Modern methods in protein research

Protein foot-printing and chemical cross-linking

Petr Novák

Modern methods in protein research| **Department of Biochemistry**| Charles University

Structural Mass Spectrometry…

Protein covalent labeling

Chemical cross-linking

H/D exchange

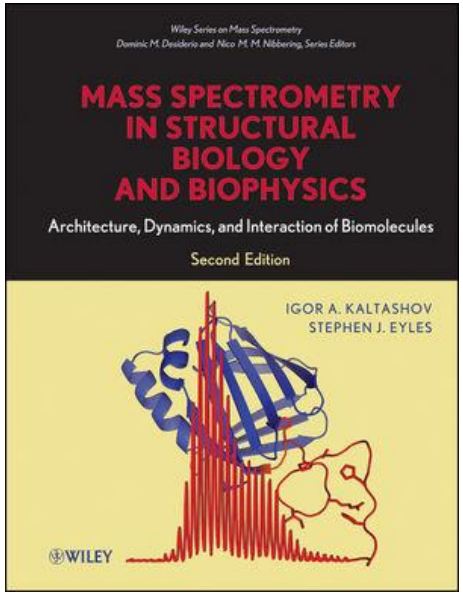
Disulfide bonds mapping

Native mass spectrometry and Ion mobility

Fast photochemical oxidation of proteins

ETD/ECD fragmentation

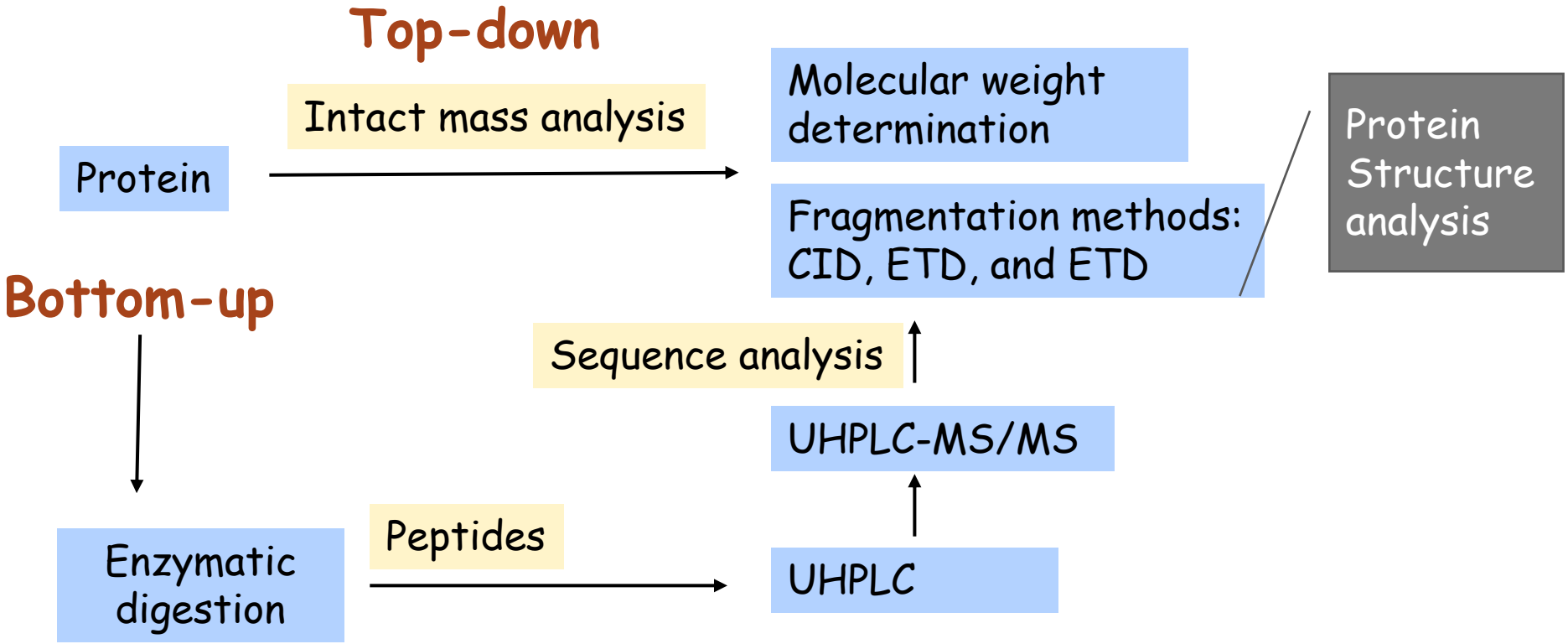
Limited proteolysis



Special Issue on Mass Spectrometry in Structural Biology (2015)
Protein Science *24*, 1173-1332

Mass Spectrometry: Goal in Protein Structure Characterization

- Sensitivity
- Analysis of complex mixtures/high MW protein
- Rapid data acquisition



Footprinting

Assay examining higher structures of biomacromolecules by monitoring solvent accessibility of their regions

- Single molecule conformation / Conformational changes
- Ligand binding / biomacromolecular interactions

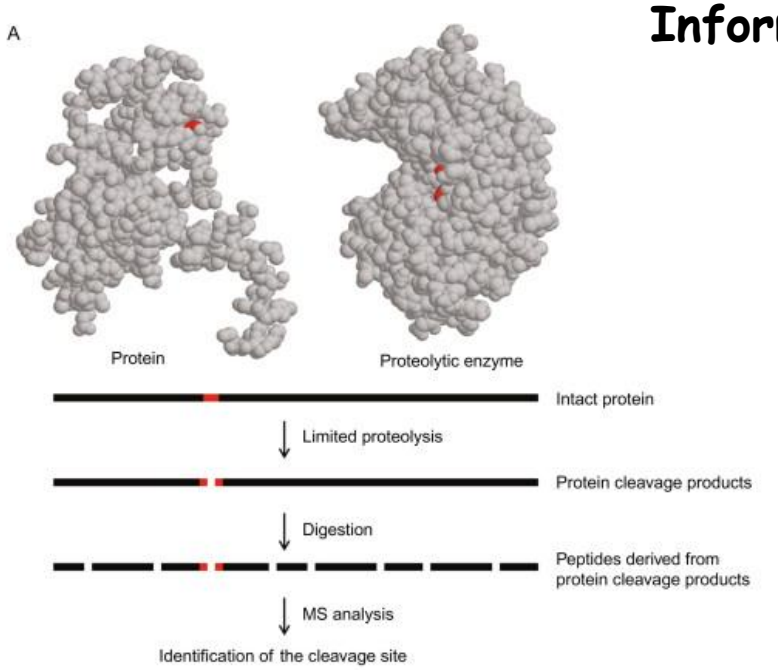
Different techniques

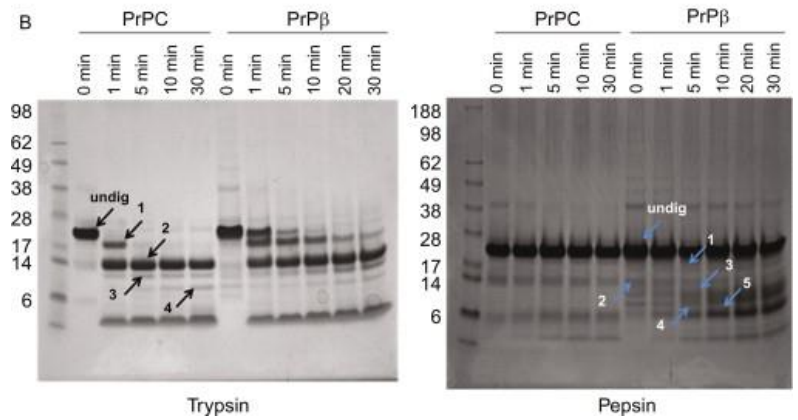
- Enzymatic / chemical cleavage
- Covalent labeling

Covalent labeling

- Hydrogen-deuterium Exchange
- Stable covalent labeling – Chemical or Radical footprinting and cross-linking

