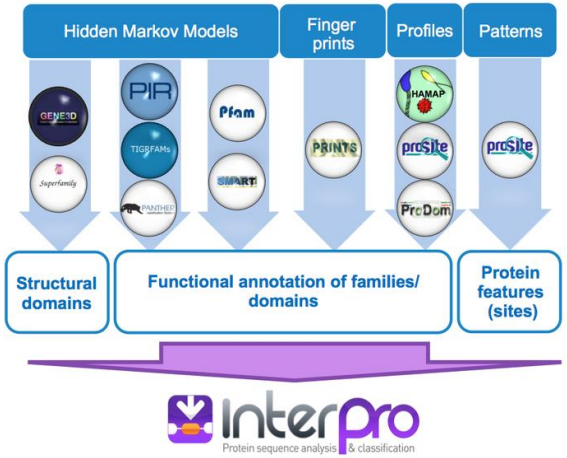
TECHNICALITIES

Protein databases

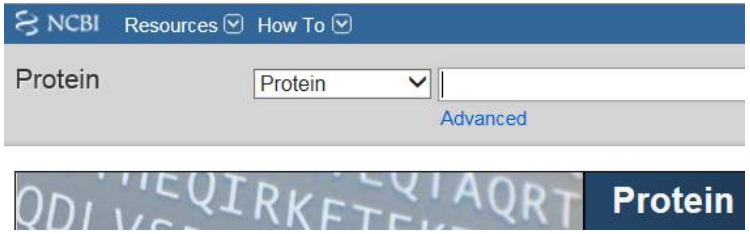
Proteomics immensely benefits from genomics and genome sequencing projects as their annotation steps include defining gene models from which protein sequences are inferred. Since more genomes are being sequenced all the time, protein databases keep on growing.



UniProt provides the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information.



InterPro is a bioinformatics resource that provides functional analysis of protein sequences by classifying them into families and predicting the presence of domains and important sites. Linked with UniProtKB.



NCBI Protein database is a collection of sequences from several sources, including translations from annotated coding regions in GenBank, RefSeq and TPA, as well as records from SwissProt, PIR, PRF, and PDB. Protein sequences are the fundamental determinants of biological structure and function.

Gene models don't inform on PTMs (phospho, glycoproteins...)

TECHNICALITIES

Automated search algorithms

Peptides are identified from mass spectra by using search algorithms which compare experimentally obtained mass spectral peaks to the theoretical masses derived from protein sequences.

Sequest was first such algorithm wherein scoring for matches is based on number of **spectral peaks that are common to theoretical and experimental spectra** (Eng *et al* 1994). **Customised libraries.**

X! tandem generates theoretical spectra for peptide sequences using information about **intensity associated with amino acids**. These spectra are compared with experimental data to generate an expectation value (Craig and Beavis 2003; Craig and Beavis 2004).

Mascot, a probabilistic algorithm, estimates the probability that a **predicted peptide sequence generated the experimentally observed peptide by chance** (Kapp *et al* 2005). **Uniprot libraries maintained by MatrixScience**.

Scaffold is a computer program that integrates search results from the **above three algorithms** (Sequest, X! tandem and Mascot) to generate peptide identification and protein identification probabilities.

Phenyx uses a scoring scheme based on signal detection theory and pattern recognition to calculate a **likelihood ratio that distinguishes true from false peptide identifications** (Colinge *et al* 2003).

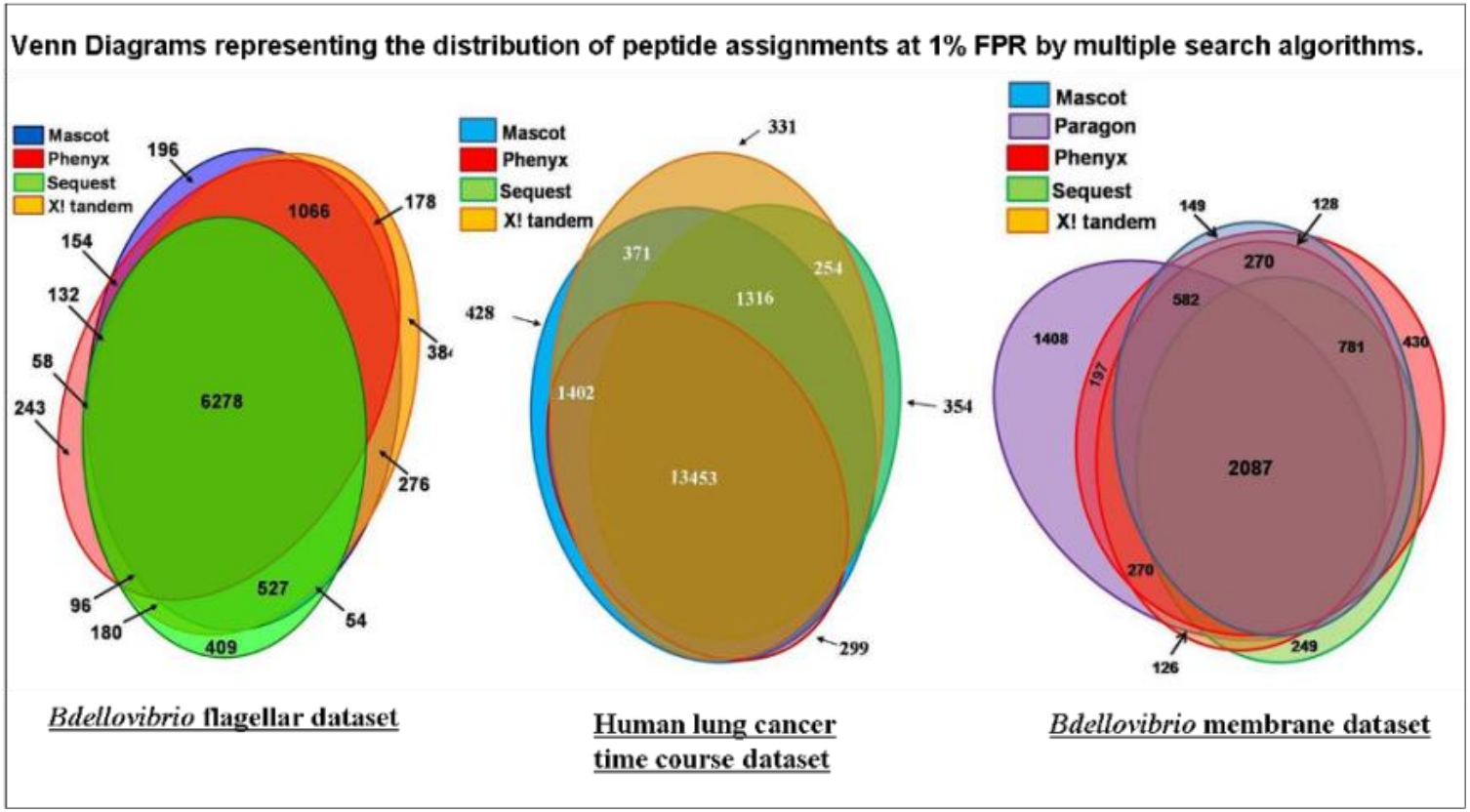
Paragon, a search algorithm that is a part of ProteinPilot suite of softwares, uses **Sequence Temperature Values along with feature probabilities** for identification of peptides from a database (Shilov *et al* 2007). **For iTRAQ experiments.**

OMSSA, Sonar and Spectrum Mill... different approach to identify peptides.

TECHNICALITIES

Search algorithms

Search algorithms all implement different methods which therefore will lead to slightly different results (different peptide hits → different proteins). Best to use several algorithms and use the common hits as the list of identified proteins. Algorithm most commonly used is Mascot.



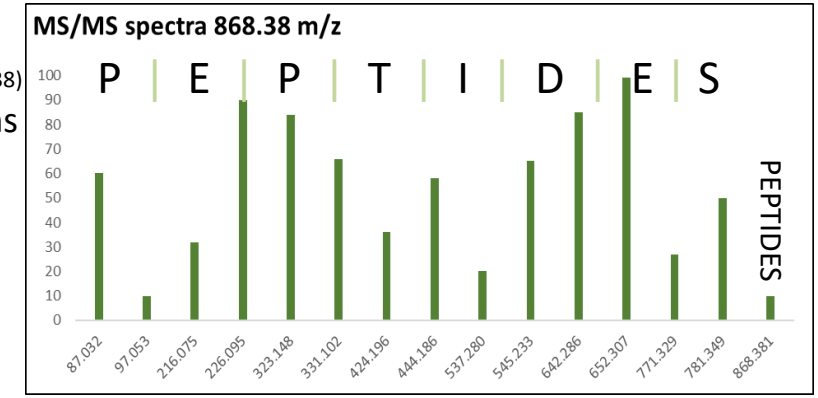
Kapp *et al* 2005 and Balgley *et al* 2007

TECHNICALITIES

Protein identification

For 1 peptide:

Accurate mass of the precursor ion (ex. m/z 868.38)
Accurate masses of the MS/MS fragment ions (y- and b-ions)
Great confidence in the peptide sequence



mass	db	ion
868.38	seditpep	precursor
868.38	eppdesit	precursor
868.38	tippeesd	precursor
868.38	peptides	precursor

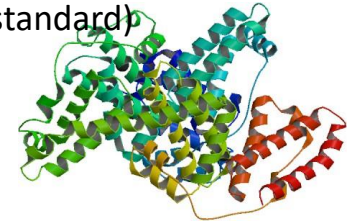
Several peptides per protein:

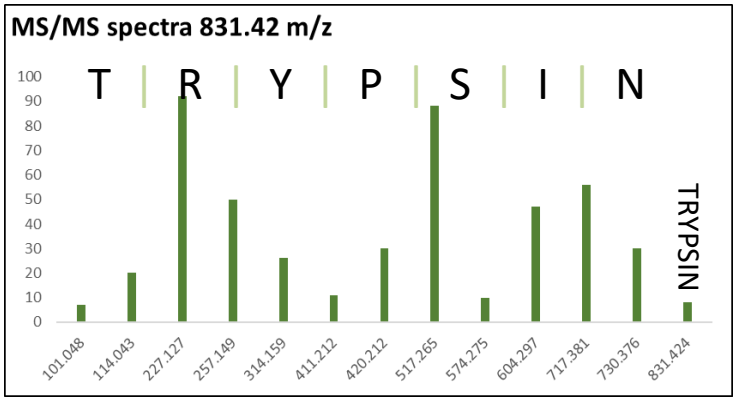
Lots of masses!
The more peptides, the greater the confidence in protein identity
Usually, at least 2 different peptides/proteins (unless proteins are of low MW).

Limitations:

Only enzyme-released peptides are searched
Assumption that protein is intact (could be truncated)

Real example: Bovine Serum Albumin (BSA, bottom-up standard)



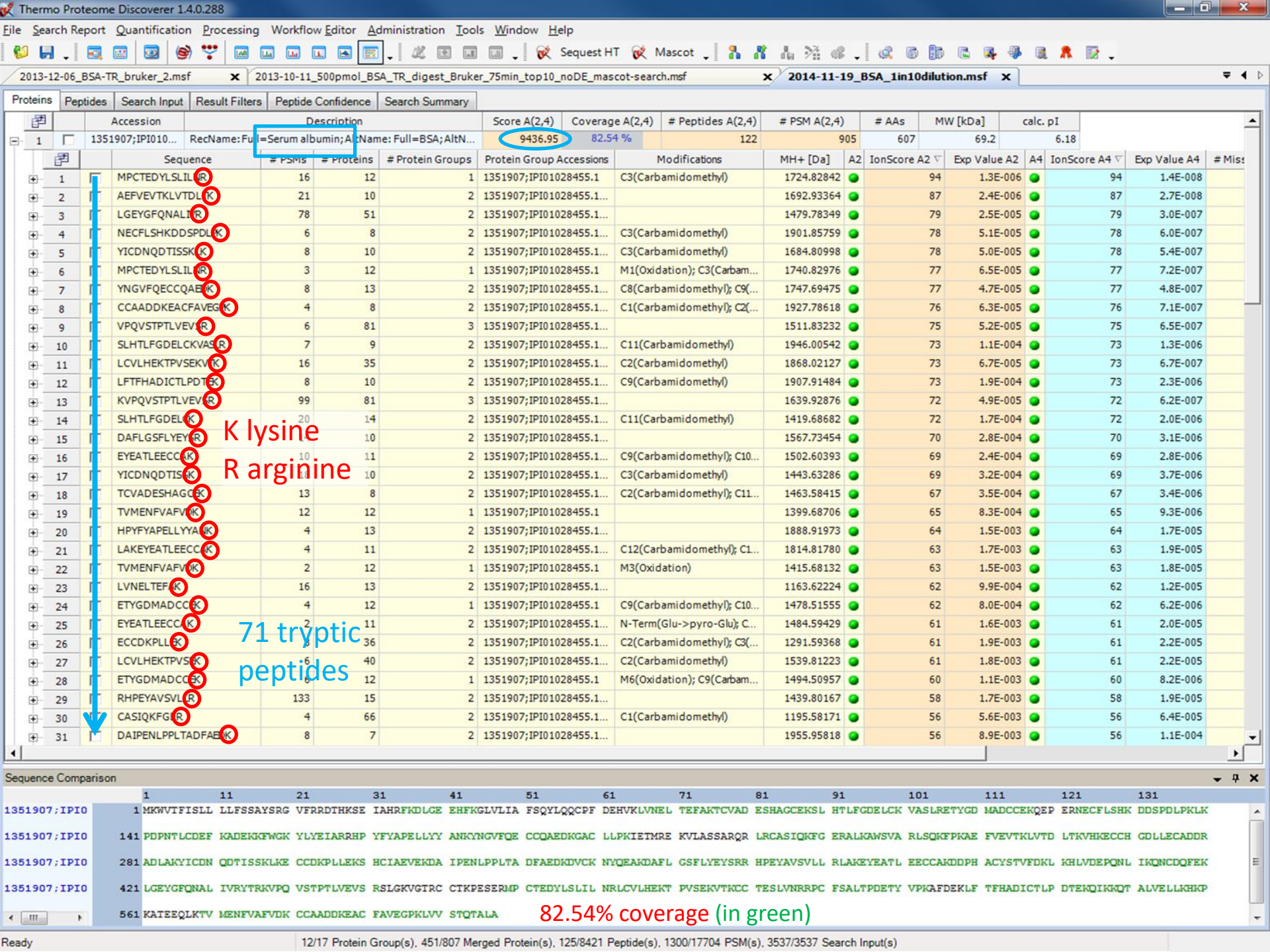


mass	db	ion
831.42	nispyrt	precursor
831.42	nytirsp	precursor
831.42	tripsyn	precursor
831.42	TRYPSIN	precursor
101.05	t	b1
114.04	n	y1
227.13	in	y2
257.15	TR	b2
314.16	SIN	y3
411.21	PSIN	y4
420.21	TRY	b3
517.26	TRYP	b4
574.28	YPSIN	y5
604.30	TRYPS	b5
717.38	TRYPSI	b6
730.38	RYPSIN	y6

Database search:

THISPROTEINISMADEOFLOTSOFPEPTIDESAMENABLETOTRYPSINDIGESTION

15/60 = 25% sequence coverage


Thermo Proteome Discoverer 1.4.0.288
File Search Report Quantification Processing Workflow Editor Administration Tools Window Help
Sequest HT Mascot
2013-12-06_BSA-TR_bruker_2.msf
2013-10-11_500pmol_BSA_TR_digest_Bruker_75min_top10_noDE_mascot-search.msf
2014-11-19_BSA_1in10dilution.msf
Proteins Peptides Search Input Result Filters Peptide Confidence Search Summary
Accession Description Score A(2,4) Coverage A(2,4) # Peptides A(2,4) # PSM A(2,4) # AAs MW [kDa] calc. pI
1351907;IPI010... RecName: Full=Serum albumin; AltName: Full=BSA; AltN... 9436.95 82.54 % 122 905 607 69.2 6.18
Sequence # PSMs # Proteins # Protein Groups Protein Group Accessions Modifications MH+ [Da] A2 IonScore A2 Exp Value A2 A4 IonScore A4 Exp Value A4 # Miss
1 MPCTEDYLSLILNR 16 12 1 1351907;IPI01028455.1 C3(Carbamidomethyl) 1724.82842 94 1.3E-006 94 1.4E-008
2 AEFVEVTKLVTDLTK 21 10 2 1351907;IPI01028455.1... 1692.93364 87 2.4E-006 87 2.7E-008
3 LGEYGFQNALIVR 78 51 2 1351907;IPI01028455.1... 1479.78349 79 2.5E-005 79 3.0E-007
4 NECFLSHKDDSPDLPK 6 8 2 1351907;IPI01028455.1... C3(Carbamidomethyl) 1901.85759 78 5.1E-005 78 6.0E-007
5 YICDNQDTISSKLK 8 10 2 1351907;IPI01028455.1... C3(Carbamidomethyl) 1684.80998 78 5.0E-005 78 5.4E-007
6 MPCTEDYLSLILNR 3 12 1 1351907;IPI01028455.1 M1(Oxidation); C3(Carbam... 1740.82976 77 6.5E-005 77 7.2E-007
7 YNGVFQECCQAEDK 8 13 2 1351907;IPI01028455.1... C8(Carbamidomethyl); C9(... 1747.69475 77 4.7E-005 77 4.8E-007
8 CCAADDKEACFAVEGPK 4 8 2 1351907;IPI01028455.1... C1(Carbamidomethyl); C2(... 1927.78618 76 6.3E-005 76 7.1E-007
9 VPQVSTPTLVEVSR 6 81 3 1351907;IPI01028455.1... 1511.83232 75 5.2E-005 75 6.5E-007
10 SLHTLFGDELCKVASLR 7 9 2 1351907;IPI01028455.1... C11(Carbamidomethyl) 1946.00542 73 1.1E-004 73 1.3E-006
11 LCVLHEKTPVSEKVTK 16 35 2 1351907;IPI01028455.1... C2(Carbamidomethyl) 1868.02127 73 6.7E-005 73 6.7E-007
12 LFTFHADICTLPDTEK 8 10 2 1351907;IPI01028455.1... C9(Carbamidomethyl) 1907.91484 73 1.9E-004 73 2.3E-006
13 KVPQVSTPTLVEVSR 99 81 3 1351907;IPI01028455.1... 1639.92876 72 4.9E-005 72 6.2E-007
14 SLHTLFGDELCK 20 14 2 1351907;IPI01028455.1... C11(Carbamidomethyl) 1419.68682 72 1.7E-004 72 2.0E-006
15 DAFLGSFLYEYSR 10 2 1351907;IPI01028455.1... 1567.73454 70 2.8E-004 70 3.1E-006
16 EYEATLEECCAK 10 11 2 1351907;IPI01028455.1... C9(Carbamidomethyl); C10... 1502.60393 69 2.4E-004 69 2.8E-006
17 YICDNQDTISSK 10 2 1351907;IPI01028455.1... C3(Carbamidomethyl) 1443.63286 69 3.2E-004 69 3.7E-006
18 TCVADESHAGCEK 13 8 2 1351907;IPI01028455.1... C2(Carbamidomethyl); C11... 1463.58415 67 3.5E-004 67 3.4E-006
19 TVMENFVAFVDK 12 12 1 1351907;IPI01028455.1 1399.68706 65 8.3E-004 65 9.3E-006
20 HPYFYAPELLYYANK 4 13 2 1351907;IPI01028455.1... 1888.91973 64 1.5E-003 64 1.7E-005
21 LAKEYEATLEECCAK 4 11 2 1351907;IPI01028455.1... C12(Carbamidomethyl); C1... 1814.81780 63 1.7E-003 63 1.9E-005
22 TVMENFVAFVDK 2 12 1 1351907;IPI01028455.1 M3(Oxidation) 1415.68132 63 1.5E-003 63 1.8E-005
23 LVNELTEFAK 16 13 2 1351907;IPI01028455.1... 1163.62224 62 9.9E-004 62 1.2E-005
24 ETYGDMADCCEK 4 12 1 1351907;IPI01028455.1 C9(Carbamidomethyl); C10... 1478.51555 62 8.0E-004 62 6.2E-006
25 EYEATLEECCAK 2 11 2 1351907;IPI01028455.1... N-Term(Glu->pyro-Glu); C... 1484.59429 61 1.6E-003 61 2.0E-005
26 ECCDKPLLEK 3 36 2 1351907;IPI01028455.1... C2(Carbamidomethyl); C3(... 1291.59368 61 1.9E-003 61 2.2E-005
27 LCVLHEKTPVSEK 6 40 2 1351907;IPI01028455.1... C2(Carbamidomethyl) 1539.81223 61 1.8E-003 61 2.2E-005
28 ETYGDMADCCEK 3 12 1 1351907;IPI01028455.1 M6(Oxidation); C9(Carbam... 1494.50957 60 1.1E-003 60 8.2E-006
29 RHPEYAVSVLLR 133 15 2 1351907;IPI01028455.1... 1439.80167 58 1.7E-003 58 1.9E-005
30 CASIQKFGER 4 66 2 1351907;IPI01028455.1... C1(Carbamidomethyl) 1195.58171 56 5.6E-003 56 6.4E-005
31 DAIPENLPPLTADFAEDK 8 7 2 1351907;IPI01028455.1... 1955.95818 56 8.9E-003 56 1.1E-004
K lysine
R arginine
71 tryptic peptides
Sequence Comparison
1351907;IPIO 1 MKWVTFISLL LLFSSAYSRG VFRRDTHKSE IAHRFKDLGE EHFKGLVLIA FSQYLQQCPF DEHVKLVNEL TEFAKTCVAD ESHAGCEKSL HTLFGDELCK VASLRETYGD MADCCEKQEP ERNECFLSHK DDSPDLPKLK
1351907;IPIO 141 PDPNTLCDEF KADEKKFWGK YLYEIARRHP YFYAPELLYY ANKYNGVFQE CCQAEDKGAC LLPKIETMRE KVLASSARQR LRCASIQKFG ERALKAWSVA RLSQKFPKAE FVEVTKLVTD LTKVHKECCH GDLLECADDR
1351907;IPIO 281 ADLAKYICDN QDTISSKLKE CCDKPLLEKS HCIAEVEKDA IPENLPPLTA DFAEDKDVCK NYQEAKDAFL GSFLYEYSRR HPEYAVSVLL RLAKEYEATL EECCAKDDPH ACYSTVFDKL KHLVDEPQNL IKQNCDQFEK
1351907;IPIO 421 LGEYGFQNAL IVRYTRKVPQ VSTPTLVEVS RSLGKVGTRC CTKPESERMP CTEDYLSLIL NRLCVLHEKT PVSEKVTKCC TESLVNRRPC FSALTPDETY VPKAFDEKLF TFHADICTLP DTEKQIKKQT ALVELLKHKP
561 KATEEQLKTV MENFVAFVDK CCAADDKEAC FAVEGPKLVV STQTALA
82.54% coverage (in green)
Ready
12/17 Protein Group(s), 451/807 Merged Protein(s), 125/8421 Peptide(s), 1300/17704 PSM(s), 3537/3537 Search Input(s)

TECHNICALITIES

In silico tools for data mining

"The mapping of complex proteomics data to biological processes has become impossible by manual means, and the need for computer-aided data analysis is essential for further progress of the field." (Kumar & Mann, 2009). Following protein identification using a bottom-up approach (peptides), further prediction analyses can be performed using amino acid sequences of the identified proteins (FASTA format usually). Plentiful! Some examples below…

http://www.genome.jp/kegg/pathway.html

National Center for Biotechnology Information
NCBI

https://www.ncbi.nlm.nih.gov/

PubMed

https://www.ncbi.nlm.nih.gov/pubmed/

http://www.expasy.org/proteomics

http://string-db.org/

UniProt

http://www.uniprot.org/help/uniprotkb

PHOSIDA
Posttranslational Modification Database

http://www.phosida.com/

http://david.abcc.ncifcrf.gov/

PeptideAtlas

http://www.peptideatlas.org/#

http://geneontology.org/

http://www.cbs.dtu.dk/index.shtml

PRIDE

https://www.ebi.ac.uk/pride/archive/

EMBL-EBI

http://www.ebi.ac.uk/

MassIVE
Mass Spectrometry Interactive Virtual Environment

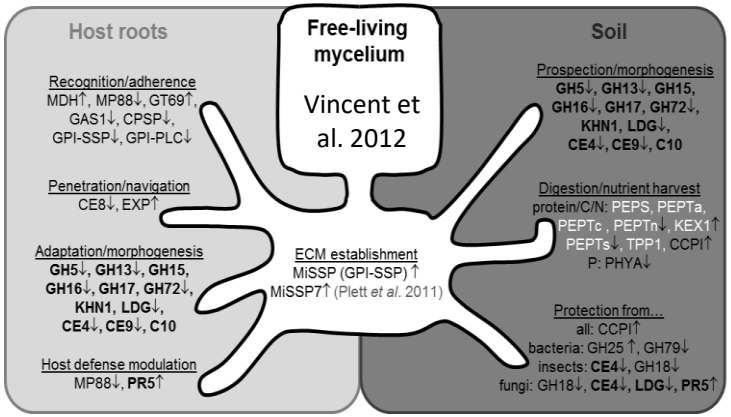
https://massive.ucsd.edu/ProteoSAFe/static/massive.jsp

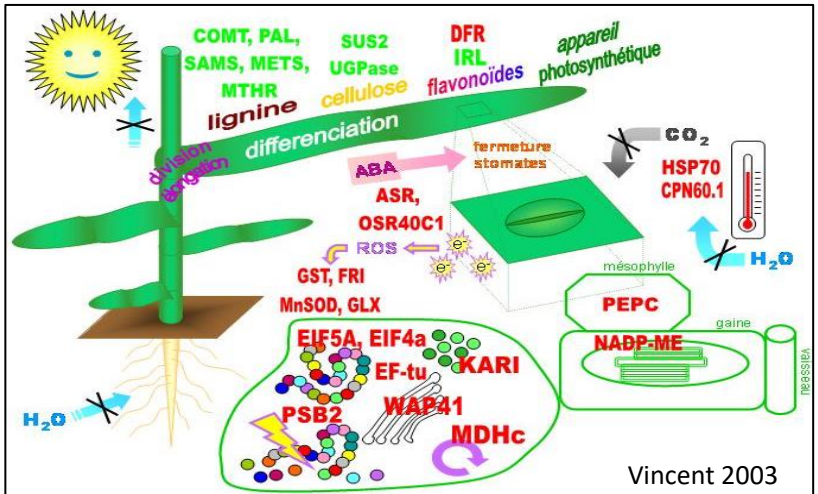
TECHNICALITIES

Search results

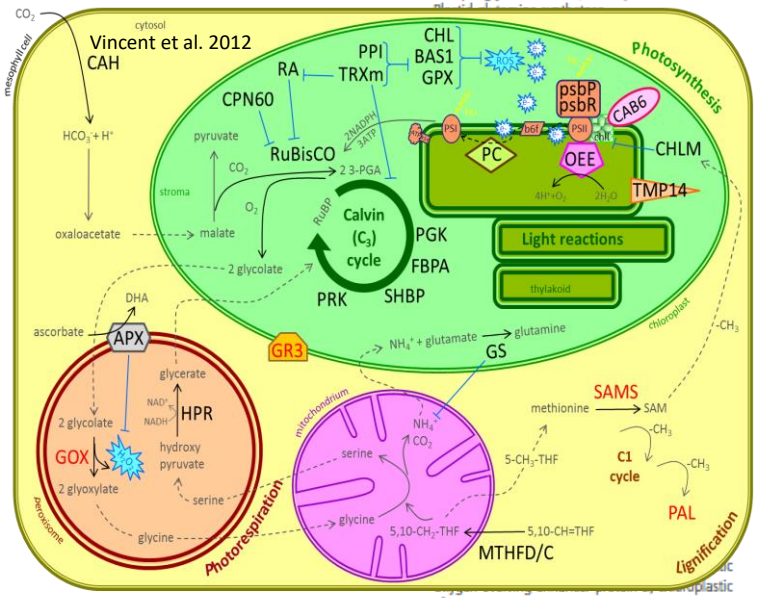
The final output of bottom-up experiments is a list of proteins that can be quite long (hundreds to thousands of proteins).

Biologically speaking, how do we make sense of it? Data mining!





Vincent 2003



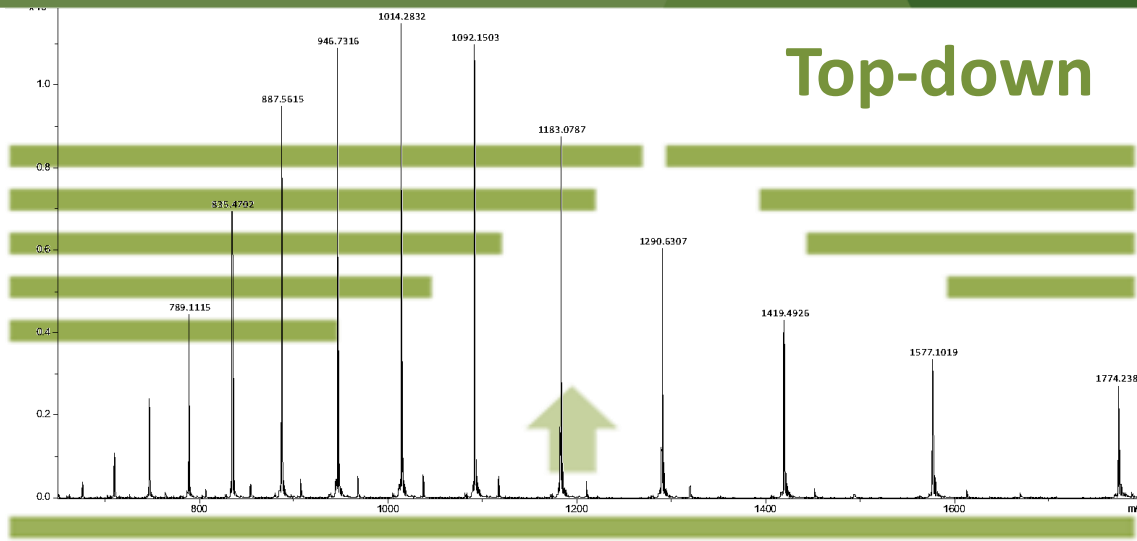
Description	Accession	MASCOT score	Coverage (%)
Heat shock protein (90 kDa)	Q43638	385	14
Putative uncharacterized protein Sb06g023840	C5YCZ2	664	16
Translation elongation factor G	C5YCZ2	719	18
Heat shock cognate protein 70	Q6QUX5	361	16
Heat shock cognate protein 70	Q6QUX5	370	15
Heat shock protein 70	Q5MGA8	259	12
70-kDa heat shock protein	C7ENF7	632	14
Heat shock protein 70	D2D320	95	7
2,3-Bisphosphoglycerate-mutase	A8QPL0	163	25
Ribulose bisphosphate carboxylase/oxygenase (RuBisCO) large chain	B6GUT8	94	5
RuBisCO large subunit-binding protein subunit β	Q43831	335	15
RuBisCO large subunit-binding protein subunit β	Q43831	686	21
RuBisCO large subunit-binding protein subunit β	Q43831	679	20
RuBisCO large subunit-binding protein subunit α	P08823	1217	29
RuBisCO large subunit-binding protein subunit α	P08823	1267	31
RuBisCO large subunit-binding protein subunit α	P08823	975	29
Ribulose bisphosphate carboxylase large chain (fragment)	Q6VW42	66	4
S-Adenosylmethionine synthase 1	A6XMY9	808	39
S-Adenosylmethionine synthase 3	Q4LB22	768	28
Elongation factor Tu	C0P699	280	11
S-Adenosylmethionine synthase 3	Q4LB22	193	15
1-Deoxy-D-xylulose 5-phosphate reductoisomerase, putative	Q4H1G4	129	11
S-Adenosylmethionine synthase 1	Q70EZ8	153	6
RuBisCO activase A, chloroplastic	Q40073	620	20
Plastid glutamine synthetase	C7DPL0	350	19
Plastid glutamine synthetase	C7DPL0	479	23
Phosphoglycerate kinase, chloroplastic	P12782	976	26
	C7DPL0	284	16
	Q40073	624	22
	Q40073	590	23
	Q40073	607	22
	P46285	219	26
	A2XG06	277	16
	P26302	739	34
	P26302	249	15
	C0KTA6	392	16
	C0KTA6	535	23
	C0KTA6	169	11
	C0KTA6	459	21
	B4FR47	283	8
	C5YXG6	166	8
	D5LXX6	103	19
	A5JV93	396	30
	Q9XEN5	114	11
	P29305	223	24
	Q08G36	344	29
	O23798	421	30
	Q38JB6	96	10
	A2XZK1	109	9
	P29546	79	13
	P29546	108	11
	A2YNZ1	96	6
	C3V134	473	59
	O81988	195	15
	P34937	247	17
	Q00434	790	47
	B8B3P0	188	11
Chaperonin 21, putative	Q00434	76	6
2-Cys peroxiredoxin BAS1, chloroplastic	A2YWS7	180	9
Oxygen-evolving enhancer protein 2, chloroplastic	P80602	194	25
Thylakoid lumenal 19-kDa protein	Q00434	166	27
2-Cys peroxiredoxin BAS1, chloroplastic	P80602	670	44
50S ribosomal protein L10	B4G1Q5	131	8
2-Cys peroxiredoxin BAS1, chloroplastic	P80602	486	53
Chlorophyll *a*–*b*-binding protein 6A	C1K5B4	156	9
Nucleoside diphosphate kinase	Q0IMS5	162	17
50S ribosomal protein L12-1, chloroplastic	Q06030	91	14
Histone H2B	B4FYZ0	73	16
Ribulose bisphosphate carboxylase small chain	Q5NDA6	453	38

