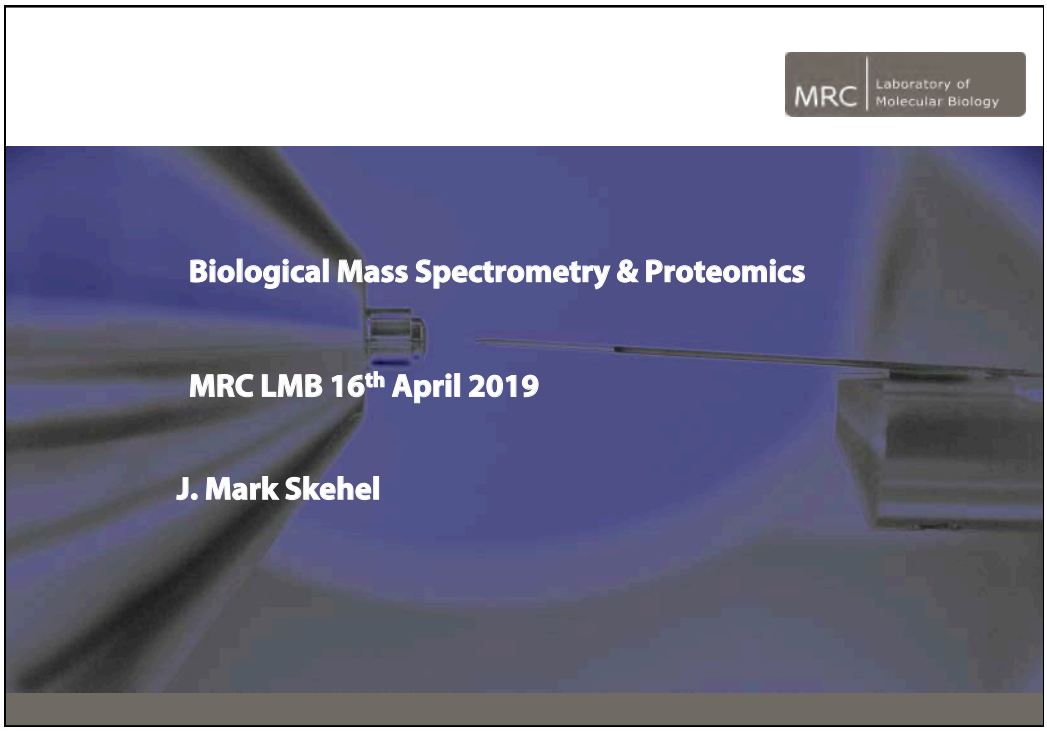


Mass Spectrometry

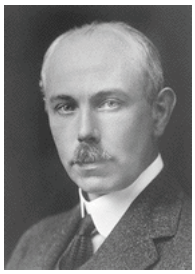
Very powerful technique in biology
Extremely sensitive (≤ femtomolar)
Measures mass/charge (m/z) ratio of ions

Things to remember:

- Good sample preparation is key!
- Not inherently quantitative
- Different compounds ionise differently.

A little history…

Joseph John Thomson
1906 Nobel Prize for Physics
"in recognition of the great merits of his theoretical and experimental investigations on the conduction of electricity by gases"

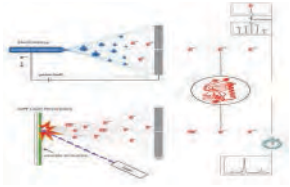


Francis William Aston
1922 Nobel Prize for Chemistry
"for his discovery, by means of his mass spectrograph, of isotopes, in a large number of non-radioactive elements, and for his enunciation of the whole-number rule"

Wolfgang Paul
1989 Nobel Prize for Physics
"for the development of the ion trap technique"

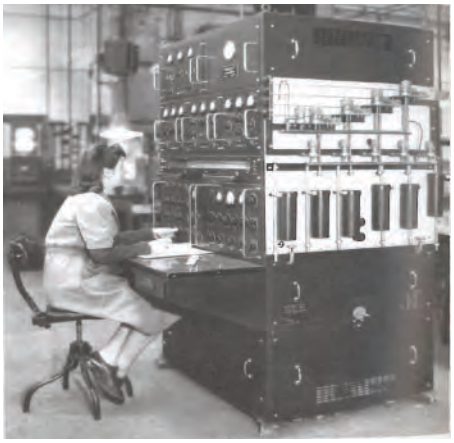
John Bennet Fenn
2002 Nobel Prize for Chemistry
"for the development of soft desorption ionisation methods (ESI) for mass spectrometric analyses of biological macromolecules"

Koichi Tanaka
2002 Nobel Prize for Chemistry
"for the development of soft desorption ionisation methods (MALDI) for mass spectrometric analyses of biological macromolecules"



16/04/19

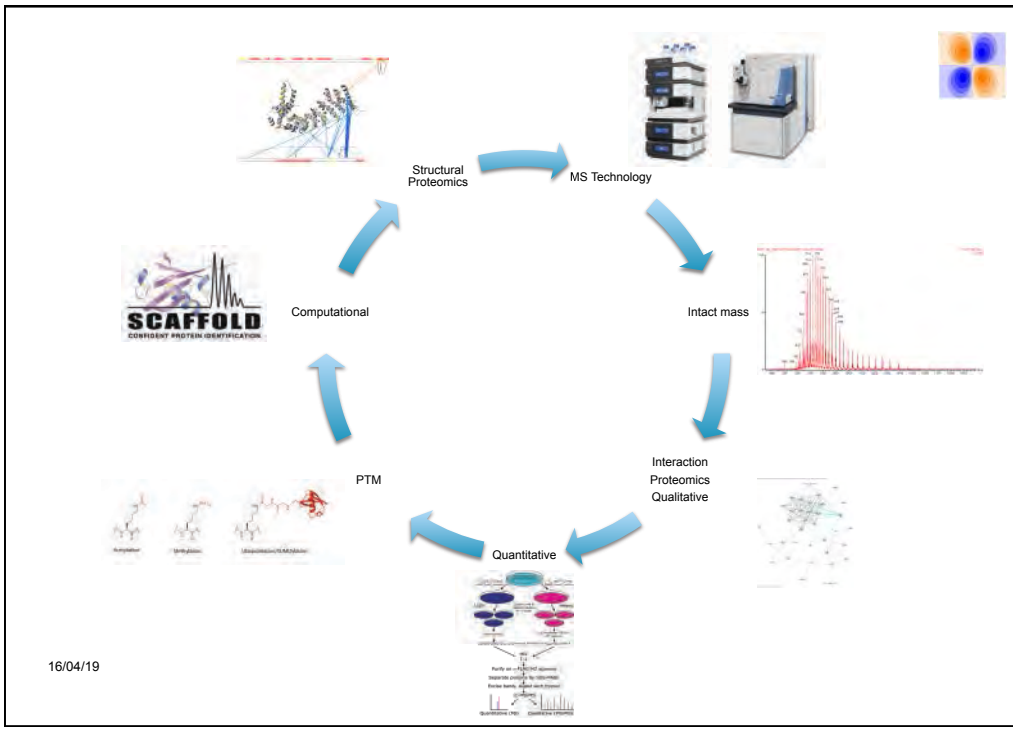
Early commercial instruments



Metropolitan Vickers MS1
Mass Spectrometer 1946

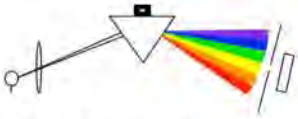
Measurement and Analysis Technologies
MAT CH4 Mass Spectrometer 1960s

16/04/19



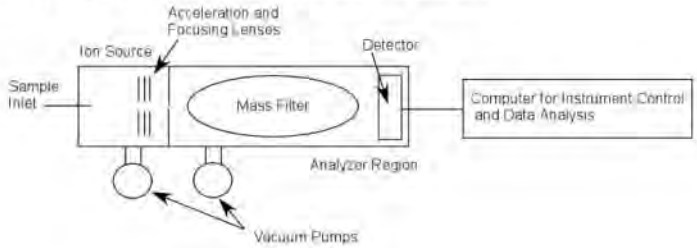
16/04/19

Mass Spectrometry



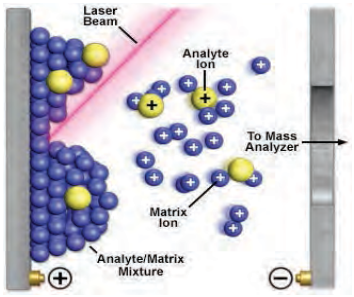
Lenses and prisms focus and refract light.
Analogous systems can focus and deflect ions in a vacuum.

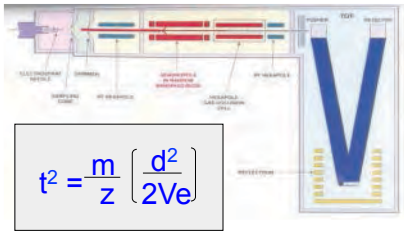
1. Get molecules into the gas phase & ionize them.
2. Give the ions a defined energy or velocity.
3. Separate or sort the ions on the basis of that defined property.
4. Detect the ions & assign their masses.



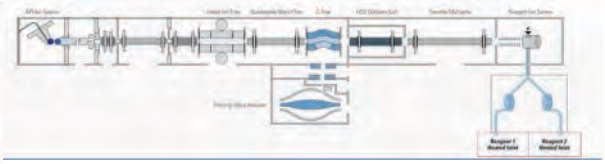
16/04/19

Mass Spectrometry





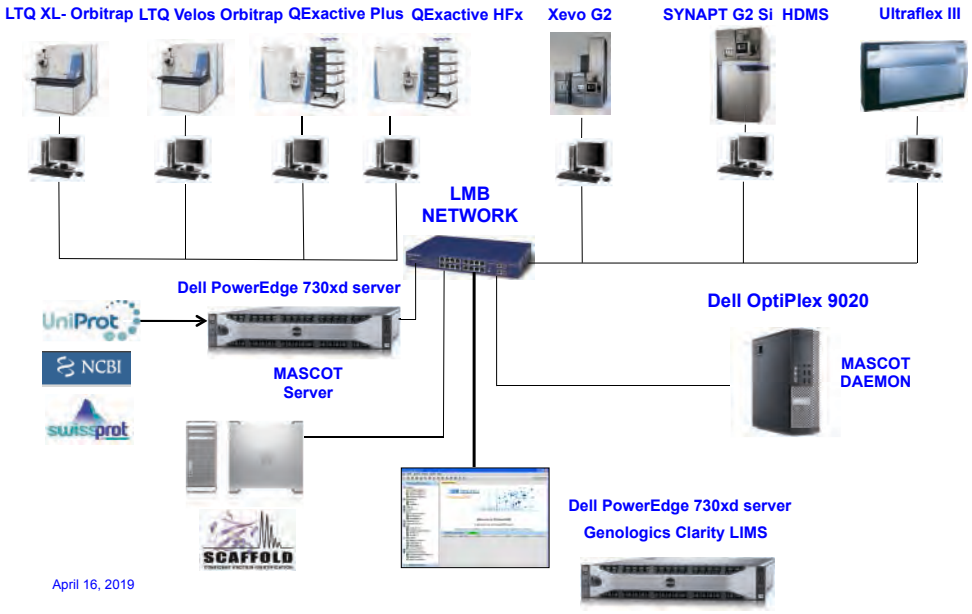
$$t^2 = \frac{m}{z}\left(\frac{d^2}{2Ve}\right)$$



$$\omega = \sqrt{k/m/z}$$

ω=Oscillation frequency
k= Instrument constant
m/z= mass to charge ratio

Local Network Configuration



April 16, 2019

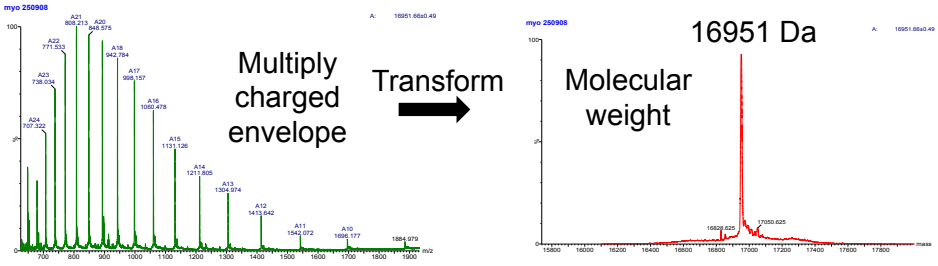
Measurement of intact mass

Why measure protein mass?