Enzymatic cleavage – Limited proteolysis

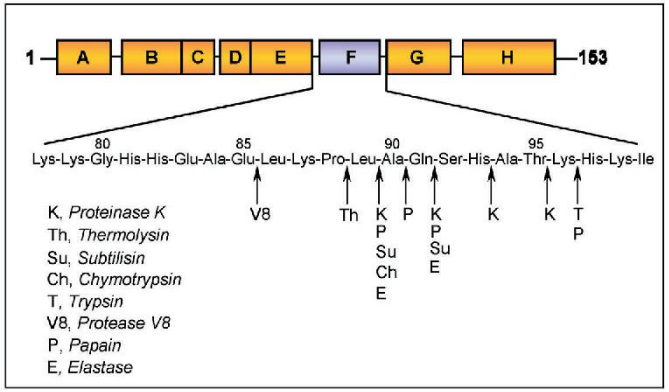


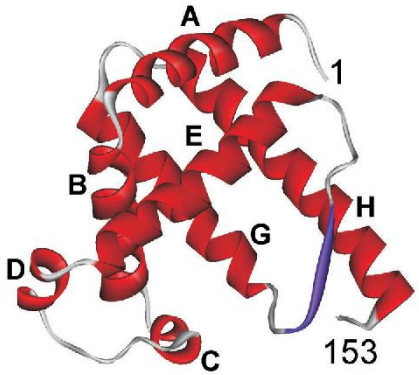


Information about the surface accessible area

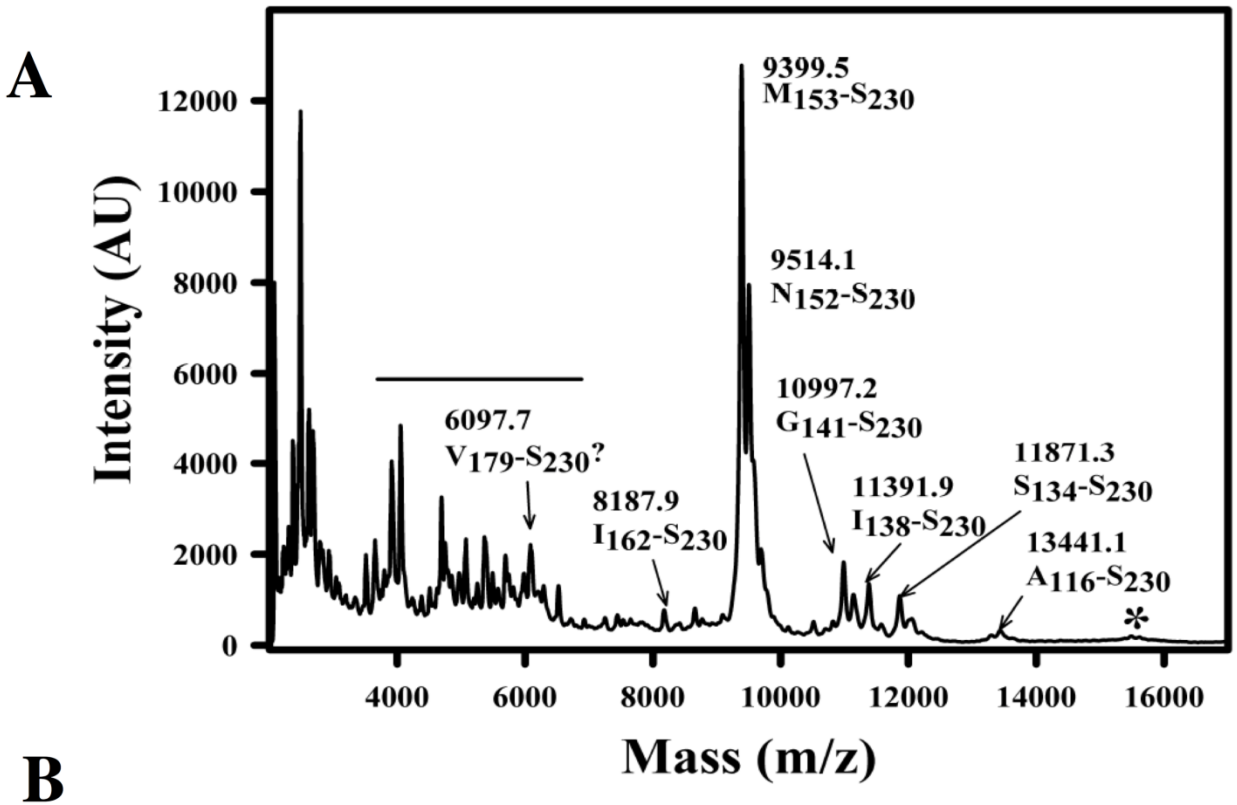
Several available proteases

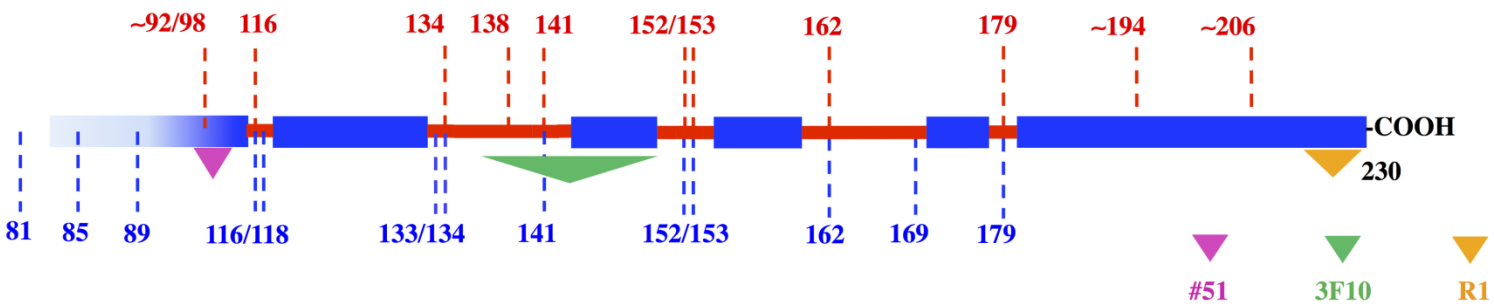
Native conditions



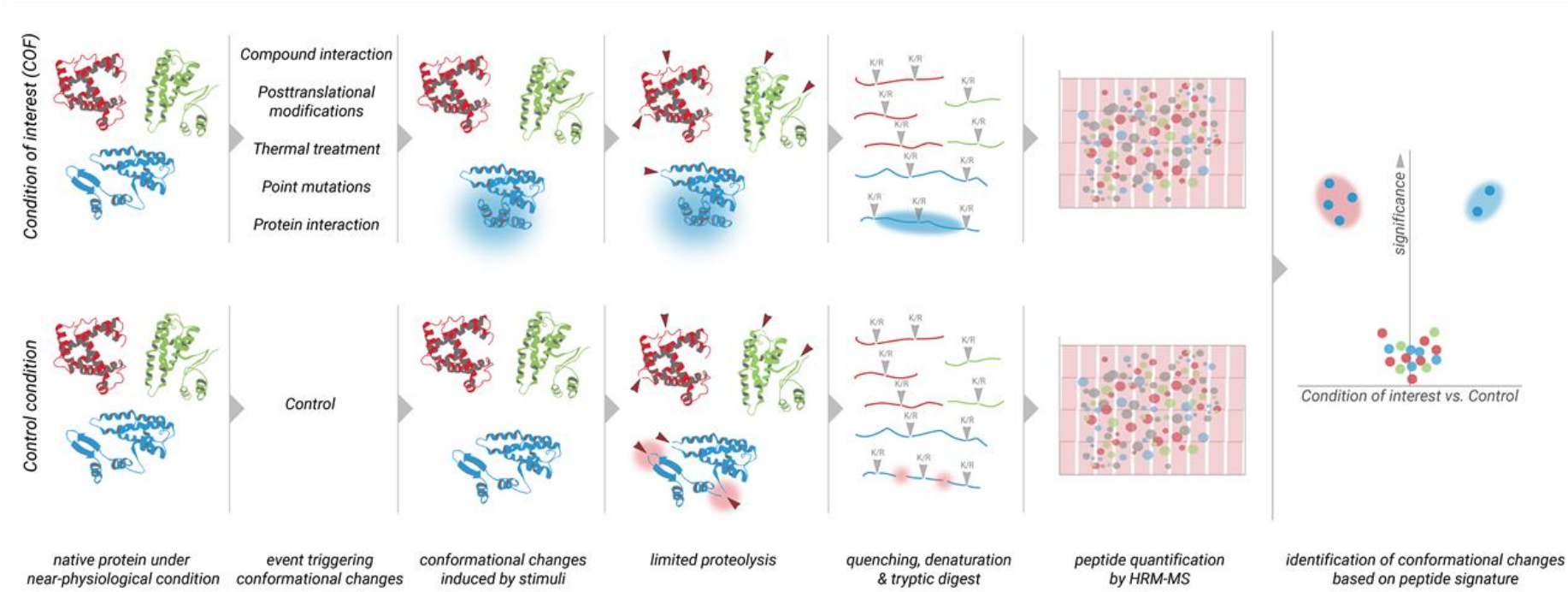


Limited proteolysis – identification of fragments (“Top down”)

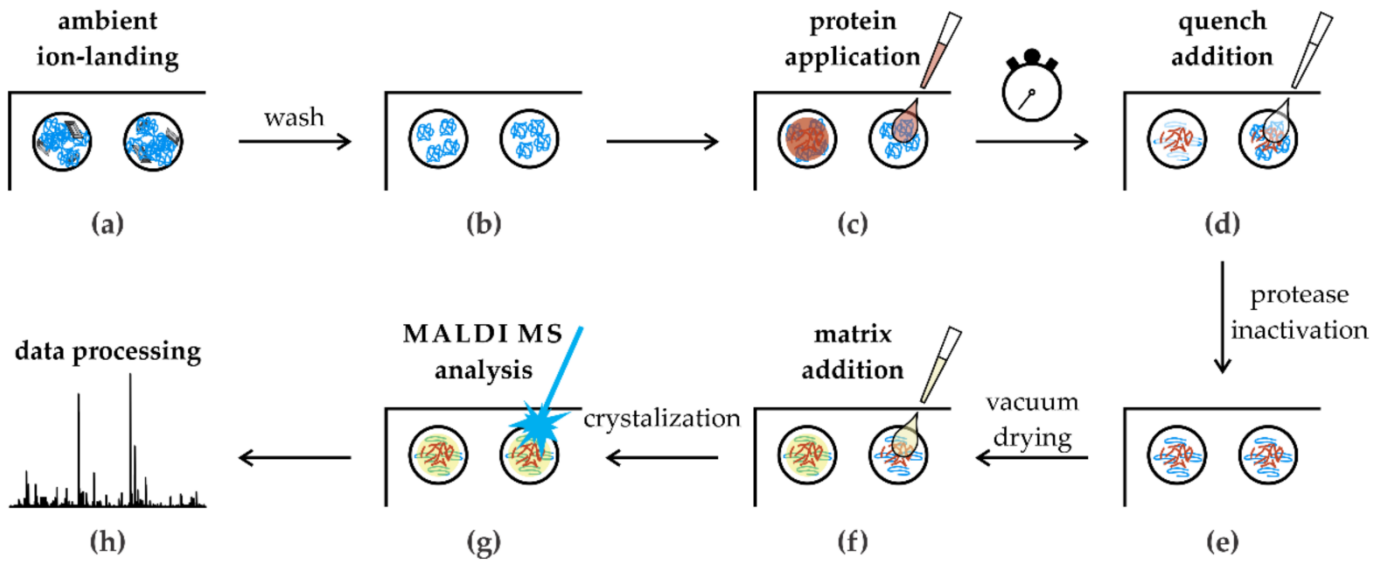


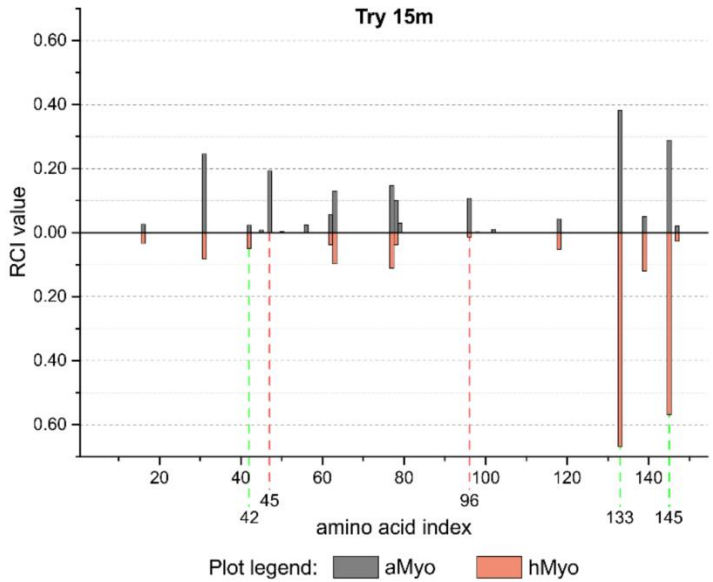


Limited proteolysis – identification of fragments (“Bottom up”)

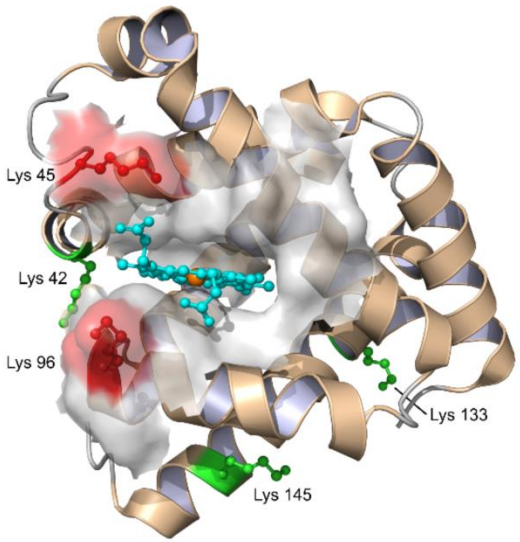


Limited proteolysis – identification of fragments ("Chip set up")





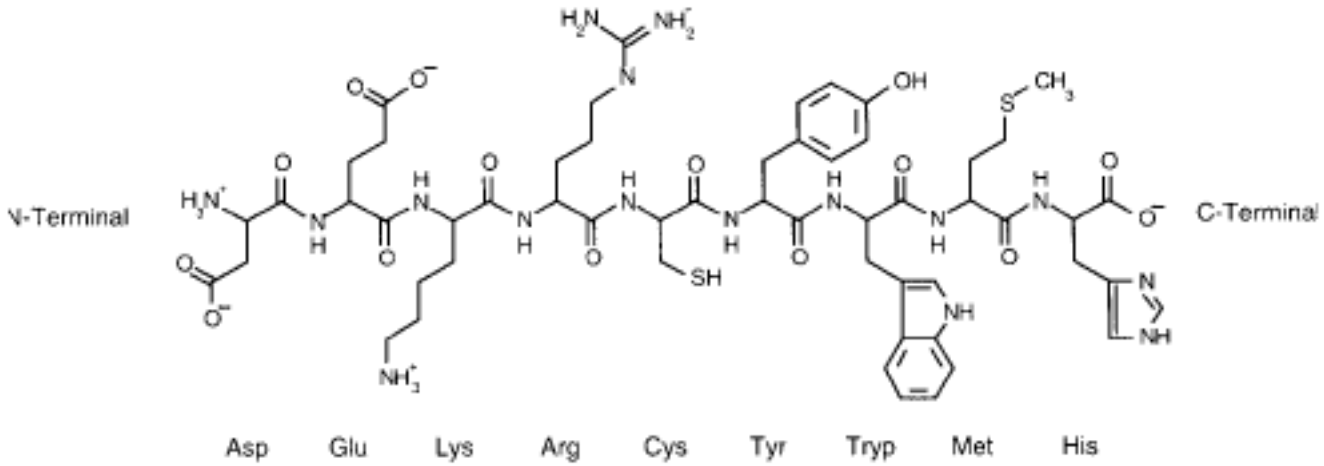
(a)



(b)

Protein covalent labeling and chemical cross-linking

Available amino acid sidechains for covalent modification



Carboxy groups – Asp, Glu, C-term, pKa (3.8, 4.3, 2.3) pH≥7 » **deprotonation**

Amino groups – Lys, Arg, His, N-term, pKa (9.4, 12, 6.8, 7.8) 7≥pH » **protonation**

Sulfhydryl groups - Cys . pKa 8.9 pH≥7 » **-SH**

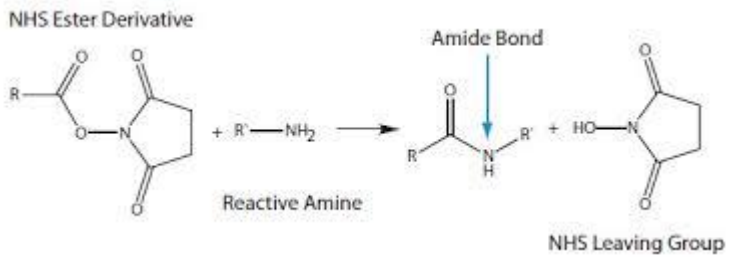
Aromatic groups– Trp (indol), Tyr (hydroxyfenyl, **pKa 9.9**)

~ 23% of amino acid can be covalently modified

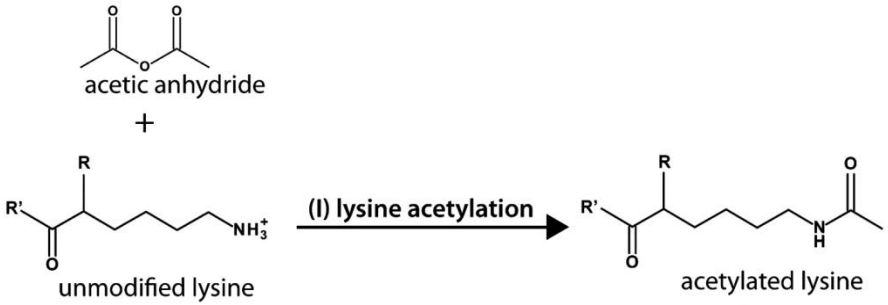
Klapper et. al. Biochem. Biophys. Res. Commun. 1977, 78, 1018.