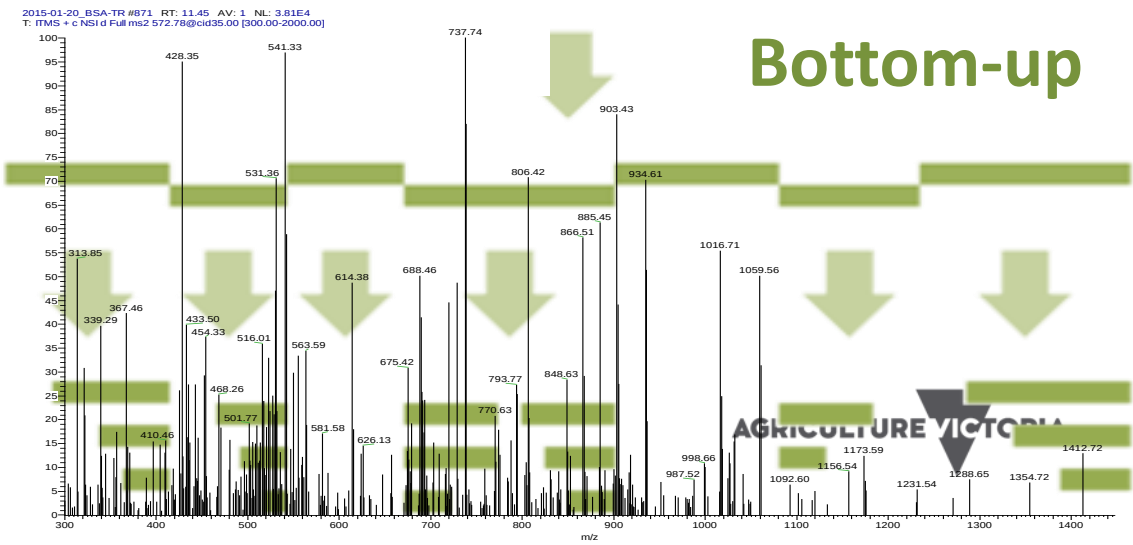
Etc…

APPLICATIONS

MILK PROTEOMICS

Bottom-up and top-down analyses

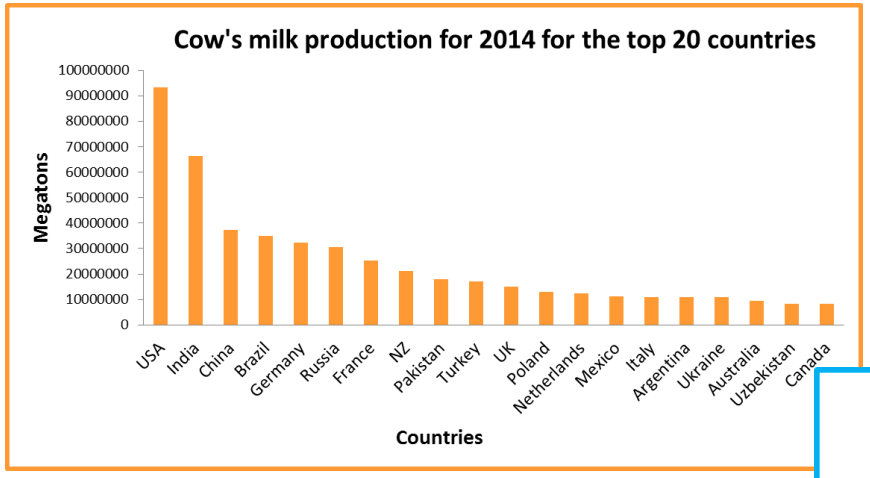


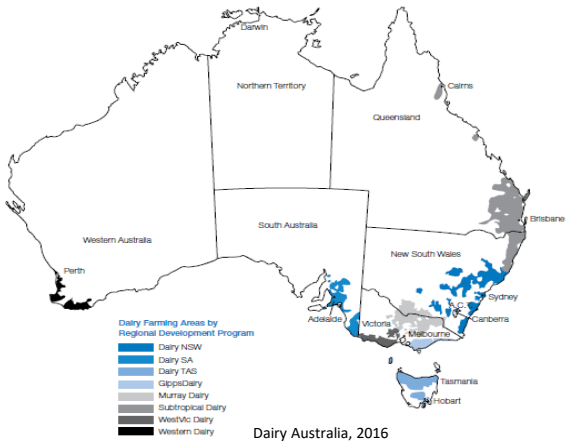


MILK PROTEOMICS

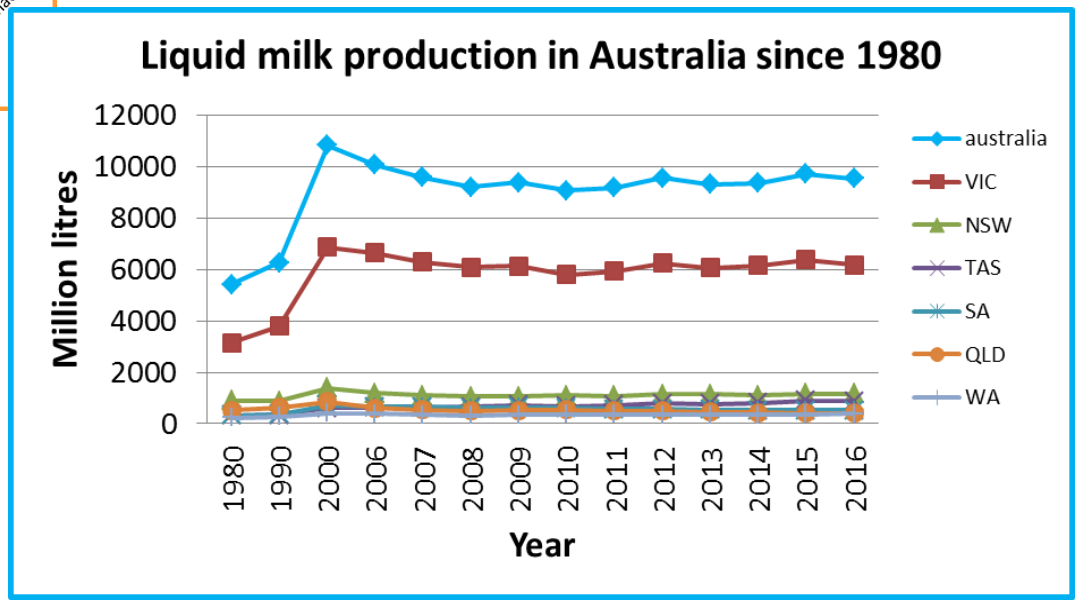
Milk production in Australia

Milk production is concentrated in the temperate zone of Australia. It is strongly seasonal in the key south-eastern dairying regions, reflecting the predominantly pasture-based nature of the industry.





Dairy Australia, 2016



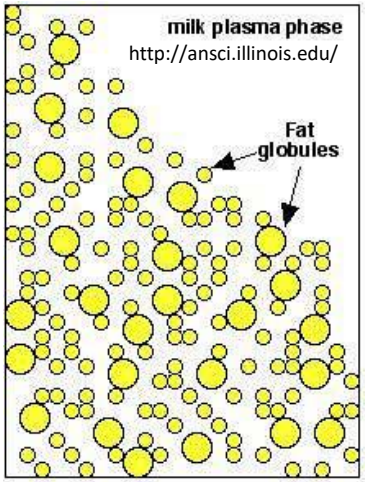
MILK PROTEOMICS

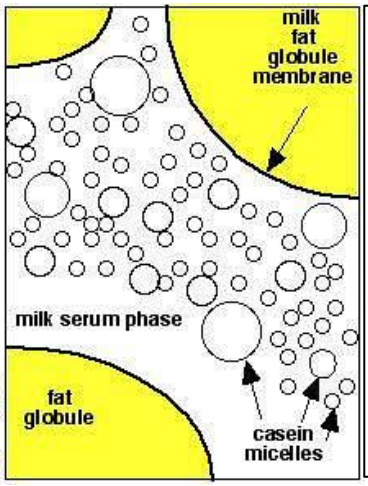
Milk composition

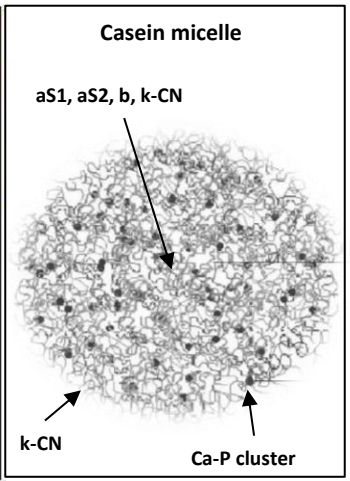
Milk: emulsion of fat globules, a suspension of casein micelles (casein, calcium, phosphorous), suspended in an aqueous phase, with solubilized lactose, whey proteins, and some minerals.

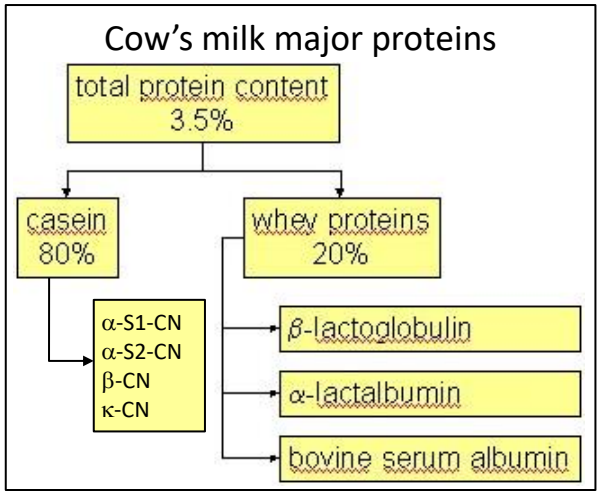
Nutrition Facts
Milk, whole, 3.25% fat

Amount Per 100 grams		
Calories 61		
Total Fat 3.2 g		
Saturated fat 1.9 g		
Polyunsaturated fat 0.2 g		
Monounsaturated fat 0.8 g		
Cholesterol 10 mg		
Sodium 43 mg		
Potassium 132 mg		
Total Carbohydrate 4.8 g		
Dietary fiber 0 g		
Sugar 5 g		
Protein 3.2 g		
Vitamin A	3%	Vitamin C
Calcium	11%	Iron
Vitamin D	12%	Vitamin B-6
Vitamin B-12	6%	Magnesium





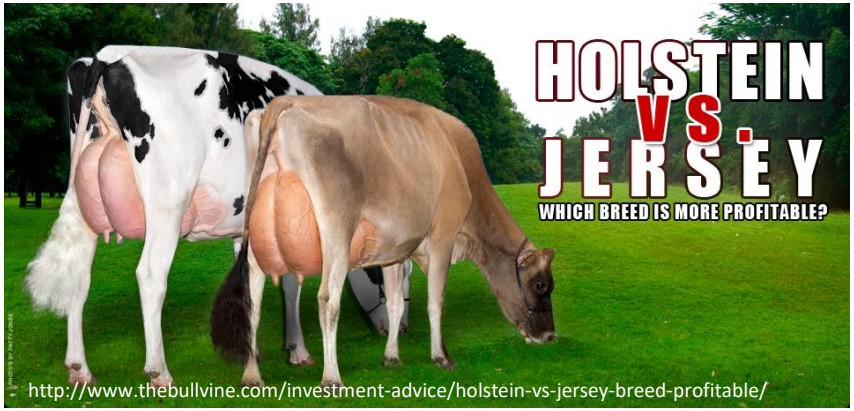




Protein	Concentration (g/L)
aS1-CN	10.0
b-CN	9.3
k-CN	3.3
b-LG	3.2
aS2-CN	2.6
a-LA	1.2
g-CN	0.8
Ig	0.8
PP/8F/8S	0.5
BSA	0.4
MFGM	0.4
PP3	0.3
Lactoferrin	0.1
Transferrin	0.1
Total	33.0

MILK PROTEOMICS

Example 1 : Cow breeds



http://www.thebullvine.com/investment-advice/holstein-vs-jersey-breed-profitable/

Study across the USA bovine herds

US dairy herd: 90% Holstein + 5% Jersey cattle (USDA, 2007)

Performance characteristic	Holstein	Jersey
Daily milk yield (kg)	29.1	20.9
Milkfat (%)	3.8	4.8
Milk protein (%)	3.1	3.7
Cheese yield (kg/kg of milk)	0.101	0.125
Calving interval (mo)	14.1	13.7
Dry period length (d)	60	60
Annual turnover (%)	34.5	30.0
Expected number of lactations	2.54	3.00
Age at first calving (mo)	26.1	25.3

Capper & Cady. J Dairy Sci. 2012, 95(1):165-76.

Australian dairy herd

Aussie dairy herd: 73% Holstein + 14% Jersey cattle

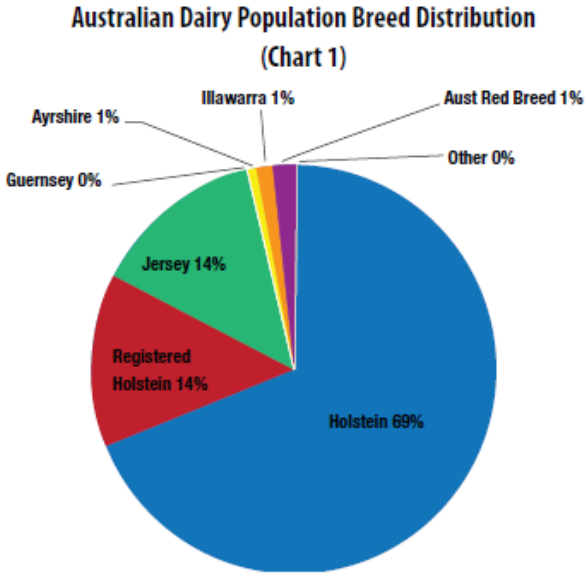


Table 1. Production Averages by Breed

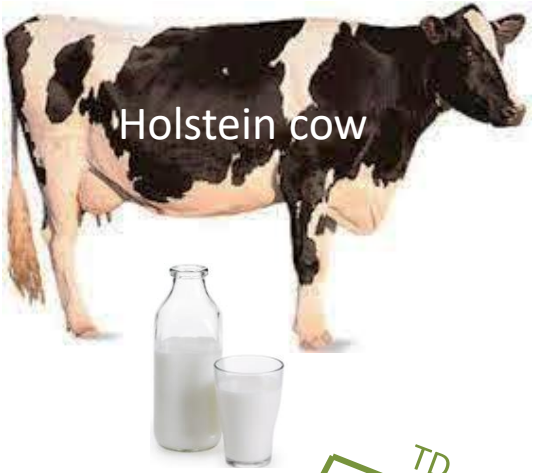
Breed	Cow Numbers	Milk Yield (L)	Fat %	Fat Yield (Kg)	Protein %	Protein Yield (Kg)	Lactation Length
Holstein	381,337	7080	3.94	279	3.29	233	322
Registered Holsteins	63,169	8005	3.84	308	3.22	258	346
Jersey	63,235	5123	4.88	250	3.74	192	306

The Australian Holstein Journal- April-May 2010

MILK PROTEOMICS

Experimental design

Bottom-up (BU):
Identifying as many milk proteins, major and minor, as possible across many samples.



Top-down (TD):
Quantifying major milk protein variants and PTMs across many samples.

Dionex nLC
Acclaim PepMap100, 75 µm x 15 cm, C18 2 µm 100 Å (Thermo)

(Vincent et al., 2015)

BU
TD
BU
TD

Aim of the study:
Compare milk proteomes across the two main bovine breeds in Australia.

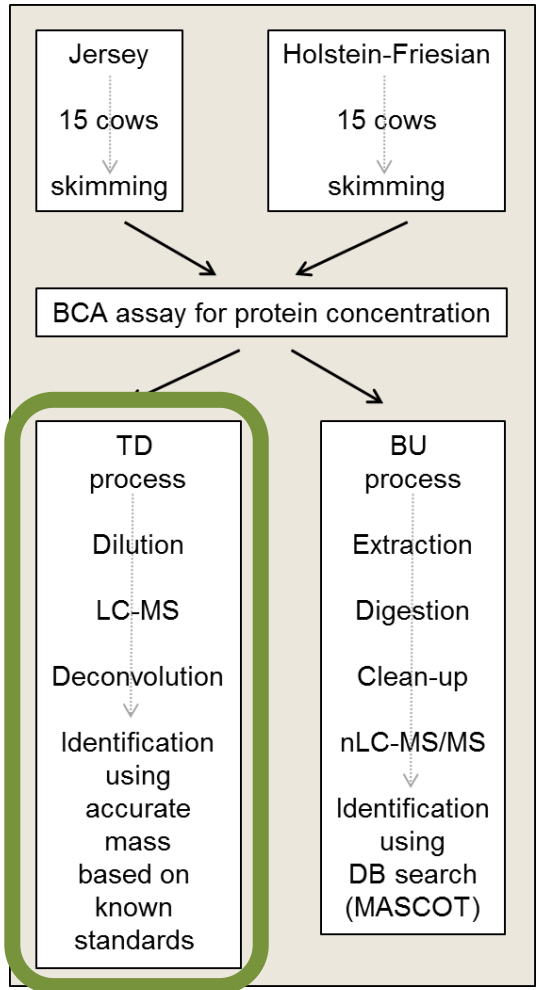
Agilent 1290 HPLC
Aeris™ XB-C8, 150 x 2.1mm, 3.6µm (Phenomenex)

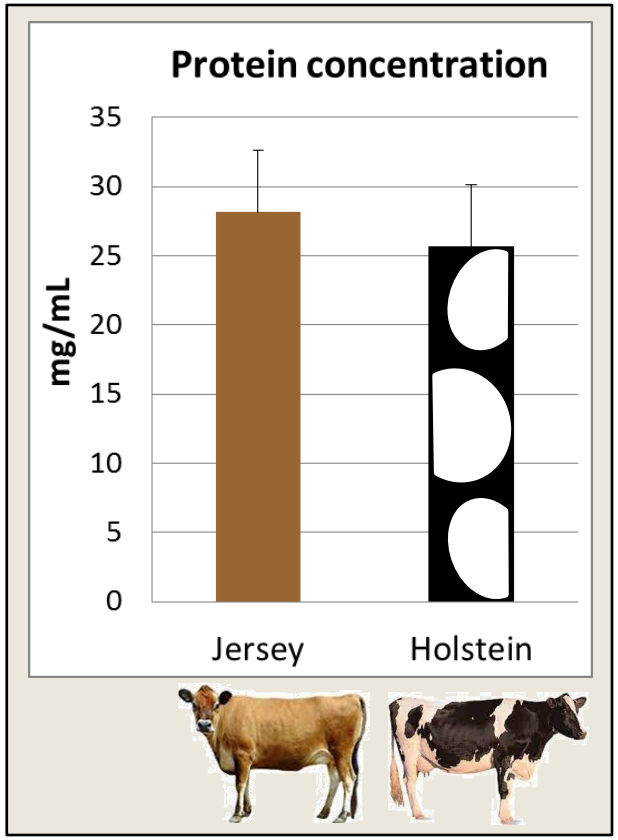
(Vincent et al., 2016)

