Proteoforms and their expanding role in biomedical research

Julia Chamot-Rooke

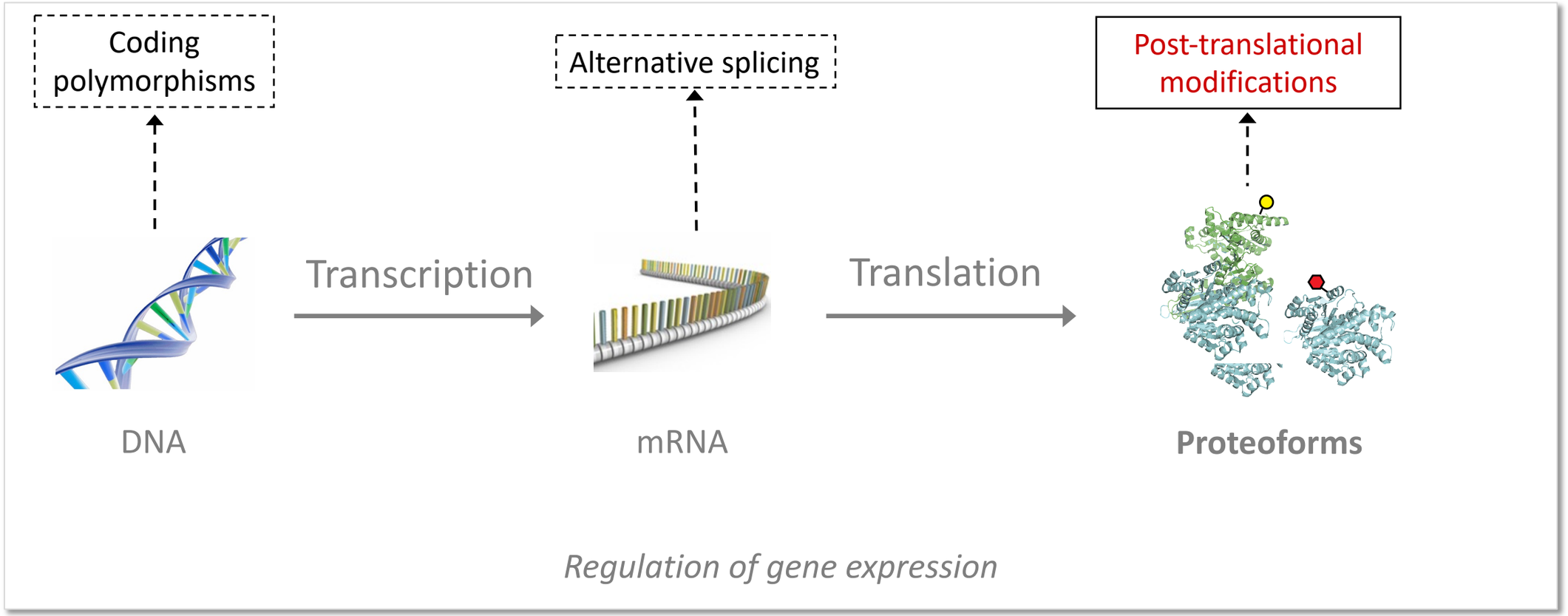
Mass Spectrometry for Biology Unit

Institut Pasteur, Paris (France)

Proteo Vilamoura

11/05/2022

Rationale behind looking at intact proteins

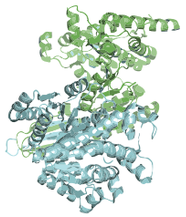


Many protein forms (proteoforms) from a single gene

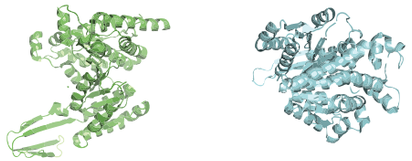
INSTITUT PASTEUR

What is a proteoform?

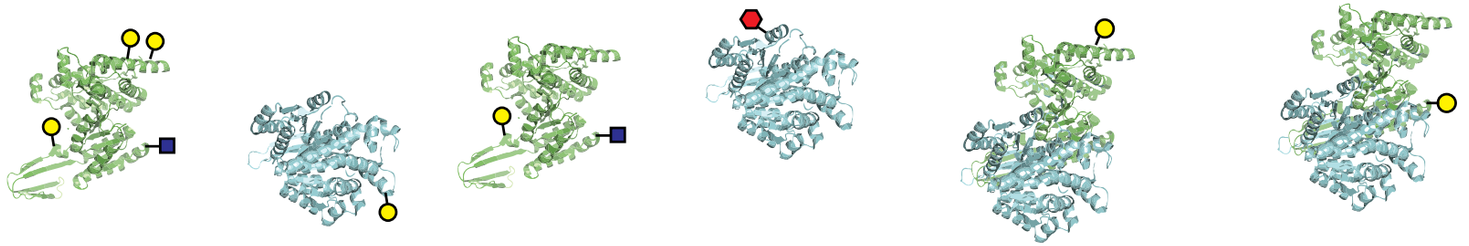
Canonical
Sequence
(UniProtKB)



Endogenous proteolysis
mRNA splicing
Mutations
SNPs



Site specific features:
Govern activity, localization, interactions, half-life



A distinct molecular form of a protein product arising from a single gene

The Consortium for Top-Down Proteomics

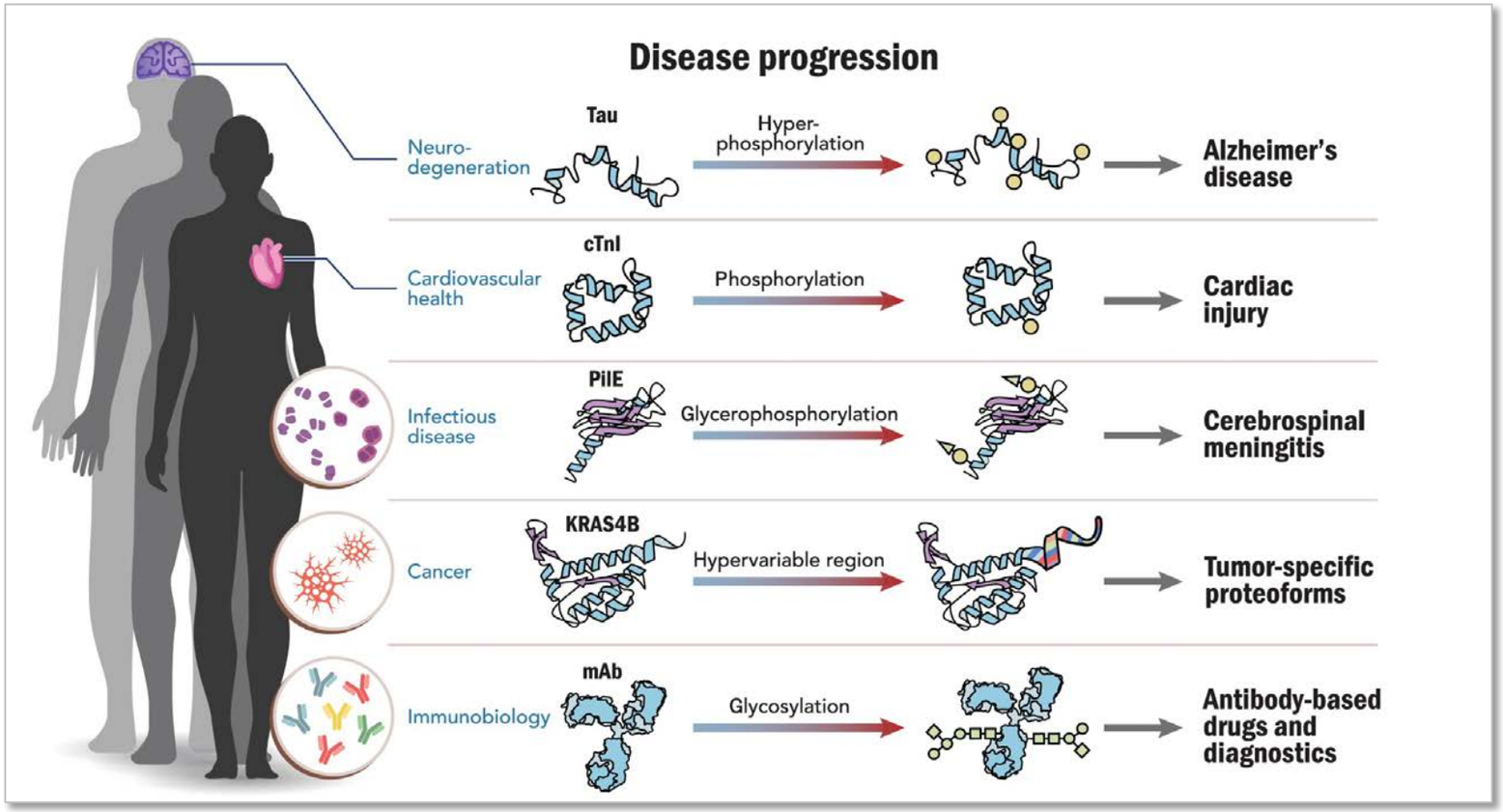
CTDP
Consortium for Top-Down Proteomics
Bringing Proteoforms to Life™

Smith et al. *Nat. Methods,* 10:186–187 (2013)

INSTITUT PASTEUR

Proteoforms in biomedical & clinical research

- Proteoform-level knowledge is essential to understand biological function
- Important clinical areas of interest where proteoforms have been identified and linked to disease progression