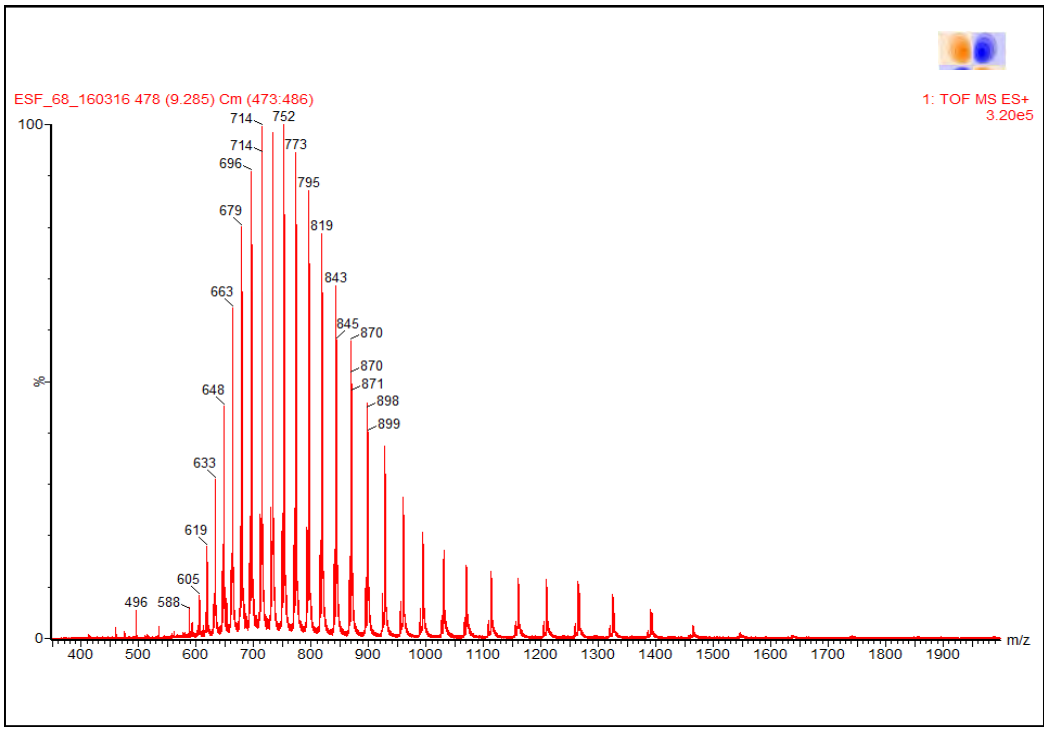
Mass analysis may reveal…
the wrong sequence was cloned/expressed
a post-translational or chemical modification
proteolytic processing (signal sequence, ragged N- and C- termini)

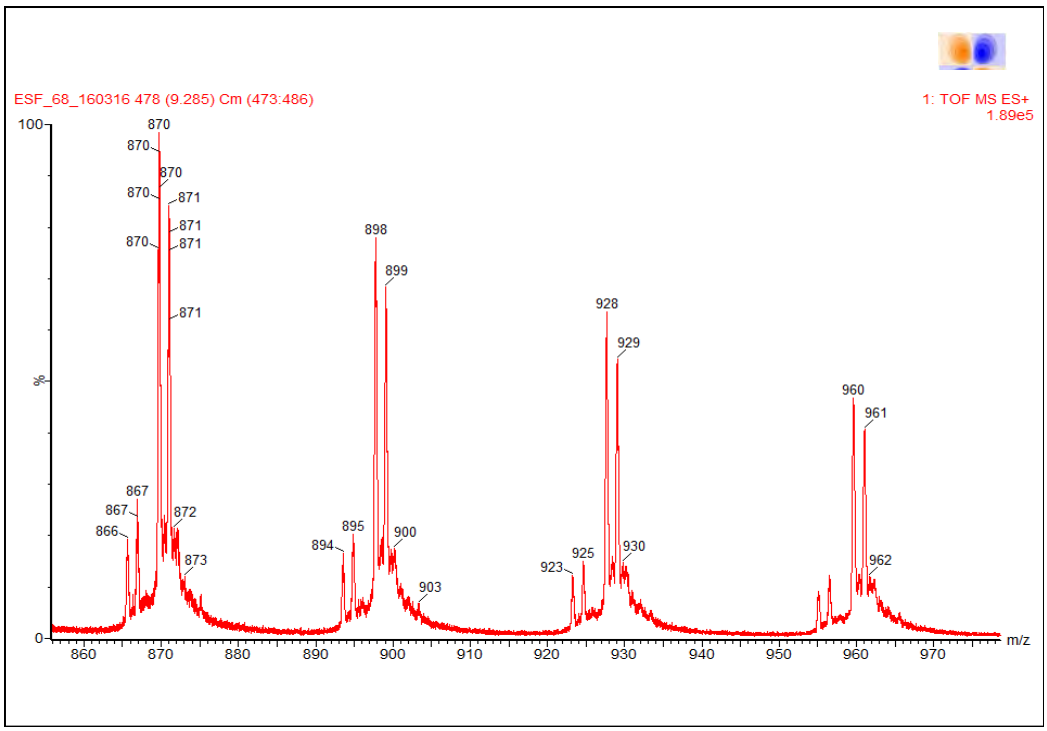
Mass analysis facilitates…
identifying unknown proteins by proteomic approaches
characterisation of stable complexes with small molecules or other proteins

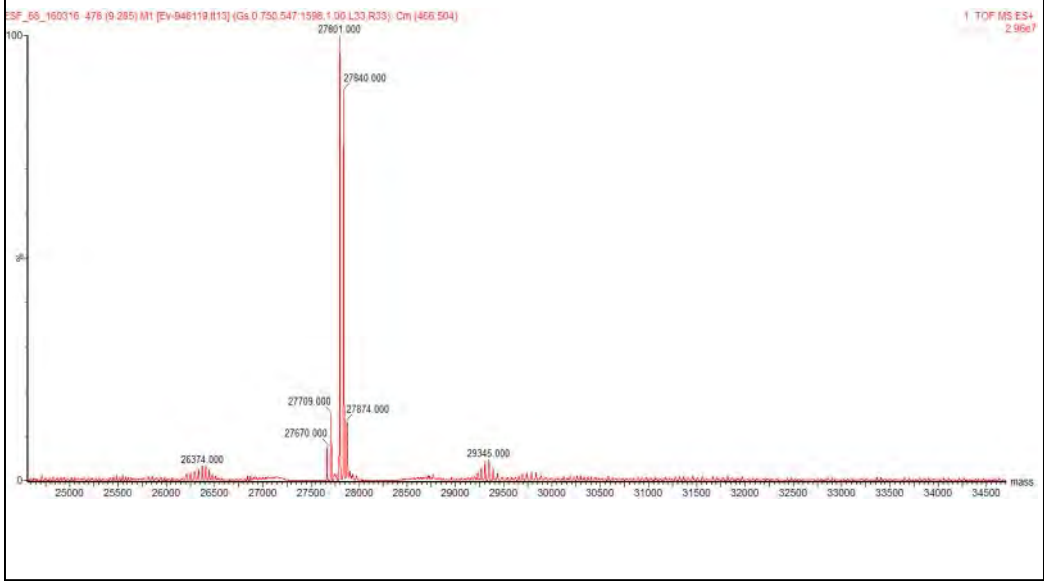
Measurement of intact proteins by ESI-TOF



$$M = \frac{(m1-1)(m2-1)}{m2-m1}$$


ESF_68_160316 478 (9.285) Cm (473:486)
1: TOF MS ES+
3.20e5
100
%
0
714
752
714
773
696
795
679
819
843
663
845
870
870
871
648
898
899
633
619
605
496
588
400
500
600
700
800
900
1000
1100
1200
1300
1400
1500
1600
1700
1800
1900
m/z


ESF_68_160316 478 (9.285) Cm (473:486)
1: TOF MS ES+
1.89e5
100
%
0
870
870
870
870
871
870
871
871
871
898
899
928
929
960
961
867
867
872
866
873
895
894
900
903
925
923
930
962
860
870
880
890
900
910
920
930
940
950
960
970
m/z



Buffer compatibility

Surfactant,Buffer and Salt	Mw (g/mol)	MALDI (mM)	MALDI (wt.%)	ESI (mM)	ESI (wt.%)
TRIS	121	100	1.0	N.C	N.C
HEPES	238	100	2.4	N.C	N.C
BICINE	163	50	0.8	N.C	N.C
Urea	60	500	3.0	N.C	N.C
Guanidine, HCl	96	250	2.4	N.C	N.C
Dithiothreitol	154	500	7.7	N.C	N.C
Glycerol	92	130	1.2	N.C	N.C
N-Octyl-β-glucopyranoside	292	3.4	0.1	3.1	0.1
n-Octyl sucrose	468	N.C	N.C	2.1	0.1
n-Dodecyl sucrose	524	N.C	N.C	1.9	0.1
n-Dodecyl maltoside	511	N.C	N.C	2.0	0.1
Cctyl thioglucoside	308	N.C	N.C	3.2	0.1
n-Hexyl glucoside	264	N.C	N.C	3.8	0.1
n-Dodecyl glucoside	348	N.C	N.C	2.9	0.1
PEG1000	1000	N.C	N.C	N.C	N.C
PEG 2000	2000	0.5	0.1	N.C	N.C
Triton X-100	628	1.6	0.1	N.C	N.C
NP-40	603	1.7	0.1	N.C	N.C
Zwittergent 3-16	392	2.6	0.1	N.C	N.C
Tween 20	1228	N.C	N.C	0.81	0.1
Thesit	583	N.C	N.C	<1.7	<0.1
SDS	288	0.35	0.01	N.C	N.C
LDAO	229	4.4	1.0	<4.4	<0.1
CTAB	284	N.C	N.C	<3.5	<0.1
CHAPS	615	0.16	0.01	1.6	0.1
Sodium Cholate	431	N.C	N.C	2.3	0.1
Sodium Taurocholate	538	N.C	N.C	<1.9	<0.1
Sodium Azide	65	15	0.1	3.1	0.02
NH_4HCO_3	79	50	0.4	50	1
NaCl	58	50	0.29	N.C	N.C
Sodium Acetate	82	50	0.41	50	1
$NaHPO_4$	120	10	0.12	N.C	N.C
TFA	114	N.C	N.C	4.4	0.05

16/04/19

Sample Preparation for LCT (ESI-TOF)

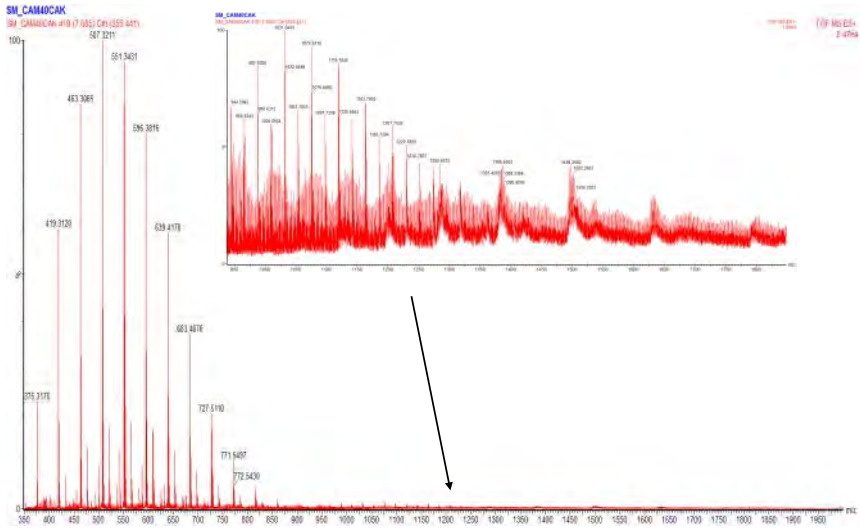
- Concentration ~ 5-10 pmol/ul
- Remove salts, detergents and glycerol
- Where possible, desalt by Dialysis, HPLC or ZipTips

Solvents:
Peptides - 0.2-1% Formic acid / 50% MeOH or MeCN
Proteins - 1-5% Formic acid / 50% MeCN, MeOH or propanol.

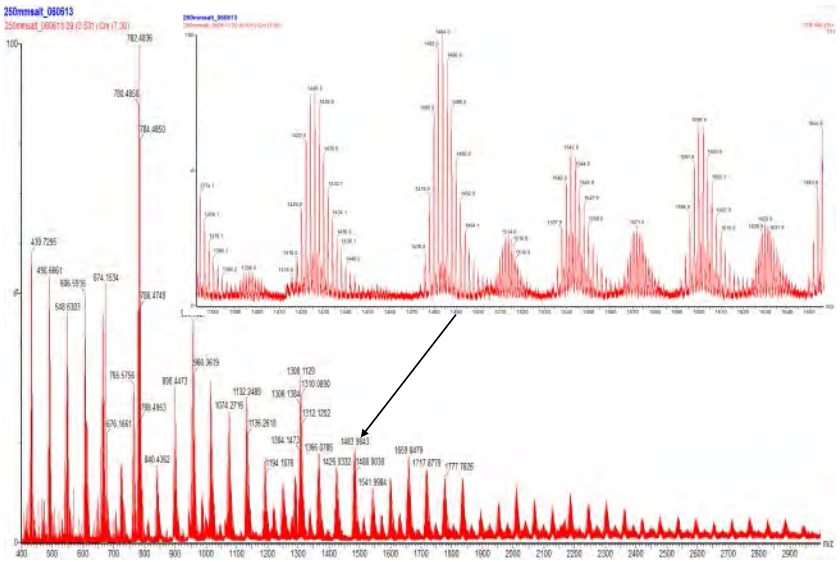
Do NOT use Trifluoroacetic acid - signal suppression

Micromass LCT. Available to trained users.

Contamination (PEG)



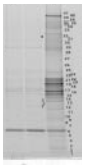
250mM NaCl



16/04/19

Proteomics – Fundamental Analysis Process for Complex Mixtures

Separation



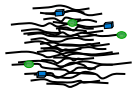
or

SDS-PAGE, 2DE
or solution

Liquid handling

proteolysis

peptides

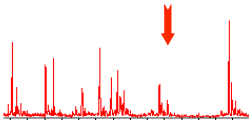


LC-MS/MS or
MALDI-MS/MS

Why is this needed?