MILK PROTEOMICS

Protein concentrations

Workflow





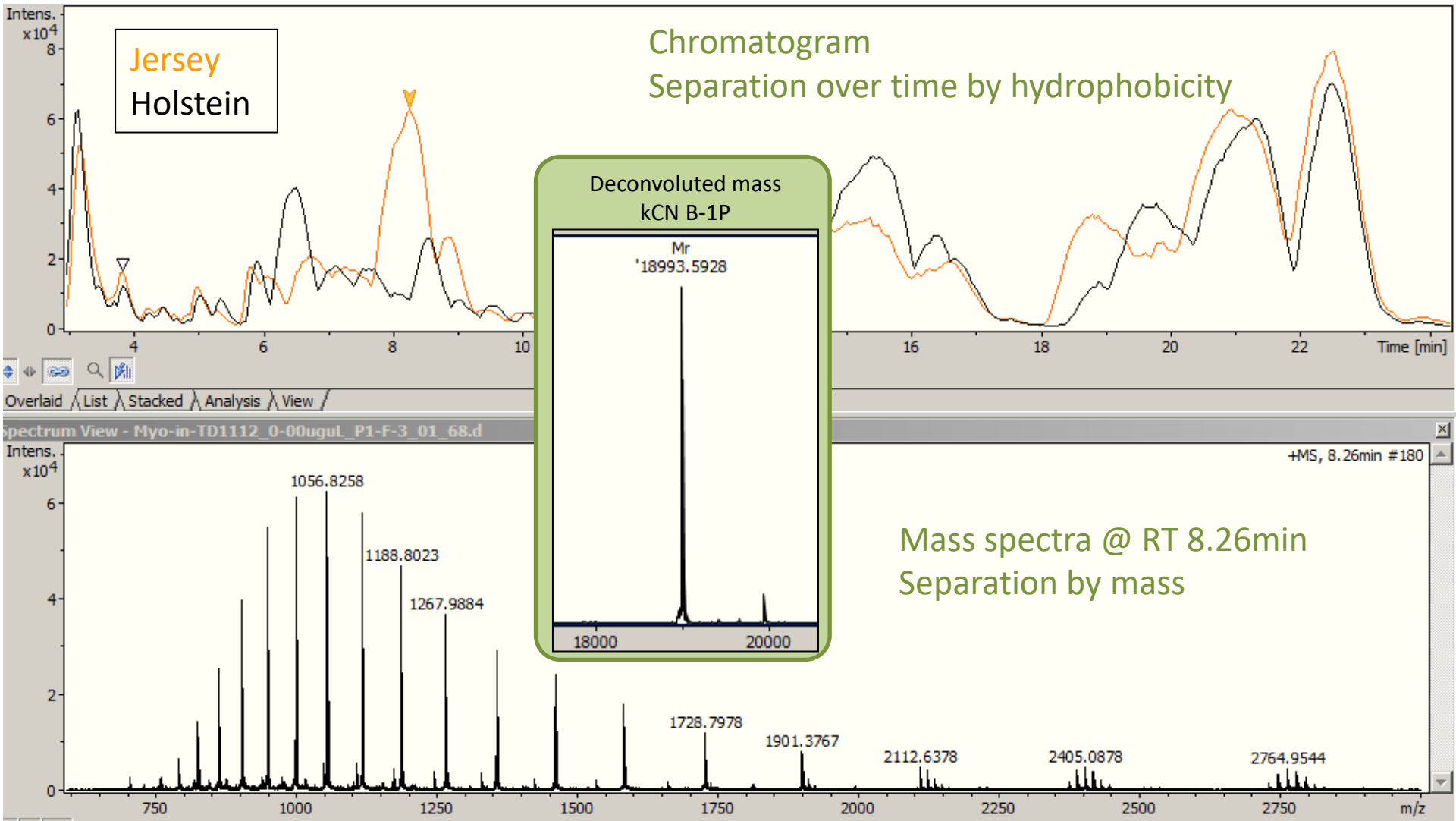
Protein concentration was on average higher in skim milk samples from Jersey cows than those from Holstein-Friesian cows.

Protein concentrations varied greatly among cows as demonstrated with the large standard deviation.

Top-down (intact proteins)

MILK PROTEOMICS

Top-down analysis – Charge state deconvolution



MILK PROTEOMICS

Top-down analysis – Holstein milk

Genedata display
Holstein

MILK PROTEOMICS

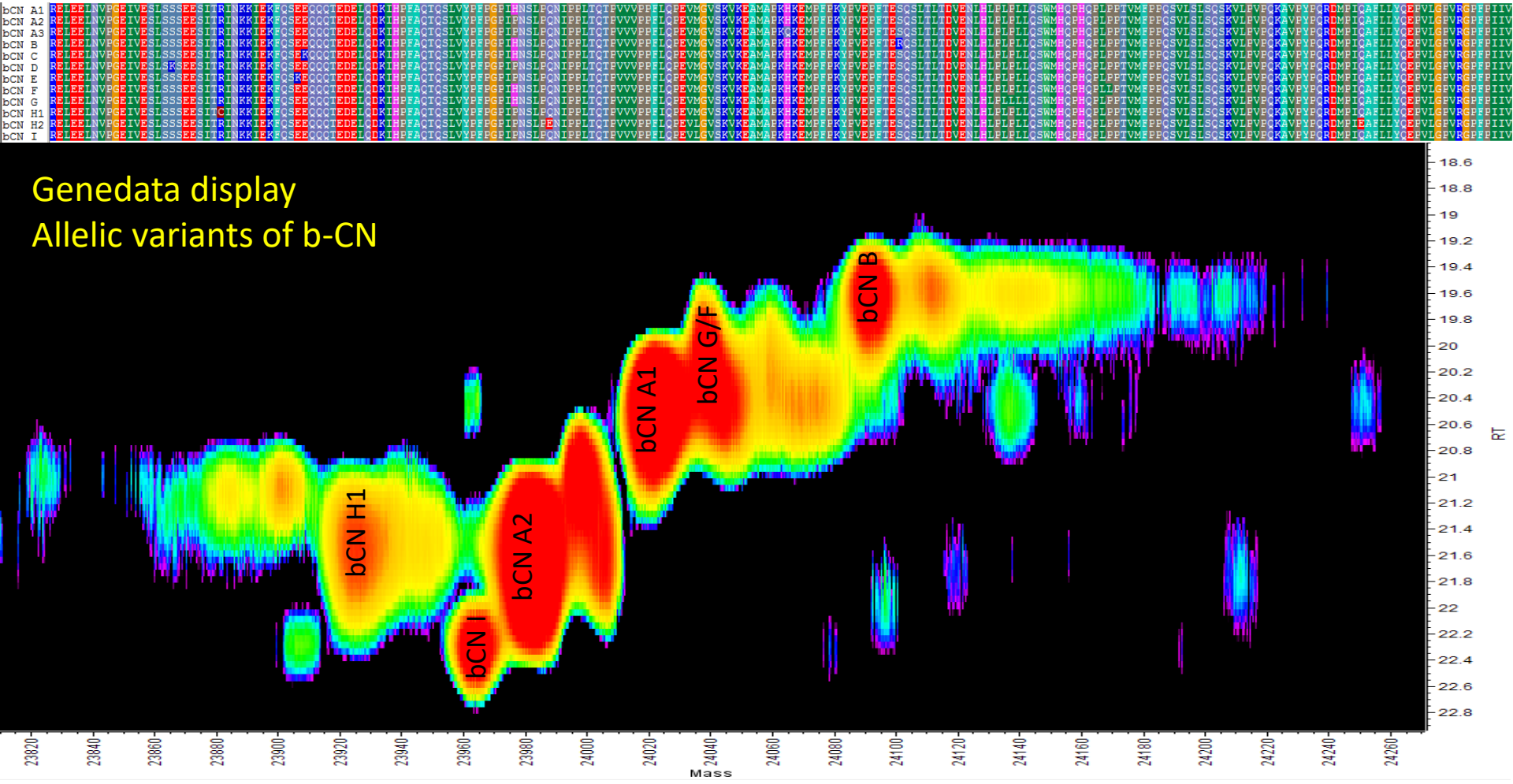
Top-down analysis – Jersey milk

Genedata display
Jersey

MILK PROTEOMICS

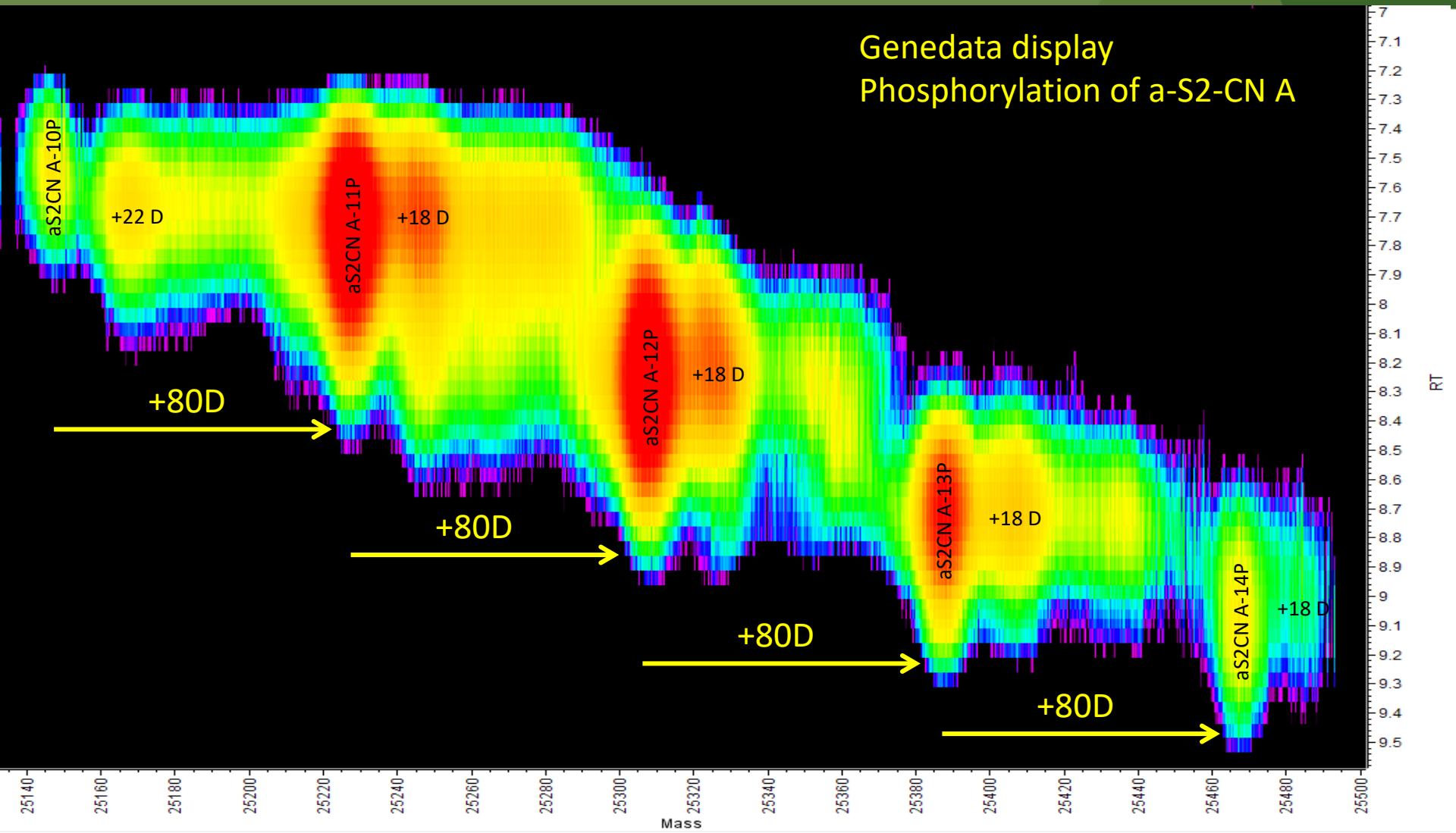
Allelic variation

There are 12 known allelic variants of bovine beta-caseins, differing by 1 or 2 AA at a time (out of 209).



MILK PROTEOMICS

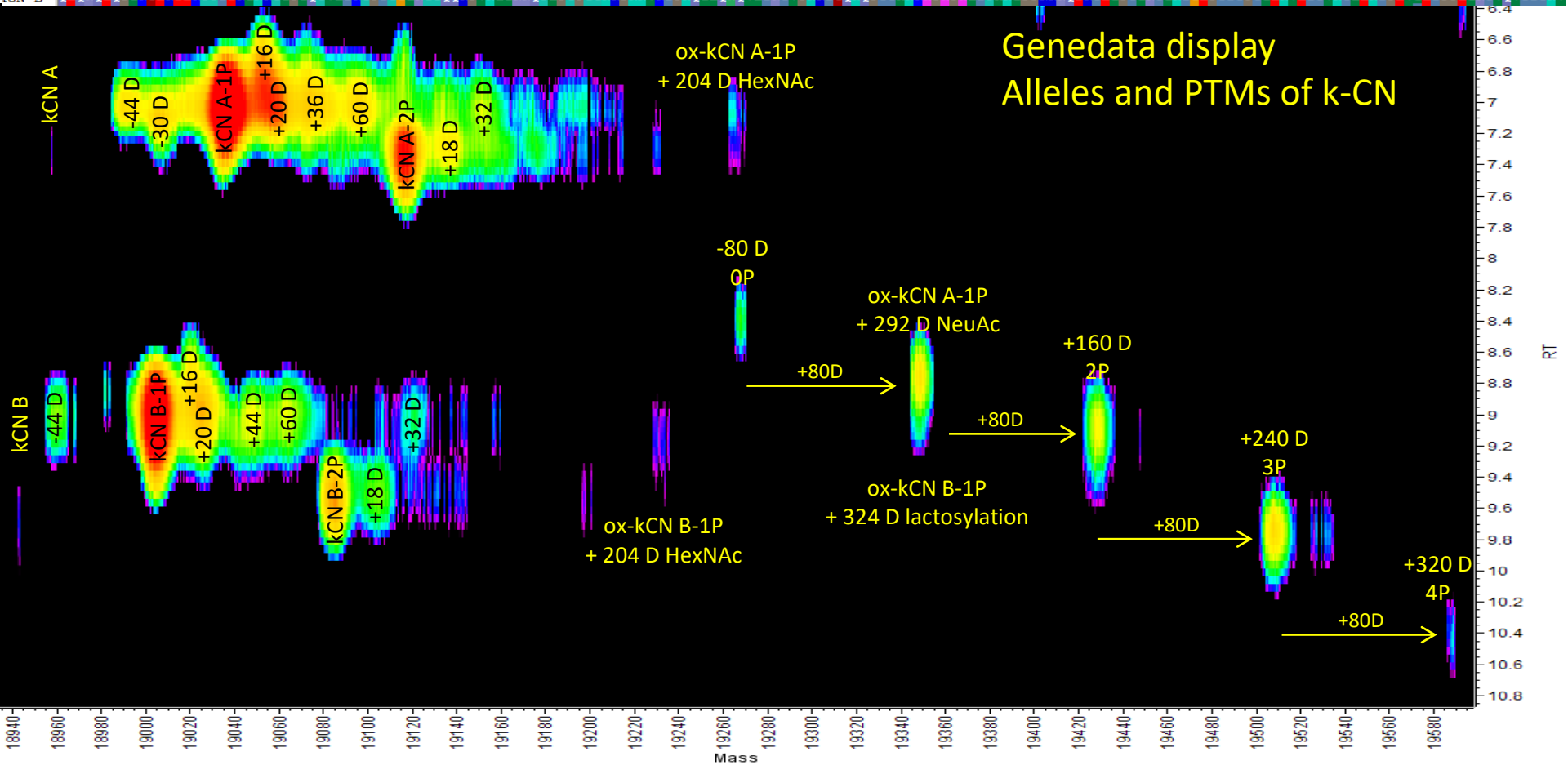
PTM - Phosphorylation



MILK PROTEOMICS

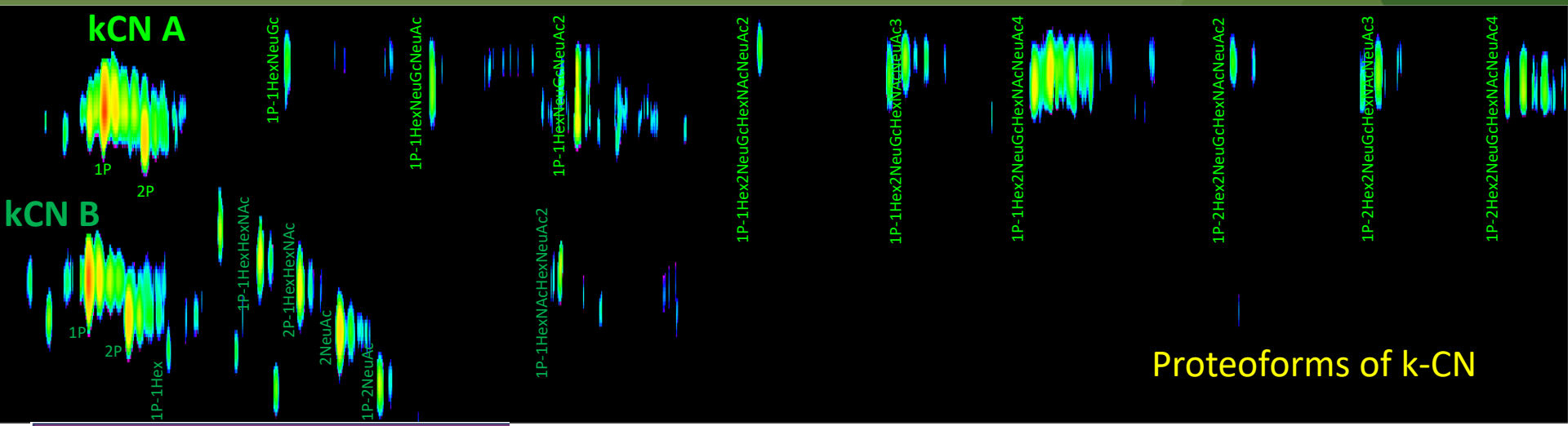
PTM – Phosphorylation and glycosylation

There are 12 known allelic variants of kappa-caseins, the most prominent ones being variants A and B. Kappa-proteins are phosphorylated (1-3P) and glycosylated.