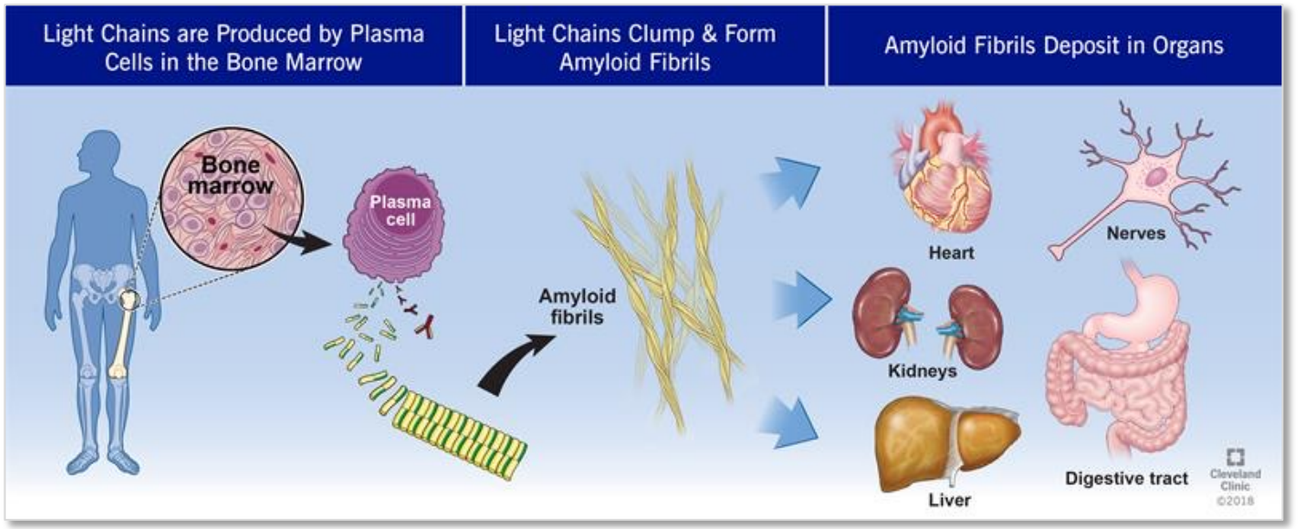
Smith *et al*., *Science Advances* 7, eabk0734 (2021)

CTDP
Consortium for Top-Down Proteomics
Bringing Proteoforms to Life™

INSTITUT PASTEUR

Multiple Myeloma & Light Chain Disease

- Multiple myeloma: malignancy of plasma cells characterized by a clonal expansion of abnormal B-cells
- B-cells accumulate in the bone marrow and secrete large amounts of monoclonal light chains that can deposit in organs such as kidneys leading to:
  - Light Chain Deposition Disease (LCDD) for aggregates of amorphous nature
  - AL-amyloidosis, where aggregates consist of amyloid fibrils



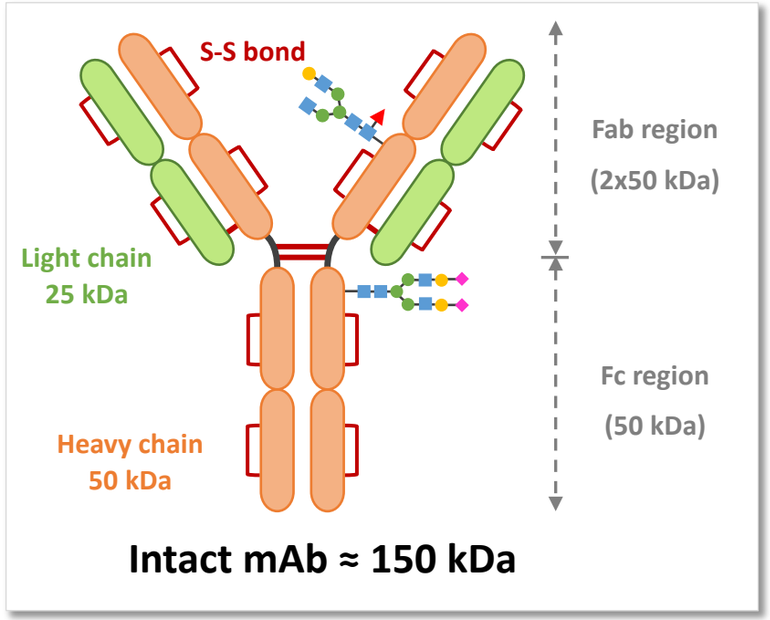
Amyloid light-chain (AL) amyloidosis

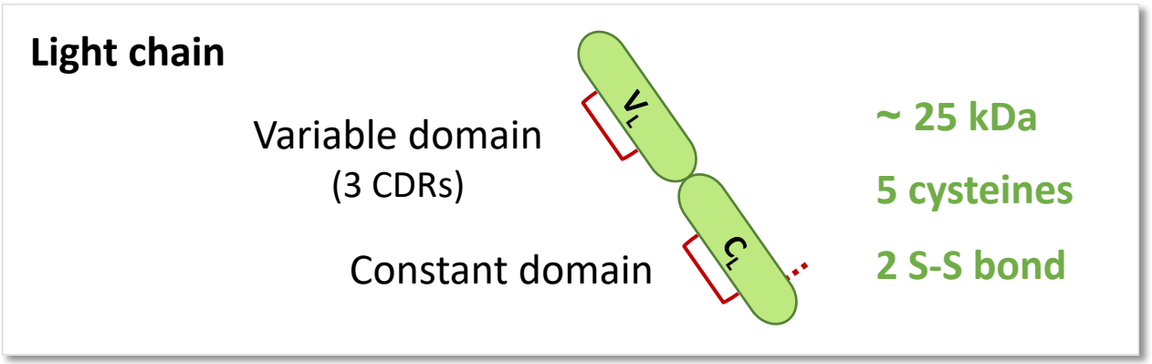
https://my.clevelandclinic.org/health/diseases/15718-amyloidosis

INSTITUT PASTEUR

Light Chain (LC) analysis

- Currently no possibility to predict the *in vivo* solubility and deposition behavior of a particular LC
- No link between LC abundance and disease severity
- To understand the factors affecting solubility and develop diagnostic tools, LC sequence is essential
  - RNA sequencing of B-cell clones
  - Mass Spectrometry





INSTITUT PASTEUR

Sequencing of mAbs by MS

- MS extensively used for the characterization of recombinant mAbs (BUP with different enzymes)
- For unknown mAbs:
  - Combination of BUP and intact mAb mass profiling
  - Combination of BUP and middle-down proteomics on Fab (50 kDa) after sample fractionation
  - Top-down proteomics on LCs extracted from human sera for classifying plasma cell disorders (21T FT-ICR MS)

Limitation of current methods

- Gaps in sequences
- 70% sequence coverage for LC study (TDP)
- No I/L differentiation, S-S assignment

Our objective

Develop a complete *de novo* MS sequencing workflow for patient-derived LC proteoforms

Collaboration A. Büll (Univ. Düsseldorf)

EPIC-XS

Peng *et al.,* J. *Proteome Res.* 20(7):3559-3566 (2021)
He *et al., Anal. Chem*. 91, 3263−3269 (2019)

Bondt *et al., Cell Systems* 12(12):1131-1143 (2021)

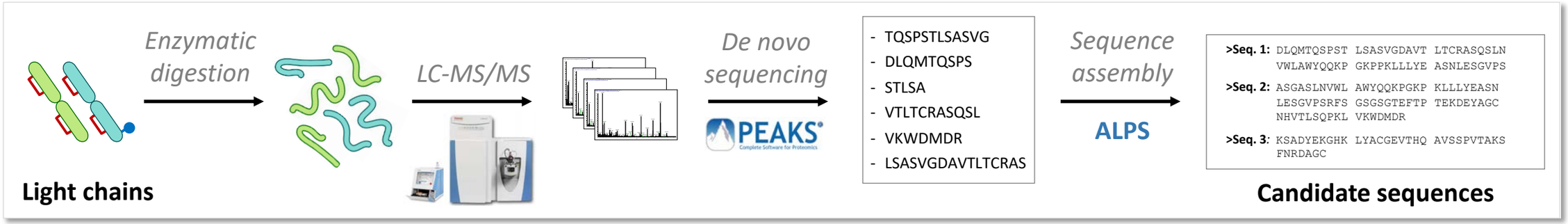
INSTITUT PASTEUR

De novo sequencing: step 1 (multiple digestions)


M. Dupré


M. Duchateau


M. Rey



- Enzymatic digestions
  - Specific cleavage: Trypsin (K, R), Lys C (K)
  - Non-specific cleavage: Pepsin, *Nepenthes* fluid

- Sequence assembly

Tran N.H. *et al.*, *Scientific Reports* 6:31730 (2016)

Rey M *et al.*, *Mol. Cell. Proteomics* 12 (2), 464–472 (2013)

INSTITUT PASTEUR

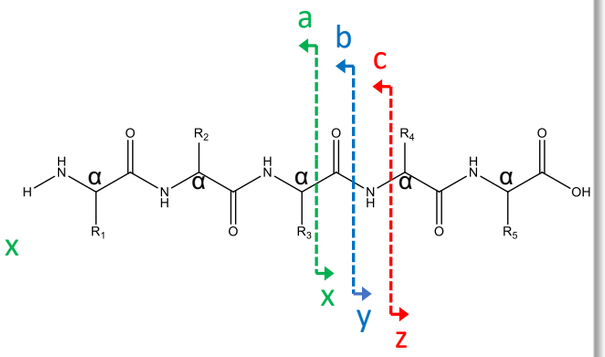
De novo sequencing: step 2 (analysis of intact proteins)

- Sample preparation: **no digestion!**
  +/- [reduction, alkylation] of S-S bonds

- LC-MS/MS on Orbitrap Fusion Lumos
  Multiple fragmentation techniques

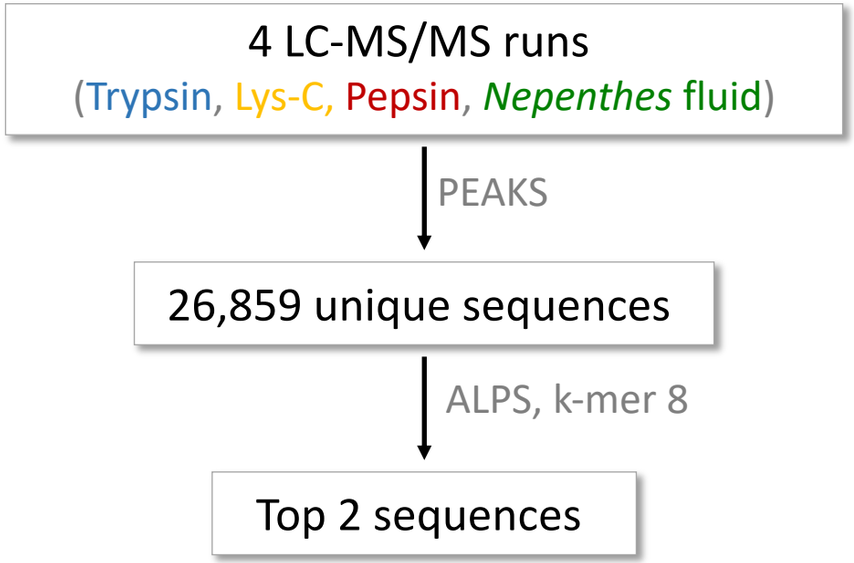
- Data analysis
  Xtract (FreeStyle), ProsightLite, 5 ppm