- Sample clean-up
- Reduce sample complexity that mass spectrometer has to deal with per unit time
- Automation

Velos Orbitrap



Data dependant MS/MS: mass *and* sequence information

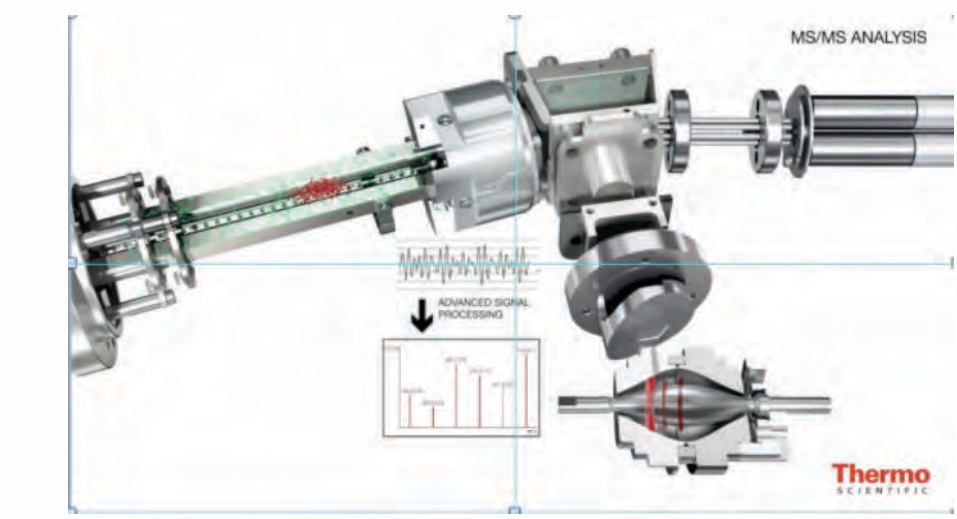
MASCOT
database search

Scaffold
results collation

Genologics Clarity
LIMS

16/0

Mass Spectormeter Fly Through

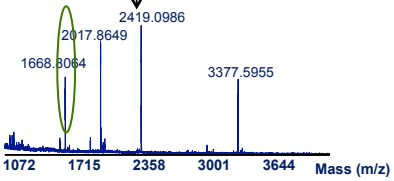


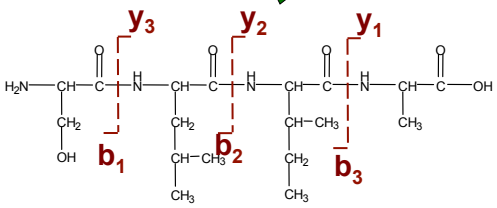
Protein/peptide identification by Mass Spectrometry

Proteolysis (trypsin)

MS/MS

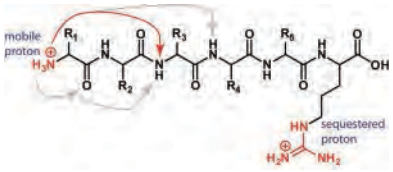
MS





Peptide mapping

Mobile Proton Theory



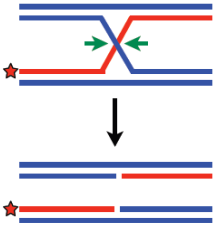
MAWPCITRACCIARFWNQLDKADIAVPLVFTKYSETEHPG
APPQPPAPLQPALAPPSRAVAIETQPAQGESDAVARATG
PAPGPSVDRETVAAPGRSGLGLGAASASTSGSGPADSV
MRQDYRAWKVQRPEPSCRPRSEYQPSDAPFERETQYQ
KDFRAWPLPRRGDHPWIPKPVQIPATSQPSQPVLGVPKR
RPQSQERGPMQ

Purification of human HJ resolvases (MUS81-EME1 & ResA)

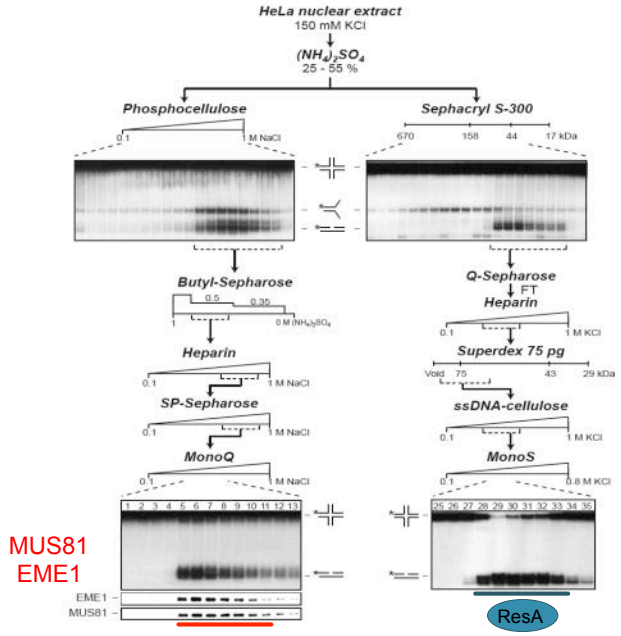
Nuclear extracts
200 litres HeLa cells

Extensive fractionation

Assay for HJ resolution

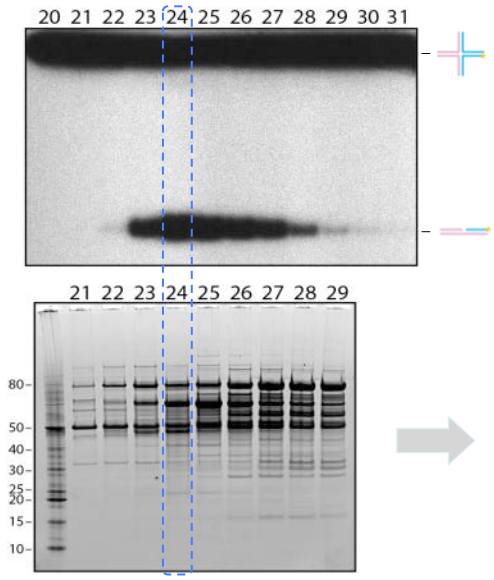


Ip et al., *Nature*, 2008



Identification of ResA by mass spectrometry, GelcMS

200 L
Nuclear extract
25-55% $(NH_4)_2SO_4$
Sephacryl S-300
Q sepharose
FT
Heparin
Superdex75 pg
ssDNA cellulose
MonoS



Mass spec analysis

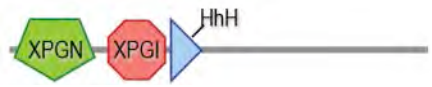
April 16, 2019

GEN1

5 peptides:

Identified an N-terminal fragment of an uncharacterised XPG-family protein known as GEN1

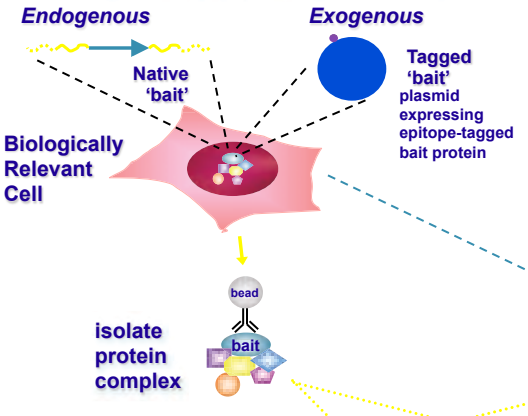
GEN1 has Rad2/XPG nuclease domains



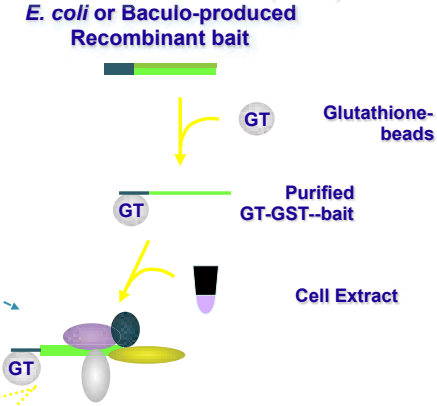
April 16, 2019

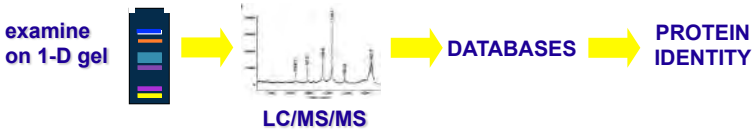
Interaction Proteomic Approaches

Endogenous or Exogenous IP



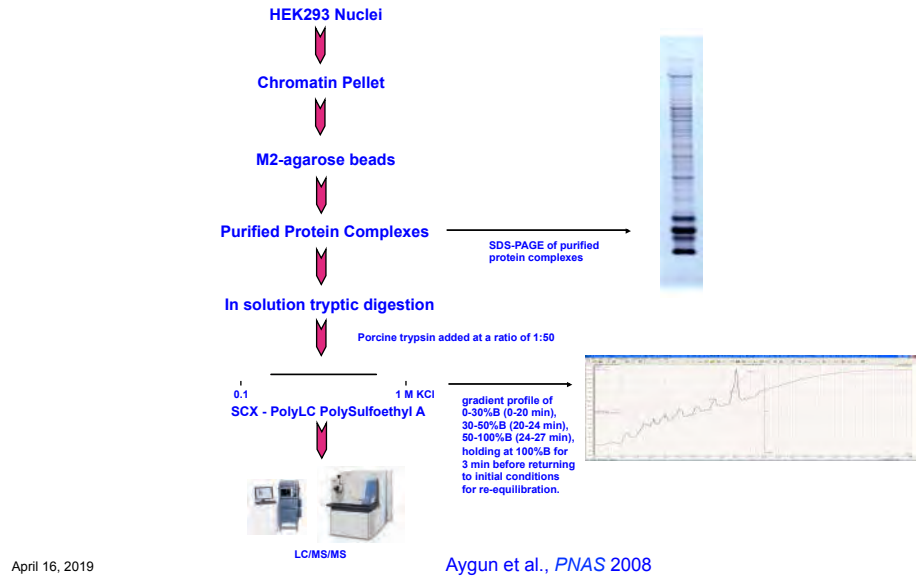
In vitro Marination (IVM)





16 April 2019

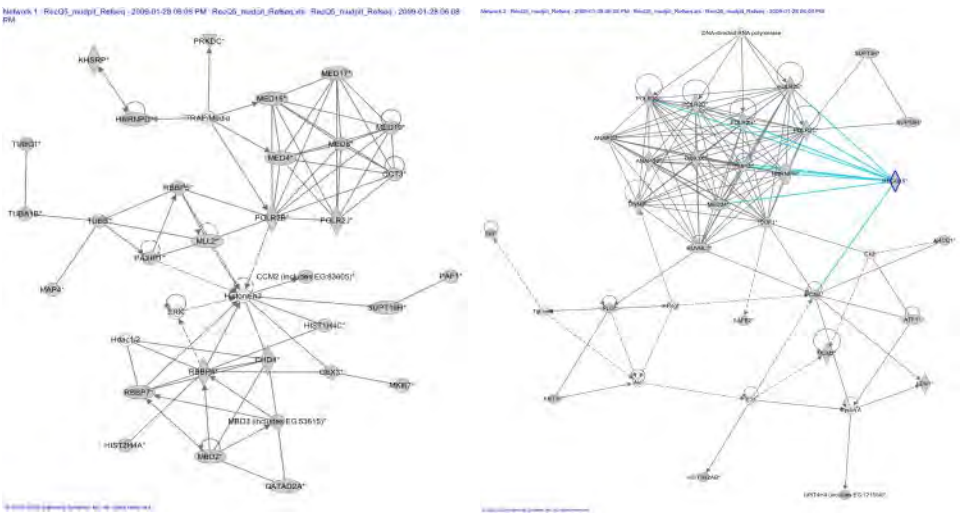
Outline MudPIT experiment



April 16, 2019

Aygun et al., *PNAS* 2008

Ingenuity Data

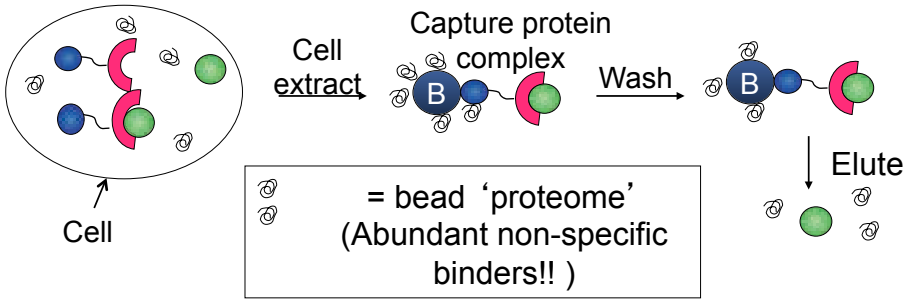


April 16, 2019

http://www.ingenuity.com/

Problems with Pull Down Assays

Non-specific binding – washing is required.
Correct controls are essential!
Weak and transient interactions may not be detected
Conditions may not simulate real biological situations



BirA or BioID Proximity-dependent biotin identification

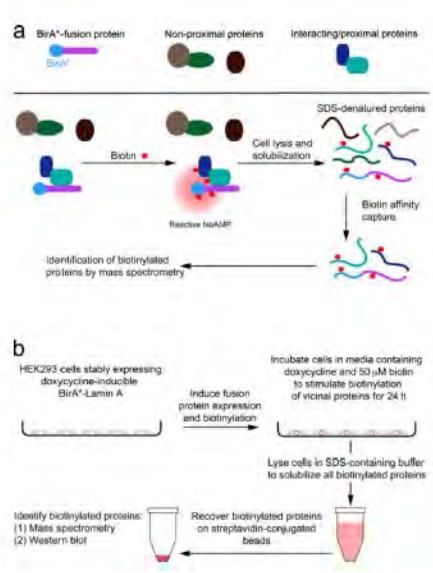
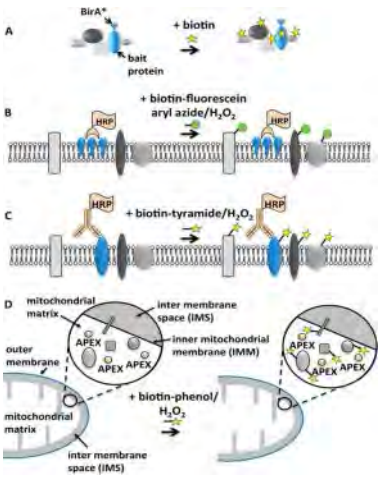


Figure 1. **Model for application of BioID method.** (a) Expression of a promiscuous biotin-ligase fusion protein in live cells leads to the selective biotinylation of proteins proximate to that fusion protein. After stringent cell lysis and protein denaturation, biotinylated proteins are affinity purified. These candidate proteins can be identified by mass spectrometry or immunoblot analysis. (b) In our application of BioID to LaA to identify candidate proteins we used HEK293 cells stably expressing inducible mycBirA*-LaA. 24 h before lysis, cells were induced to express mycBirA*-LaA with doxycycline and to biotinylate endogenous proteins with 50 μM biotin. Cells were lysed under stringent conditions and biotinylated proteins collected on streptavidin-conjugated beads for subsequent analysis and identification.