Covalent modification of amino acid side chains

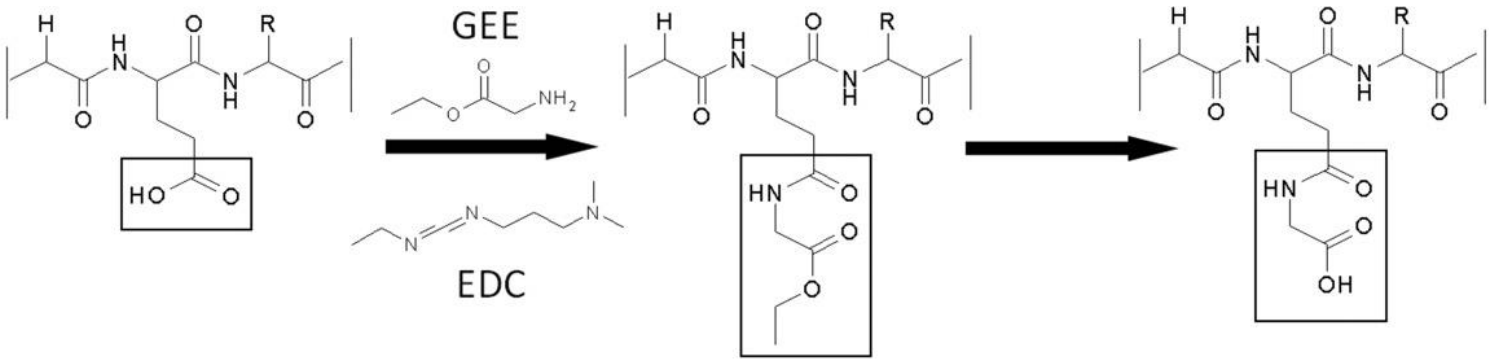
Acetylation
(N-hydroxysuccinimide esters)



Acetylation
(Anhydrides)

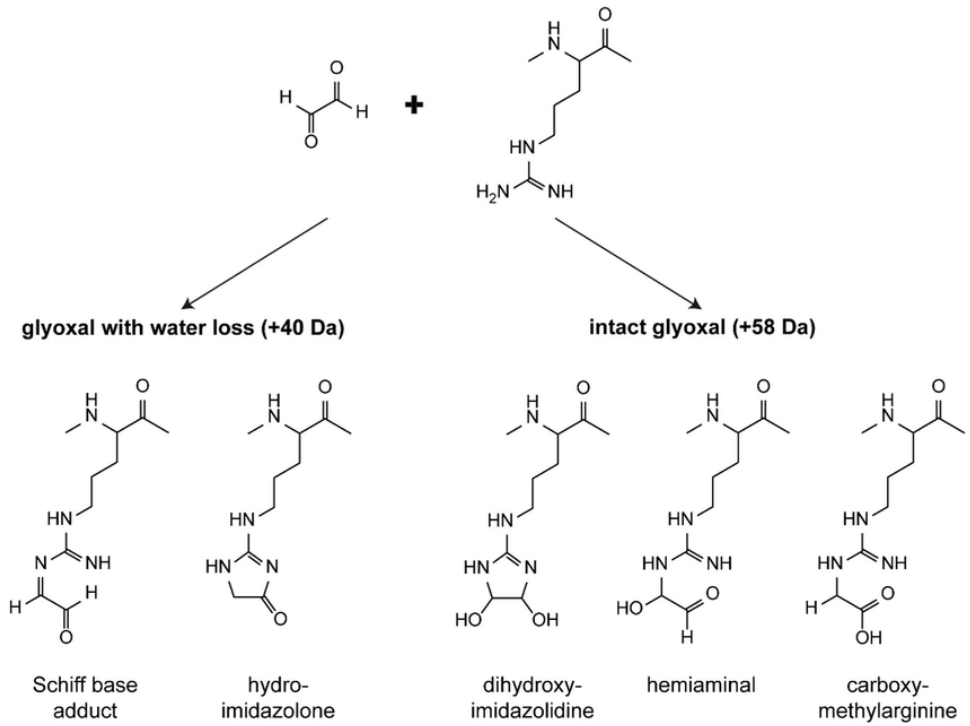


Modification of carboxylic acids
(1-Ethyl-3-(3-dimethylaminopropyl) carbodiimide
and glycine ethyl ester)

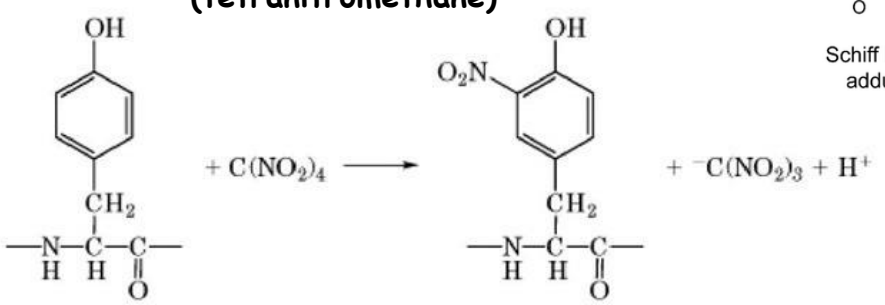


Covalent modification of amino acid side chains

Modification of arginine
(diketones)

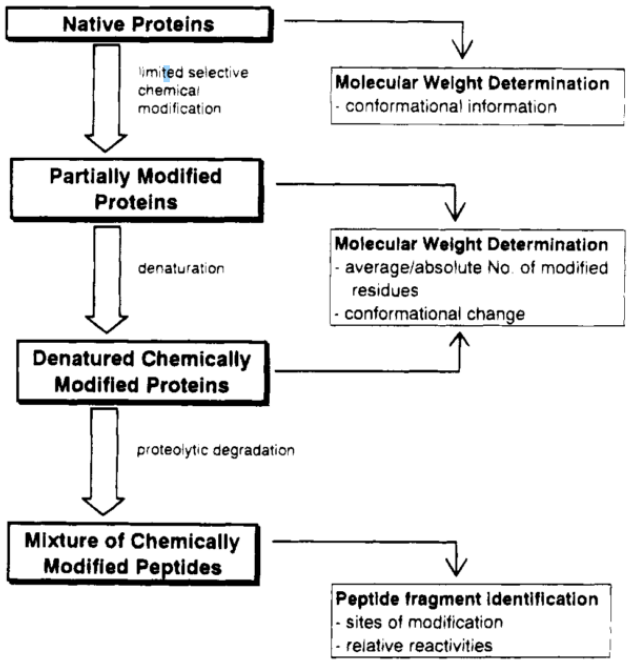


Nitration
(tetranitromethane)

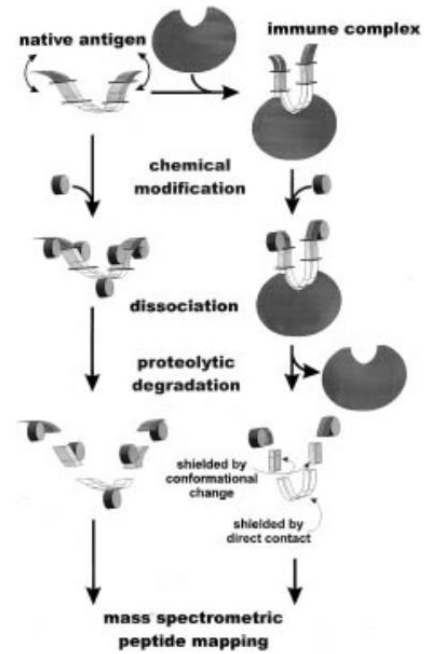


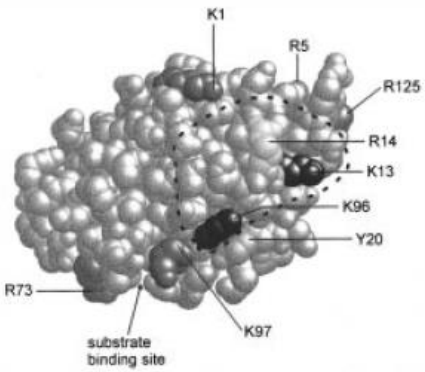
Tyrosine [I] Tetranitromethane [II] 3-Nitrotyrosine

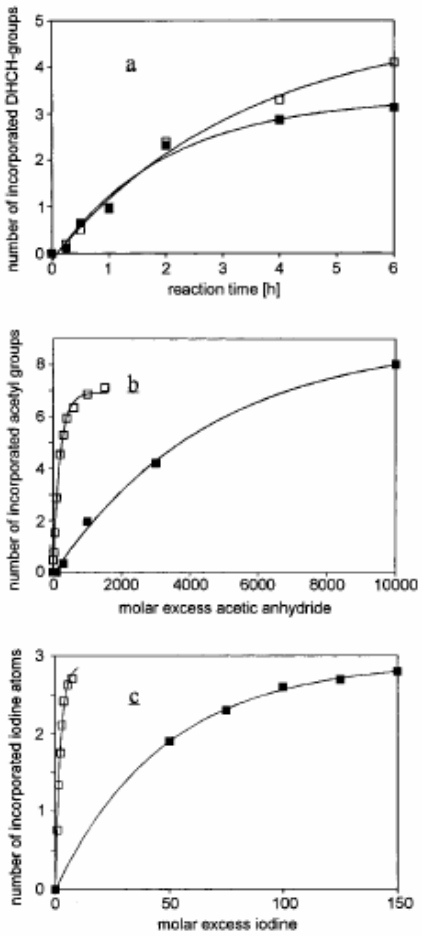
Protein covalent labeling: Lys, Tyr, Arg



relative reactivity NH_2 groups[a]	lysine residues, ϵ-amino groups		
	RNAse A	HEL	myoglobin
	α-NH_2		α-NH_2
1	41, 104	97, 33 α-NH_2	45, 63, 77, 79, 145, 147
2	1, 7, 37	1	16, 42, 87
3	31, 61, 91	13, 116	56, 50, 62, 78, 102
4	66, 98	96	96, 47, 87, 133,[b] 118[b]