Frag. modes	Ion types
CID / HCD	b, y
ETD	c, z
EThcD	b, y, c, z
UVPD	b, y, c, z, a, x



INSTITUT PASTEUR

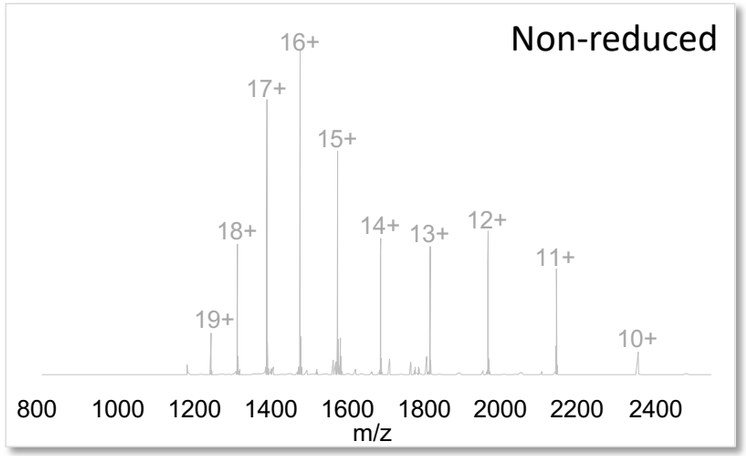
Example for the P15 sample



Candidate seq. 1 M_{theo}: **23,461.50 Da**	DLQMTQSPST LSASVGDAVT LTCRASQSLN VWLAWYQQKP GKPPKLLLYE AS**N**LESGVPS RFSGSGSGTE FTLTLSSLQP DDFATYYCQQ YNSYPYTFGQ GAKLELKRTV AAPSVFLFPP SDEQLKSGTA SVVCLLNNFY PREAKVQWKV DNALQSGNSQ ESVTEQDSKD STYSLSSTLT LSKADYEKHK LYACEVTHQG LSSPVTKSFN RGEC
Candidate seq. 2 M_{theo}: **23,462.47 Da**	DLQMTQSPST LSASVGDAVT LTCRASQSLN VWLAWYQQKP GKPPKLLLYE AS**D**LESGVPS RFSGSGSGTE FTLTLSSLQP DDFATYYCQQ YNSYPYTFGQ GAKLELKRTV AAPSVFLFPP SDEQLKSGTA SVVCLLNNFY PREAKVQWKV DNALQSGNSQ ESVTEQDSKD STYSLSSTLT LSKADYEKHK LYACEVTHQG LSSPVTKSFN RGEC

INSTITUT PASTEUR

P15: Intact mass analysis



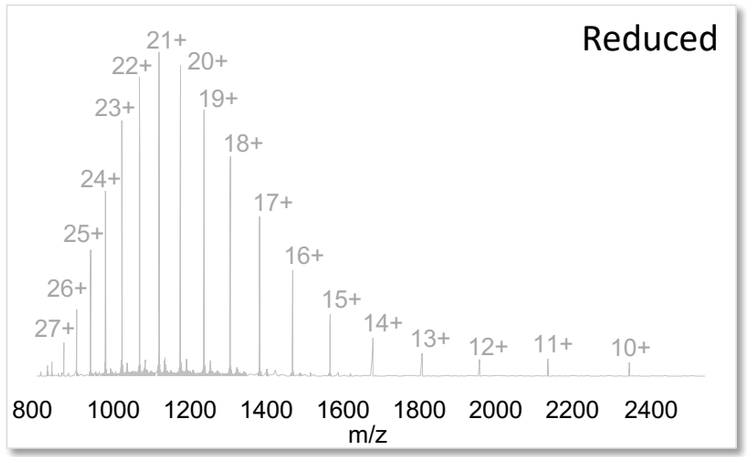
M=23,576.48 Da

Reduction

Δm= -114.93 Da

Δm = +4.03 Da ➔ 2 disulfide bonds

Δm = -118.96 Da ➔ cysteinylation (119.00 Da)



M=23,461.55 Da

Match with sequence 1

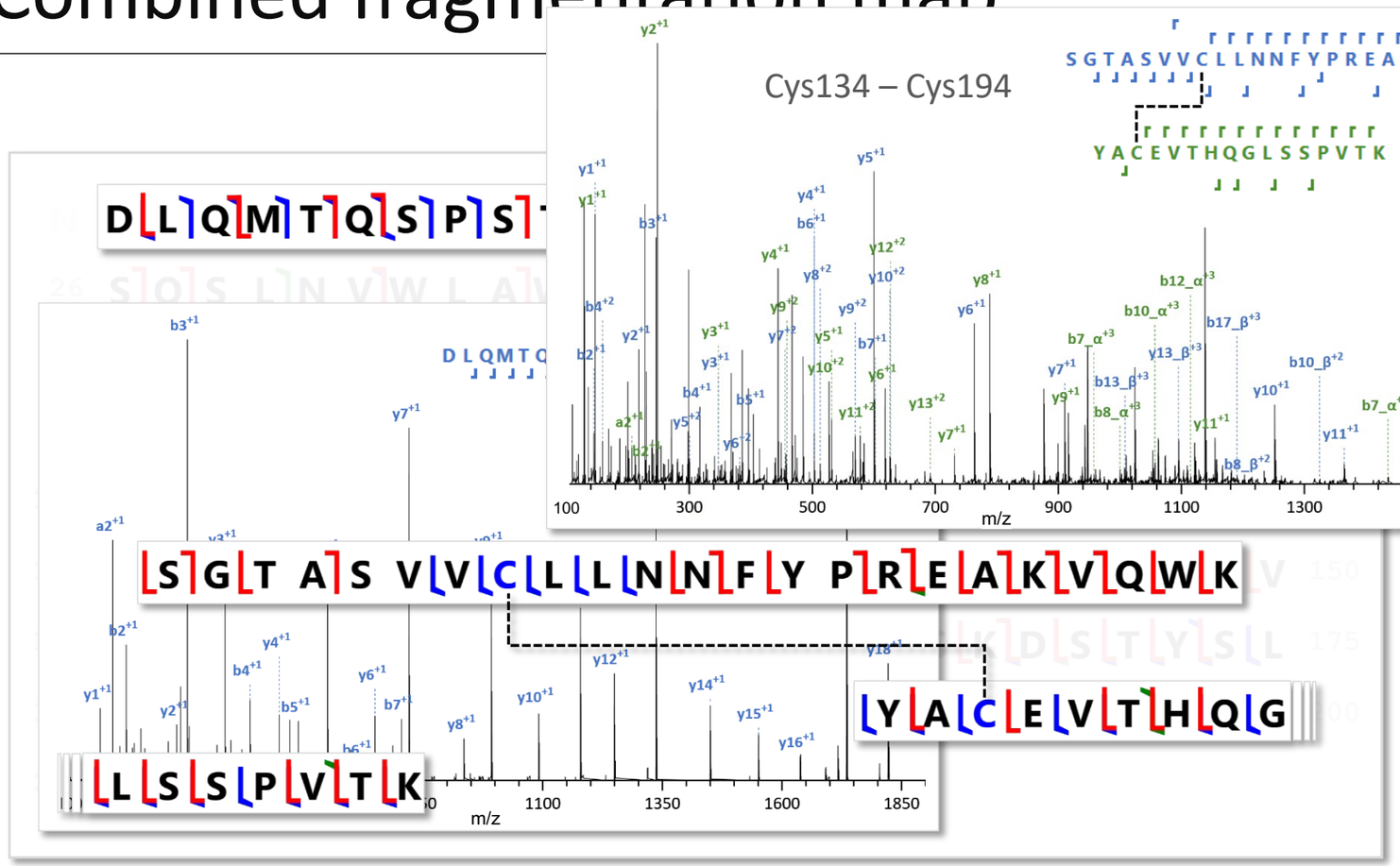
Sequence 1
+ 2 disulfide bonds
+ 1 cysteinylation

INSTITUT PASTEUR

P15: Top-down fragmentation maps

INSTITUT PASTEUR

P15: Combined fragmentation map



Sequence 1

2 S-S + 1 cysteinylation

Δmass (Exp. / Theo): **0.95 ppm**
nr fragments: **375**
Residue cleavage: **85%**

c/z fragment ions
b/y fragment ions
a/x fragment ions

5 ppm fragment mass tolerance

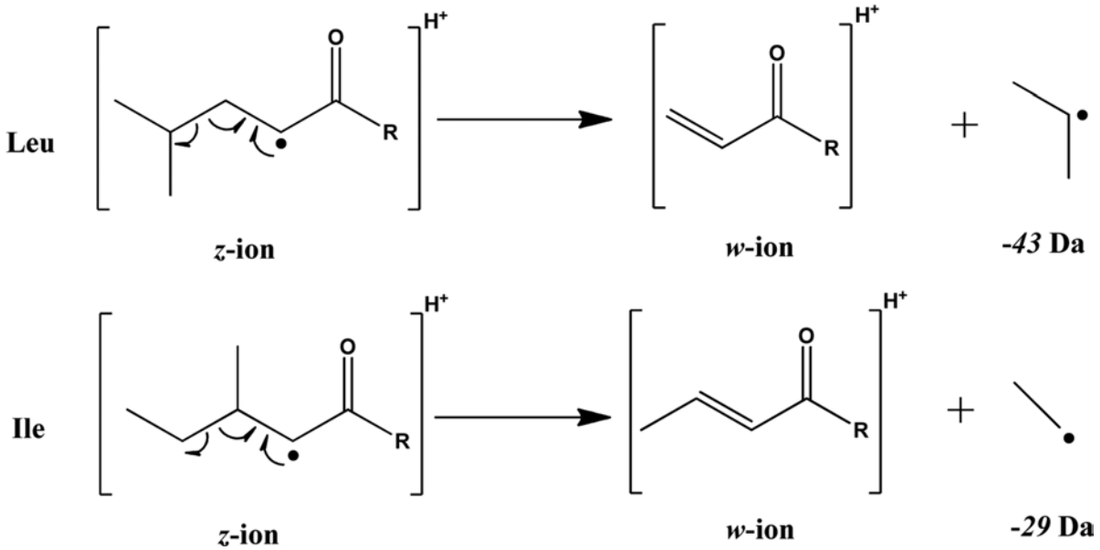
- Analysis of trypsin digest (+/- reduction/alkylation)
  - 100% sequence coverage
  - Disulfides: Cys23 – Cys88, Cys134 – Cys194
  - Cysteinylation on Cys214