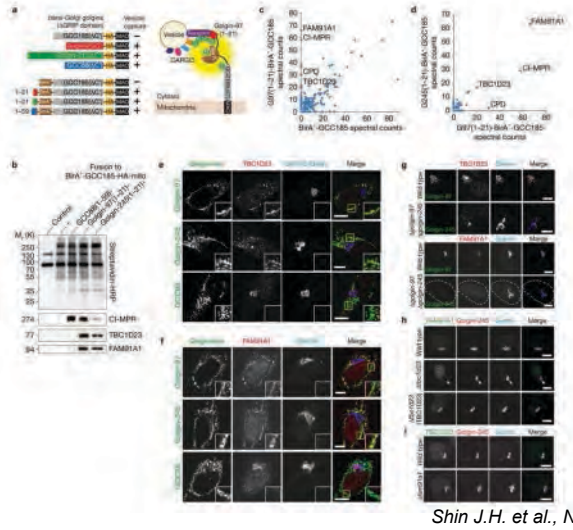
Kim DI, *et al.*, Proc Natl Acad Sci USA. 2014 Jun 17;111(24):E2453-61.

Summarized outline of the major published enzyme-catalyzed proximity labeling assays.



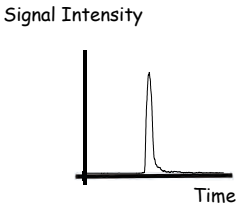
Johanna S. Rees et al. Mol Cell Proteomics 2015;14:2848-2856

Example of BioID application



Shin J.H. et al., Nature Cell Biology 2017

Quantitative Proteomics (Relative)



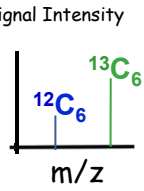
Integration of peak area in extracted ion chromatograms

Label free:

- Peak area of peptides ions
- Spectral Counting
- Protein abundance index

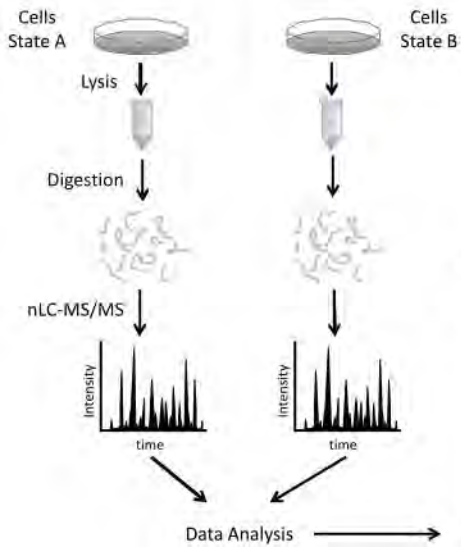
Isotopic labelling:

- Several methods are available:
- Chemical
  - ICAT - cysteine specific labelling reagents
  - iTRAQ/TMT - amine-specific labelling reagents
- Metabolic
  - SILAC – isotopic incorporation during cell culture

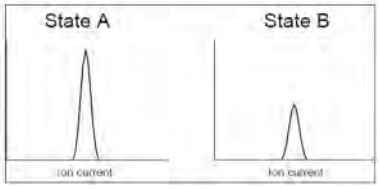


Compare isotopic intensities in mass spectra

Label free: Total Ion count



The **ion current** (intensity/time) of each peptide is extracted and compared between two states



Label Free: Spectral Counting

Abundant protein = Abundant peptides = more ms/ms

- Number of peptides identifying a protein increases with increasing protein amount.
- Larger proteins generate more measurable peptides than smaller ones

$$PAI = \frac{\text{\# Observed Peptides}}{\text{\# Observable Peptides}}$$

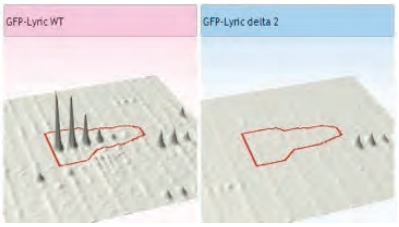
$$emPAI = 10^{PAI} - 1$$

16/04/19

Rappsilber J et al. ***Genome Res*** *(2002)12:1231-1245*
Ishihama Y. et al. ***MCP*** *(2005) 4:1264*

Label free

Advantages
- Simple biochemical workflows
- Whole proteome analysis
- Comparison of multiple states
- Linear dynamic range



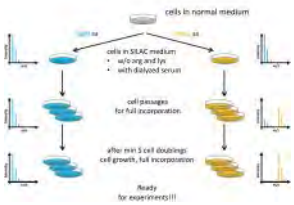
Disadvantages
- Sample prep – problematic for multistage protocols
- Chromatography (EMRTs)

16/04/19

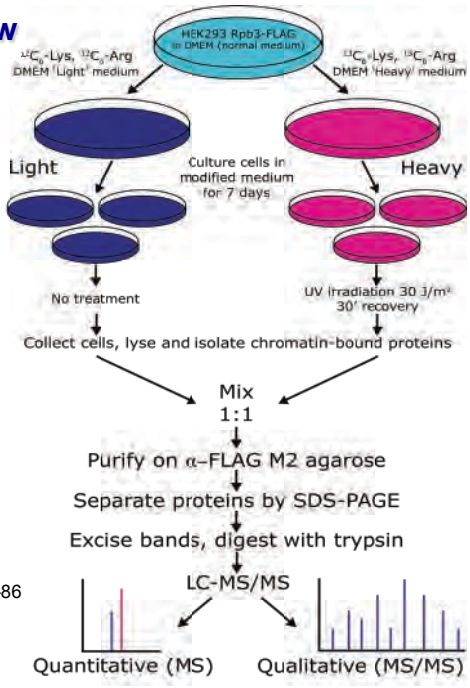
SILAC Applications

SILAC can be employed for:

- Investigating regulation of gene expression
- Finding biomarkers for diseases
- Signalling pathways
  - Quantitative phosphoproteomics
- Identifying protein interactors



SILAC labelling workflow

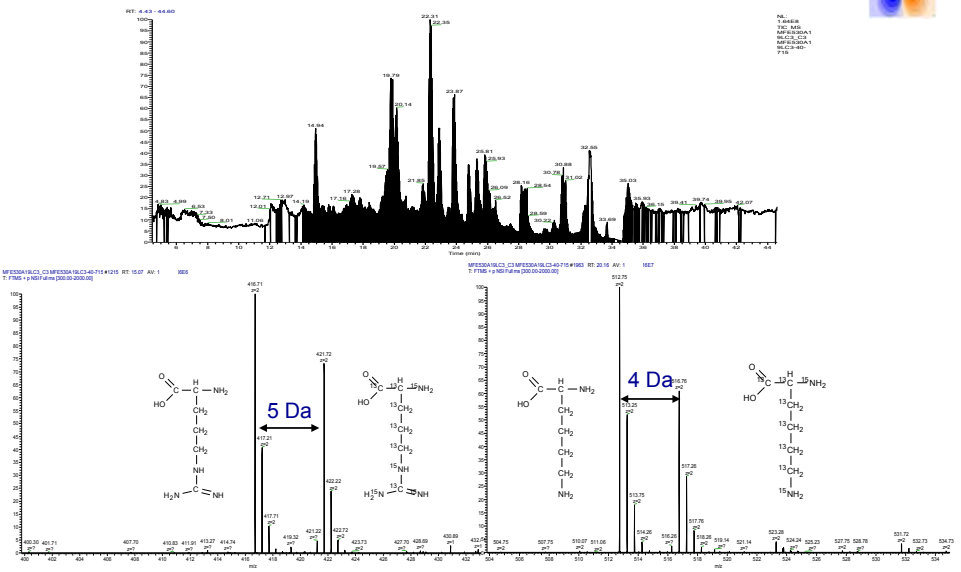


Data analysis:
MaxQuant, Mascot Distiller, Scaffold Q+S

Ong et al. (2002) *Mol & Cell Prot.* **1**: 376–86

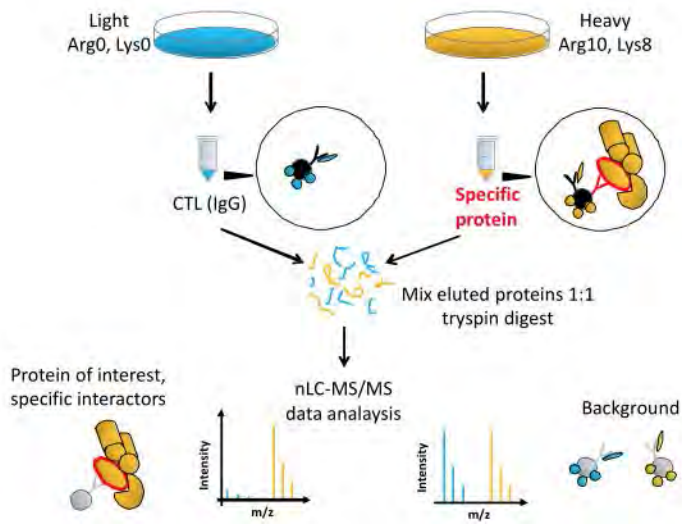
16/04/19

LC-MS showing SILAC peptides



16/04/19 Arginine 10 containing peptide Lysine 8 containing peptide

SILAC for Protein-Protein Interactions



16/04/19

Sara Zanivan, CRUK, Beatson Institute

SILAC Advantages/Disadvantages

Advantages over other quantitation methods:

- $^{13}C_6$ labelling is done before protein purification
- Accurate relative quantitation

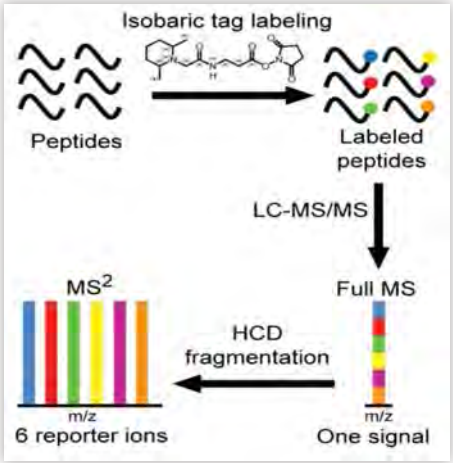
Disadvantages over other quantitation methods:

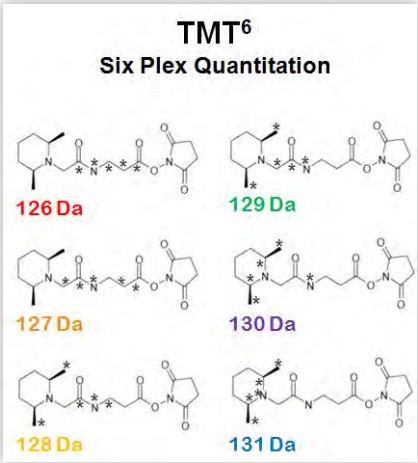
- Is limited to cell culture?
- Use of dialysed FCS. 10Kmwco

SILAC mouse

Kruger et al. (2008) Cell, 134, 353–364

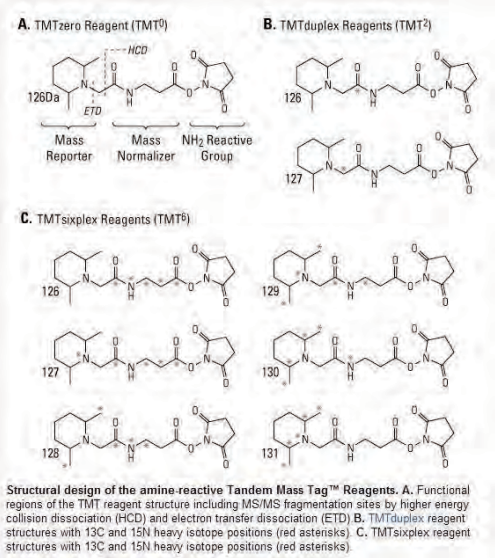
Reporter Ion-Based Quantitation





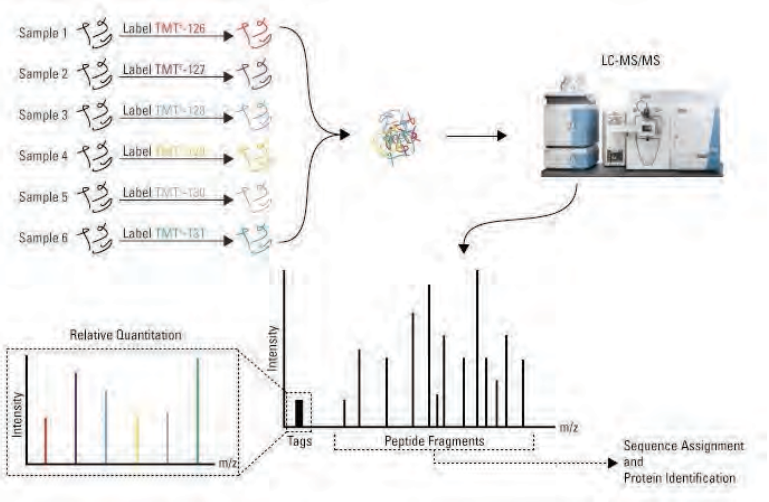
16/04/19

TMT reagents



Structural design of the amine-reactive Tandem Mass Tag™ Reagents. A. Functional regions of the TMT reagent structure including MS/MS fragmentation sites by higher energy collision dissociation (HCD) and electron transfer dissociation (ETD) **B.** TMTduplex reagent structures with 13C and 15N heavy isotope positions (red asterisks). **C.** TMTsixplex reagent structures with 13C and 15N heavy isotope positions (red asterisks).