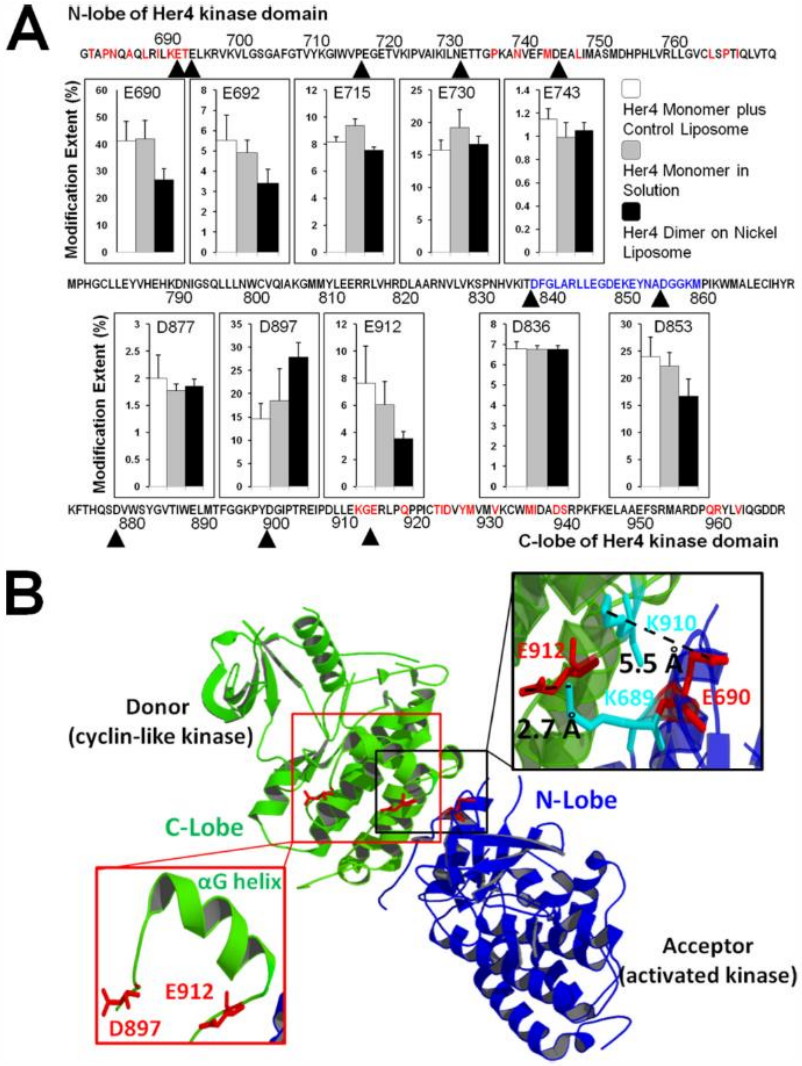




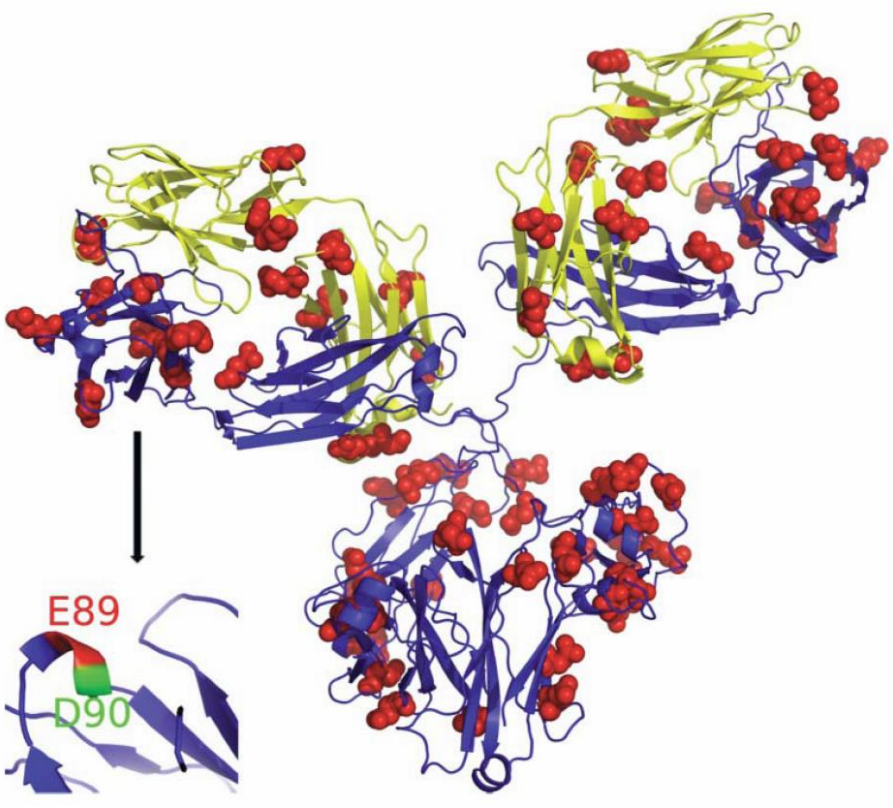
Suckau et. al. *PNAS 1992, 89, 5630*
and ***Glocker et. al.*** *Bioconj. Chem. 1994, 5, 583*

Fiedler et. al. *Bioconj. Chem. 1998, 9, 236*

Protein covalent labeling: Asp, Glu



Zhang H. et al. MCP 2011



Kaur P. et al. mAb 2015

Hydroxyl Radical Footprinting

Products of water or hydrogen peroxide molecule homolytic bond cleavage

Hydroxyl radicals can be generated by various means:

- **Fenton reaction**
- **Irradiation of water by x-rays or electron beams**
- **Photolysis of hydrogen peroxide FPOP (fast photochemical oxidation of proteins)**
- **Other radicals available $^{\bullet}OH$, $^{\bullet}I$, $^{\bullet}CF3$**

The relative reactivity of the amino acid side chains

Cysteine, Methionine, Tryptophan

Tyr > Phe > His > Leu ~ Ile > Arg ~ Lys ~ Val > Ser ~ Thr ~ Pro > Gln ~ Glu > Asp

Alanine, Glycine

- **Reactive species**
- **React efficiently with most AA side chains**
- **Form STABLE oxidation products**

Takamoto K . et al. *Annu Rev Biophys Biomol Struct.* 2006, 35, 251-276

Conditions for radical labeling

ELECTRON PULSE RADIOLYSIS:

- reproducible 1-100 ns pulses; MeV energy range on linear accelerator

SYNCHROTRON RADIOLYSIS:

- X-ray; 3-30 keV @ beam current ~ 250 mA

LASER H_2O_2 PHOTOLYSIS:

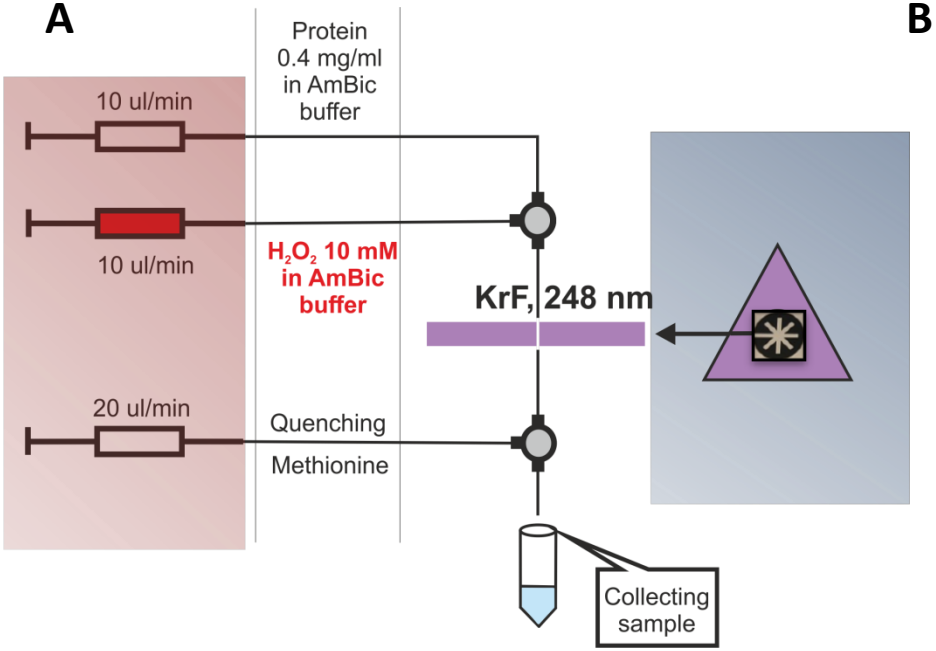
1% - 0,04% H_2O_2 (mixing by stopped-flow device or just before irradiation); quench and removal of residual peroxide is vital

- Nd:YAG; 2 mJ/pulse @ 266 nm; 3-5 ns pulse; 1-100 shots
- 17 ns KrF excimer laser; 50 mJ/pulse @ 248 nm
- 18 ns KrF excimer laser; 62,5 mJ/pulse @ 248 nm; 16 Hz

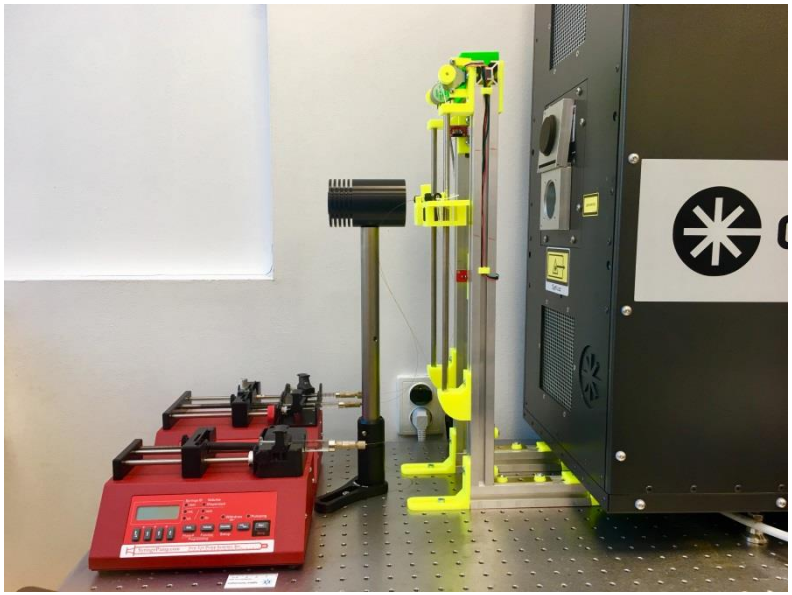
Fast photochemical oxidation of proteins

- **Covalent modification ➡ preserves the primary sequence of modified residues**
- **High reactivity of •OH ➡ the modifications of more than half of amino acid side-chains, providing a higher coverage**

A



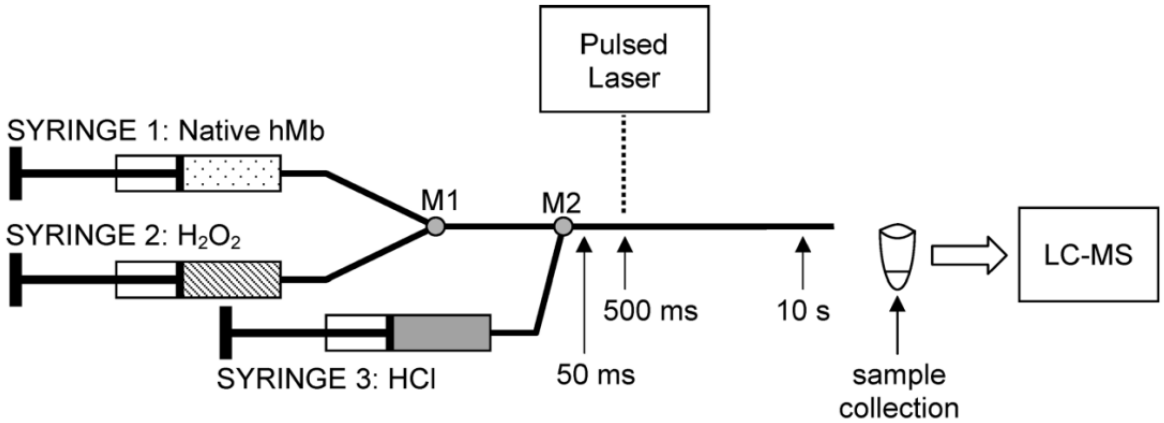
B



Experimental setup

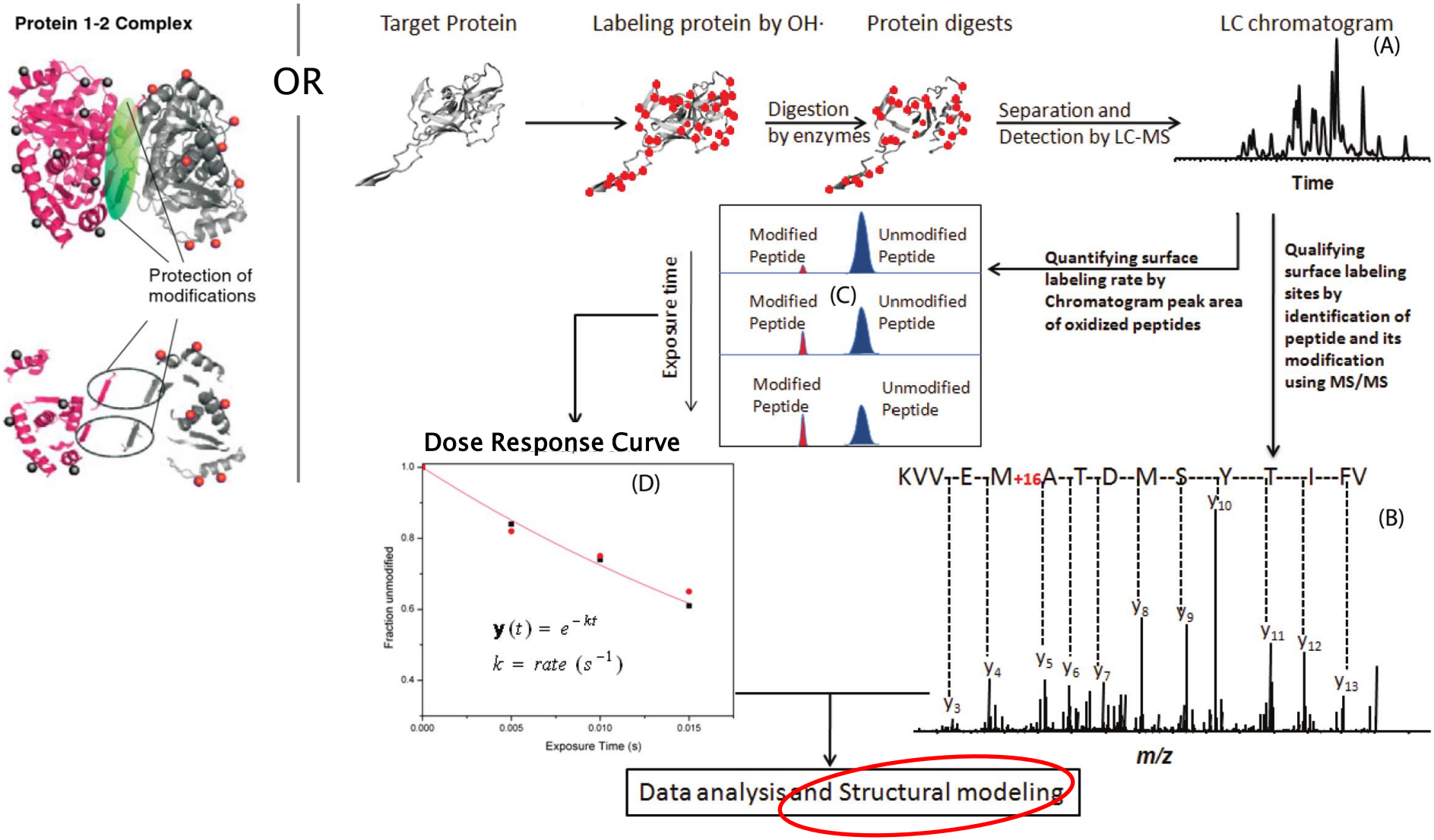
SAMPLE MIXED AND IRRADIATED IN

- **µtubes** (sample volume ~ 15 µl) or in
- **stopped-flow microfluidic mixing device** - essential for folding / kinetic studies (capillary flow



• Short pulses with high energy are needed to create sufficient concentration of radicals on very short (sub-microsecond) timescales to avoid conformational changes of protein during labeling.
• Possible protein conformational changes occur mostly on a longer than milisecond timescale.