Sarpe V, *et al.*, *Mol. Cell. Proteomics* 15,3071-80 (2016)

INSTITUT PASTEUR

P15: Ile/Leu Differentiation

z to w ion
EThcD

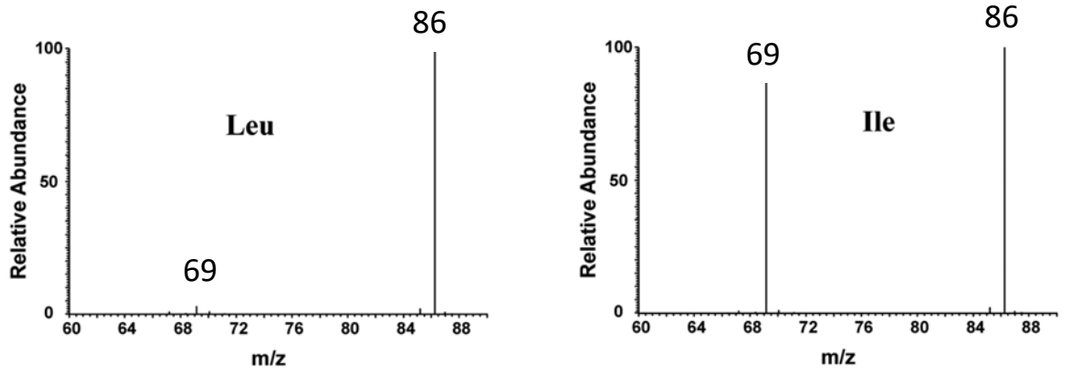


EThcD to generate z and w ions

$\Delta w\text{-}z_L$ = - 43 Da

$\Delta w\text{-}z_I$ = - 29 Da

Immonium ion
HCD



Small b or y ion containing I or L

↓ HCD (MS3)

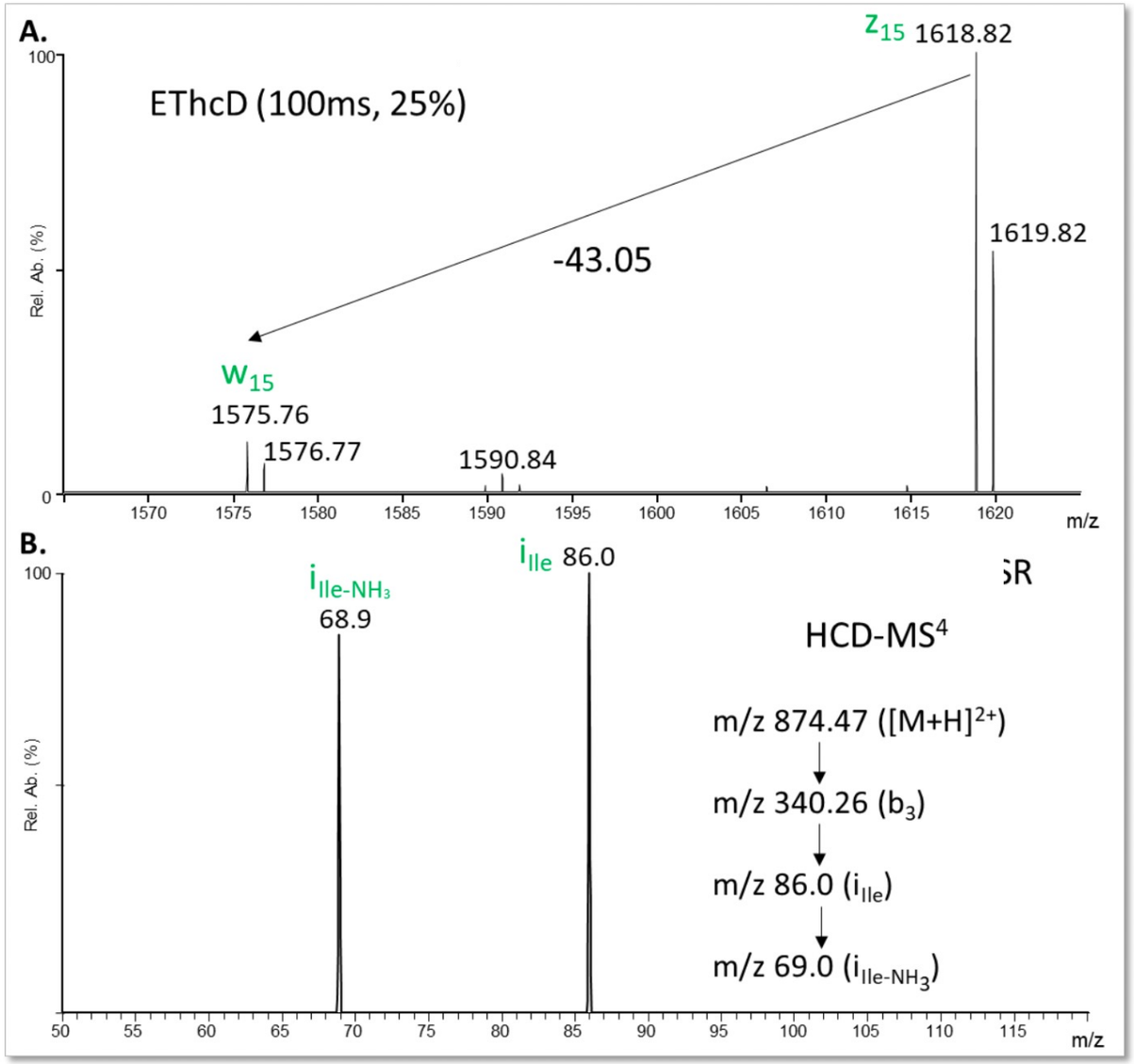
m/z 86

↓ HCD (MS4)

m/z 69

INSTITUT PASTEUR

P15: Ile/Leu Differentiation



LLIYEASNLESGVPSR

LLIYEASNLESGVPSR

INSTITUT PASTEUR

P15 Final Sequence

κ isotype

CDR1 CDR2

DIQMTQSPSTLSASVGDAVTITCRASQSLNVWLAWYQQKPGKPPKLLIYEASNLESGVP

CDR3

SRFSGSGSGTEFTLTISSLQPDDFATYYCQQYNSYPYTFGQGAKLEIKRTVAAPSVFIF

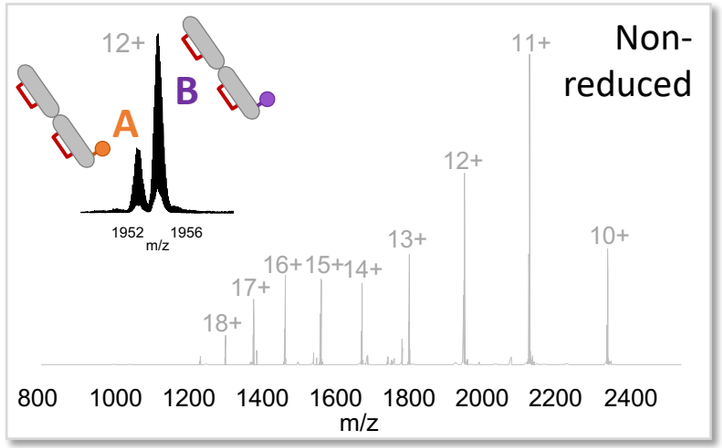
PPSDEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVTEQDSKDSTYSLSS

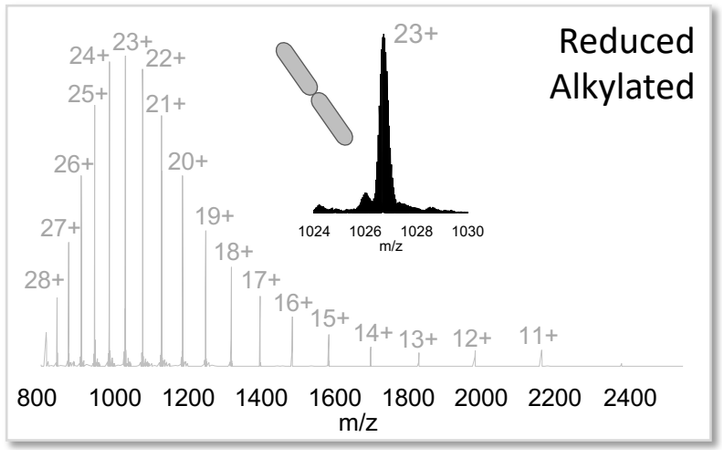
TLTLSKADYEKHKLYACEVTHQGLSSPVTKSFNRGEC*

* Cysteinylation

P8 Analysis

Intact mass analysis





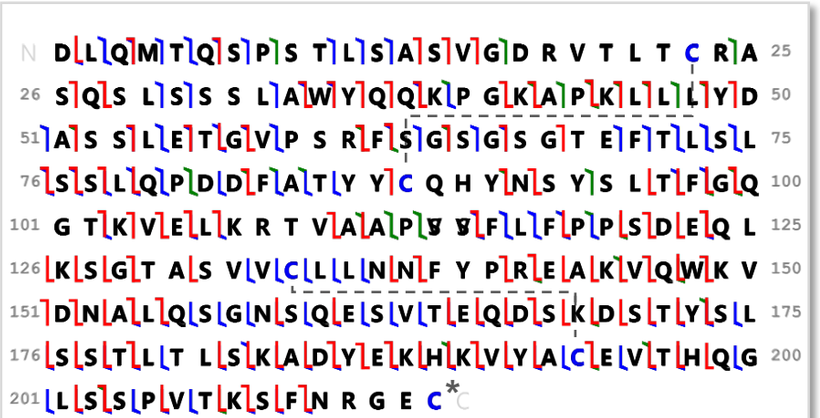
Two proteoforms @:
23,407.35 Da
23,428.39 Da

Δ= +170.13 Da *Reduction*
Δ= +149.09 Da *Alkylation*

One proteoform @:
23,577.48 Da

A single sequence with modified Cys

De novo top-down fragmentation analysis



ΔMass (Exp. / Theo): **-1.03 ppm**
nr fragments: **468**
Residue cleavage: **86%**