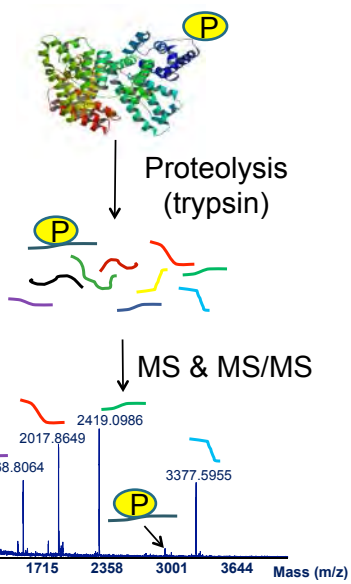
TMT workflow



Characterisation of post-translational modifications (PTMs)

MS can characterise many PTMs including:

- Phosphorylation
- Acetylation
- Methylation
- Sites of ubiquitination
- Glycosylation
- GPI anchors

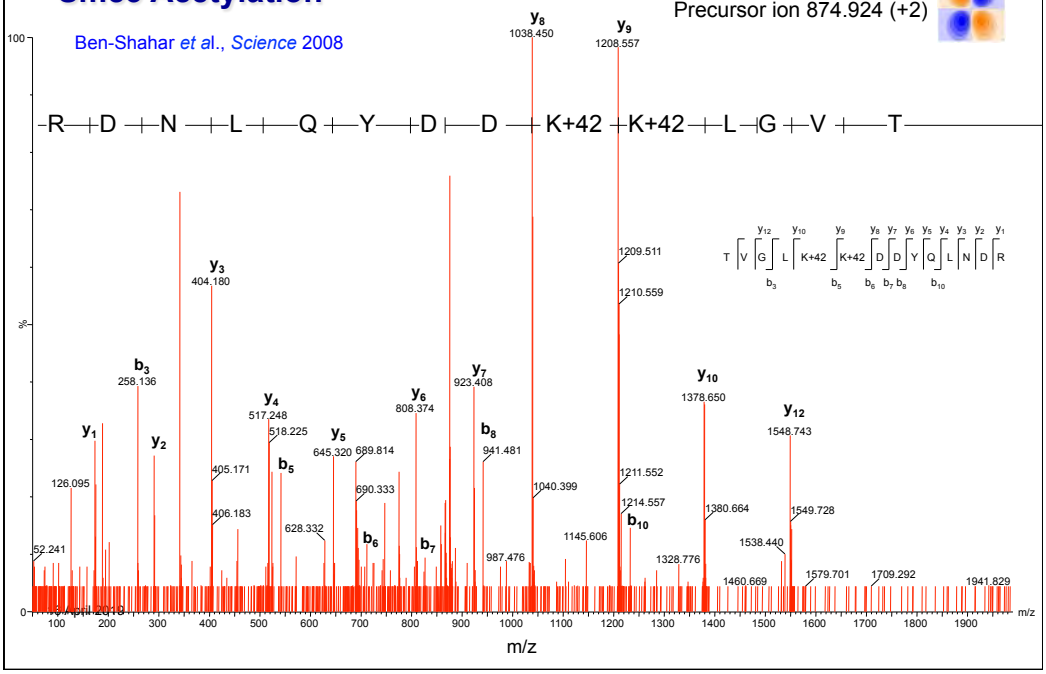


Identify PTMs by searching the protein database using Mascot (not large glycans & GPI anchors)

Smc3 Acetylation

Ben-Shahar *et al.*, *Science* 2008

Precursor ion 874.924 (+2)

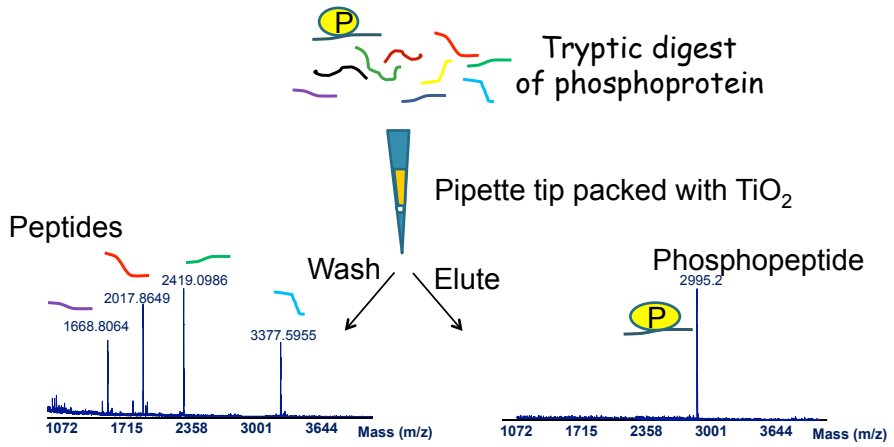


Characterisation of Phosphoproteins

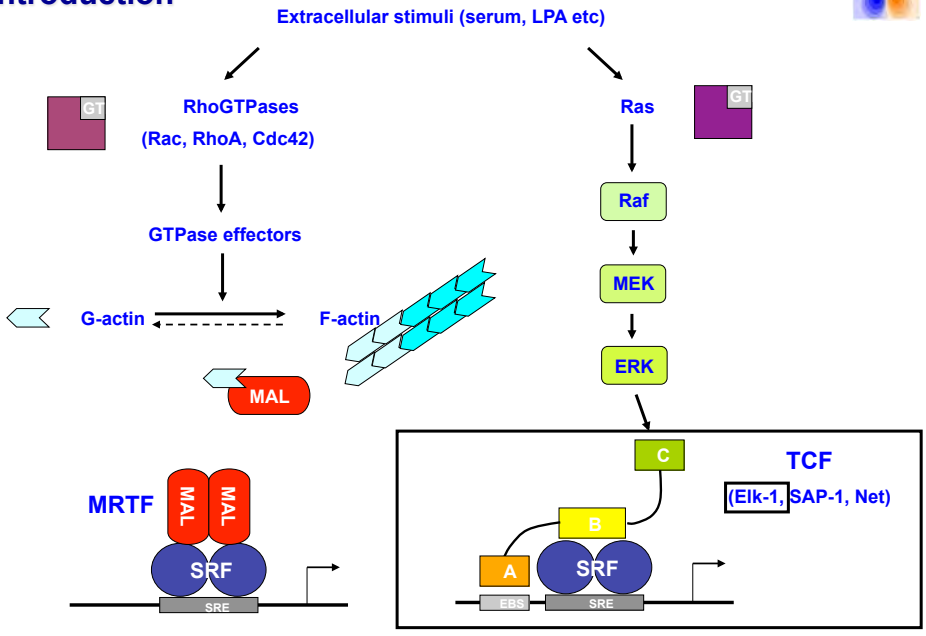
Phosphopeptides are often difficult to detect

- substoichiometric phosphorylation.

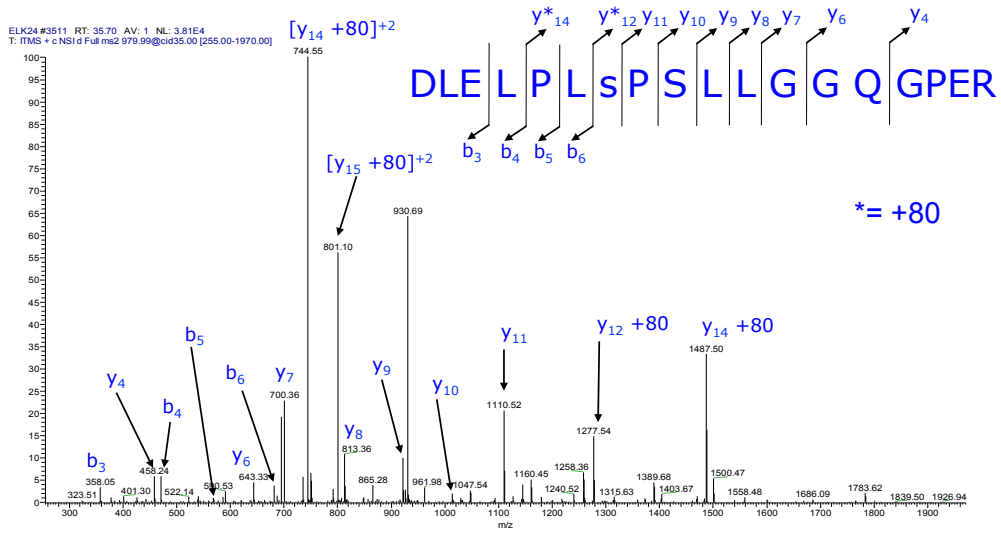
Enrichment of phosphopeptides using metal affinity (TiO_2 or Gallium (III)):



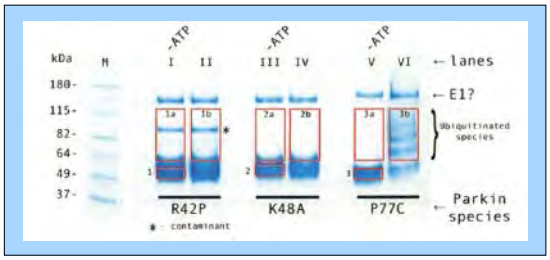
Introduction



Collision Induced Dissociation

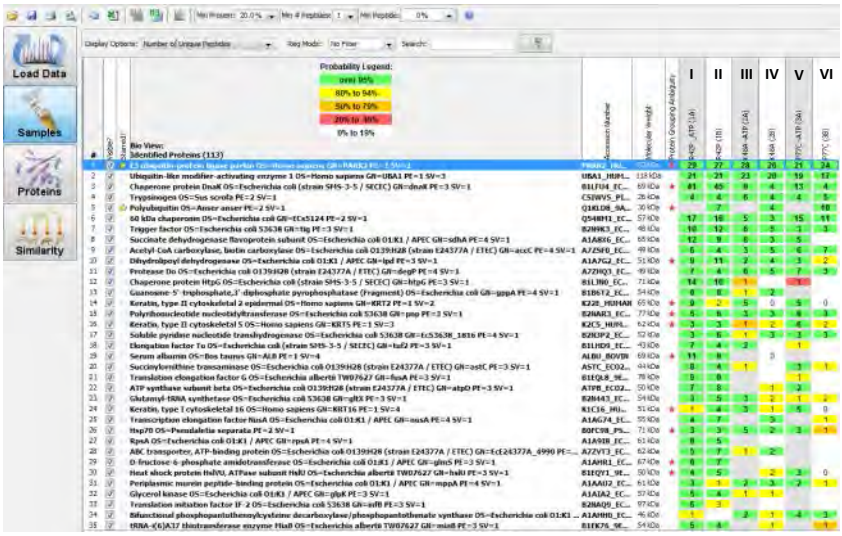


Ubiquitination of Parkin

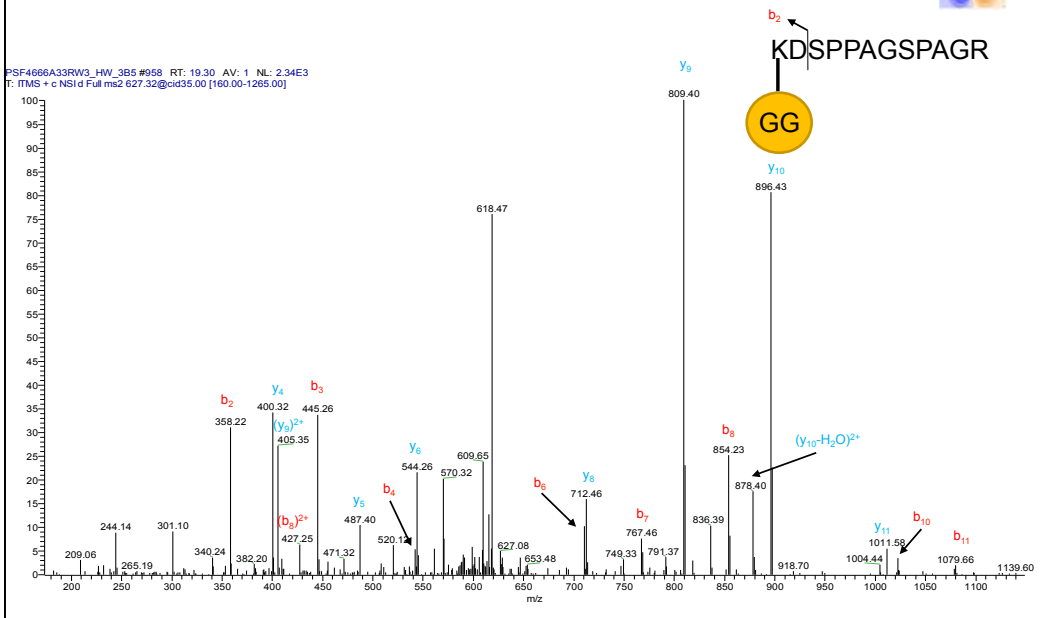


- gel sliced into 5
- reduced (DTT) and alykalated (iodoacetamide)
- digested with trypsin
- extracted with 50% AcN 0.1% formic acid
- analysed by LC-MS/MS (LTQ Orbitrap-XL)
- 116 proteins were identified using mascot
- dat file loaded into scaffold viewer