MILK PROTEOMICS

PTM – Phosphorylation and glycosylation



Common Monosaccharides

Symbols	Monosaccharide	Abbreviation	Residue Mass
	N-Acetylgalactosamine	GalNAc	203.0794
	N-Acetylglucosamine	GlcNAc	203.0794
	Mannose	Man	162.0528
	Glucose (Hex)	Glc	162.0528
	Galactose	Gal	162.0528
	Fucose	Fuc	146.0579
	Xylose	Xyl	132.0423
	N-Acetylneuraminic Acid	NeuAc	291.0954
	N-Glycolylneuraminic Acid	NeuGc	307.0903
	Glucuronic acid	GlcA	176.0321
	Iduronic acid	IdoA	176.0321

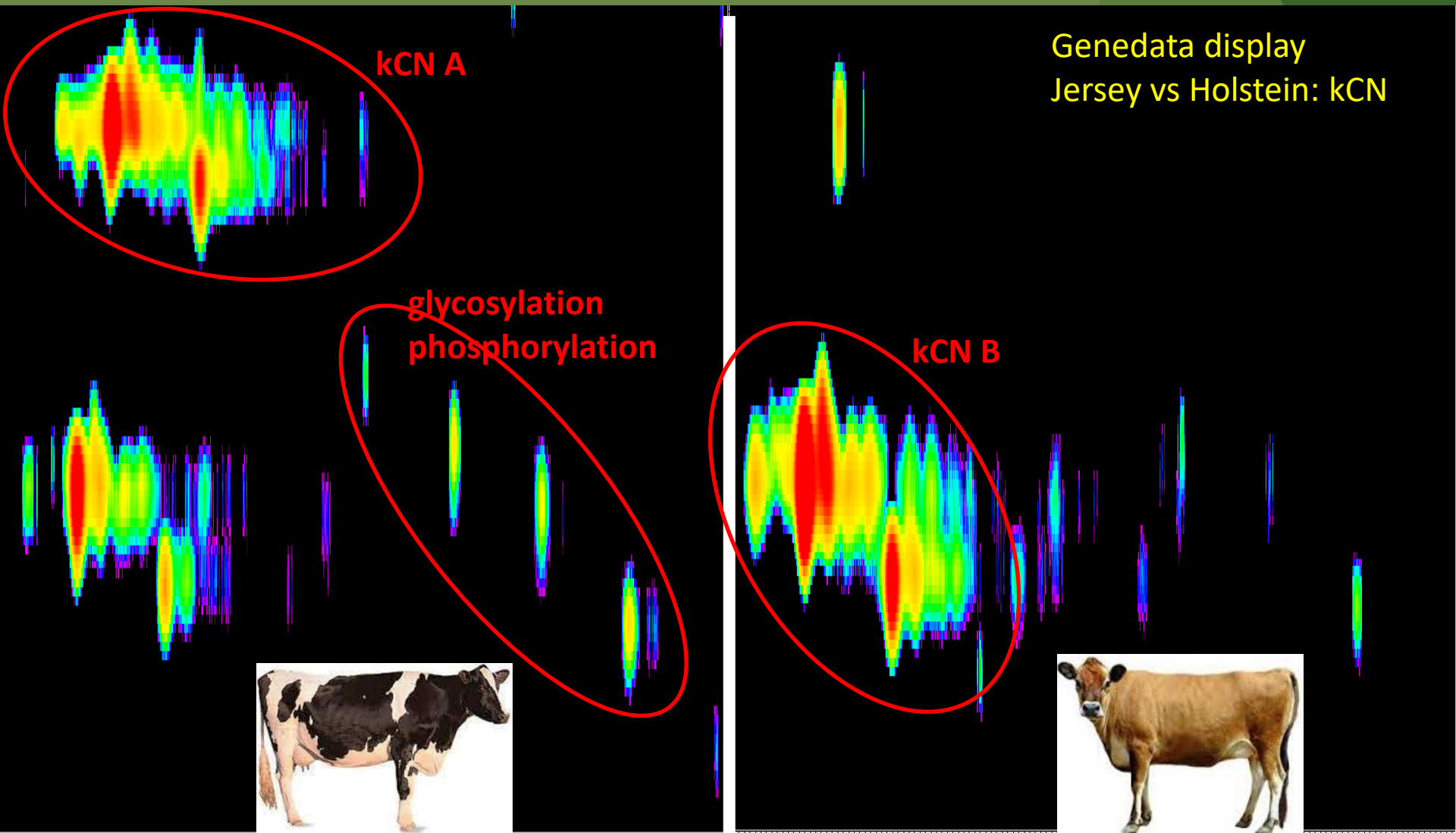
http://www.functionalglycomics.org/static/consortium/Nomenclature.shtml

Glycans suspected to associate with kCN in cow milk sample:

1P-1HexHexNAc = 80+162+203 = 445 Da
2P-1HexHexNAc = (2x80)+162+203 = 525 Da
2NeuAc = 2x291 = 582 Da
1P-2NeuAc = 80+(2x291) = 662 Da
1P-1HexNeuGc = 80+162+307 = 549 Da
1P-1HexNeuGcNeuAc = 80+162+307+291 = 840 Da
1P-1HexNeuGcNeuAc2 = 80+162+307+(2x291) = 1131 Da
1P-1Hex2NeuGcHexNAcNeuAc2 = 80+162+(2x307)+162+203+(2x291) = 1803 Da
1P-1Hex2NeuGcHexNAcNeuAc3 = 80+162+(2x307)+162+203+(3x291) = 2094 Da
1P-1Hex2NeuGcHexNAcNeuAc4 = 80+162+(2x307)+162+203+(4x291) = 2385 Da
1P-2Hex2NeuGcHexNAcNeuAc2 = 80+(2x162)+(2x307)+162+203+(2x291) = 1965 Da
1P-2Hex2NeuGcHexNAcNeuAc3 = 80+(2x162)+(2x307)+162+203+(3x291) = 2256 Da
1P-2Hex2NeuGcHexNAcNeuAc4 = 80+(2x162)+(2x307)+162+203+(4x291) = 2547 Da

MILK PROTEOMICS

Comparison Holstein vs. Jersey milk



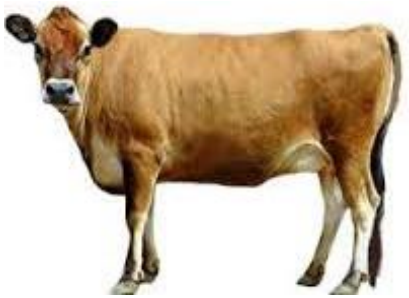
MILK PROTEOMICS

Top-Down results - Summary

Intact protein proteomics readily detects allelic variation and PTMs in a quantitative manner. Kappa caseins display the greatest difference across breeds. Gurses and coll. (2016) have shown B allele of kCN to be positively associated with increased protein and solids-no-fat contents in milk.

	Jersey milk			Holstein milk			T-test		
Protein	Average RT (min)	Average Response	CV Response (%)	Average RT (min)	Average Response	CV Response (%)	p-value Response	significance	**Fold-difference**
kCN-B-1P	8.50	2.7025	3.0301	8.73	0.9285	3.3471	0.0000	***	2.9
bCN-B-5P	19.22	4.7576	4.2614	19.36	2.0470	1.7351	0.0000	***	2.3
kCN-B-2P	9.12	0.1667	8.2451	8.96	0.0920	12.5636	0.0060	**	1.8
aS2CN-A-14P	8.31	0.5240	2.1923	8.66	0.3464	6.9380	0.0000	***	1.5
aS2CN-A-11P	7.32	0.9330	6.4905	7.43	0.6549	4.4493	0.0004	***	1.4
aLA-B-G	15.61	0.2639	13.5836	15.42	0.1970	6.2683	0.0236	*	1.3
aS2CN-A-13P	8.13	1.1416	2.4312	8.43	0.8545	5.0902	0.0001	***	1.3
aS2CN-A-12P	7.82	2.2673	1.9770	7.95	1.6973	4.0134	0.0000	***	1.3
aS1CN-B-9P	16.54	0.4291	6.2196	16.68	0.3363	3.3860	0.0015	**	1.3
bLG-A	22.87	5.3522	5.7464	22.98	4.2130	4.3259	0.0015	**	1.3
bLG-D	23.31	0.1020	5.9428	23.62	0.0847	10.1951	0.0353	*	1.2
aLA-B	16.83	3.1430	2.1924	17.00	2.6865	3.7495	0.0006	***	1.2
bCN-A2-5P	21.04	12.8962	2.8863	21.19	11.2693	5.2077	0.0067	**	1.1
BSA	15.44	0.9757	4.4888	15.62	0.9283	6.7009	0.3224	n.s.	1.1
aS1CN-B-8P	15.46	1.5917	3.9823	15.61	1.5511	7.5522	0.6165	n.s.	1.0
aS2CN-A-10P	7.12	0.5305	3.8224	6.92	0.5573	6.3023	0.2962	n.s.	1.0
bCN-I-5P	23.50	0.0932	2.4644	23.42	0.0987	4.2572	0.2000	n.s.	0.9
bLG-B	20.85	2.4886	3.4503	20.99	2.8132	4.8285	0.0128	*	0.9
bCN-A1-5P	20.12	5.1872	1.8887	20.16	6.9617	4.9996	0.0001	***	0.7
kCN-B-1P-G	7.30	0.0917	7.0158	7.48	0.1254	12.8383	0.7929	n.s.	0.7
kCN-A-2P	6.41	0.0999	8.1087	6.65	0.1424	4.8685	0.0005	***	0.7
kCN-A-1P	6.72	0.3971	3.3511	6.67	1.5740	5.4925	0.0000	***	0.3

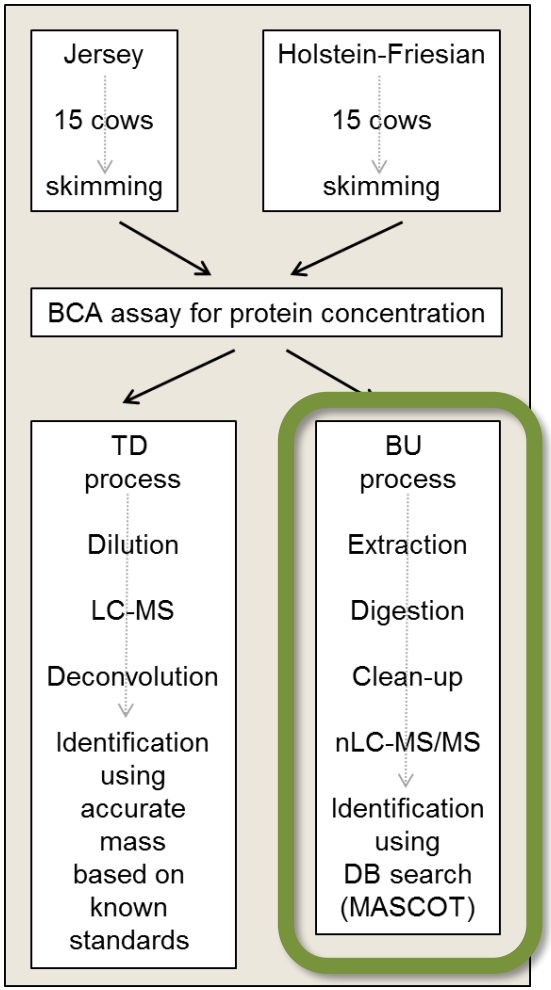
n.s. not significant, * p-value < 0.1, ** p-value <0.01, *** p-value < 0.001

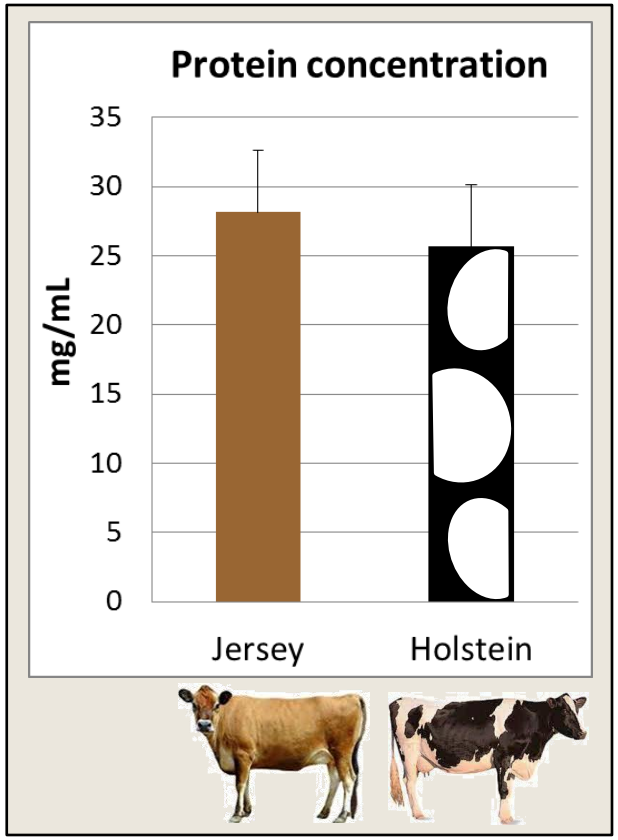


MILK PROTEOMICS

Protein concentrations

Workflow





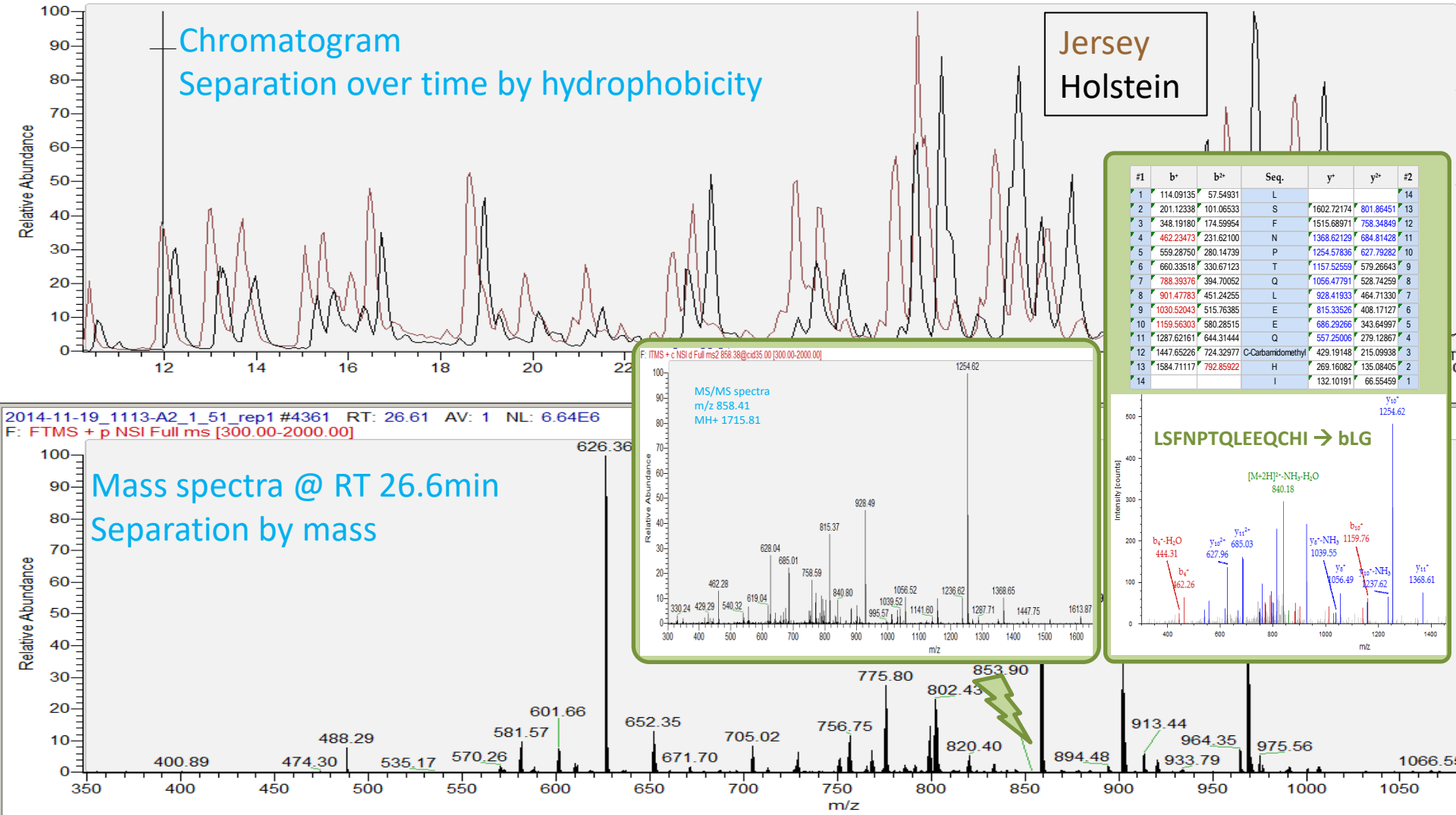
Protein concentration was on average higher in skim milk samples from Jersey cows than those from Holstein-Friesian cows.

Protein concentrations varied greatly among cows as demonstrated with the large standard deviation.