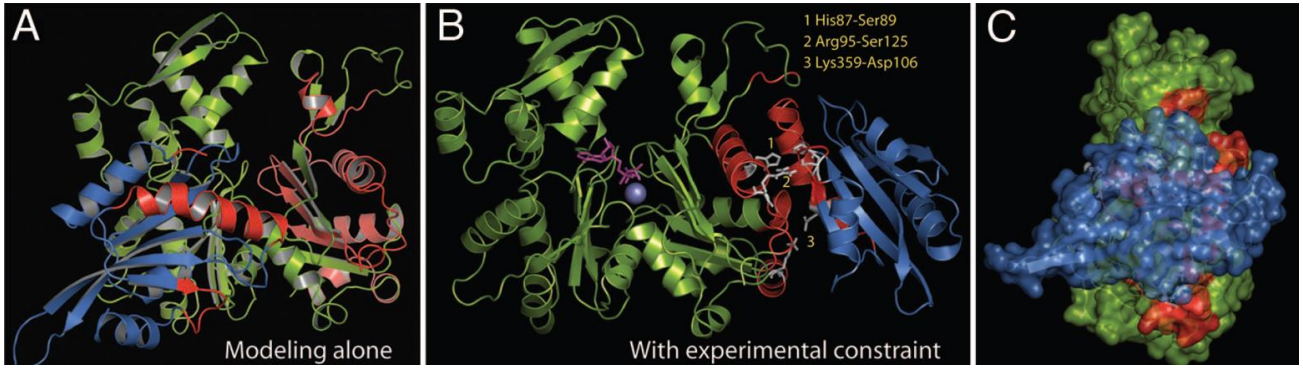
MS Protein footprinting - workflow



Some (recent) publications

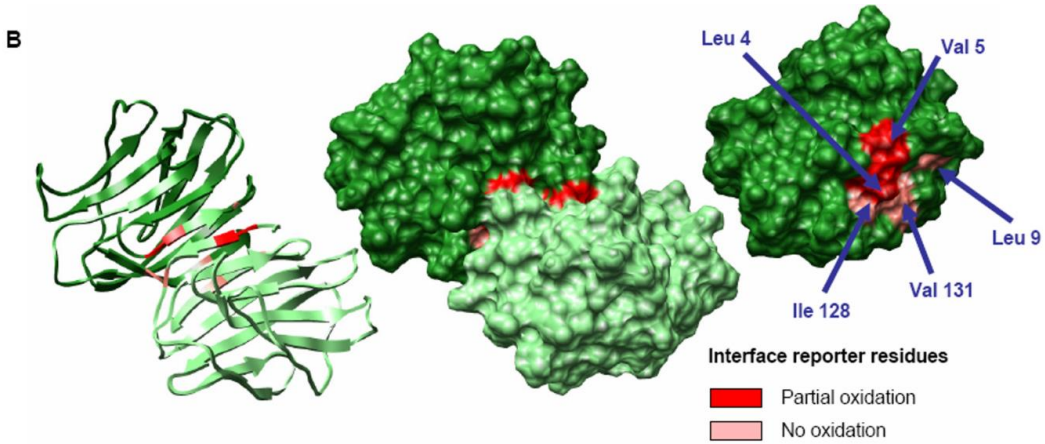
PROTEIN-PROTEIN INTERACTION

- Actin – coffilin interaction:



Kamal JK et al. PNAS, 104, 19, 7910-7915 (2007)

- Dimerization interface of galectin-1:

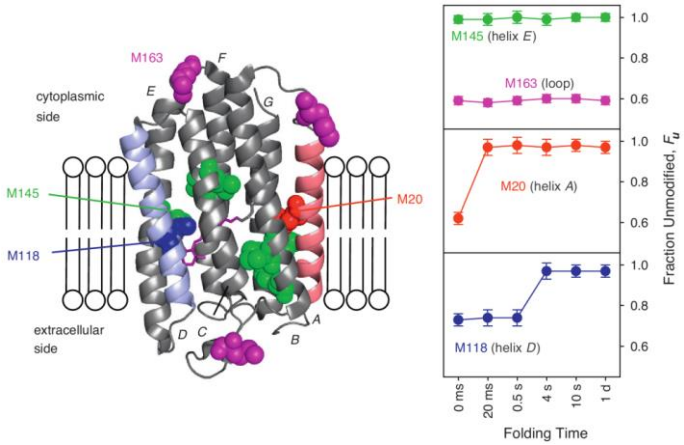


Charvatova O et al. JASMS, 19, 11, 1692-1705 (2008)

Some (recent) publications

PROTEIN STRUCTURE

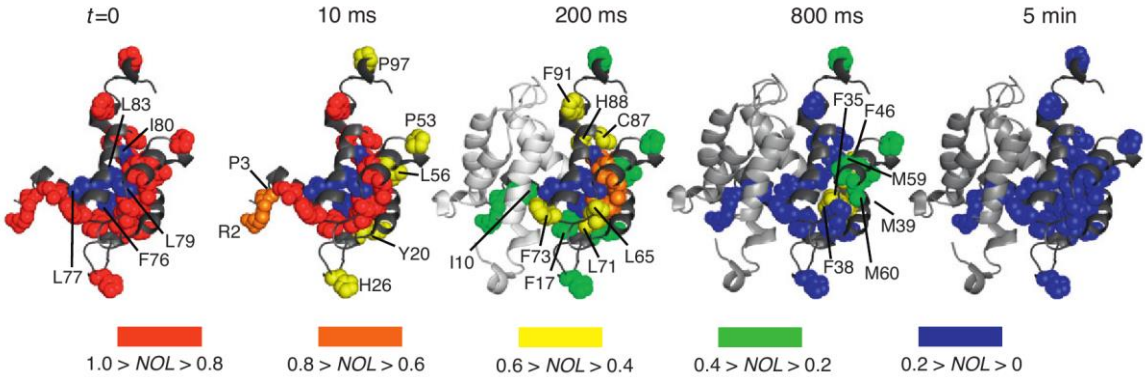
- Integral membrane protein (BR) structure in its natural lipid environment



Pan Y et al. J Mol Biol, 410, 146-158 (2011)

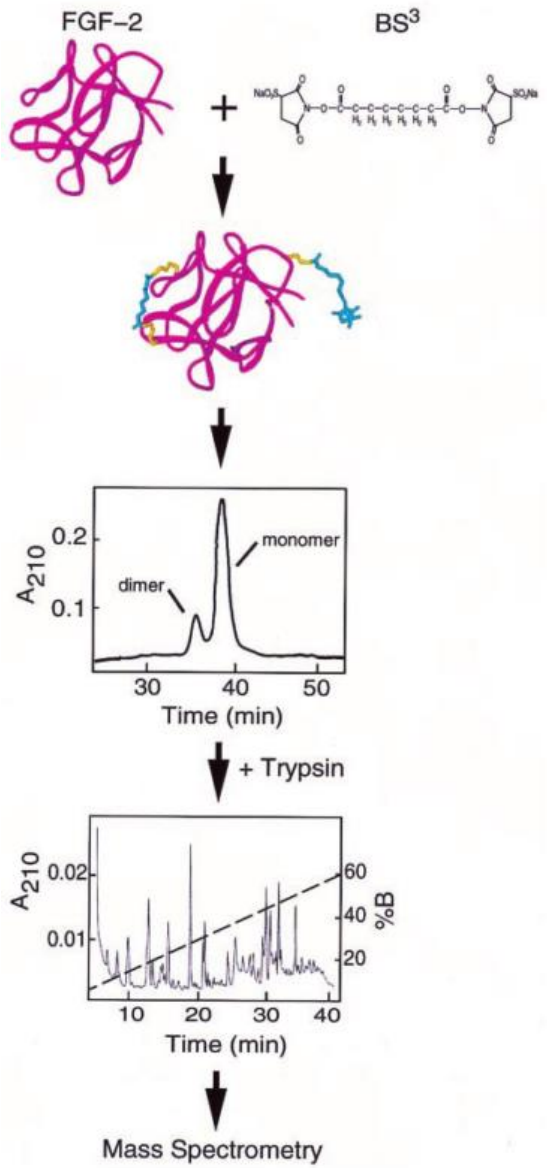
PROTEIN FOLDING KINETICS

- Time-resolved folding and dimerization changes in Ca-binding protein

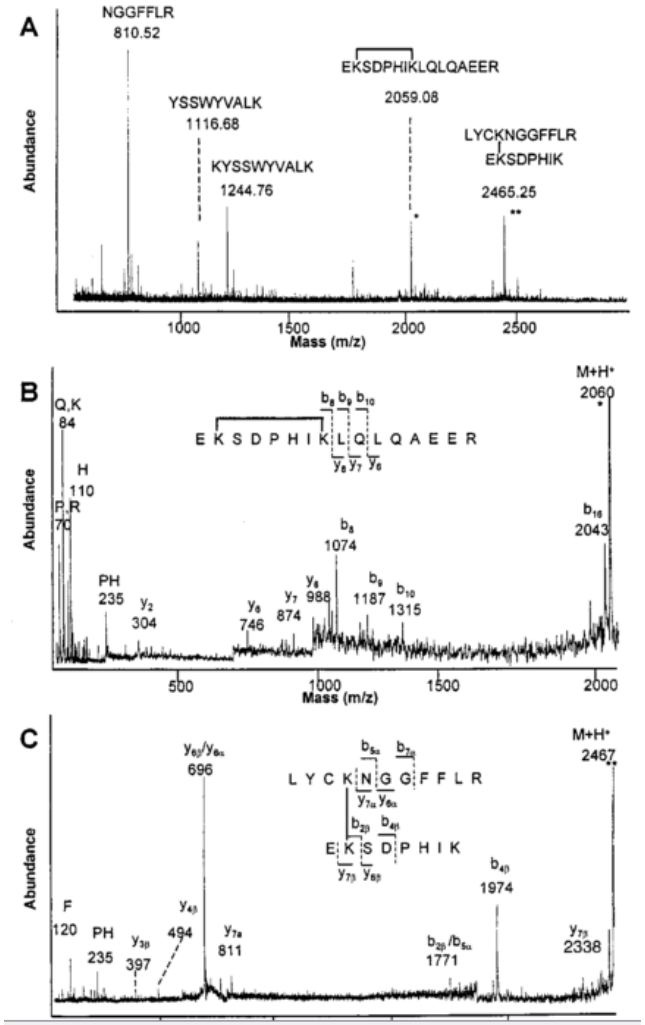


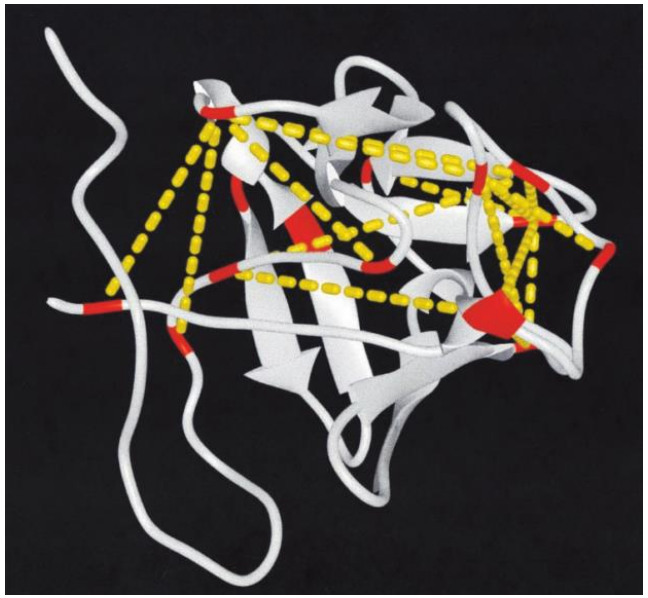
Stocks BB et al. J Mol Biol, 409, 669-679 (2011)