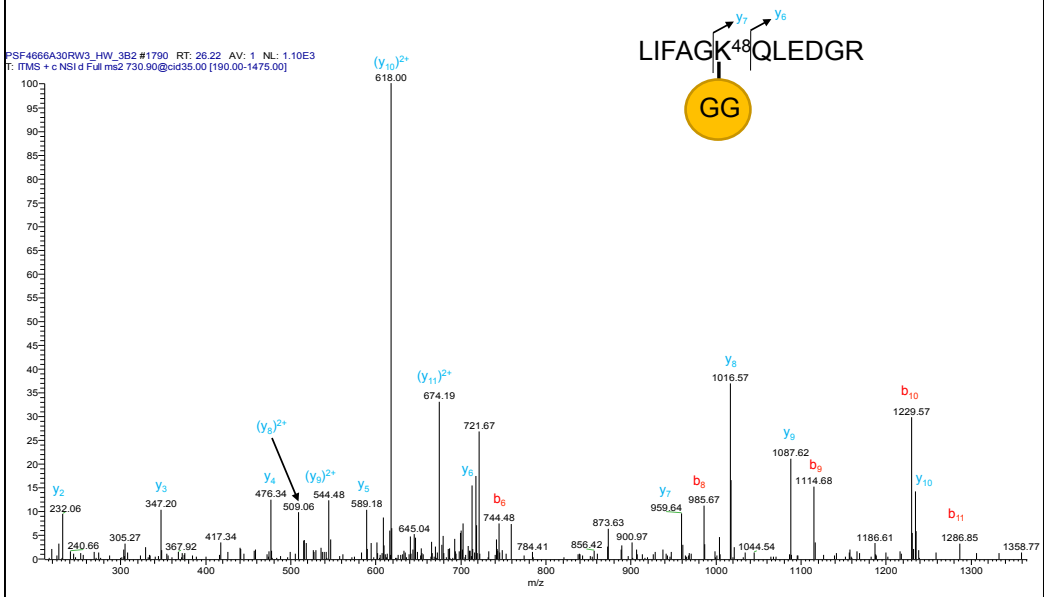
Ubiquitination of Parkin

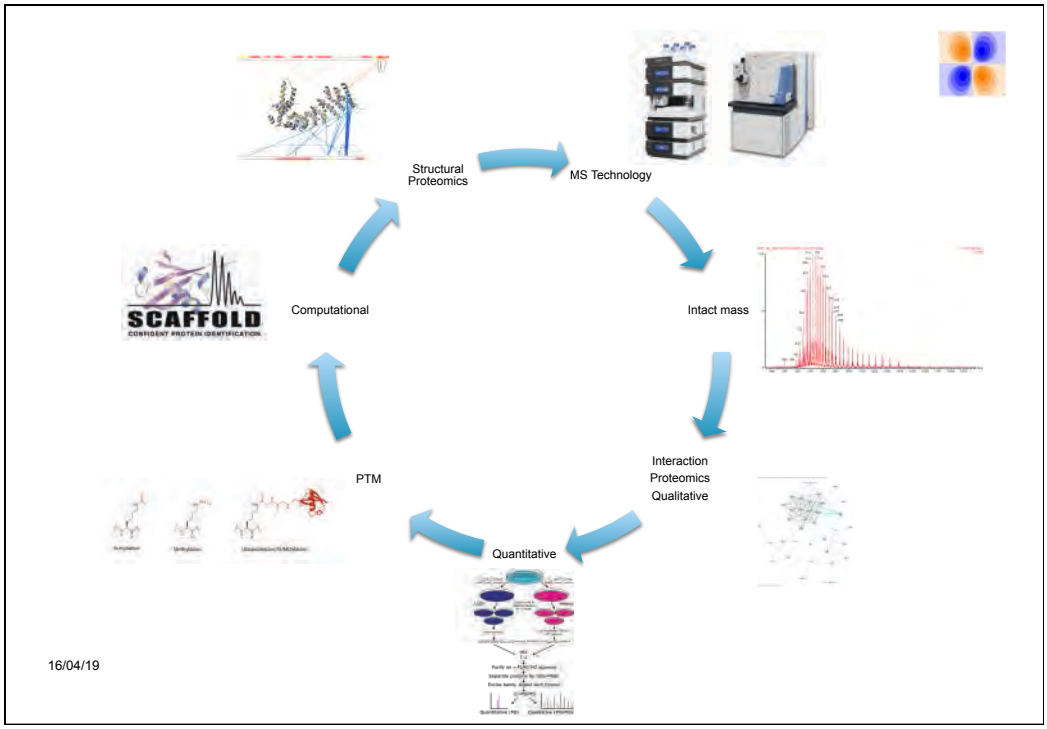


MS/MS of peptide from Parkin showing site of ubiquitination



MS/MS of peptide from ubiquitin showing linkage

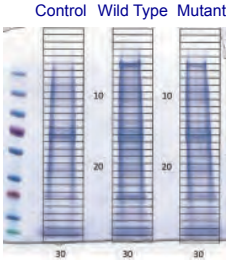


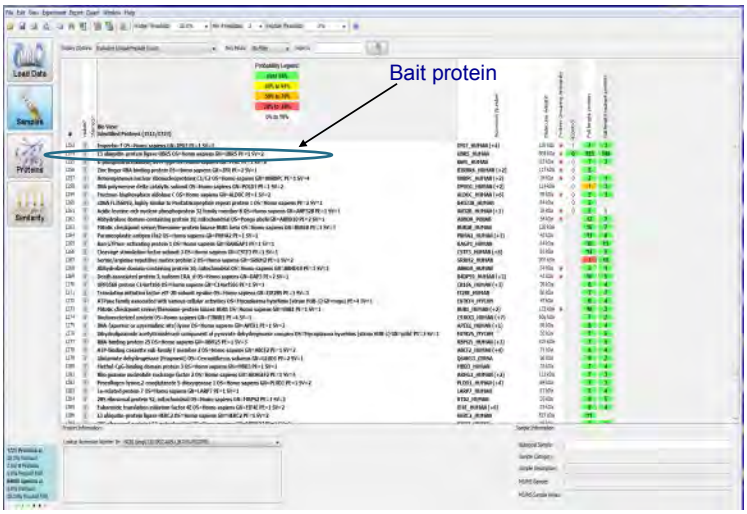


GelcMS/MS

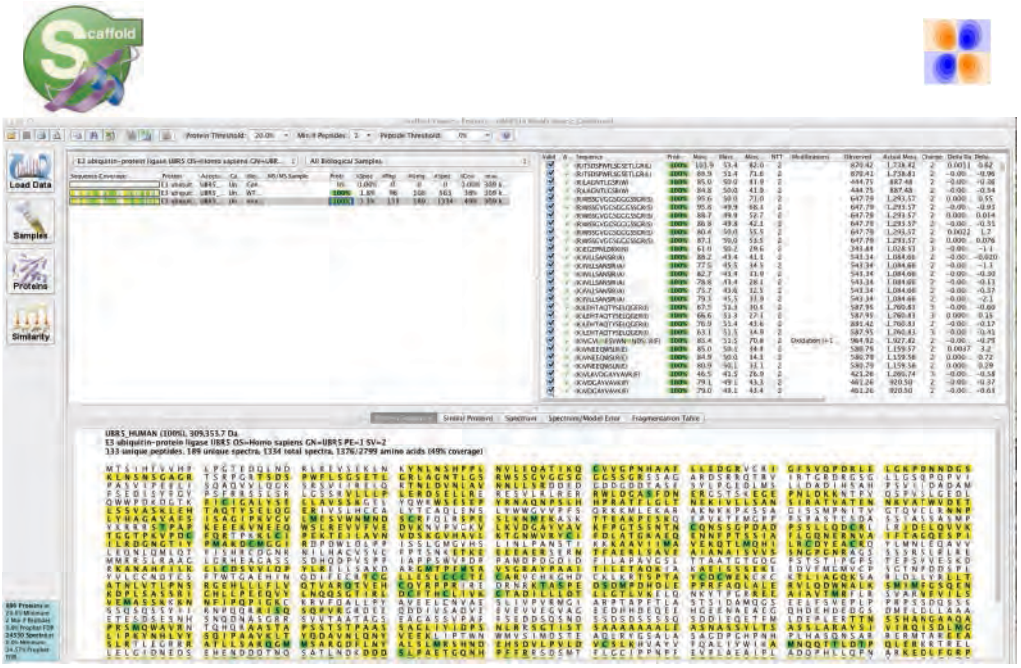
Nikola Novcic
PNAC

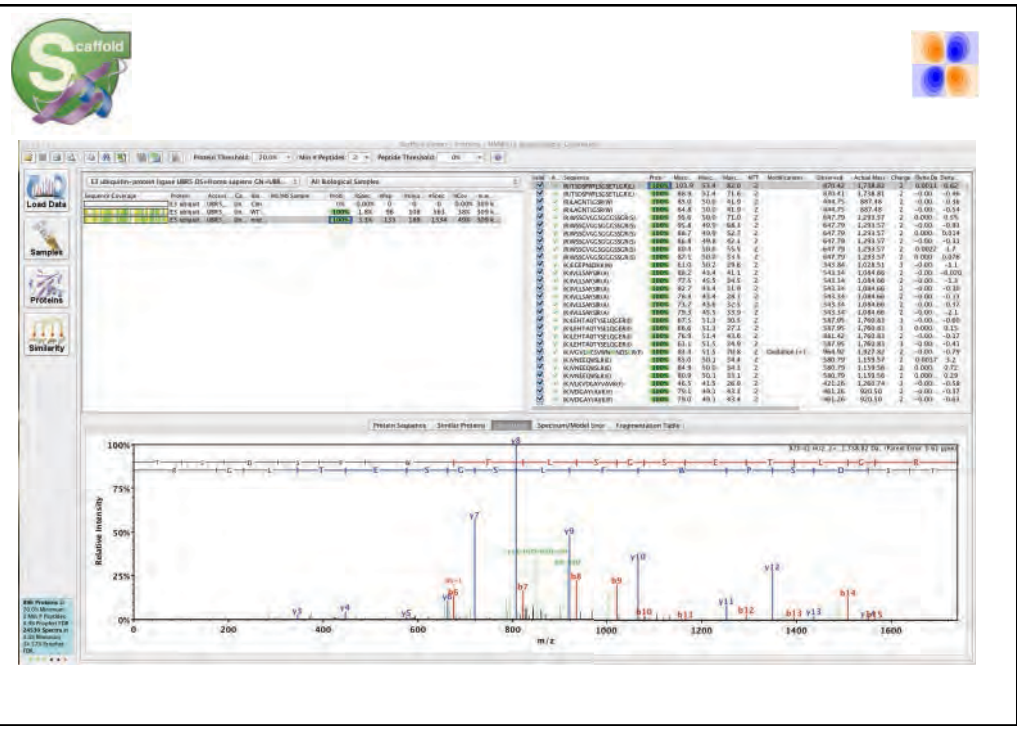
Immunprecipitation using a Flag-tag

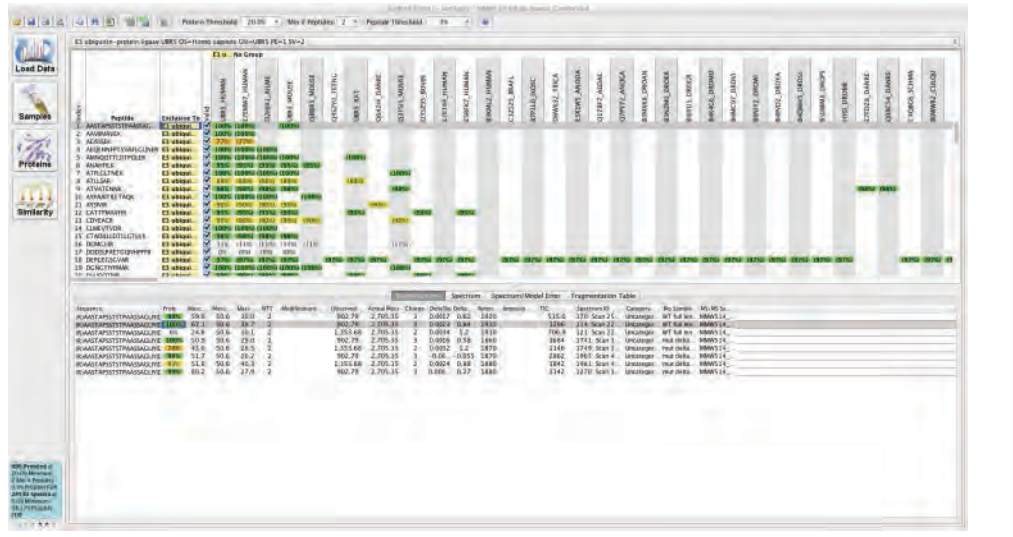




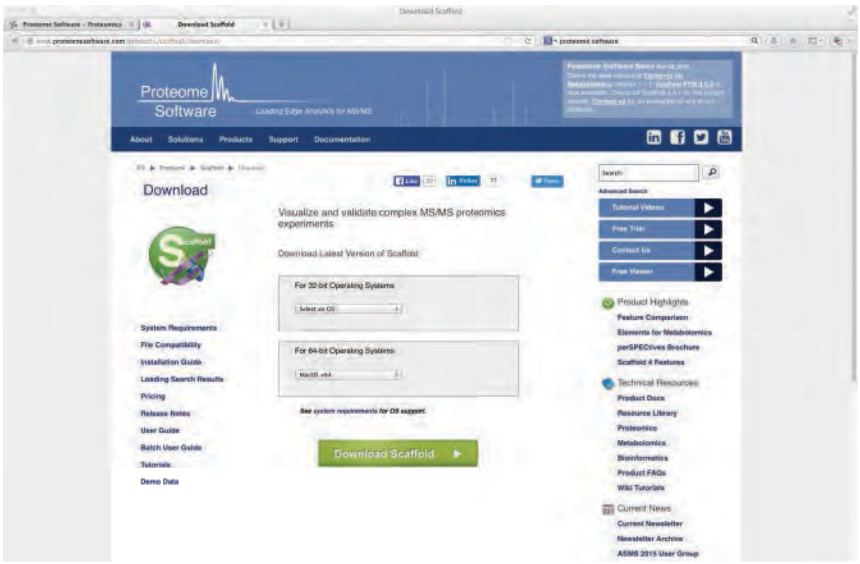
16/04/19




Load Data
Samples
Proteins
Similarity
Relative Intensity
m/z

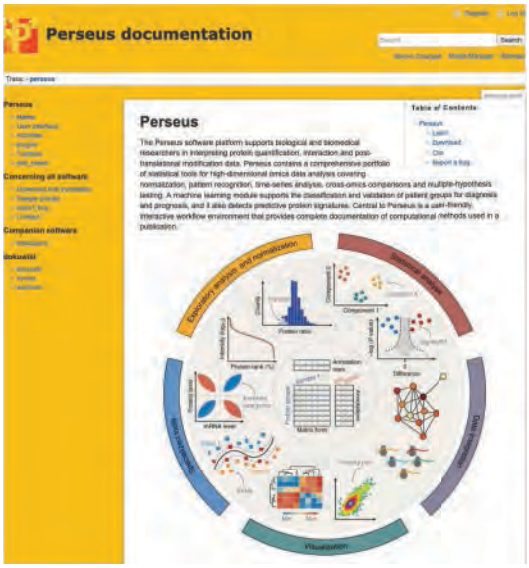

Load Data
Samples
Proteins
Similarity

16/04/19



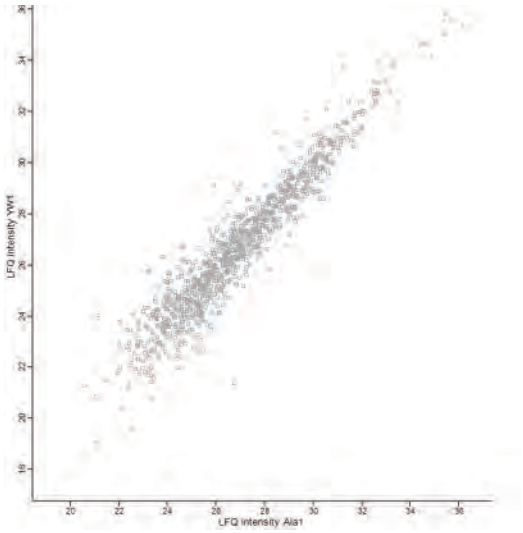
16/04/19

Statistical tools



16/04/19

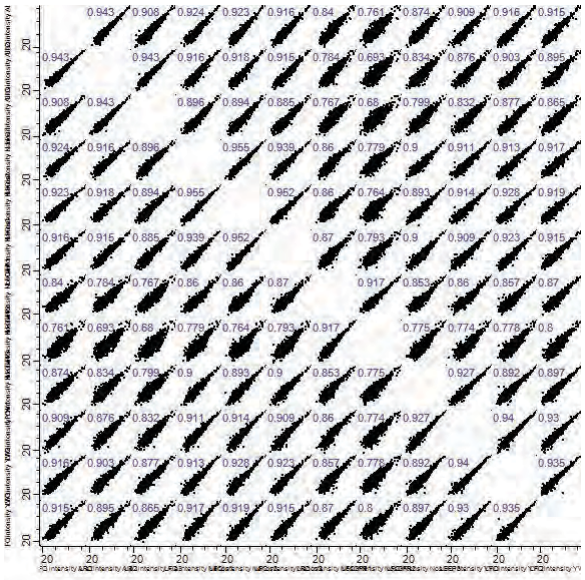
Perseus generated Scatter Plot



16/04/19

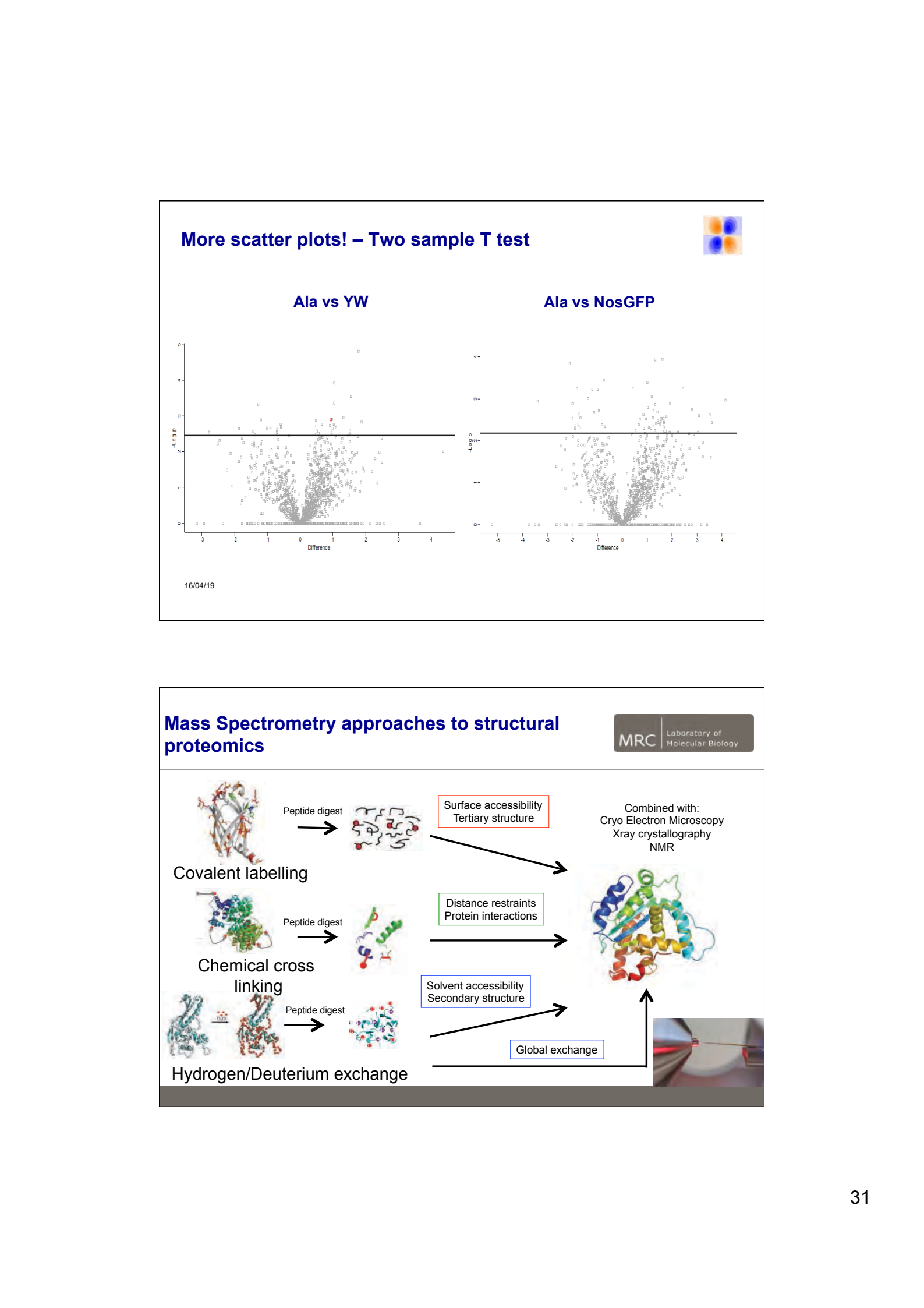
Tyanova, S., et al., Nature Methods, 2016.