Trypsin digest

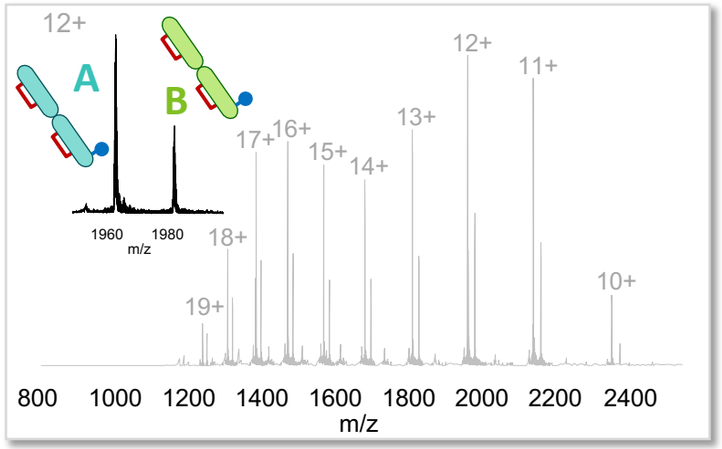
Disulfides: Cys23 – Cys88 & Cys134 – Cys194
Cys214*: Cysteinylation or CoenzymeM (CoM)

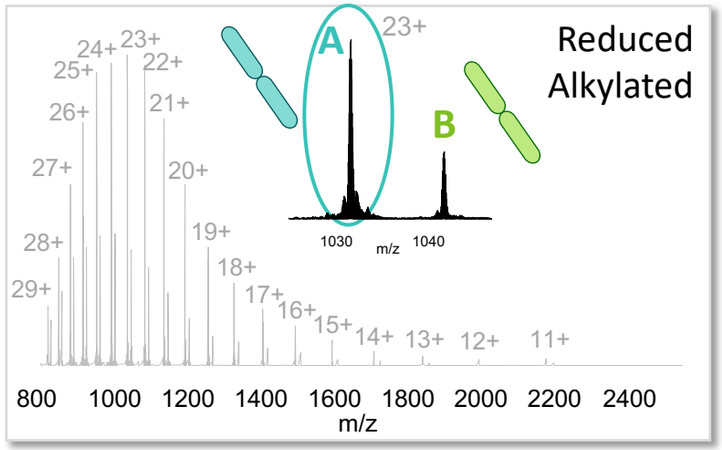
CoM: adjuvant in chemotherapy

INSTITUT PASTEUR

P5 Analysis

Intact mass analysis





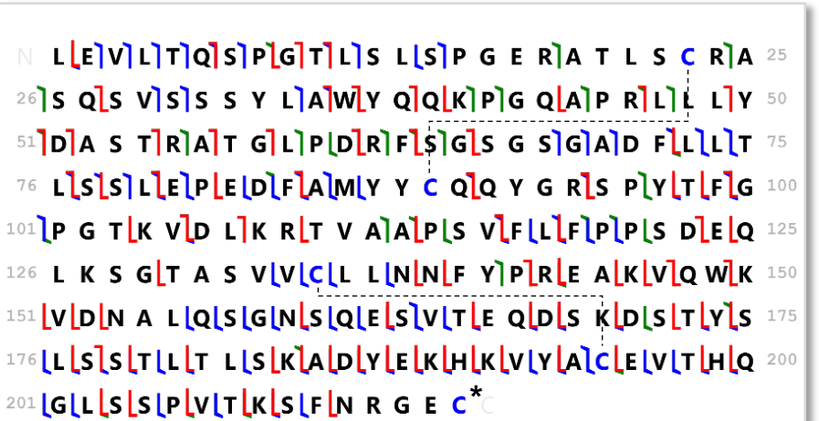
Two proteoforms @:
23,543.49 Da
23,776.61 Da

Reduction
Alkylation

Two proteoforms @:
23,713.62 Da
23,946.81 Da

Two different sequences modified Cys

Top-down fragmentation map (5A)



Δmass (Exp. / Theo): **0.46 ppm**
Residue cleavage: **83%**

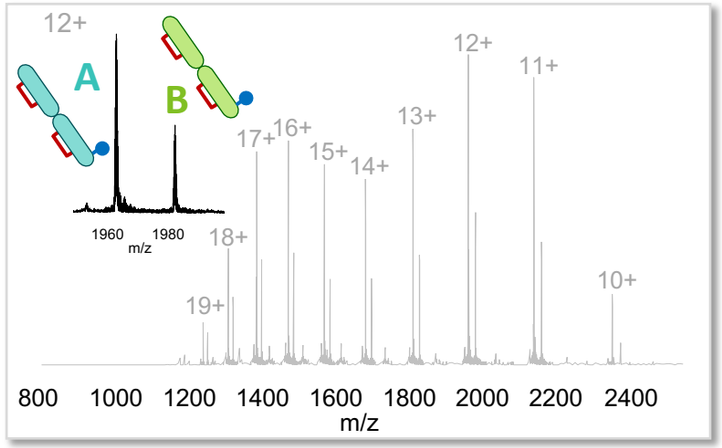
Trypsin digest

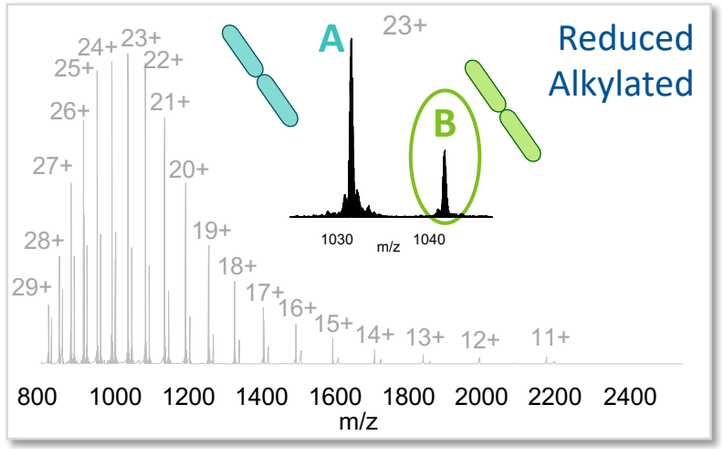
Disulfides: Cys23 – Cys89 & Cys135 – Cys195
Cys215*: Cysteinylation

INSTITUT PASTEUR

P5 Analysis

Intact mass analysis





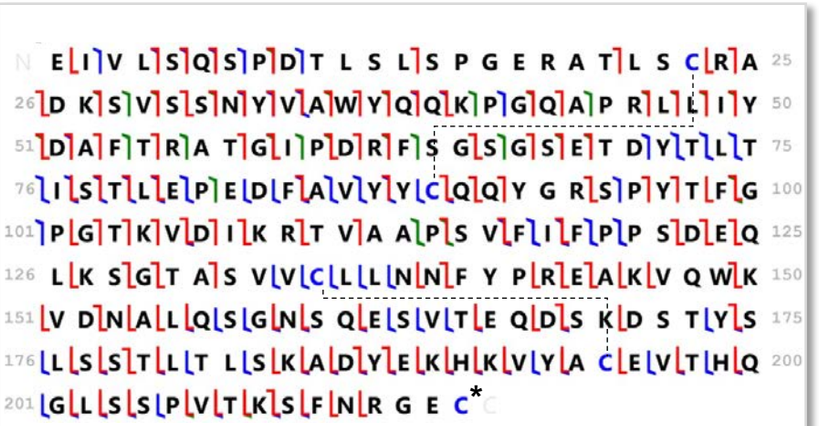
Two proteoforms @:
23,543.49 Da
23,776.61 Da

Reduction
Alkylation

Two proteoforms @:
23,713.62 Da
23,946.81 Da

Two different sequences modified Cys

Top-down fragmentation map (5B)



Δmass (Exp. / Theo): **0.6 ppm**
Residue cleavage: **80%**

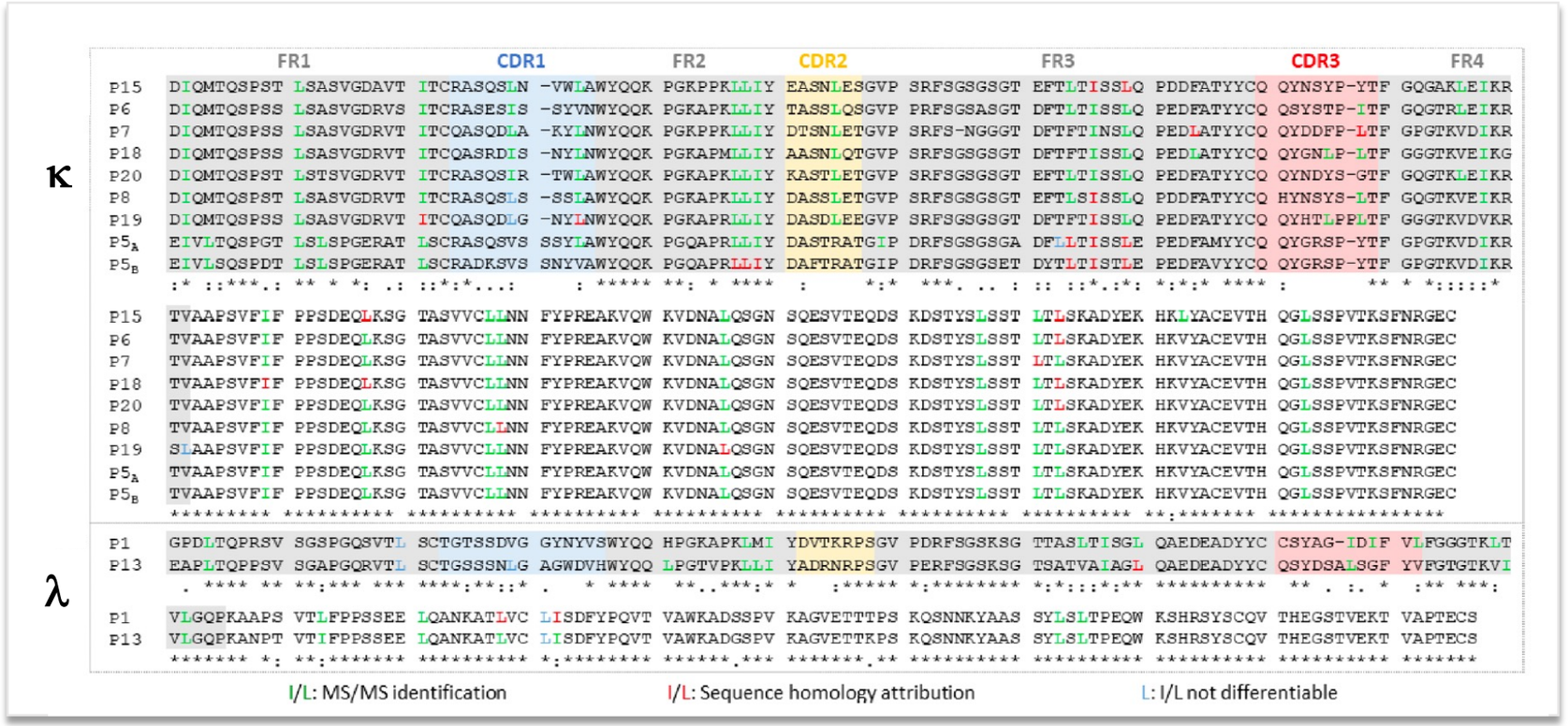
Trypsin digest

Disulfides: Cys23 – Cys88 & Cys134 – Cys194
Cys215*: Cysteinylation
94% identity between 5A and 5B