Chemical cross-linking: the first 3-D structure



Fibroblast grow factor 2





Top 20 threading models ranked by constraint error

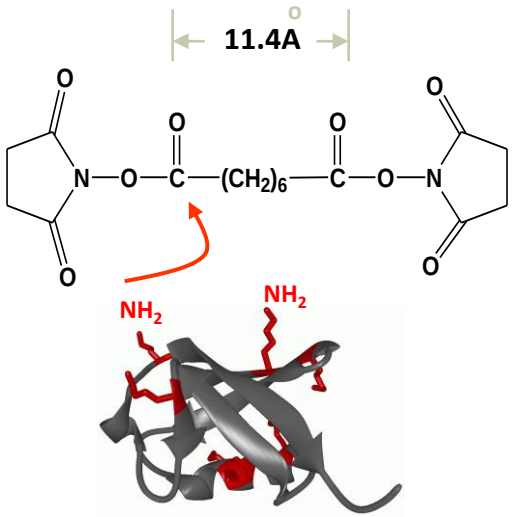
Name	Fold family	% Sequence identity	Threading rank	Constraint error, Å*	Number of violations†
FGF-2	β-Trefoil	98.6	1	0.0	0
IL-1β	β-Trefoil	12.7	5	0.0	0
Gastrotropin	Lipocalin	7.1	8	2.9	1
Hisactophilin	β-Trefoil	8.6	12	5.5	2
Guanylate kinase	P-loop	12.4	9	7.4	4
NTP pyrophosphohydrolase	NTP pyrophosphohydrolase	9.3	6	14.5	3
Glutathione peroxidase	Thioredoxin	11.1	14	16.6	5
Retinol-binding protein	Lipocalin	9.1	18	17.1	3
Nucleoside diphosphokinase	Ferridoxin-like	8.8	20	18.6	2
Cytochrome c_4	Cytochrome *c*	12.6	11	21.4	5
Aspartate carbamoyltransferase	Ferridoxin-like	9.8	13	22.6	4
D-UTPase	β-Clip	7.8	2	27.5	7
Disulfide bond formation protein	Thioredoxin	8.4	15	28.1	8
ASV integrase	Ribonuclease H-like	7.8	19	28.6	5
Endoglucanase C	Galactose binding	11.6	4	33.8	6
TATA-box-binding protein	TATA-box-binding protein-like	10.3	7	40.0	8
Phospholipase A2	Phospholipase A2	9.5	16	55.4	7
PRD paired domain	3-Helix bundle	12.7	17	143.4	8

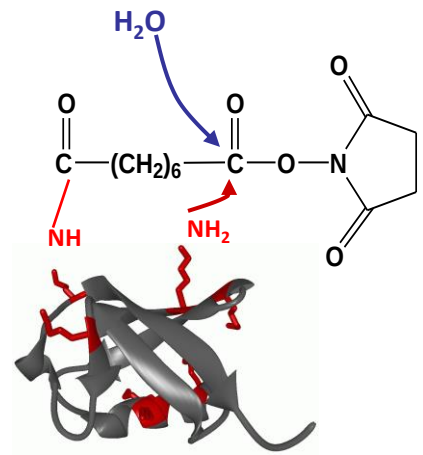
Young et. al. PNAS 2000, 97, 5802

Chemical cross-linking: the chemistry behind…

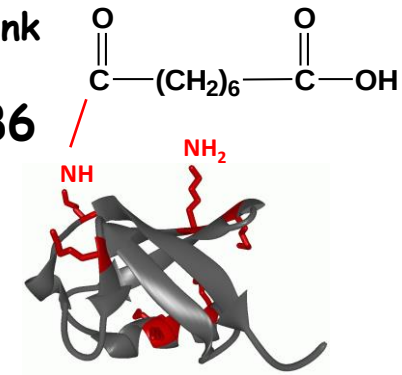
Primary amine reactive cross-linker

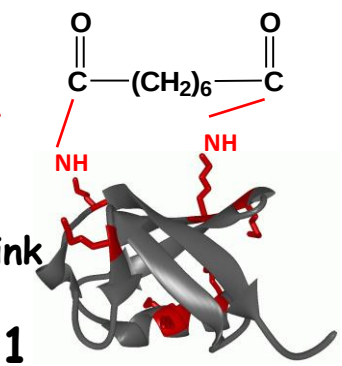
Disuccinimidyl Suberate (DSS)
Cross-linker "arm length"



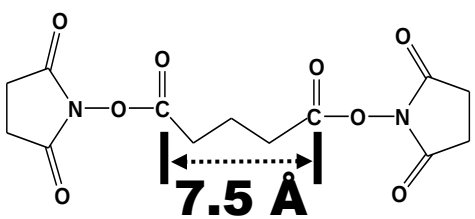


"Hanging" Cross-link
Add $C_8H_{12}O_3$,
Δm = 156.0786

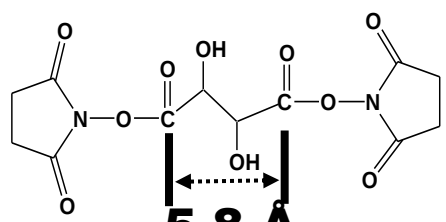




"Internal" Cross-link
Add $C_8H_{10}O_2$,
Δm = 138.0861

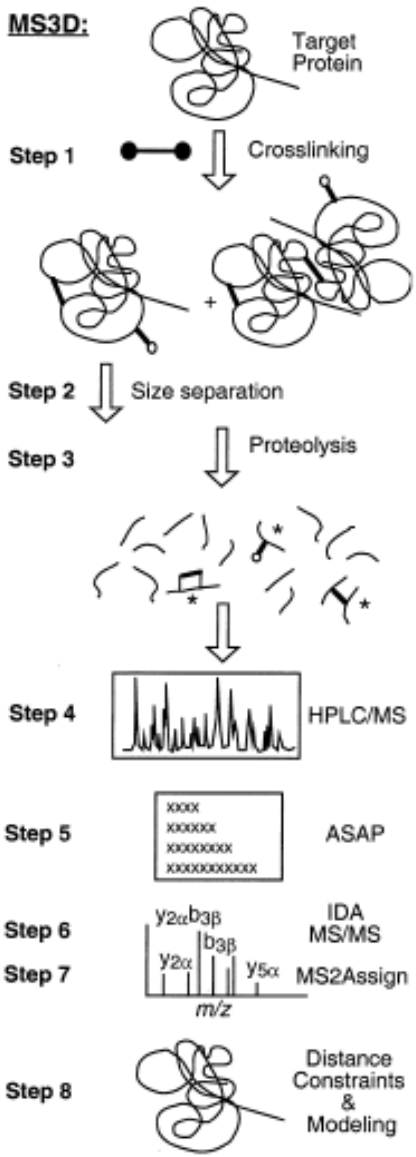


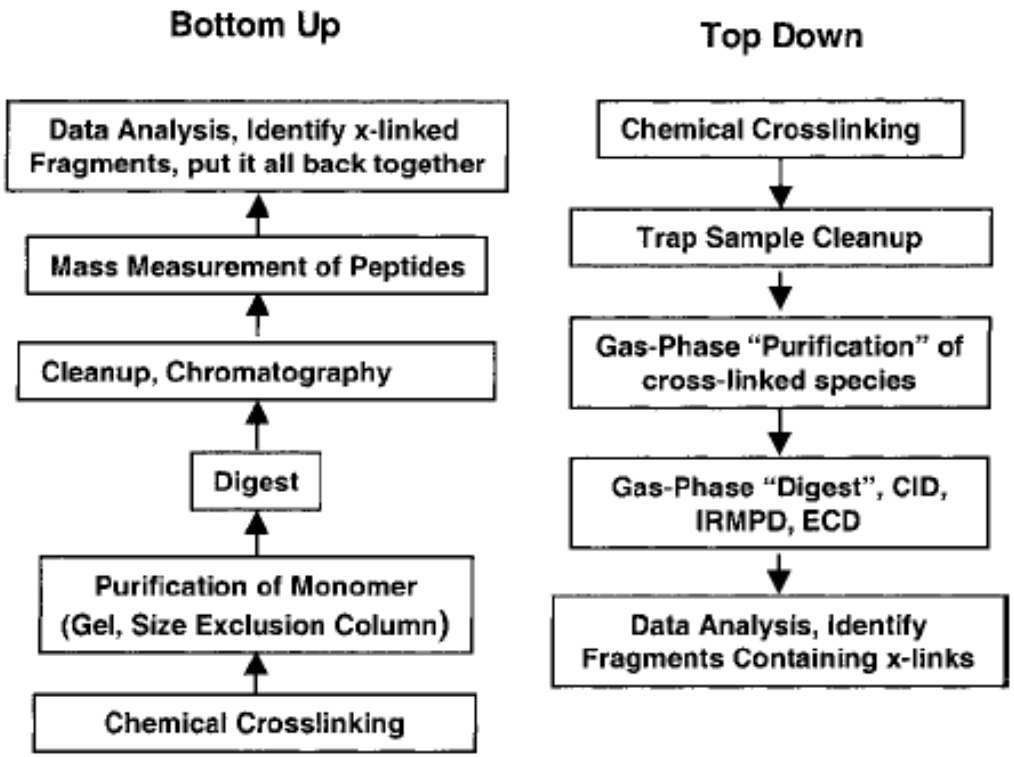
DSG



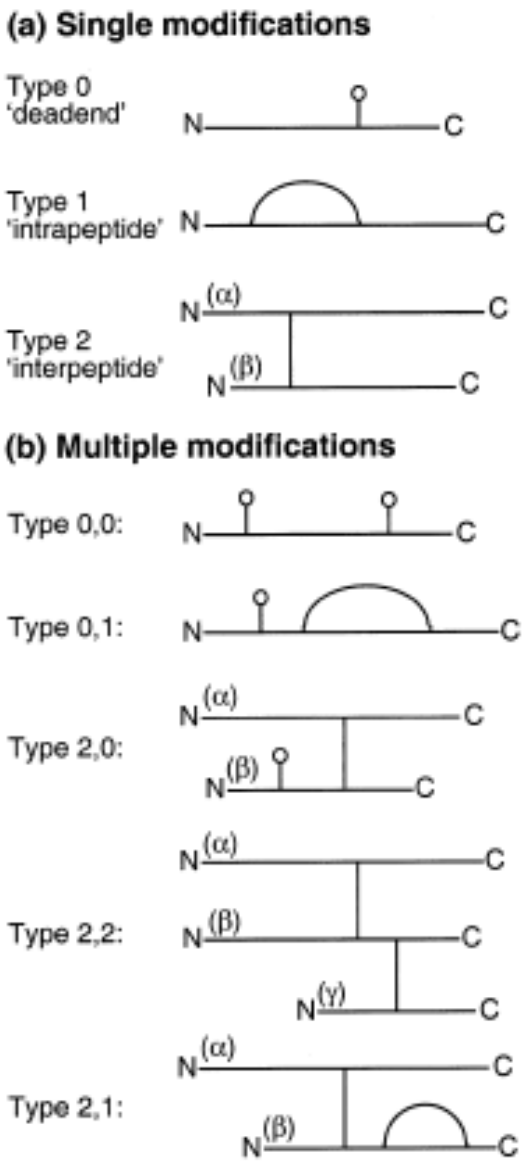
DST

CXMS experiment

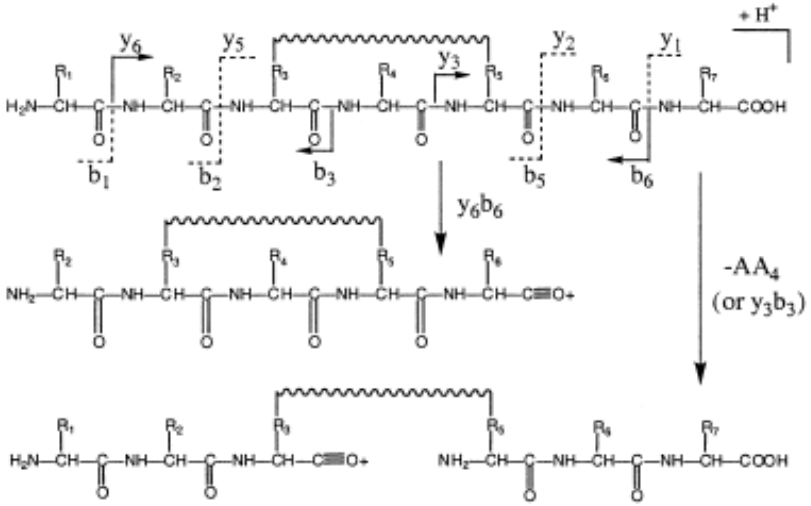




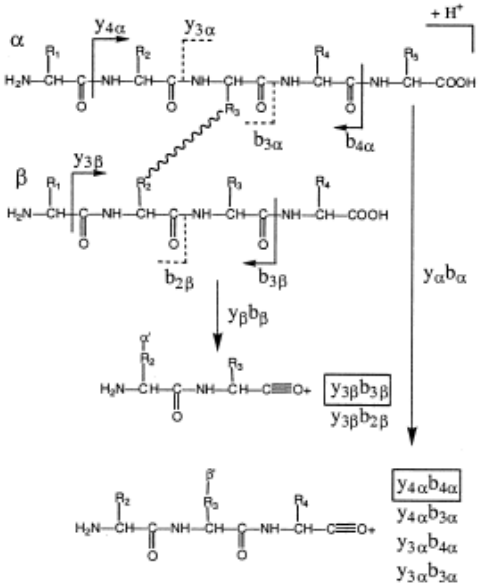
Nomenclature of peptide cross-linked fragments