Multi-Scatter Plot with R^2 values



16/04/19

More scatter plots! – Two sample T test



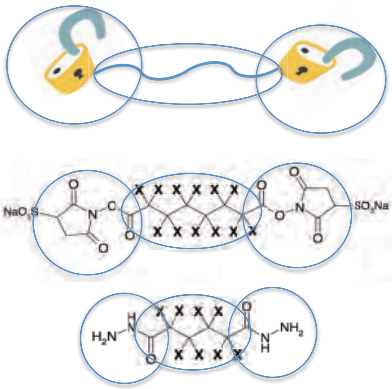
16/04/19

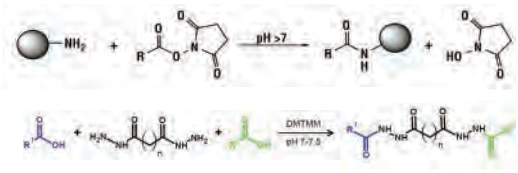
Mass Spectrometry approaches to structural proteomics

MRC Laboratory of Molecular Biology

Cross-Linking and Mass Spectrometry (XL-MS) as a tool for structural investigation

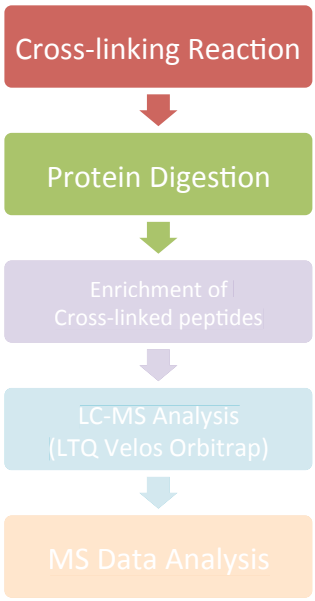
- **Chemical Cross-linking** (XL) is used to covalently fix two or more proteins
- **Mass Spectrometry** allows protein identification and peptide sequencing
- **Cross-linking** and **Mass Spectrometry** provides information about protein-protein interactions within a distance cut-off set by the cross-linker used

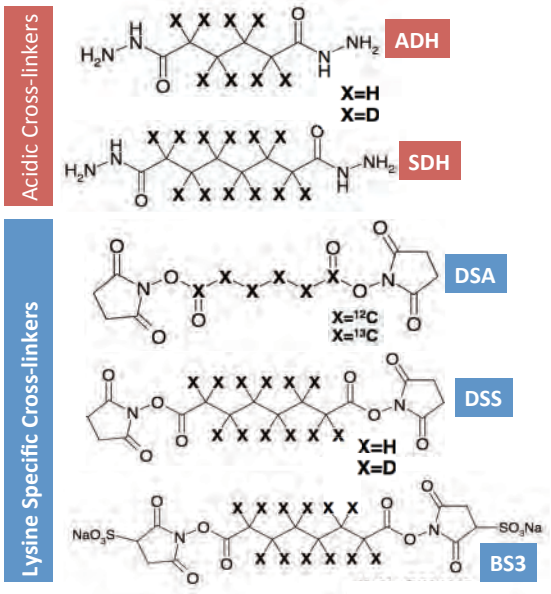


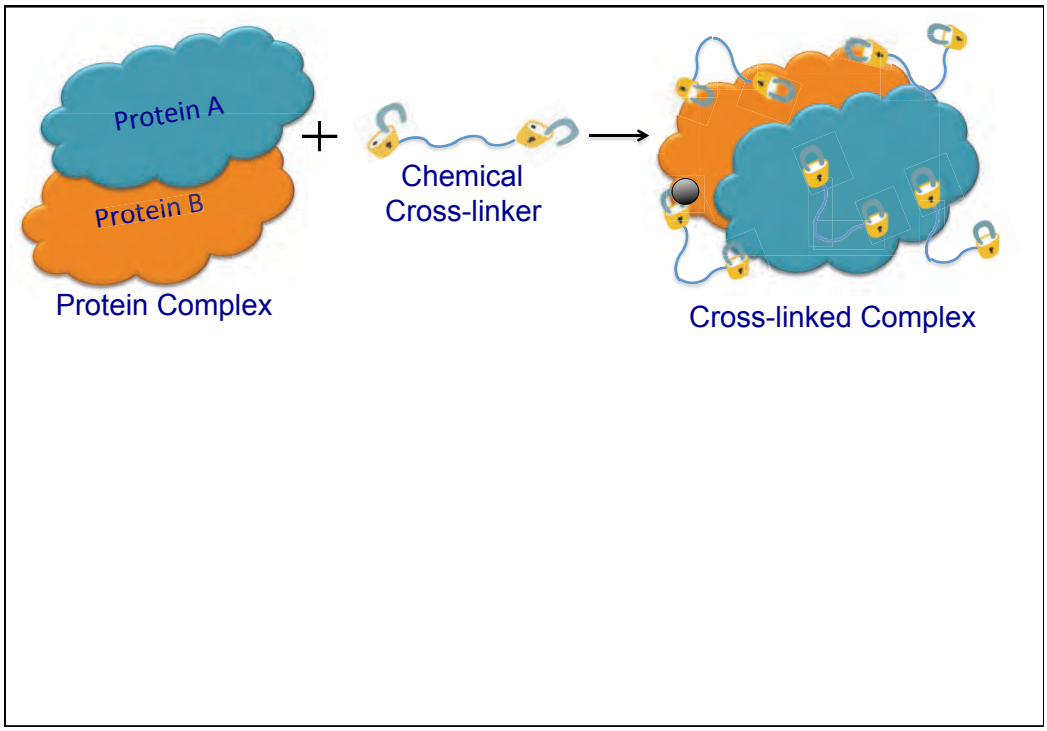


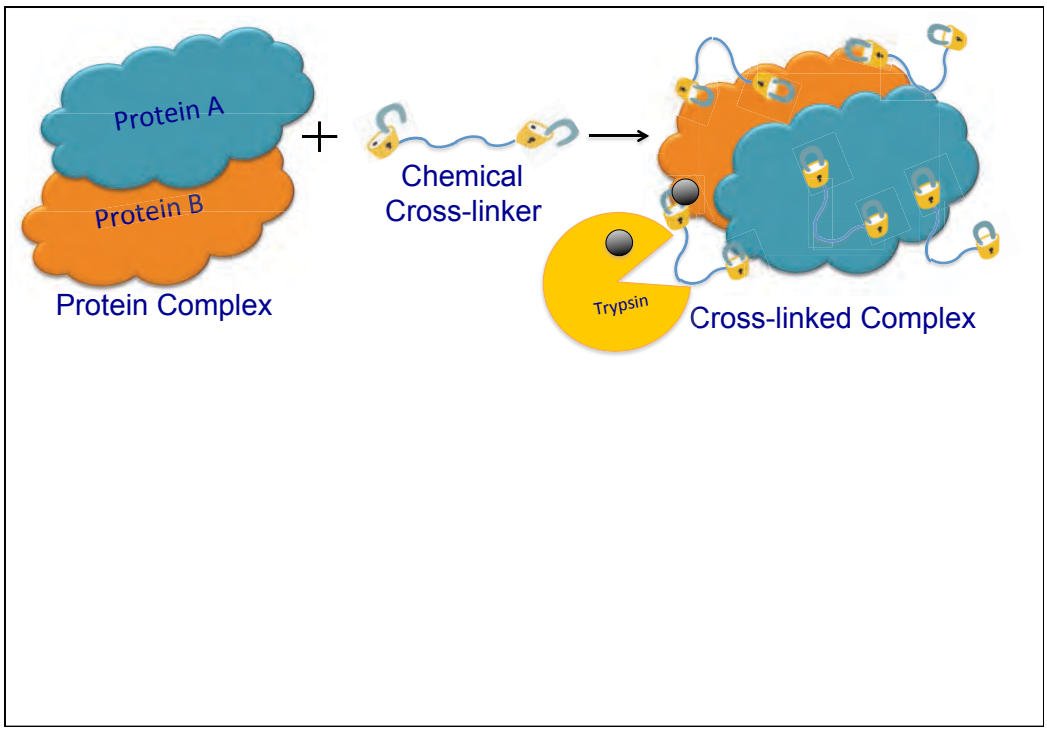
Spacer arm: sets the maximum distance cut-off between two potential reactive residues