INSTITUT PASTEUR

Summary

- Light chains extracted for the urine of 10 patients fully *de novo* sequenced (including all CDRs)



Between 66.7% (κ) and 94% (λ) sequence homology

Unexpected range of modifications
Cysteinylation, CoenzymeM, HexNAc(1)dHex(1)

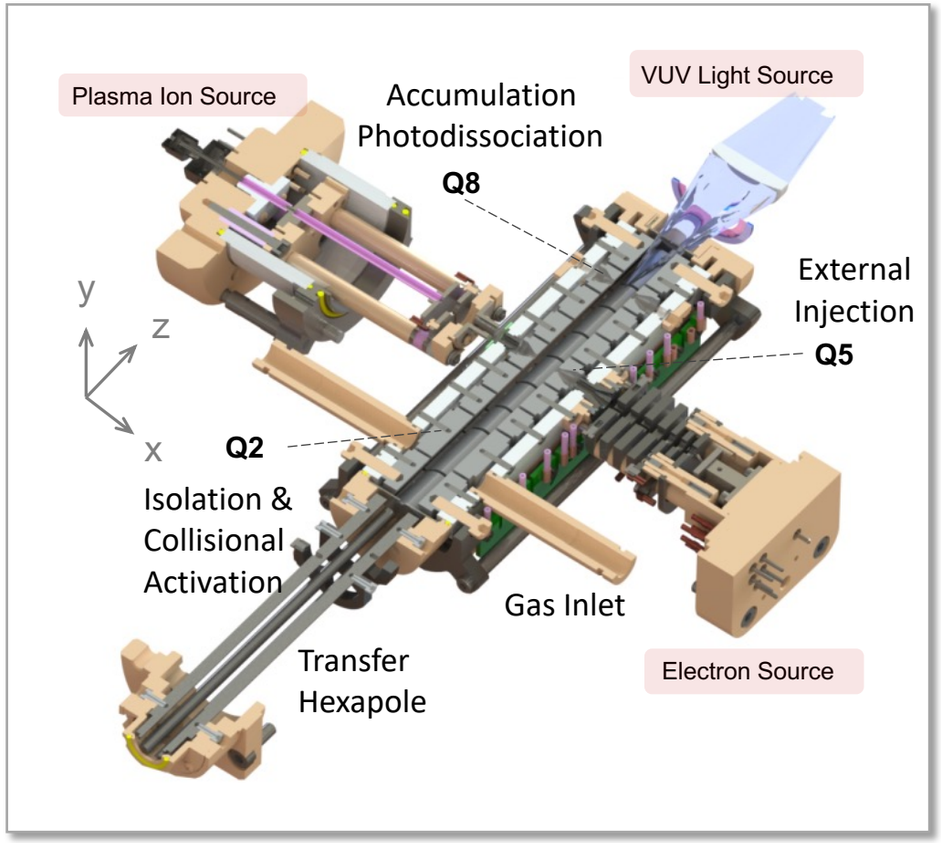
- Top-down proteomics essential to address samples with multiple proteoforms (could be improved?)

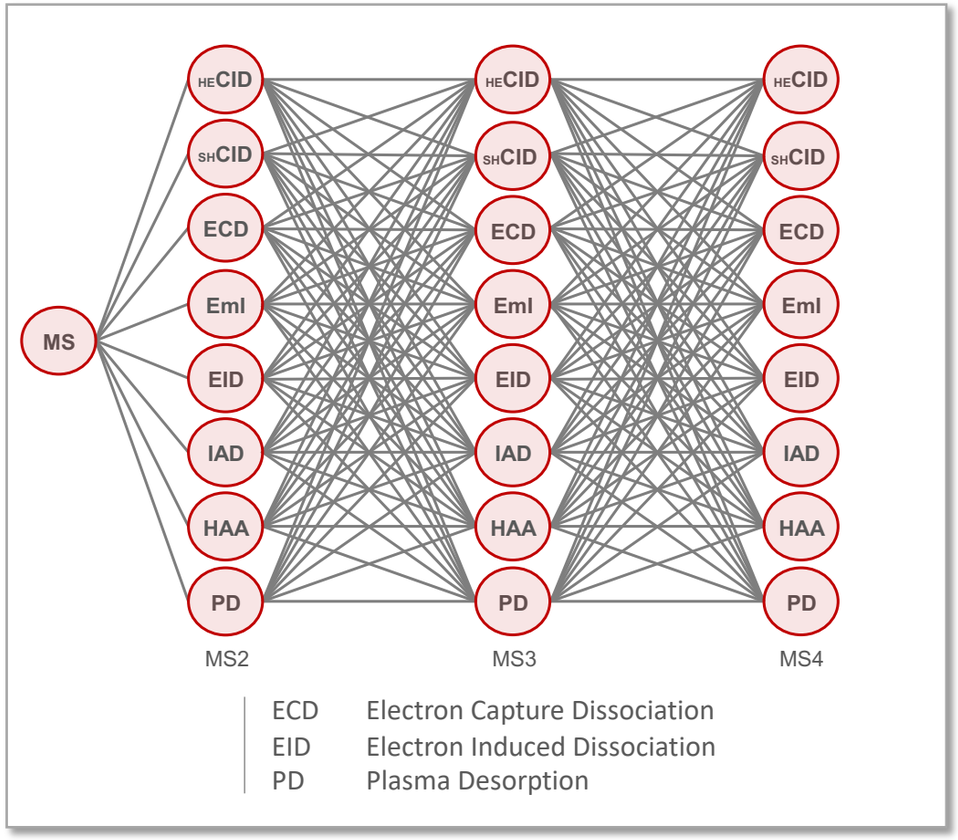
Dupré M. *et al.*, *Anal. Chem.* 93, 10627−10634 (2021)

INSTITUT PASTEUR

Omnitrap platform for improved Top-Down Proteomics

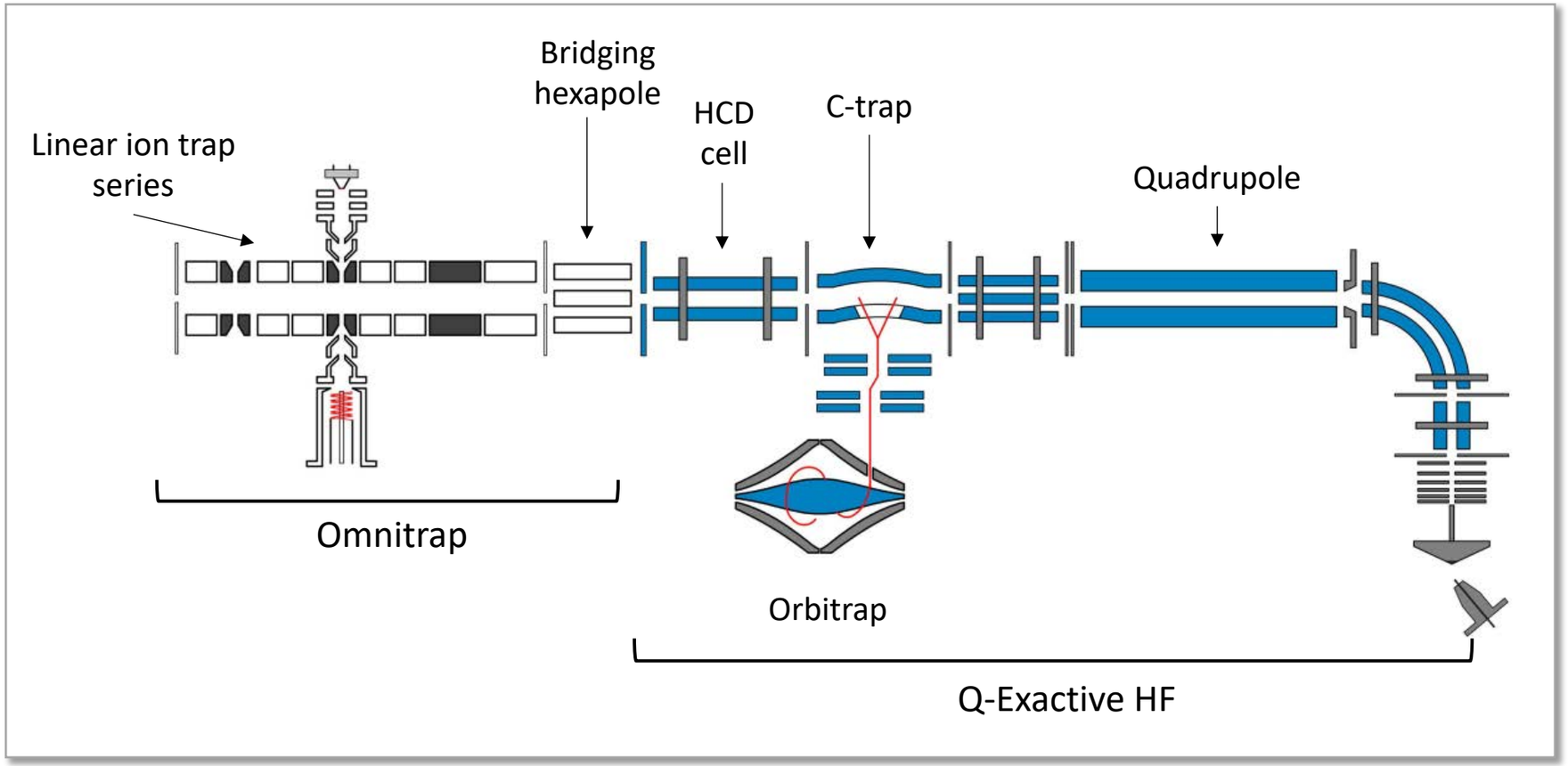
- A novel segmented linear ion trap for enhanced activation and ion storage
- Arsenal of ion activation techniques: CID, ECD, EID, photo-dissociation and others…