Type 1

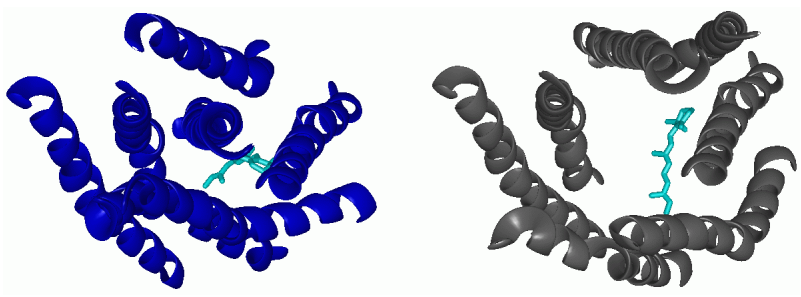


Type 2

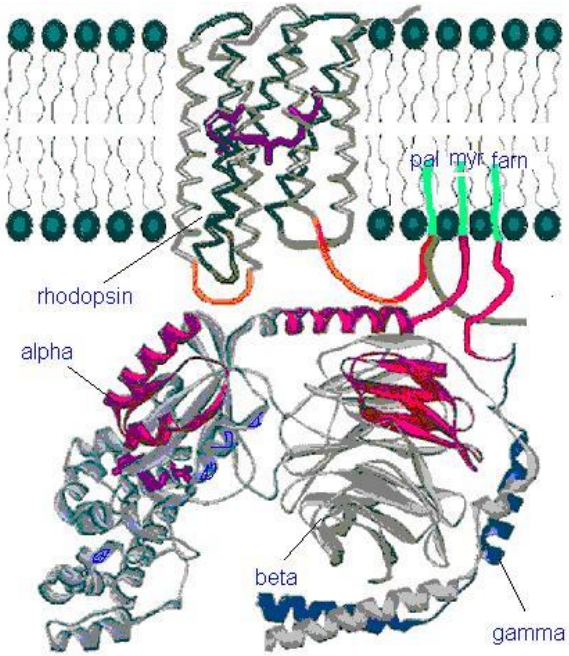


Bottom Up/Top Down - *Rhodopsin*

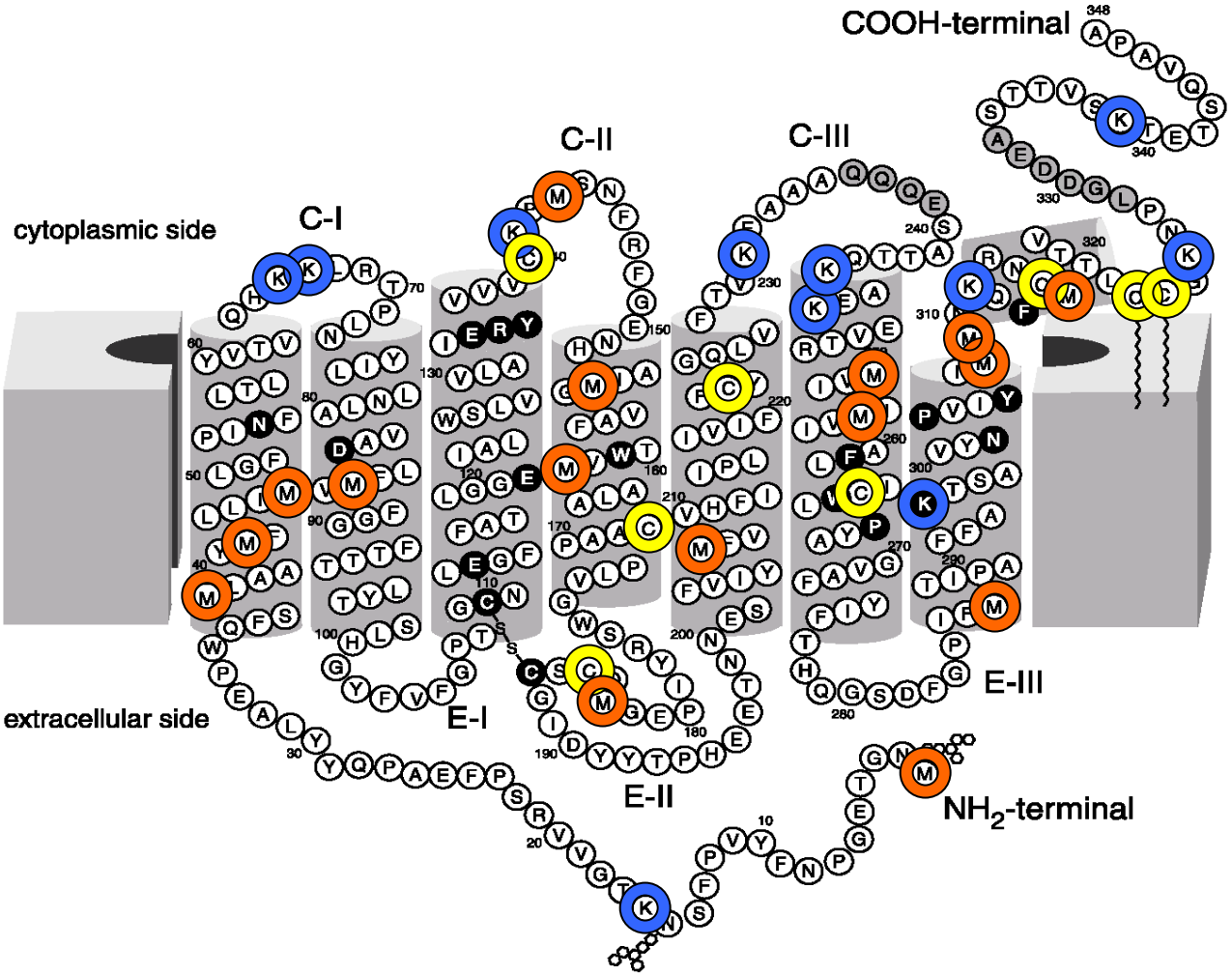
Rhodopsin has open structure/function questions



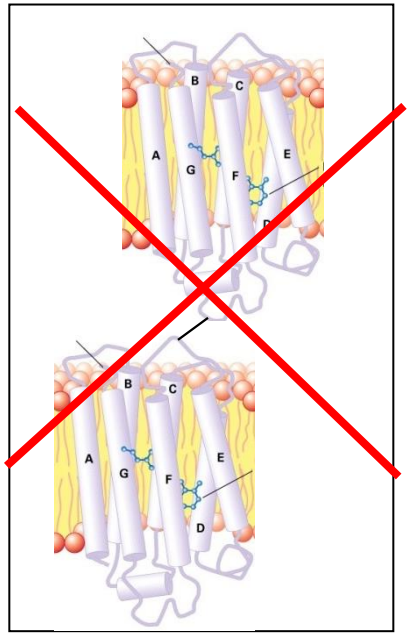
- **What is the conformational change that occurs upon light activation?**
- **What is the configuration of loops involved in Gi binding (not visible on X-ray)?**



Rhodopsin Has Many Potential Cross-links

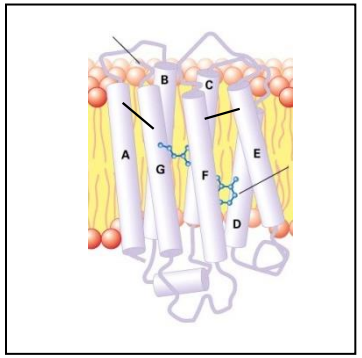


Proteoliposomes/Detergent



Crosslink Protein

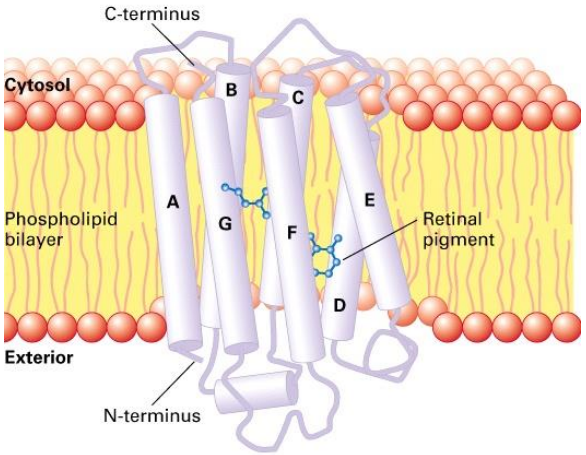
Purify Monomer



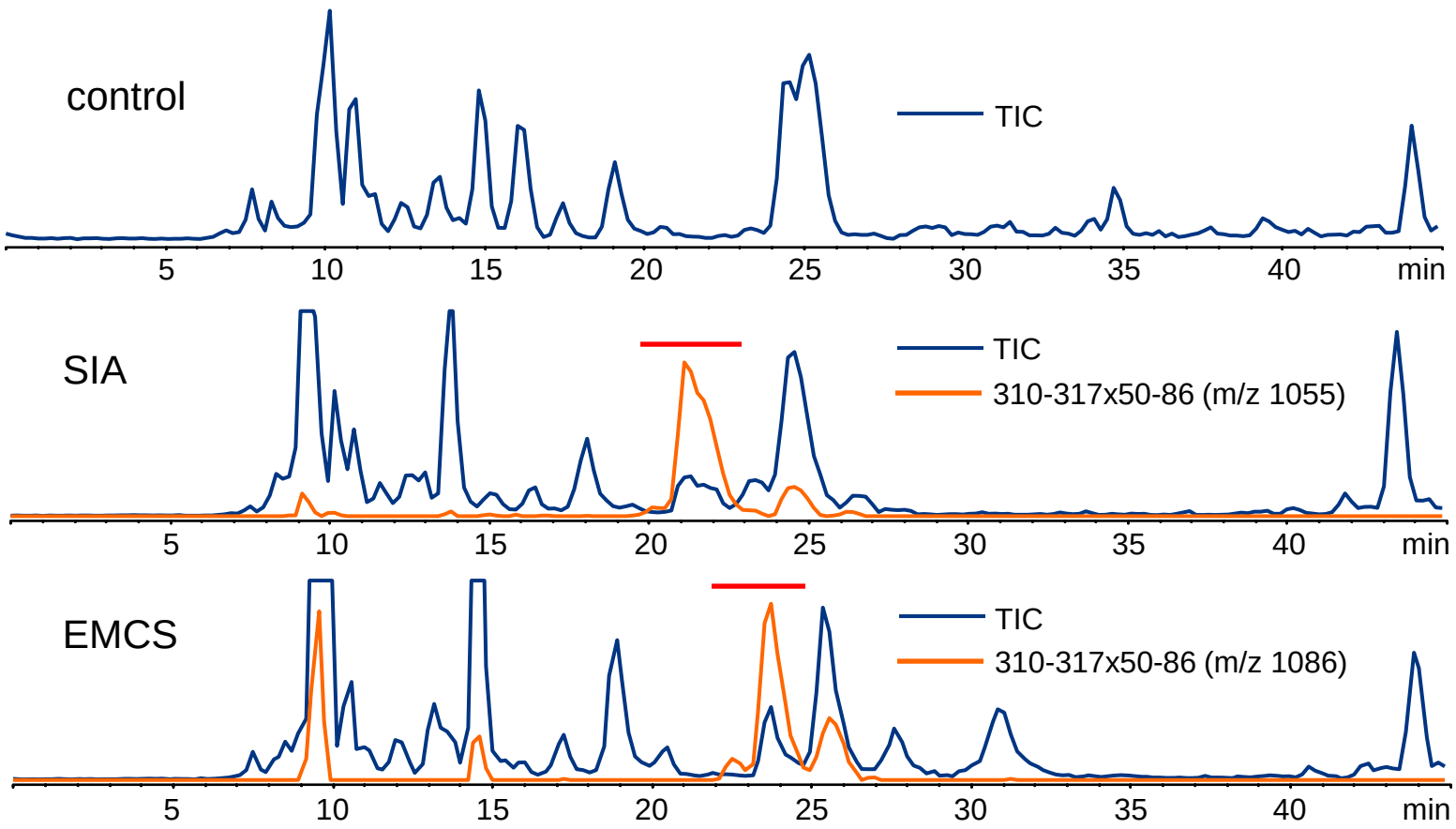
Protein Digestion

LC/MS Analysis

Data Analysis



LCMS analysis of Rhodopsine CNBr digest



Red line corresponds to extracted ion chromatograms of selected cross-linked peptides.