Bottom-up (trypsinised proteins)

MILK PROTEOMICS

Bottom-up analysis



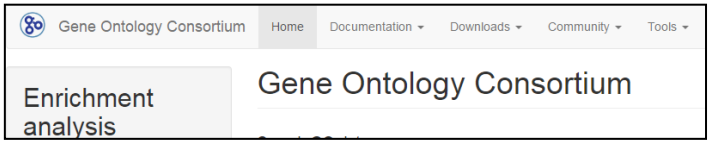
#1	b+	b2+	Seq.	y+	y2+	#2
1	114.09135	57.54931	L			14
2	201.12338	101.06533	S	1602.72174	801.86451	13
3	348.19180	174.59954	F	1515.68971	758.34849	12
4	462.23473	231.62100	N	1368.62129	684.81428	11
5	559.28750	280.14739	P	1254.57836	627.79282	10
6	660.33518	330.67123	T	1157.52559	579.26643	9
7	788.39376	394.70052	Q	1056.47791	528.74259	8
8	901.47783	451.24255	L	928.41933	464.71330	7
9	1030.52043	515.76385	E	815.33526	408.17127	6
10	1159.56303	580.28515	E	686.29266	343.64997	5
11	1287.62161	644.31444	Q	557.25006	279.12867	4
12	1447.65226	724.32977	C-Carbamidomethyl	429.19148	215.09938	3
13	1584.71117	792.85922	H	269.16082	135.08405	2
14			I	132.10191	66.55459	1

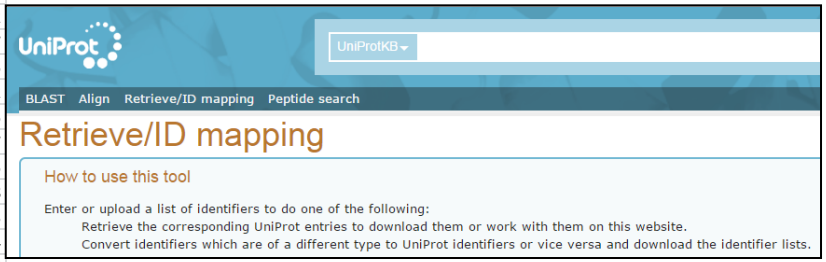
MILK PROTEOMICS

Database search

Accession	Description	ΣCoverage	Σ# Peptides
861514	pancreatic elastase inhibitor, EI=serpin {N-terminal, reactive-site loop} [Bos taurus]	66.67	1
299676	alpha-s1-casein homolog {N-terminal} [cattle, mammary glands, Peptide Partial, 17 aa]	64.71	1
1351907;IPI01028455.1	Serum albumin; AltName: Full=BSA; AltName: Allergen=Bos d 6; Flags: Precursor [Bos tau	59.64	35
251198	kappa-casein [cattle, milk, Peptide Partial, 20 aa]	55.00	1
343197026	immunoglobulin lambda light chain constant region 3 allotypic variant IGLC3c [Bos taur	52.83	5
IPI00843089.3	11 kDa protein	52.34	4
30794292;IPI00710664.1	lactotransferrin precursor [Bos taurus]	50.14	40
27805809;IPI00691946.2	fatty acid-binding protein, heart [Bos taurus]	47.37	7
229416	casein para kappaA	46.67	4
528953246	alpha-S1-casein isoform X7 [Bos taurus]	46.33	7
310893435	immunoglobulin light chain [Bos taurus]	45.54	3
27806963;IPI00698843.1	alpha-S2-casein precursor [Bos taurus]	44.14	17
223165	lactoglobulin beta	43.83	10
2194088	Chain A, Bovine Beta-Lactoglobulin, Lattice X	43.83	10
49259423	Chain X, The Cys121ser Mutant Of Beta-lactoglobulin	43.83	10
27807339;IPI00716366.1	glycosylation-dependent cell adhesion molecule 1 precursor [Bos taurus]		
162807	kappa-casein precursor, partial [Bos taurus]		
2323400	immunoglobulin light chain variable region [Bos taurus]		
2501351;IPI00690534.1	Serotransferrin; Short=Transferrin; AltName: Full=Beta-1 metal-binding globulin; AltNan		
528953238	alpha-S1-casein isoform X3 [Bos taurus]	39.81	7
87196497;IPI00699698.1	beta-lactoglobulin precursor [Bos taurus]	39.33	12
528953236	alpha-S1-casein isoform X2 [Bos taurus]	38.83	7
315143016	kappa casein [Bos indicus]	36.81	5
528958213;IPI01017444.1	perilipin-2 isoform X2 [Bos taurus]	34.23	12
8099324	kappa-casein precursor, partial [Bos taurus x Bos indicus]	33.75	5
IPI00712994.3	Uncharacterized protein	33.54	7
428755219	k-casein, partial [Bison bonasus]	33.33	5
54037712	Beta-lactoglobulin; Short=Beta-LG	33.33	8
2494285;IPI00689035.1	Lactadherin; AltName: Full=BP47; AltName: Full=Components 15/16; AltName: Full=MFG	33.26	13
27806881;IPI00711862.1	epididymal secretory protein E1 precursor [Bos taurus]	31.54	4
124056491;IPI00713505.2	Complement C3; Contains: Complement C3 beta chain; Contains: Complement C3 alpha	31.19	46
27805979;IPI00717424.1	alpha-lactalbumin precursor [Bos taurus]	30.99	9
119393699	alpha lactabumin [Bos taurus]	30.77	1
528912092;IPI00685784.3	neutrophil gelatinase-associated lipocalin isoform X3 [Bos taurus]	30.50	4
343197004	immunoglobulin lambda light chain constant region 2 allotypic variant IGLC2b [Bos taur	30.19	3
562890035	beta lactoglobulin, partial [Bos indicus]	30.00	1
528953244	alpha-S1-casein isoform X6 [Bos taurus]	29.15	4
3914346;IPI00696714.1	Polymeric immunoglobulin receptor; Short=PIgR; Short=Poly-Ig receptor; Contains: Secre	28.67	23
3183510;IPI00708535.1	Butyrophilin subfamily 1 member A1; Short=BT; Flags: Precursor	28.14	14
94966811;IPI00691212.1	alpha-1-acid glycoprotein precursor [Bos taurus]	27.23	5
75832056;IPI00715548.1	apolipoprotein A-I preproprotein [Bos taurus]	27.17	6
528952550;IPI00691887.2	osteopontin isoform X1 [Bos taurus]	26.62	7

Ultimately, a list of proteins is produced. It's up to the end-user to make sense of it.
Some in silico tools exist but the literature is the best source of information.





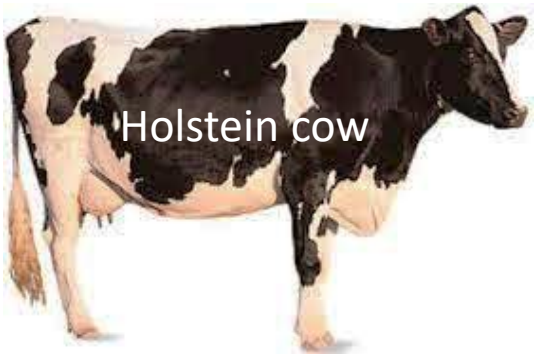
MILK PROTEOMICS

Bottom-Up results - Summary

Description	Holstein	Jersey
fatty acid-binding protein	30%	
fibrinogen alpha chain	28%	
alpha-2-HS-glycoprotein	28%	
alpha-1B-glycoprotein	27%	
lactadherin	25%	
epididymal secretory protein E1	24%	
beta-1,4-galactosyltransferase 1	24%	
peptidoglycan recognition protein 1	24%	
fibroblast growth factor-binding prote	24%	
protein HP-25 homolog 2	23%	
kininogen-2	22%	
Fab Pgt123 Hiv-1 Neutralizing Antibod	21%	
antibody Blv5b8	20%	
uncharacterized protein	19%	
IgG2a heavy chain constant region	19%	
serpin A3-1	16%	
histatherin	16%	
vitamin D-binding protein	15%	
folate receptor alpha	15%	
kininogen-1	14%	
fibrinogen beta chain	12%	
angiogenin-1	12%	
alpha-S1-casein	11%	
apolipoprotein A-II	11%	
cathelicidin-4	11%	
alpha-1-antiproteinase	11%	
uncharacterized protein		10%
Ig lambda-1 variable region		10%
Ig anti-HIV-1		12%
beta-lactoglobulin-like		12%
mammary serum amyloid A3.2		13%
CD5L protein		23%
actin 1	?	26%

forms soluble complexes with Ca and $PO4^{3-}$, Holstein milk contains less total Ca than Jersey milk

facilitates the transfer of FAs between extra- and intracellular membranes. Jersey milk fat containing higher concentrations of saturated FAs, especially of FAs with short and medium carbon chains.



These breeds have evolved different milk qualities and immune responses.

The searched DB doesn't inform on the variants.

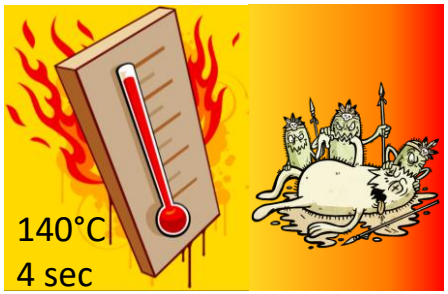
BU approach allows the detection of minor proteins and greatly increase proteome coverage.

Highly complementary to TD analysis.

MILK PROTEOMICS

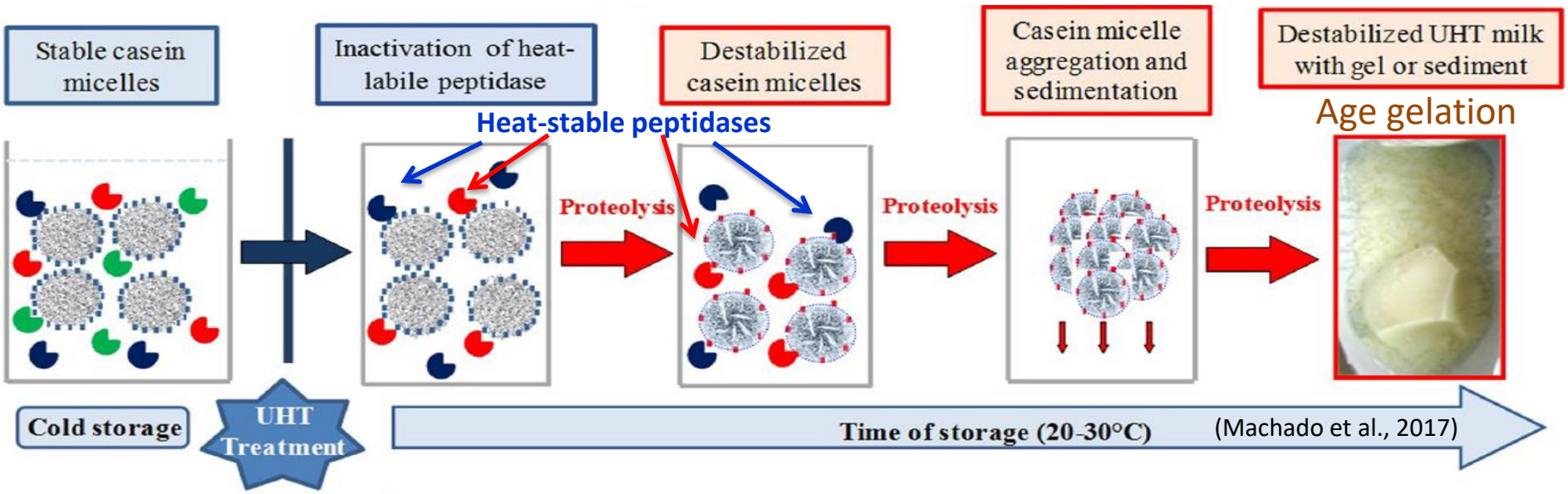
Example 2 : UHT milk

Milk manufacture → collection → UHT treatment → Storage < 6 months



Should be like this

But can be like this!



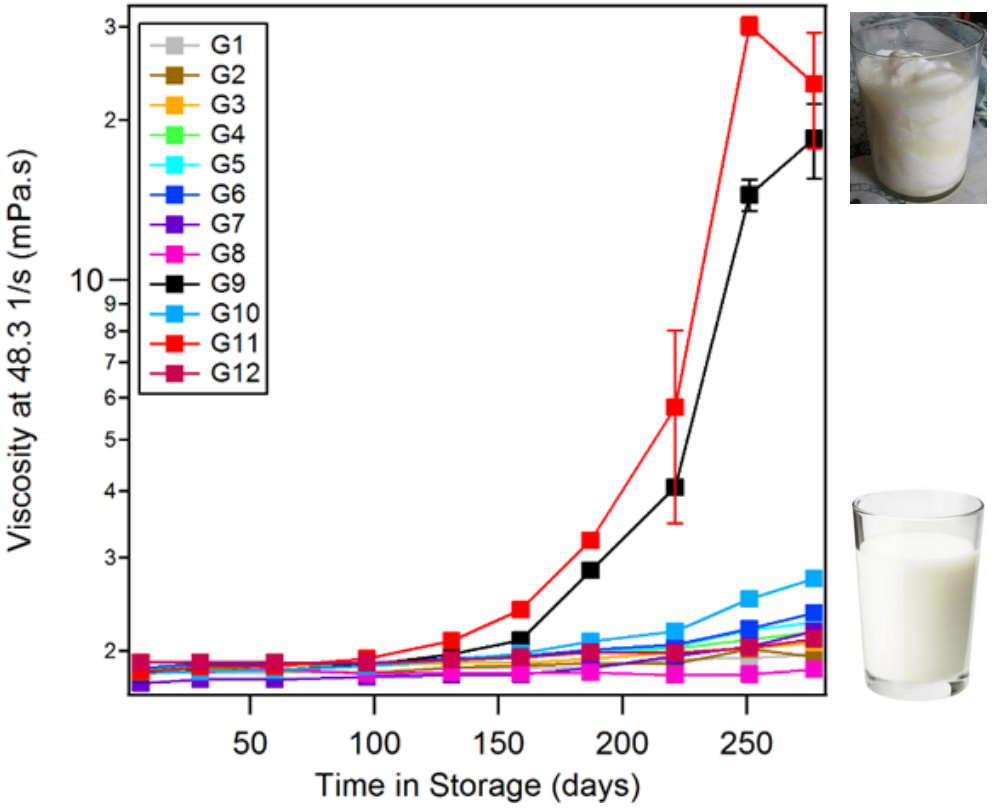
(Machado et al., 2017)

MILK PROTEOMICS

Experimental design

Cows genotyped based on their major protein variants and grouped accordingly. The 12 groups of milk were UHT-treated and stored for 9 months at room temperature. Physical, metagenomics and top-down proteomics analyses were performed.

Group	k-CN	b-CN	b-LG
G1	**AA**	A1A1	AB
G2		A1A1	BB
G3		A1A2	AB
G4		A1A2	BB
G5		A2A2	AB
G6		A2A2	BB
G7	**AB**	A1A1	AB
G8		A1A1	BB
G9		A1A2	AB
G10		A1A2	BB
G11		A2A2	AB
G12		A2A2	BB

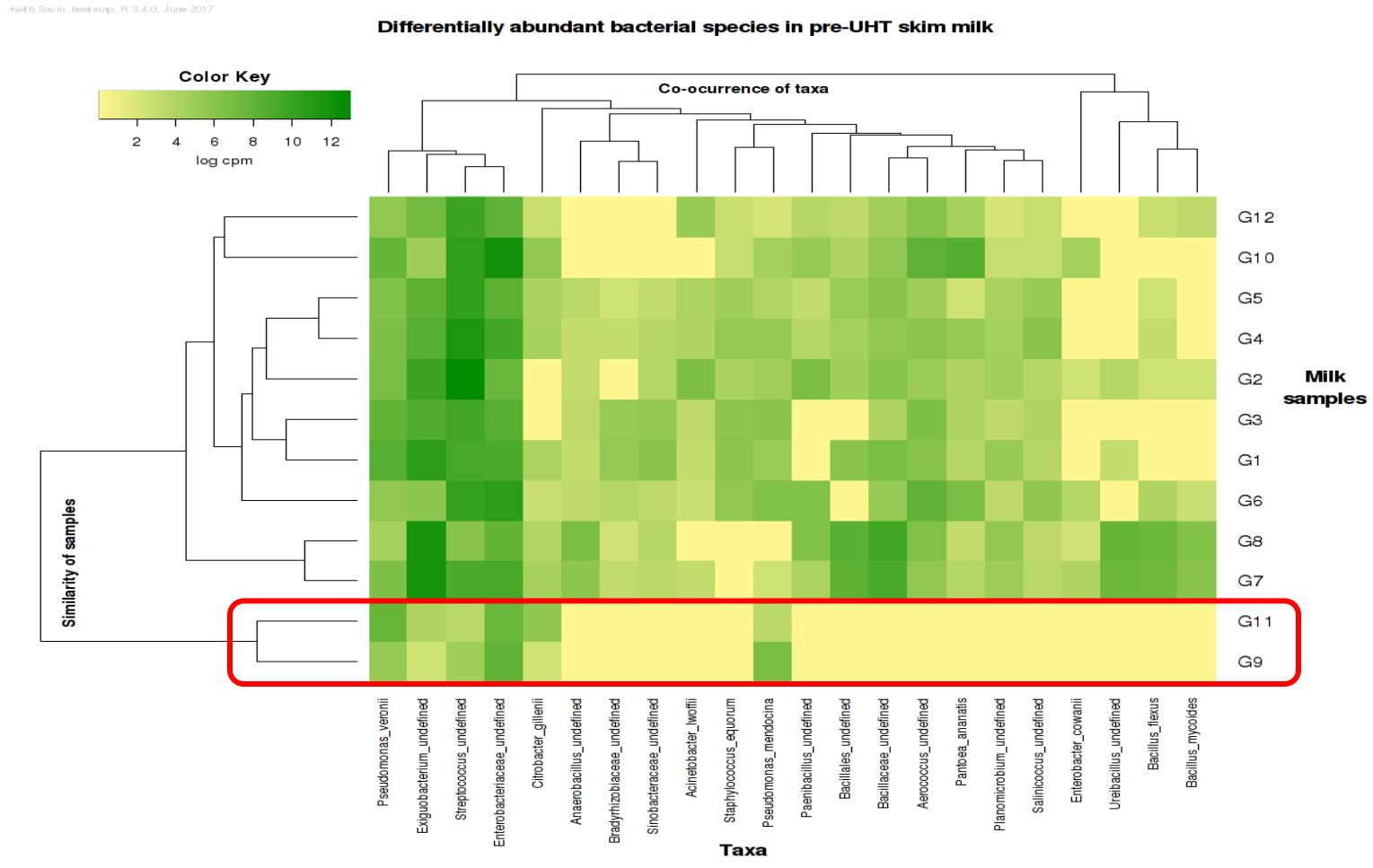


Milk viscosity in group 9 (kCN AB bCN A1A2 bLG AB) and group 11 (kCN AB bCN A2A2 bLG AB) increases from 6 months onwards.
They are the only groups displaying age gelation.