Q-Exactive HF modified with the Omnitrap platform

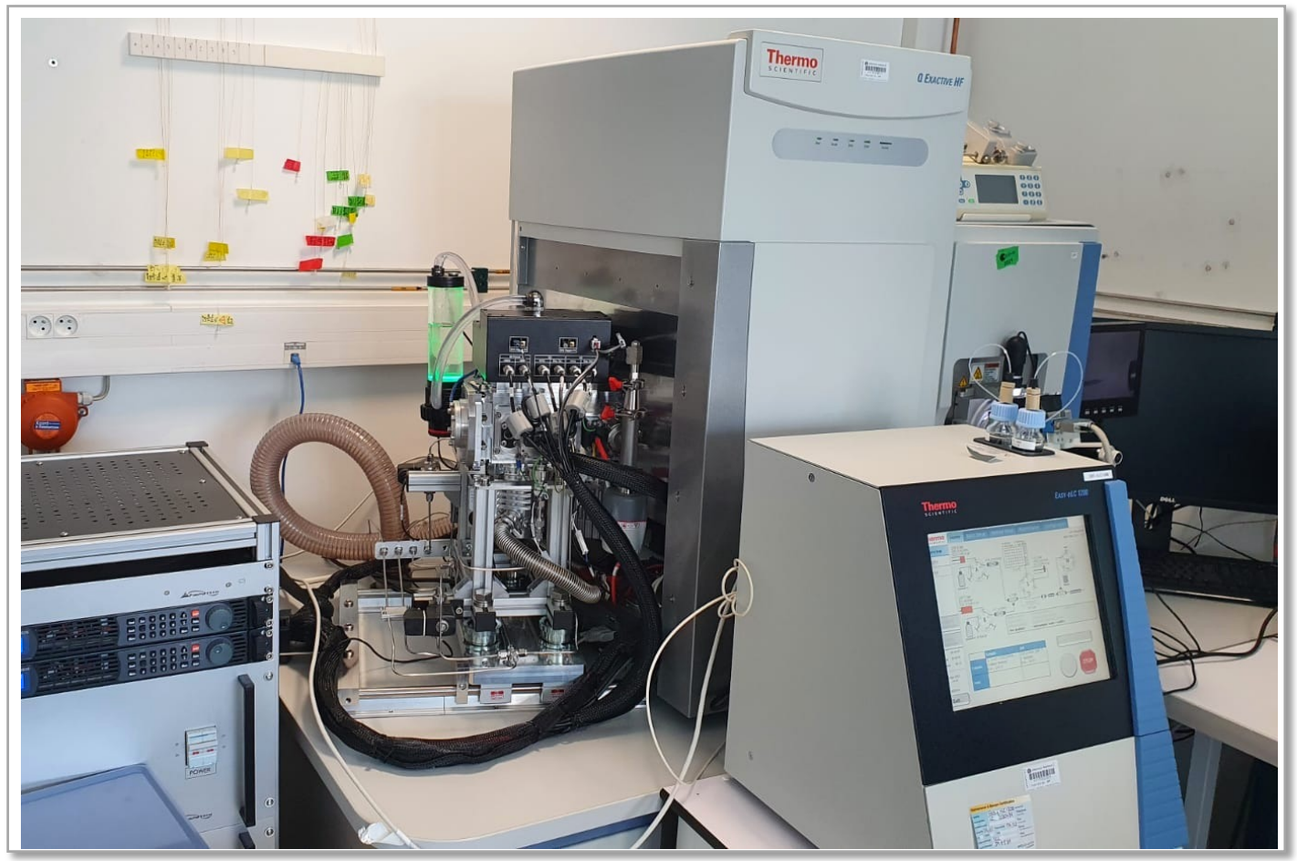
- Currently available as a retrofit to the Q-Exactive/Exploris instrument series



Application areas: top-down/bottom-up proteomics, native MS, glycomics…

Q-Exactive HF modified with the Omnitrap platform

- Smooth installation (3 days), performance of the Q-Exactive HF (HeLa digest) kept the same



TOP SPEC

FasmaTECH

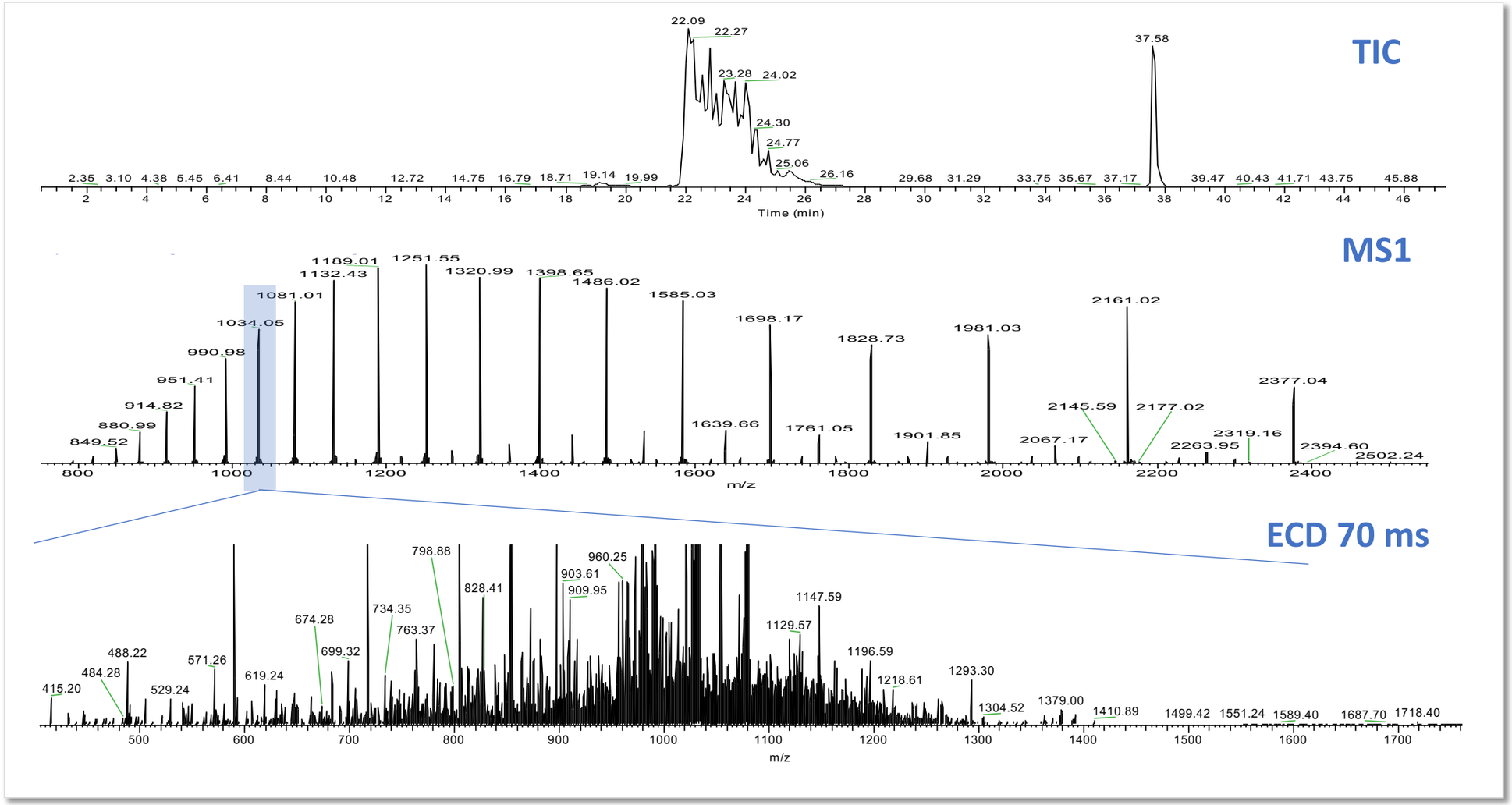
Thermo SCIENTIFIC

MSVision

Application areas: top-down/bottom-up proteomics, native MS, glycomics…

INSTITUT PASTEUR

P15: ECD 70 ms – MS^2 (nanoLC time scale)



INSTITUT PASTEUR

P15: ECD (non-reduced) – MS^2

Residue cleavage: 37%
Out of the S-S bridges: 88%

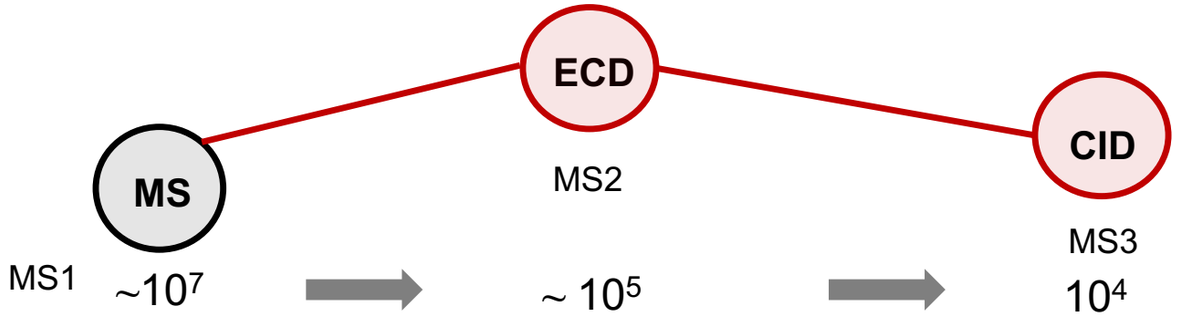
a : 18%
b : 7%
c : 23%
x : 15%
y : 16%
z : 24%