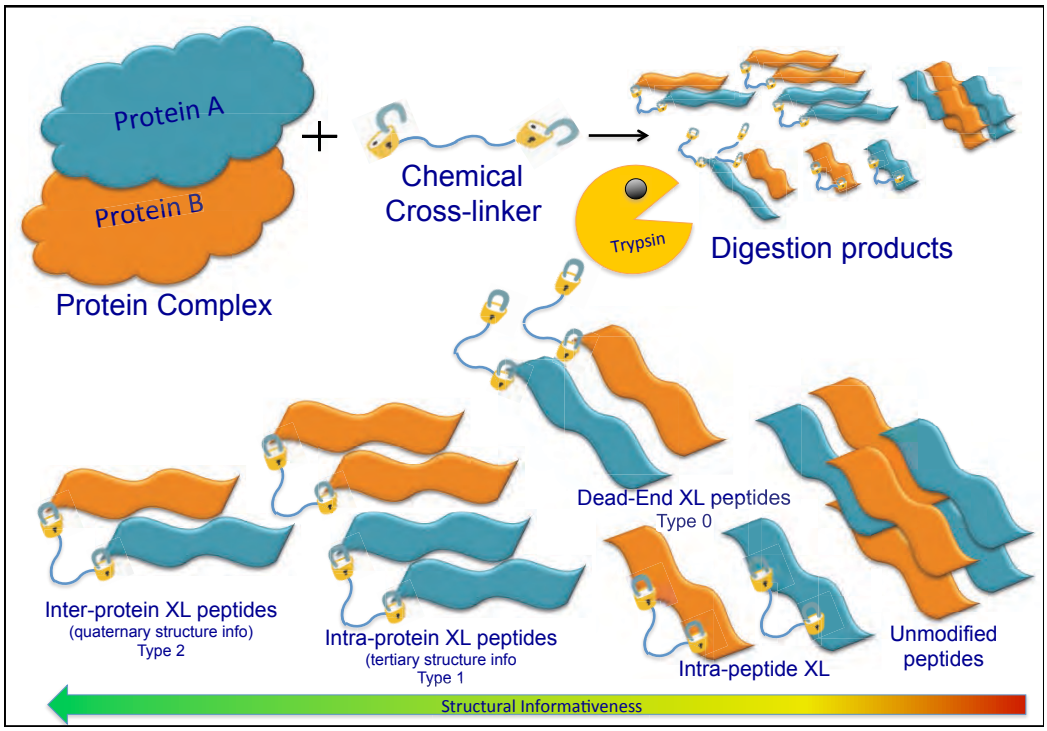
XL-MS Analysis Workflow






Protein A
Protein B
Protein Complex
Chemical
Cross-linker
Cross-linked Complex


Protein A
Protein B
Protein Complex
Chemical
Cross-linker
Trypsin
Cross-linked Complex



XL-MS Analysis Workflow

- Cross-linking Reaction
- Protein Digestion
- Enrichment of Cross-linked peptides
- LC-MS Analysis (LTQ Velos Orbitrap)
- MS Data Analysis

Raw conversion
- msconvert *.raw → *.mzXML

proteowizard

Fasta Database
- 44 Complex I sequences
- 44 Complex I reversed sequences (Decoys)

XL search
- xQuest

Identification of cross-linked peptides from large sequence databases
Aebersold
Nature Methods (2008)

Data Analysis
- Ranking of mono-links and cross-linked peptides
- False Discovery Rate (FDR) analysis
- Manual interrogation of the spectral data

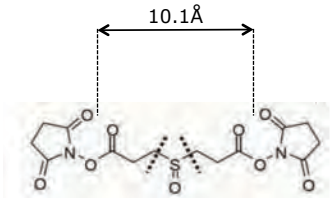
Data reporting
- Results storage
- Cross-link viewer

xiNET

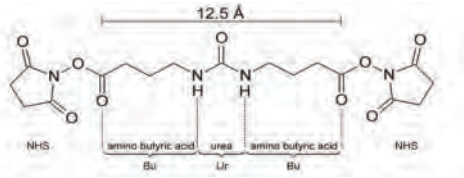
xiNET cross-link network maps with residue resolution
Rappsilber
Mol Cell Proteomics. 2015

New cross linkers

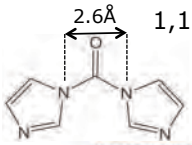
MRC Laboratory of Molecular Biology



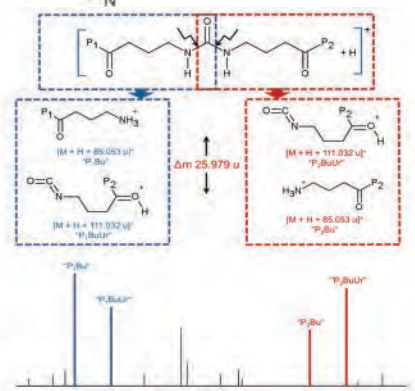
Disuccinimidyl sulphoxide



Disuccinimidyl dibutyric acid (BuUrBu)

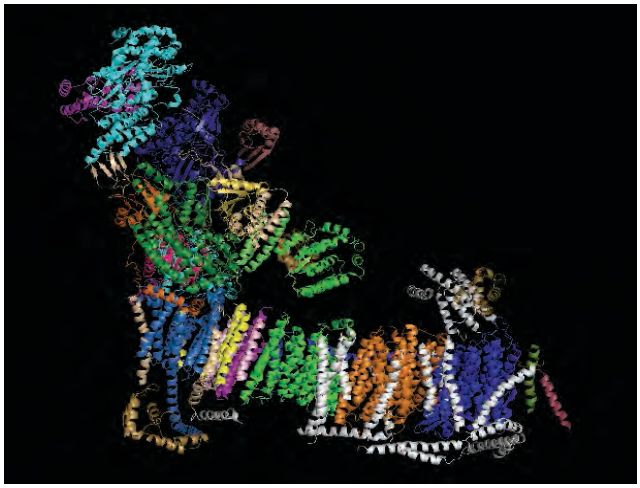


1,1'-carbonyldiimidazole



Atomic structure of the entire mammalian mitochondrial complex I

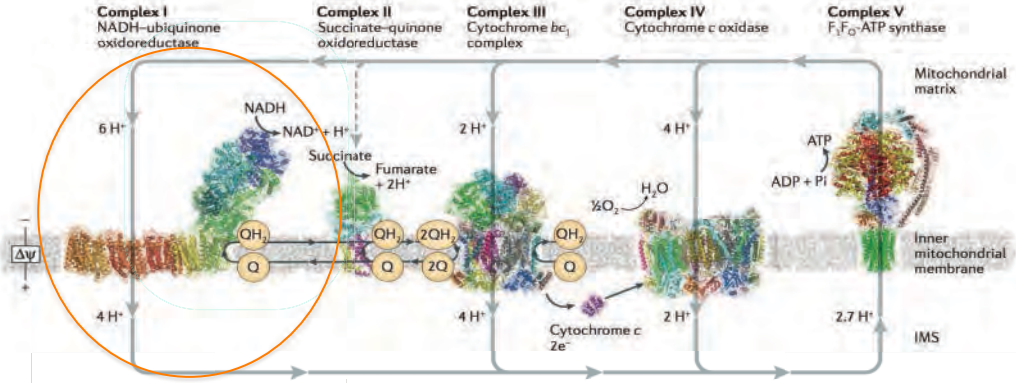
MRC Laboratory of Molecular Biology



Fiedorczuk *et al.*, Nature 2016 Sep 5. doi: 10.1038/nature19794

Mitochondrial Complex I

- **NADH–ubiquinone oxidoreductase**
- **First and largest** enzyme of the respiratory chain
- Two domains:
  - Peripheral arm: **redox** activity
  - Membrane domain: **proton translocation**
- Composed by:
  - **14 Core Subunits** extremely **conserved** during evolution from procariotes to eucariotes
  - **Accessory Subunits** acquired and conserved during the evolution by eucariotes



Supernumerary Subunits Assignment

Subunit Assignment based on

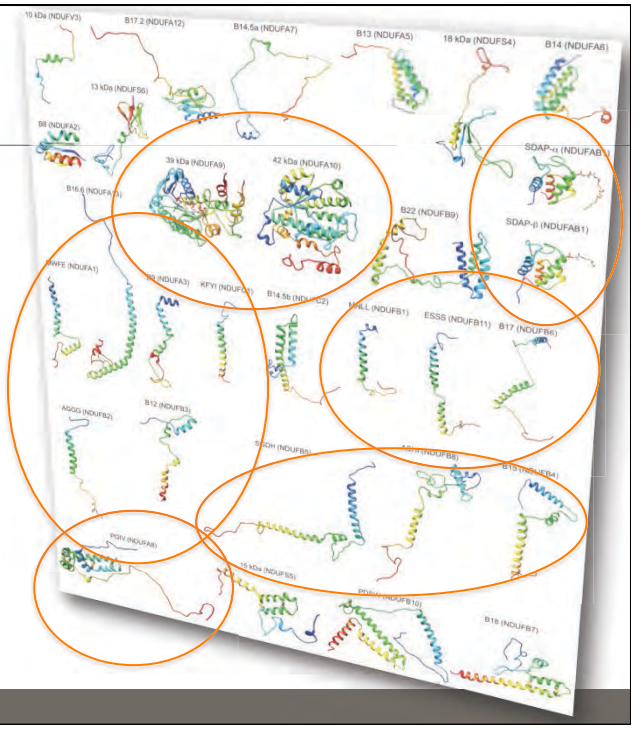
- Secondary structure prediction & modelling
- Indirect biochemical data (e.g. structural or functional interactions)

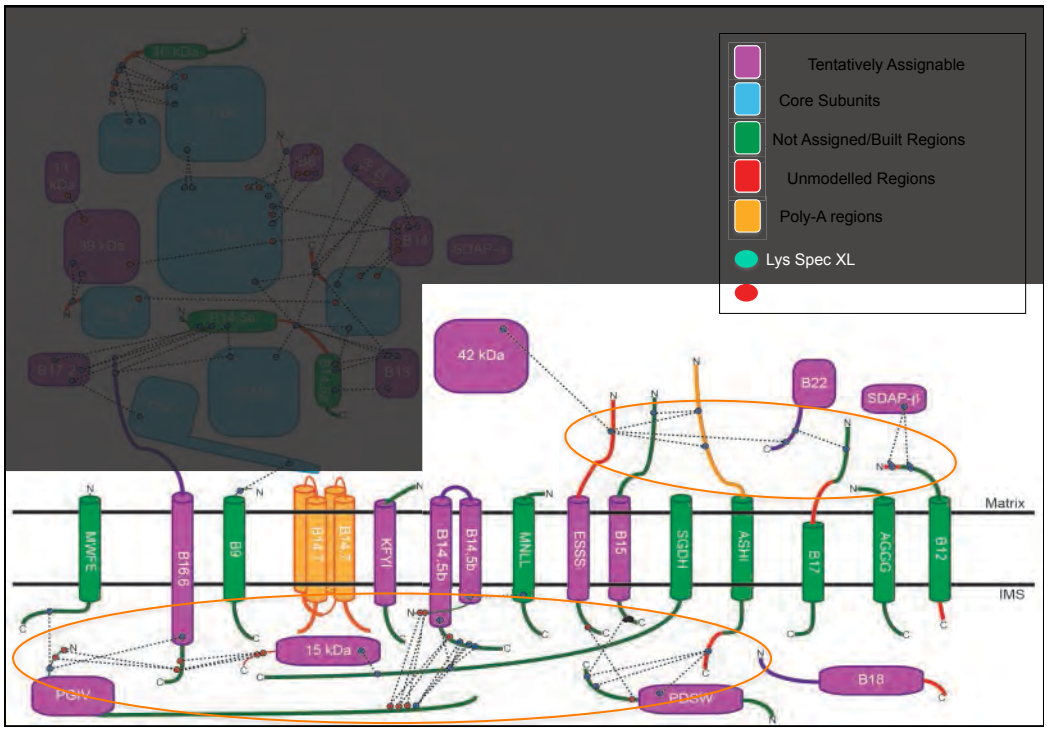
Secondary Structure Prediction & Modelling

- Most of the subunits have an extended structure containing α-helices and coils
- 12 subunits associated to Membrane Domain have single TM-helix domains

How Cross-linking MS may help?

- Confirm the interactions between the tentatively assigned subunits
- Provide insights on protein interactions
  - Orientation of STMD subunits
  - Simplify the assignment of intertwined subunits and unstructured regions





Acknowledgements

MRC Laboratory of Molecular Biology

Biological Mass Spectrometry & Proteomics

Sew Peak-Chew
Farida Begum
Sarah Maslen
Gianluca Degliesposti

Collaborations

Leo Sazanov (Vienna, CxI)
James Letts (Vienna, CxI)