Economic Development, Jobs, Transport and Resources

MILK PROTEOMICS

Metagenomics analysis

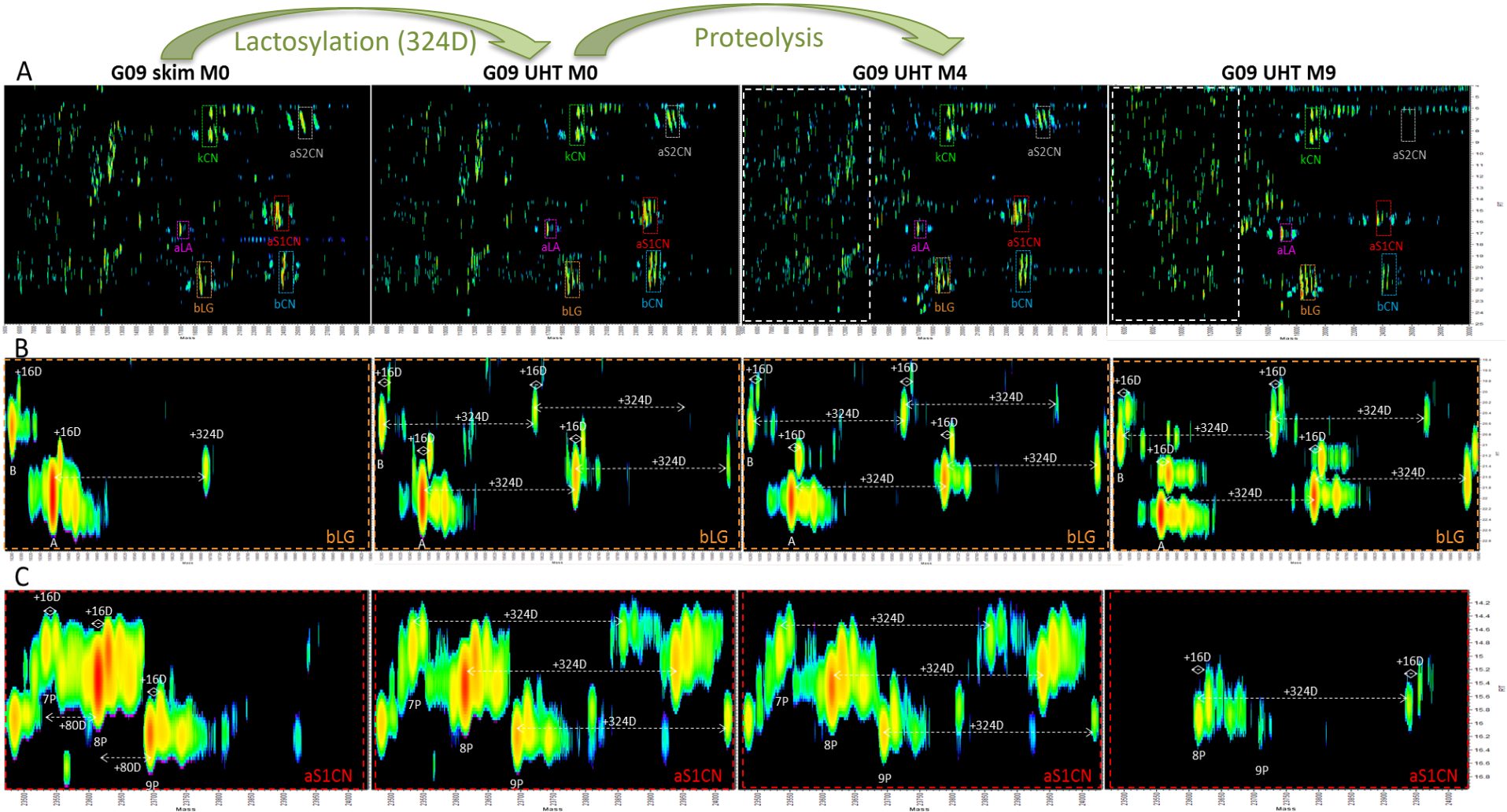


Groups 9 and 11 contain very similar types of bacteria and differ from the other 10 groups by lacking Pantoea ananatis, a Gammaproteobacterium and 4 members of the *Bacillus* genus (Keith Savin).

MILK PROTEOMICS

Top-down protein analysis – LC/MS maps

Maps of deconvoluted proteins in skim milk and UHT milk over time.



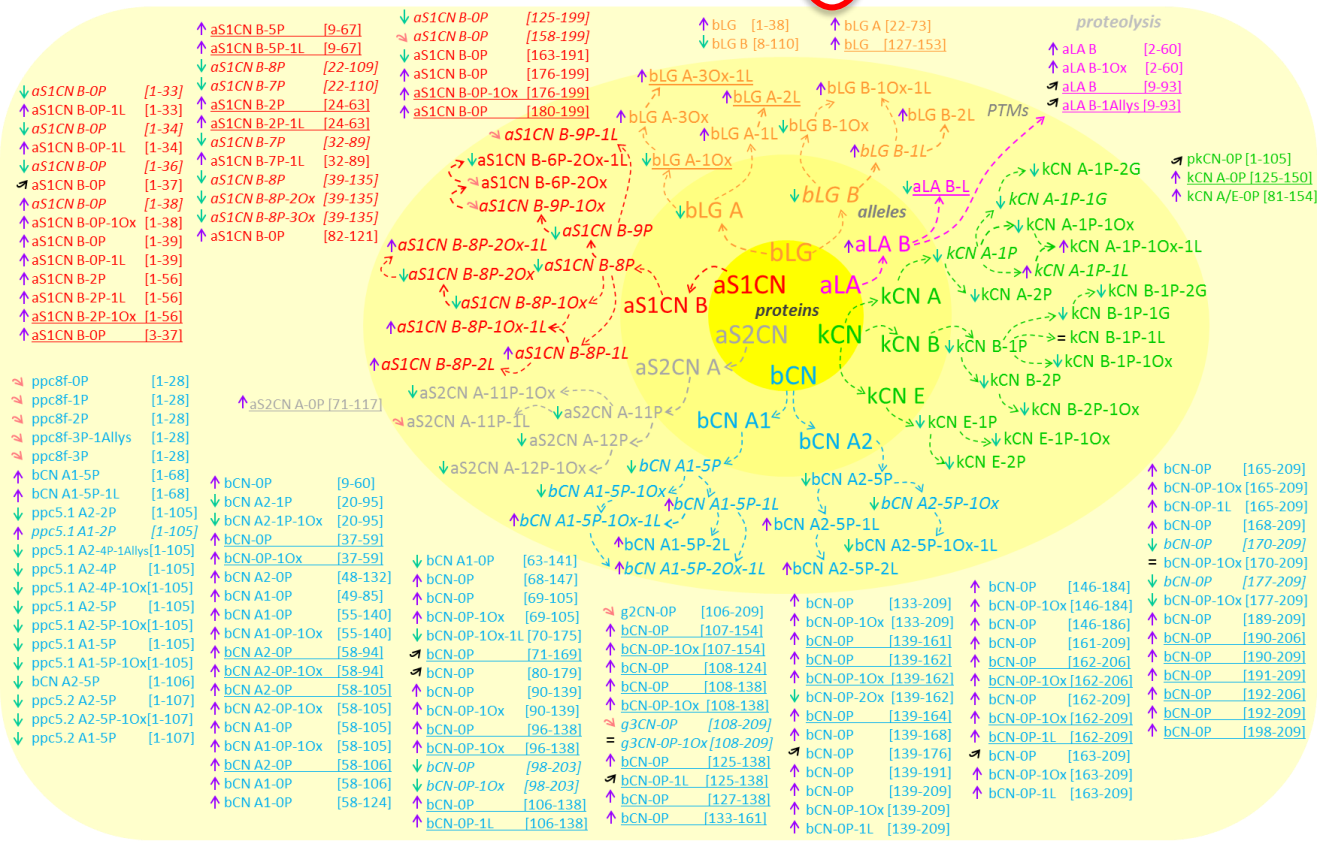
MILK PROTEOMICS

Intact protein analysis - Proteoforms

Summary of all the proteoforms identified in UHT milk:

209 protein compounds =
58 intact proteoforms
+ 151 degradation products

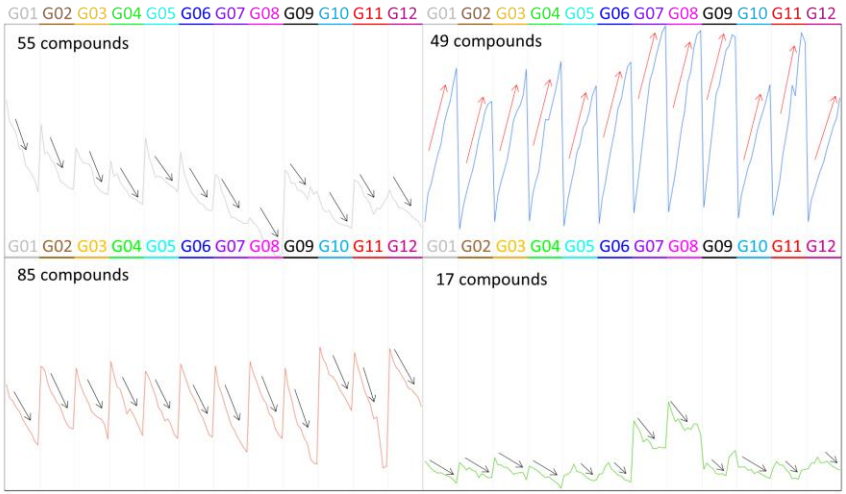
Protein type	Main proteoform	Allysine	Glycosylation	Lactosylation	Lactosylation + Oxidation	Lysis	Oxidation	Oxidation + Lysis	Phosphorylation	TOTAL
aLA	1	1	0	1	0	2	0	1	0	6
aS1CN	1	0	0	10	3	20	4	6	1	44
aS2CN	1	0	0	1	0	1	2	0	1	5
bCN	2	2	0	11	3	67	2	30	0	117
bLG	2	0	0	4	2	4	3	0	0	15
kCN	3	0	4	2	1	3	4	0	3	17

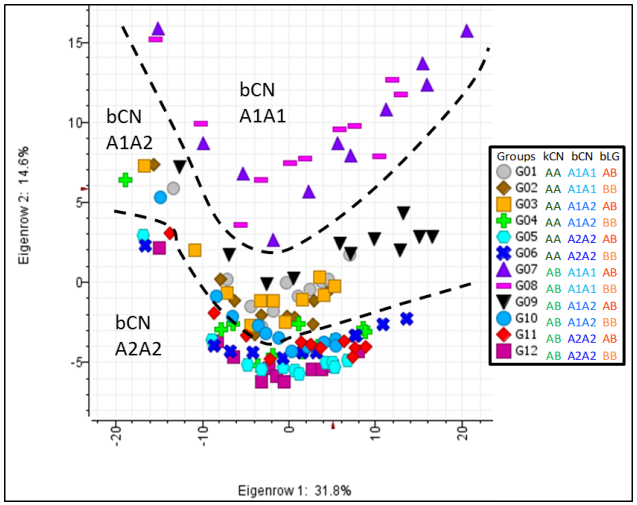


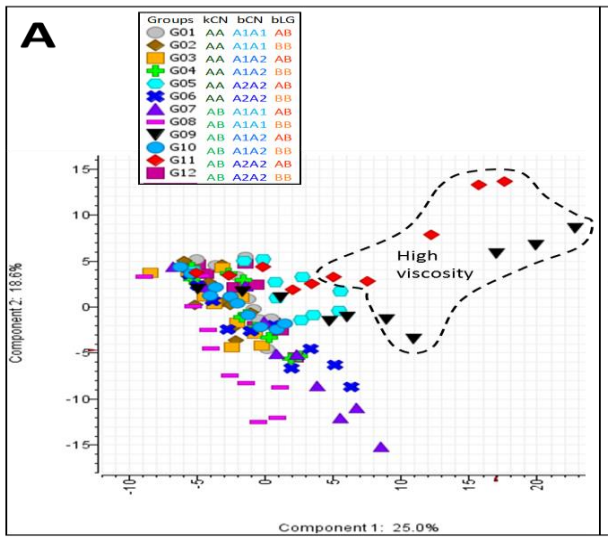
MILK PROTEOMICS

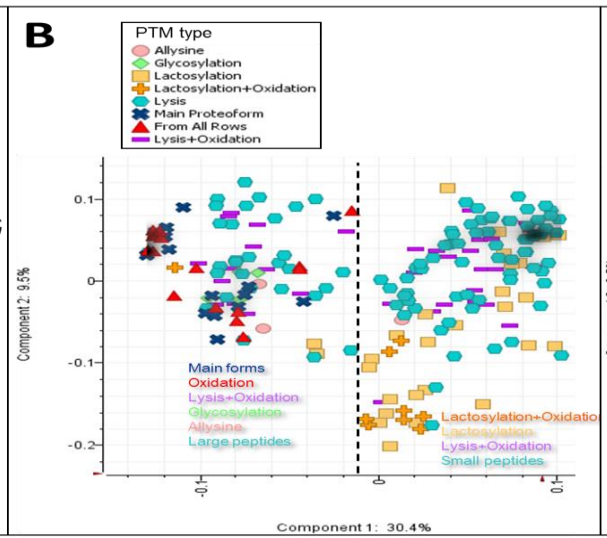
Intact protein analysis - Statistics

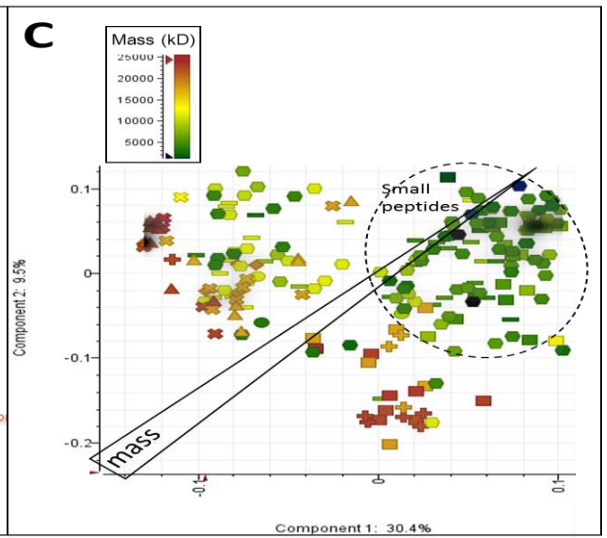
2 main profiles: decrease of intact proteins and increase of degraded proteins (SOM, PCA, PLS).









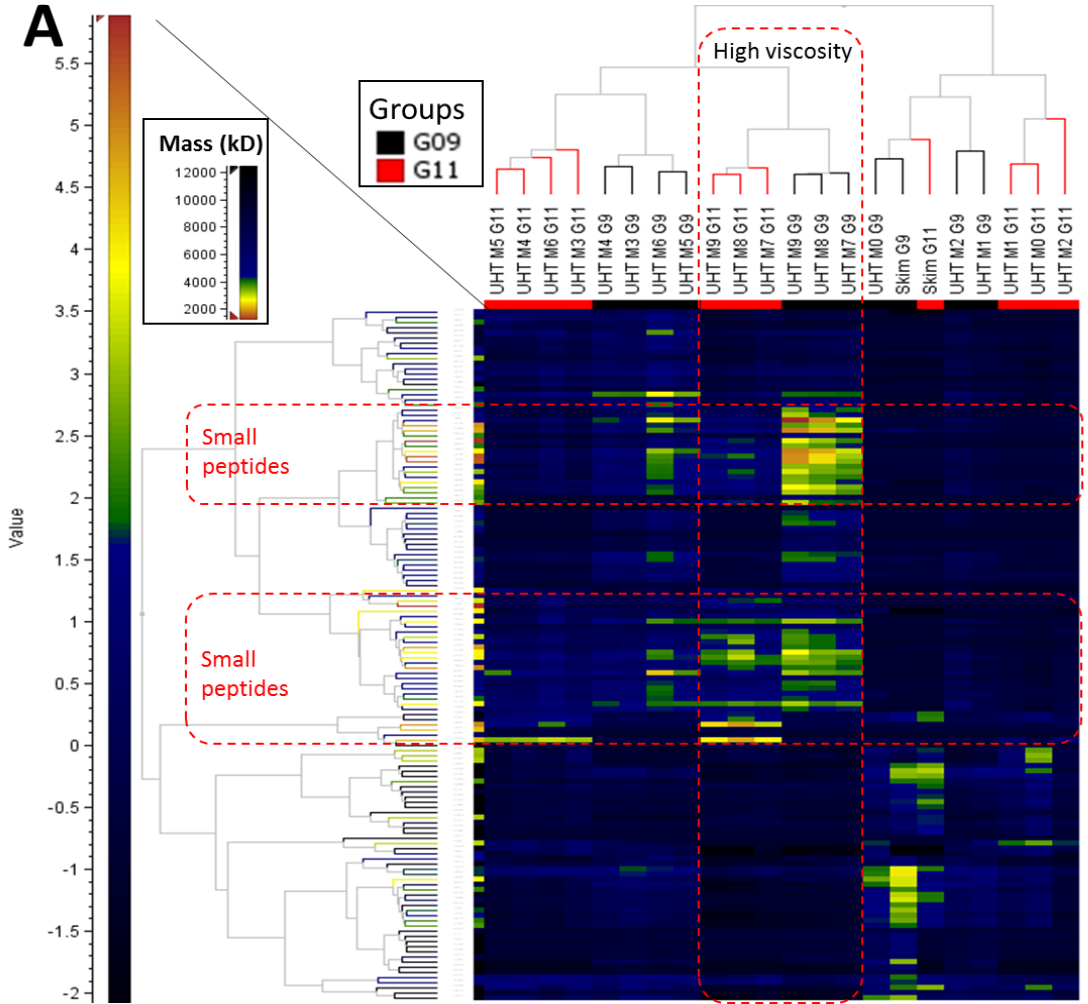


MILK PROTEOMICS

Intact protein analysis - Statistics

Confirming PLS results, HCA and correlation analyses list the biomarkers of high viscosity. They are the smallest peptides.