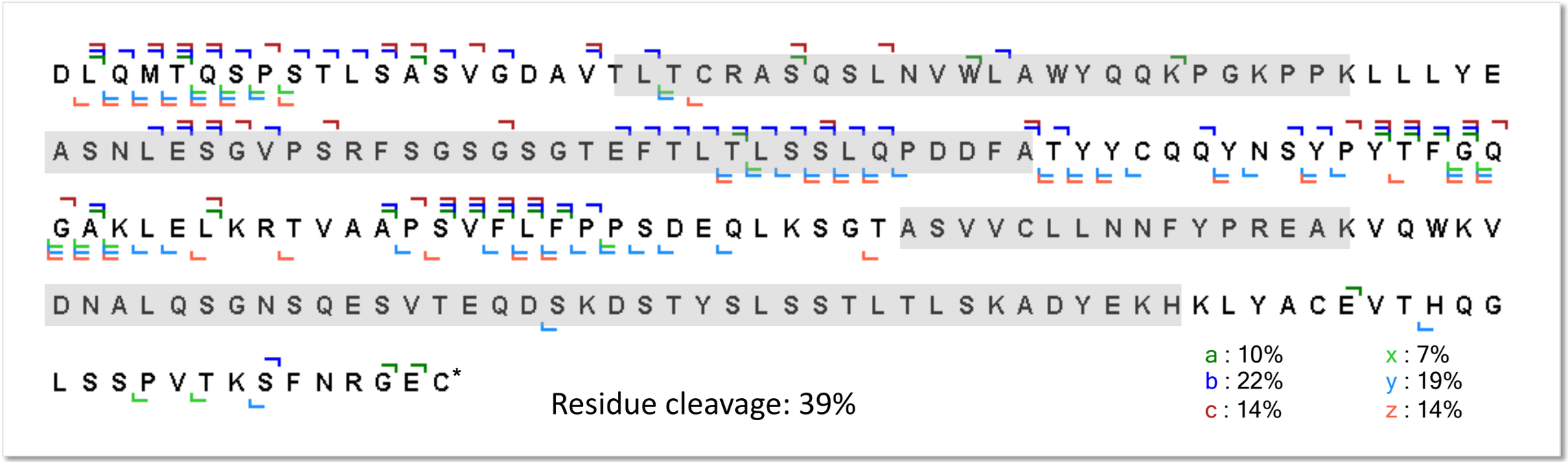
Cronus software, no deconvolution

Excellent sequence coverage out of the S-S bridges allowing their easy localization

INSTITUT PASTEUR

P15: ECD followed by CID (non-reduced) – MS^3

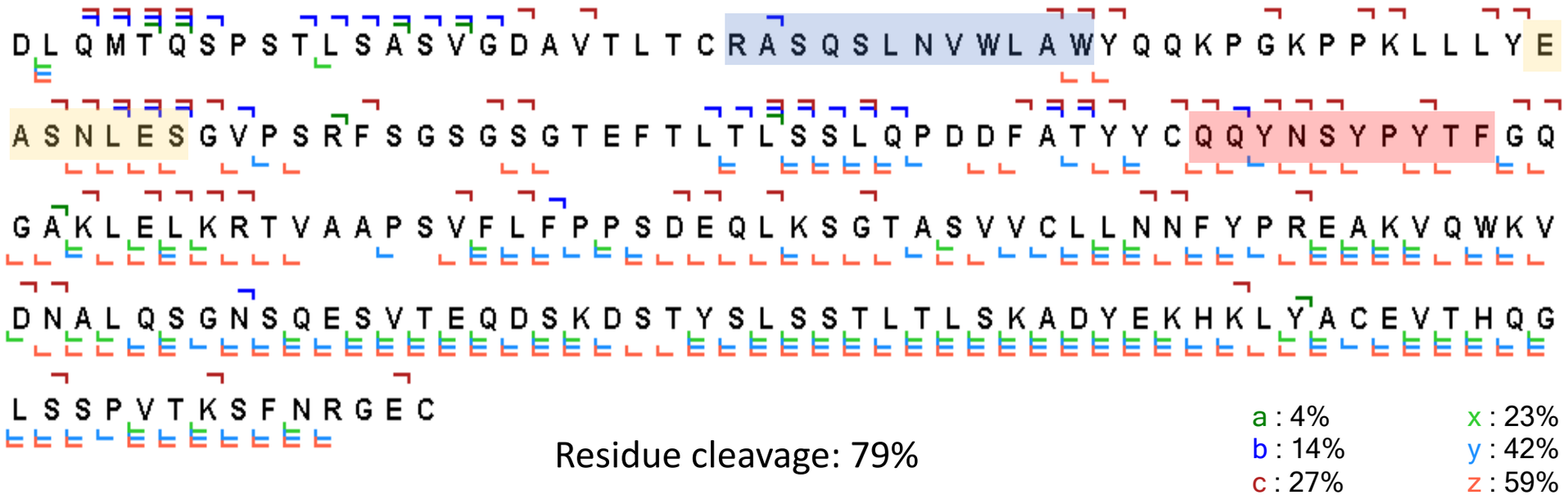




First S-S bond has been cleaved leading to new fragment ions

INSTITUT PASTEUR

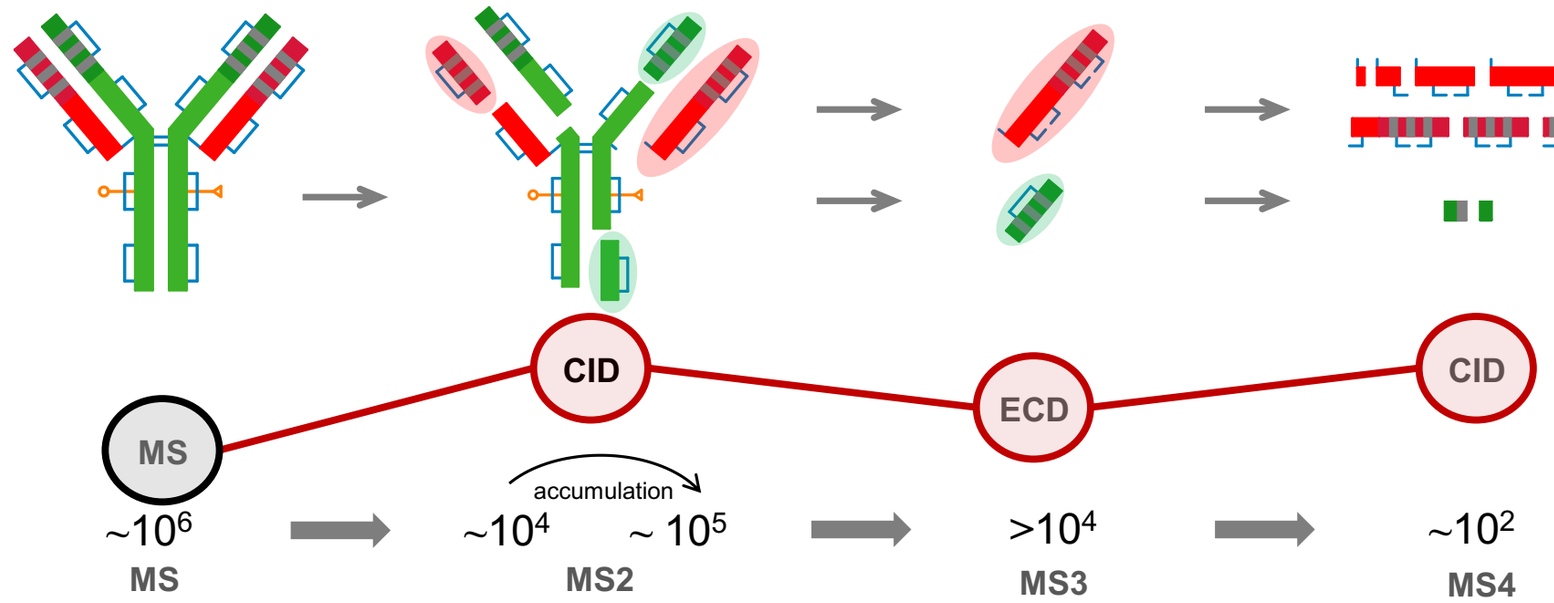
P15: ECD (reduced) – MS^2



79% sequence coverage in a single experiment

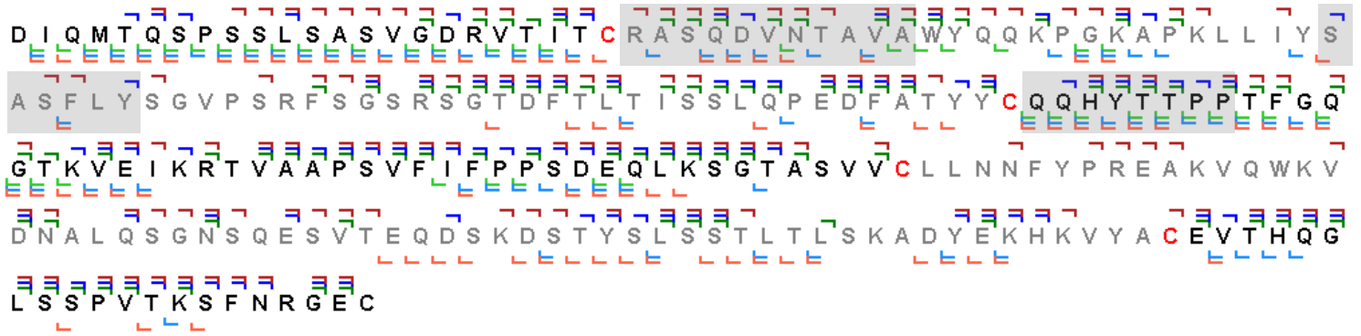
INSTITUT PASTEUR

Intact herceptin 150 kDa (infusion) – MS^4



Herceptin Light Chain

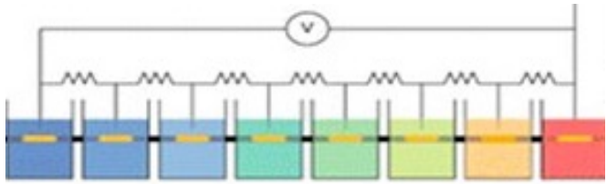
Sequence Coverage ~84%



INSTITUT PASTEUR

Conclusion & perspectives

- Efficient workflow for the *de novo* sequencing of light chains (clinical samples): combination of bottom-up and Top-Down Proteomics (TDP)
- TDP is essential for samples with multiple proteoforms
- The Omnitrap platform holds great promise for efficient fragmentation of intact proteins (even large ones)
  More to come: VUV for S-S bond cleavage followed by CID
- Importance of data analysis, new developments required (*de novo* sequencing from top-down data)
- PI trap for proteoform separation (isoelectric focusing)



INSTITUT PASTEUR

Aknowledgements

- **Mass Spectrometry for Biology Unit**
  Institut Pasteur, Paris
  J. Dhenin
  M. Dupré
  M. Duchateau
  T. Fu
  C. Garcia
  M. Rey

- **Fastmatech (Greece)**
  D. Papanastasiou
  T. Smyrnakis
  M. Kosmopoulou

- **MSVision**
  A. Langlands
  J. Commandeur

TOP SPEC

EPIC-XS

European Commission

INSTITUT PASTEUR