A



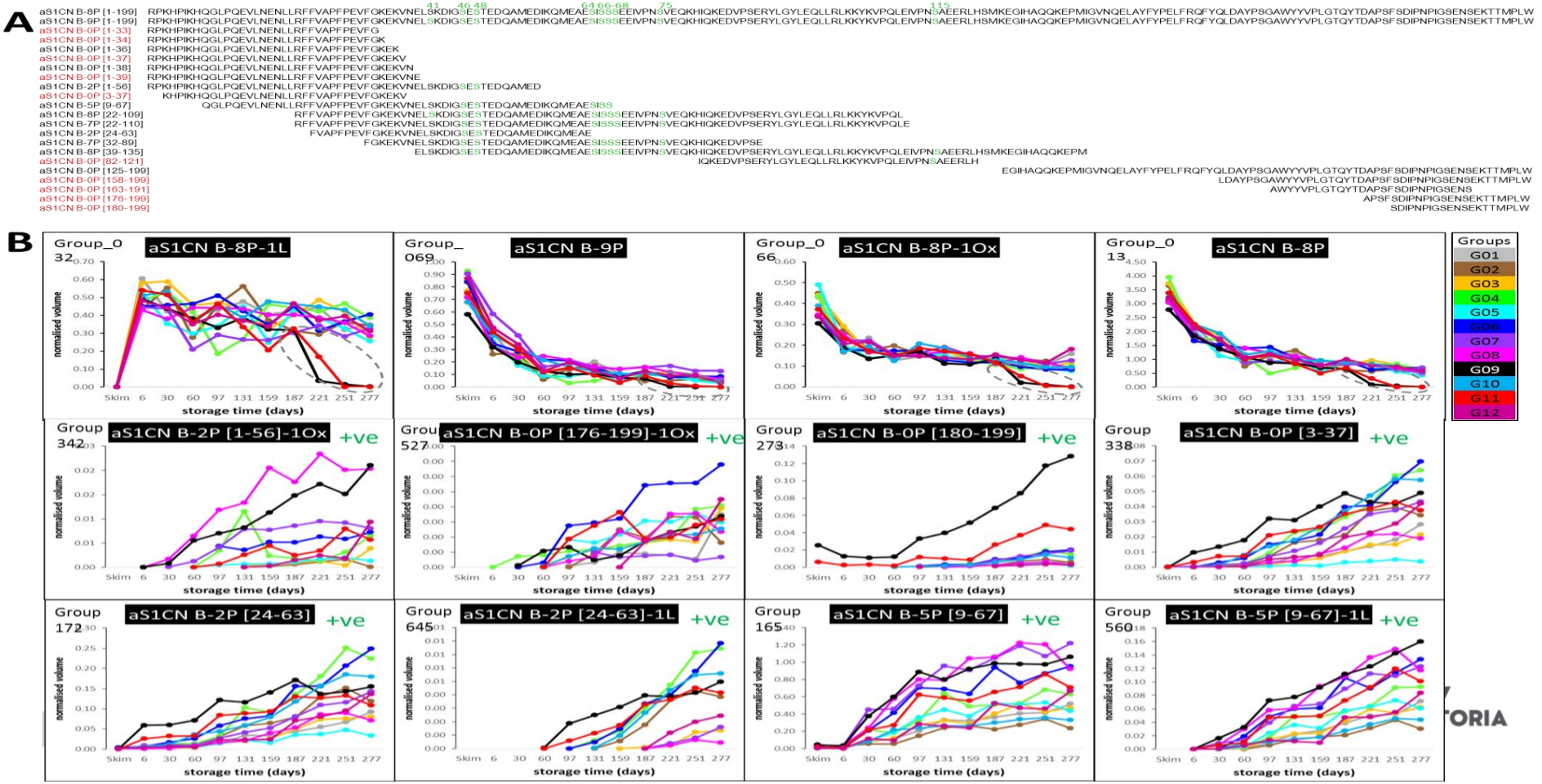
B

ID	Protein type	PTM type	Name	Mass (D)	RT (min)
Group_317	aLA	Lysis	aLA B [2-60]	6783.46	4.58
Group_342	aS1CN	Oxidation +Lysis	aS1CN B-2P-1Ox [1-56]	6592.28	5.63
Group_313	aS1CN	Lysis	aS1CN B-0P [176-199]	2617.25	4.33
Group_527	aS1CN	Oxidation +Lysis	aS1CN B-0P-1Ox [176-199]	2633.25	3.66
Group_273	aS1CN	Lysis	aS1CN B-0P [180-199]	2215.05	3.43
Group_172	aS1CN	Lysis	aS1CN B-2P [24-63]	4661.51	5.23
Group_338	aS1CN	Lysis	aS1CN B-0P [3-37]	4084.25	5.60
Group_408	aS1CN	Lysis	aS1CN B-0P [82-121]	4807.67	11.86
Group_165	aS1CN	Lysis	aS1CN B-5P [9-67]	7026.25	4.06
Group_413	aS2CN	Lysis	aS2CN A-0P [71-117]	5725.08	12.40
Group_337	bCN	Lysis	bCN-0P [106-138]	3832.95	5.53
Group_198	bCN	Lysis	bCN-0P [107-154]	5588.02	13.97
Group_422	bCN	Oxidation +Lysis	bCN-0P-1Ox [107-154]	5604.01	12.54
Group_463	bCN	Lysis	bCN-0P [108-124]	2011.93	3.01
Group_245	bCN	Lysis	bCN-0P [108-138]	3567.79	8.14
Group_764	bCN	Oxidation +Lysis	bCN-0P-1Ox [108-138]	3583.79	8.14
Group_416	bCN	Lysis	bCN-0P [125-138]	1573.90	5.24
Group_556	bCN	Lysis	bCN-0P [127-138]	1359.78	3.93
Group_315	bCN	Lysis	bCN-0P [133-161]	3364.87	4.40
Group_282	bCN	Lysis	bCN-0P [139-161]	2695.36	3.68
Group_301	bCN	Lysis	bCN-0P [139-162]	2794.42	4.11
Group_302	bCN	Oxidation +Lysis	bCN-0P-1Ox [139-162]	2810.42	4.11
Group_316	bCN	Lysis	bCN-0P [139-164]	2994.54	4.59
Group_359	bCN	Lysis	bCN-0P [162-206]	4998.77	6.61
Group_672	bCN	Oxidation +Lysis	bCN-0P-1Ox [162-206]	5014.77	5.58
Group_192	bCN	Oxidation +Lysis	bCN-0P-1Ox [162-209]	5340.01	9.20
Group_193	bCN	Oxidation +Lysis	bCN-0P-1Ox [165-209]	5040.82	9.94
Group_371	bCN	Lysis	bCN-0P [191-209]	2106.23	7.16
Group_507	bCN	Lysis	bCN-0P [192-206]	1667.92	3.49
Group_347	bCN	Lysis	bCN-0P [192-209]	1993.15	6.02
Group_621	bCN	Lysis	bCN-0P [198-209]	1263.81	4.69
Group_160	bCN	Lysis	bCN-0P [37-59]	2711.35	3.69
Group_268	bCN	Oxidation +Lysis	bCN-0P-1Ox [37-59]	2727.35	3.69
Group_196	bCN	Lysis	bCN A2-0P [58-105]	5160.83	12.07
Group_810	bCN	Oxidation +Lysis	bCN A2-0P-1Ox [58-105]	5176.83	12.08
Group_442	bCN	Lysis	bCN A2-0P [58-106]	5298.90	10.57
Group_639	bCN	Lysis	bCN A2-0P [58-94]	3992.16	18.06
Group_350	bCN	Lysis	bCN-0P [69-105]	3966.19	6.17
Group_666	bCN	Oxidation +Lysis	bCN-0P-1Ox [69-105]	3982.17	5.44
Group_619	bCN	Lysis	bCN-0P [96-138]	4902.54	4.63
Group_617	bCN	Oxidation +Lysis	bCN-0P-1Ox [96-138]	4918.54	4.62
Group_777	bLG	Lysis	bLG [127-153]	3125.65	9.14
Group_251	kCN	Lysis	kCN A-0P [125-150]	2589.26	3.53

MILK PROTEOMICS

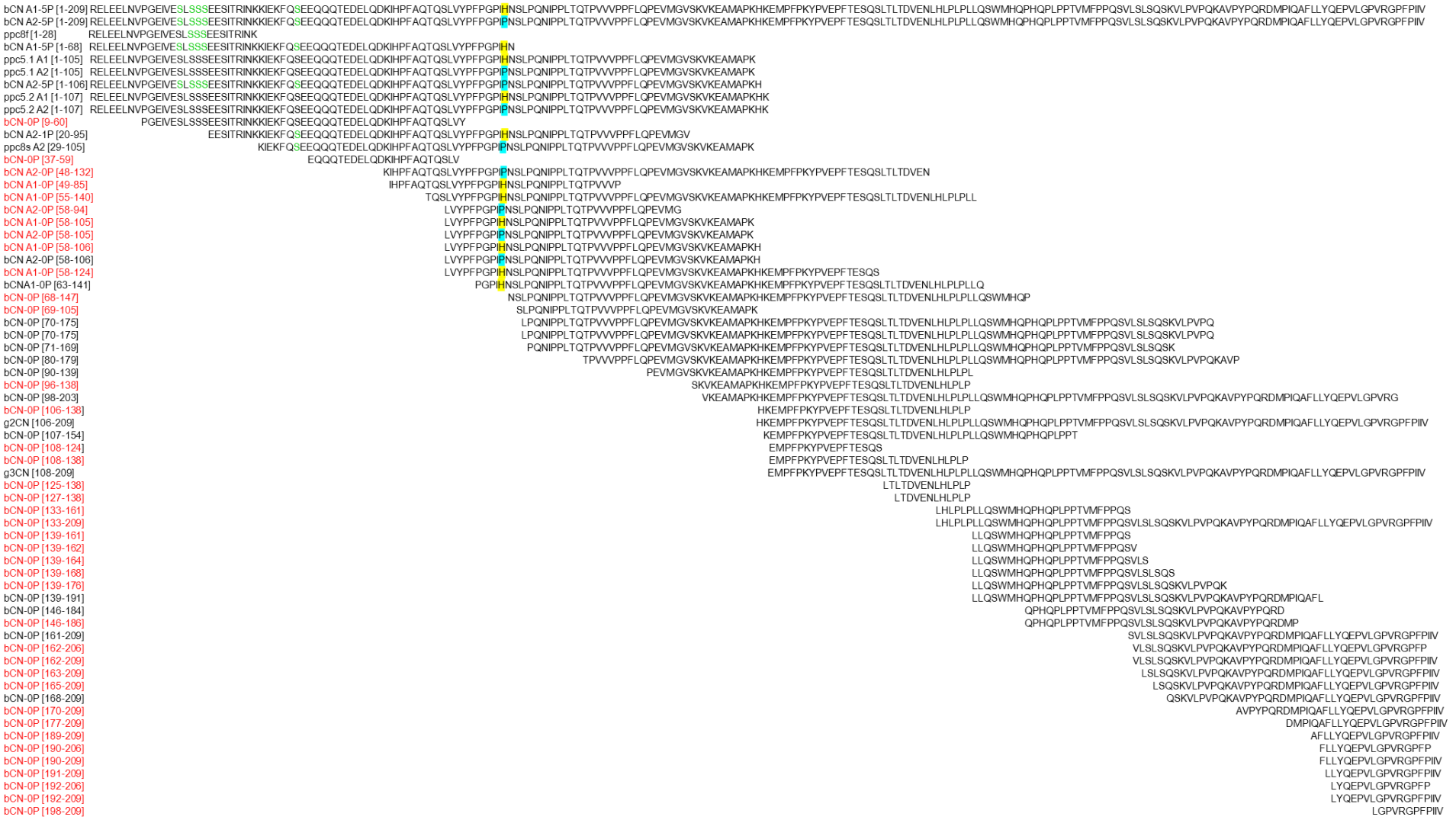
Intact protein analysis – alpha caseins

Here are examples of alpha-casein compounds associated with viscosity.



MILK PROTEOMICS

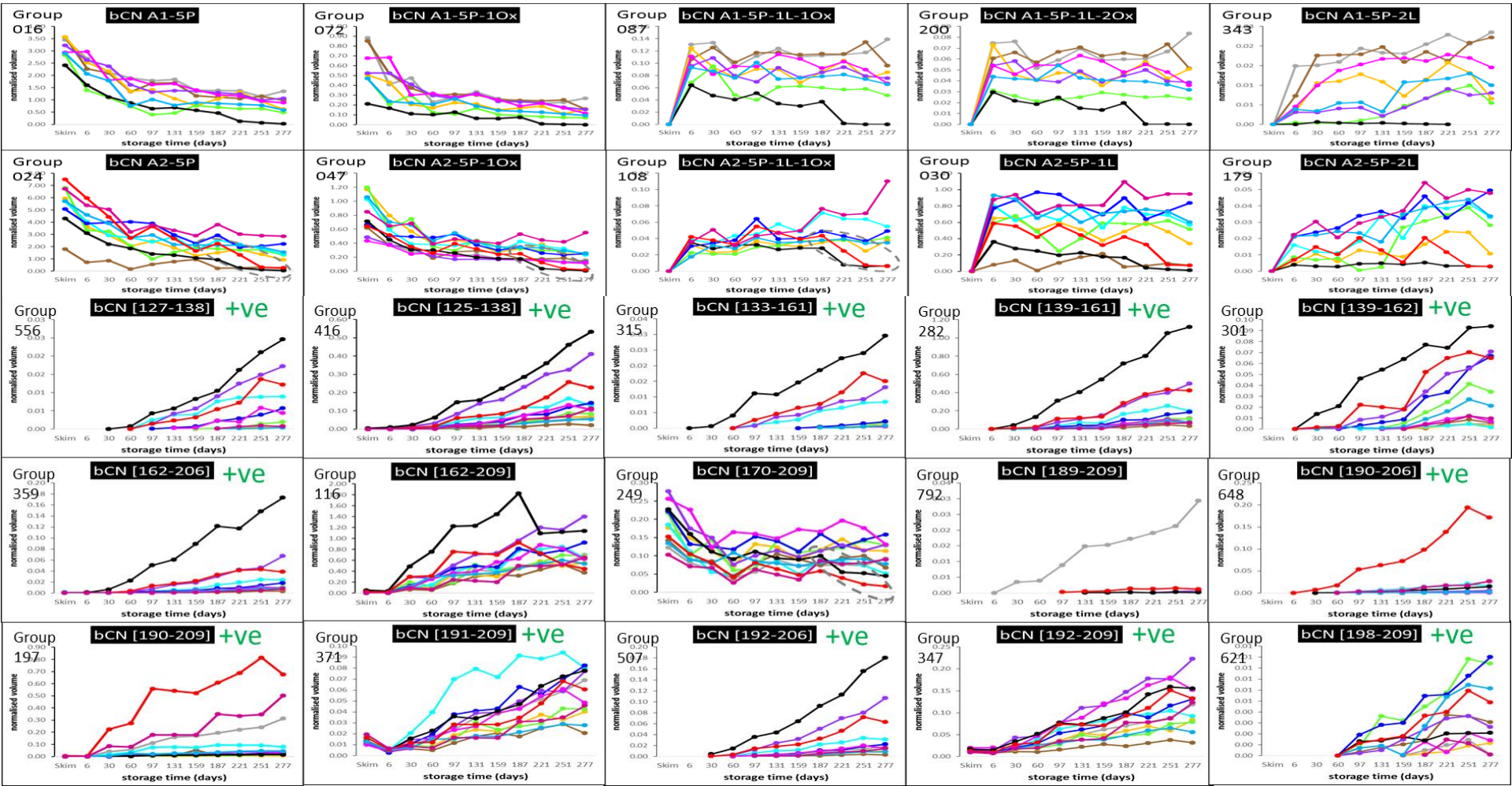
Intact protein analysis – beta caseins



MILK PROTEOMICS

Intact protein analysis – beta caseins

Here are examples of beta-casein compounds associated with viscosity.



MILK PROTEOMICS

Top-down analysis – summary

Three general trends from top-down proteomics analyses:

- Decreases in abundance all of the native intact milk proteins identified over storage time of UHT milk.

- Occurrence of lactosylated proteins following UHT treatment.

- Accumulation of peptides over storage time of UHT milk.

Proteolysis would be a leading mechanism for the instability of stored milks.

Such observations could have not been drawn using a bottom-up strategy.

CONCLUSIONS

CONCLUSIONS

Bottom-up or top-down? Which one is best?
Provided you're technically flexible enough, Both of course!