What's wrong? Too many possibilities….

α - 50 LGFPINFLTLYVTVQHKKLRTPLNYILLNLAVADLFM 86

?

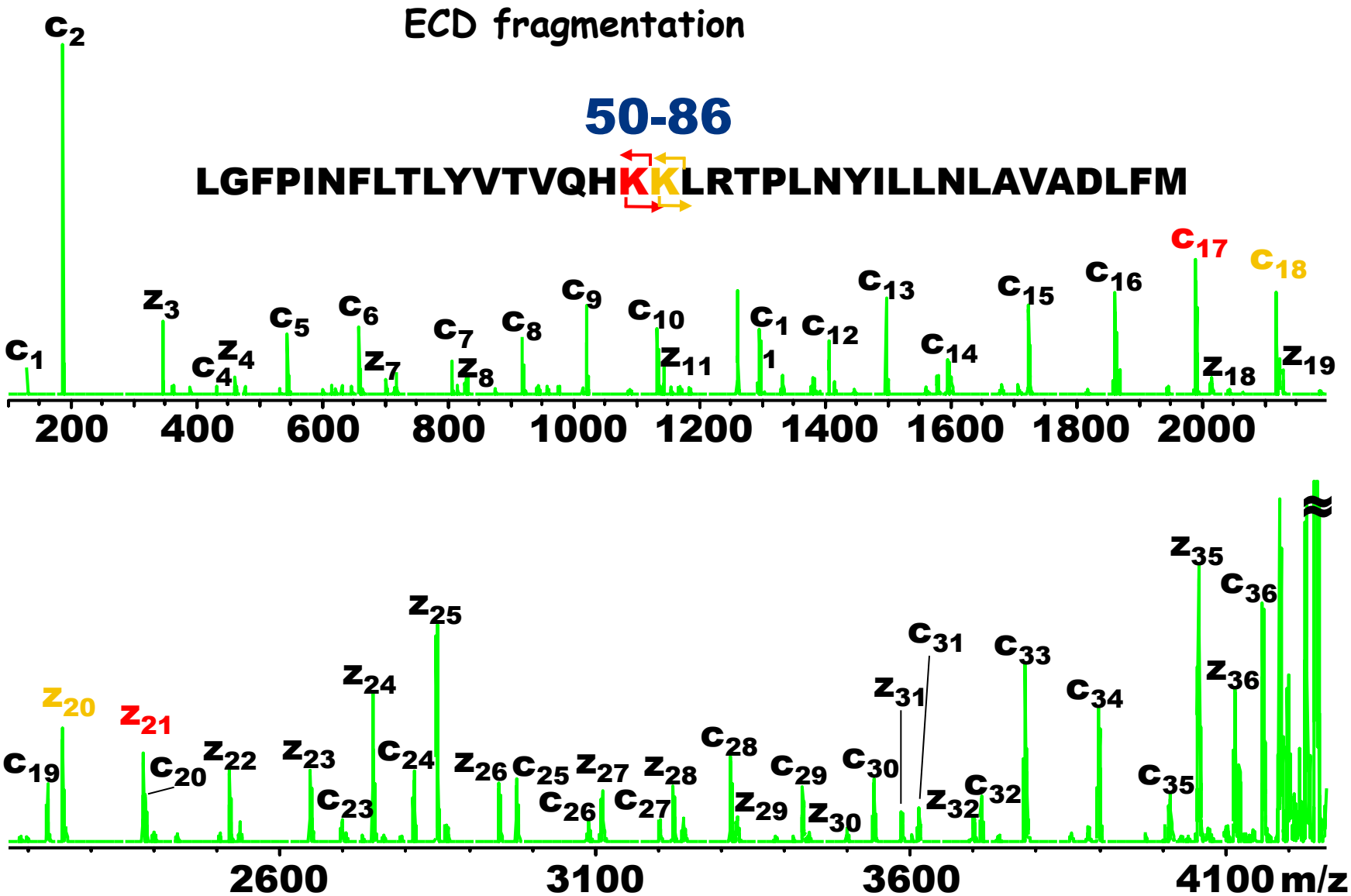
β1 - 310 NKQFRNCM 317

α - 50 LGFPINFLTLYVTVQHKKLRTPLNYILLNLAVADLFM 86

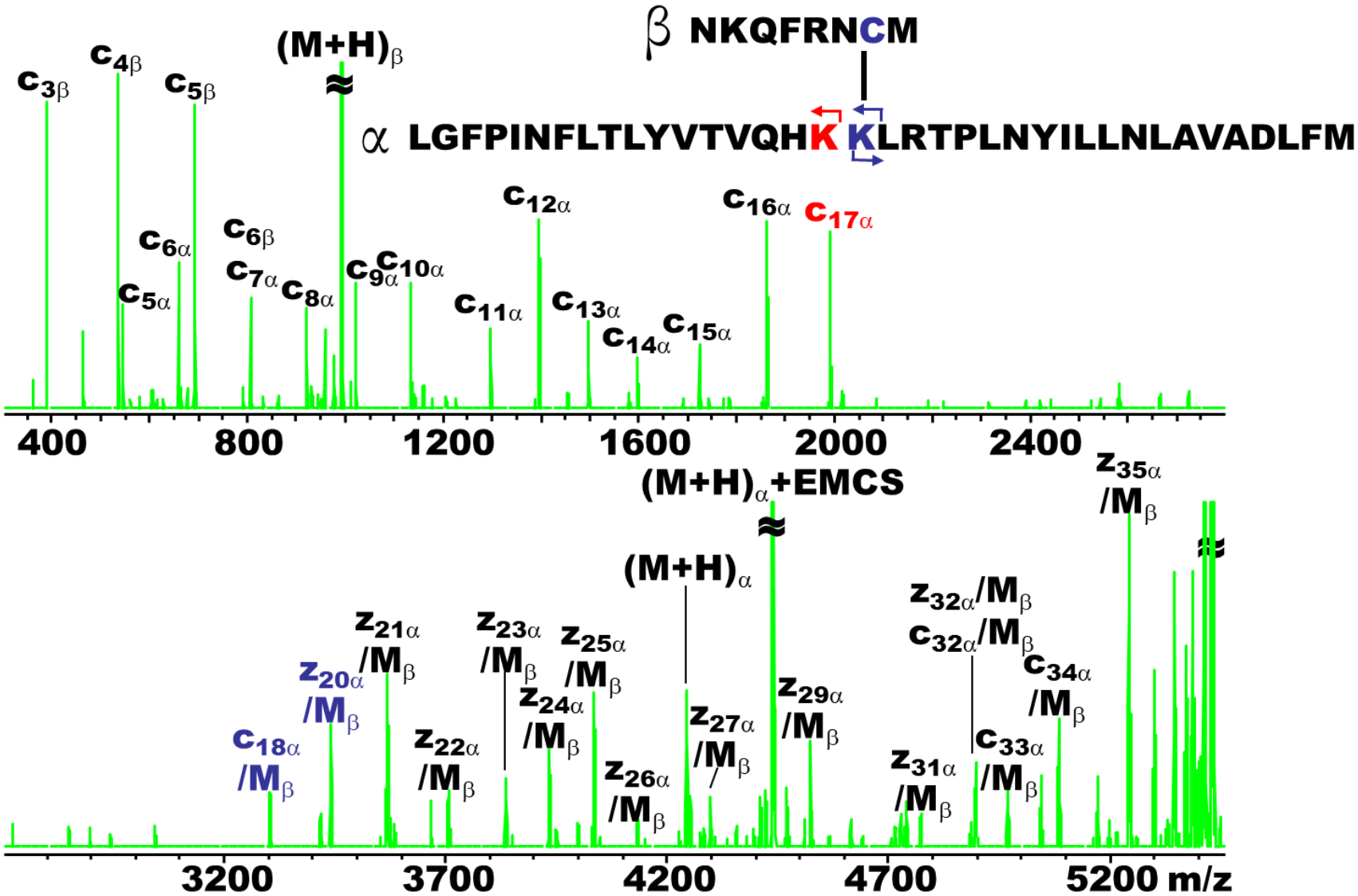
? ?

β2 - 318 VTTLCCGKNPLGDDEASTTVSKTETSQVAPA 348

Can We Resolve the Cross-link at K66/K67?



ECD fragmentation of cross-linked peptides



Problem solved!

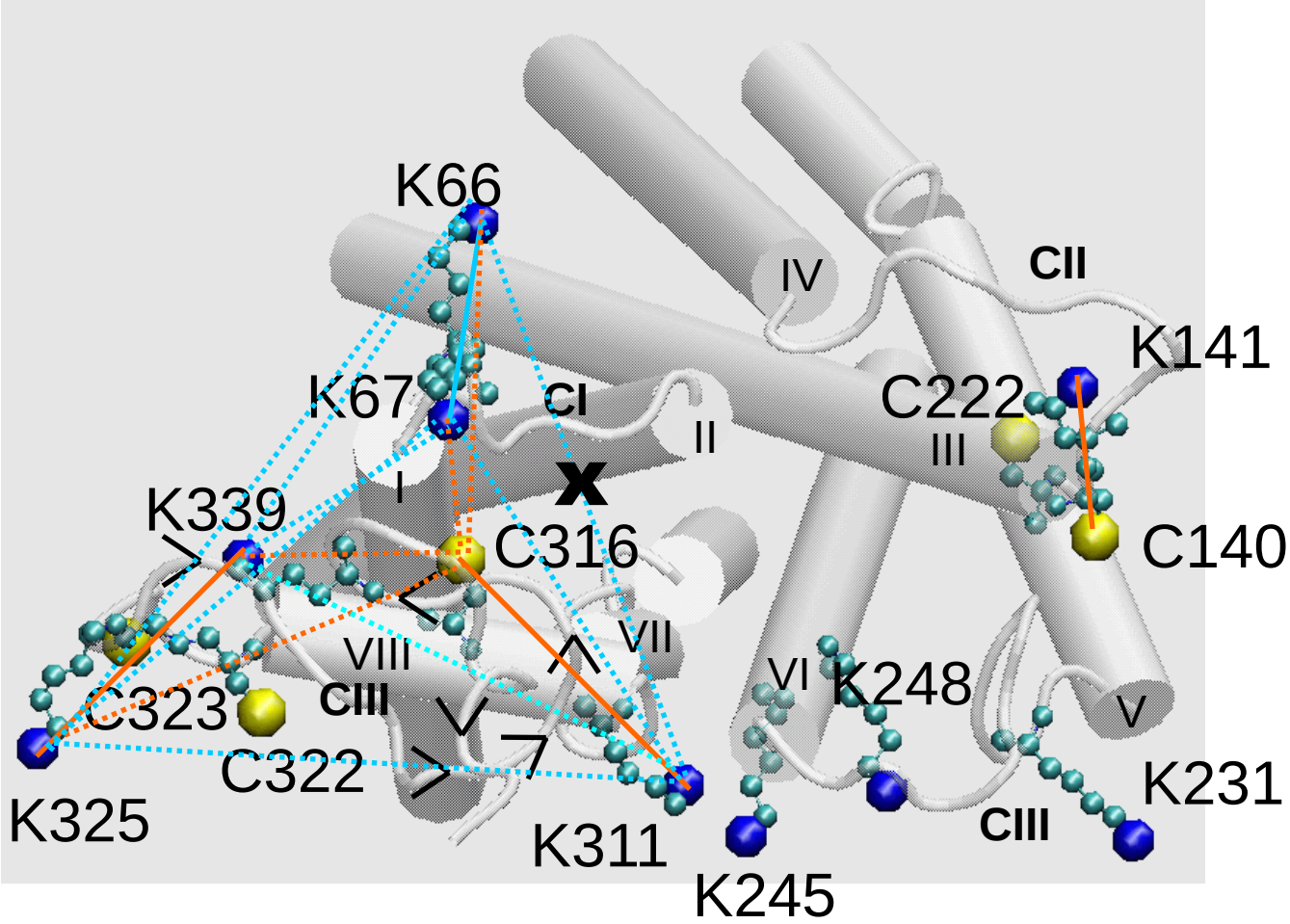
α - 50 LGFPINFLTLYVTVQH**KK**LRTPLNYILLNLAVADLFM 86

β1 - 310 NKQFRN**C**M 317

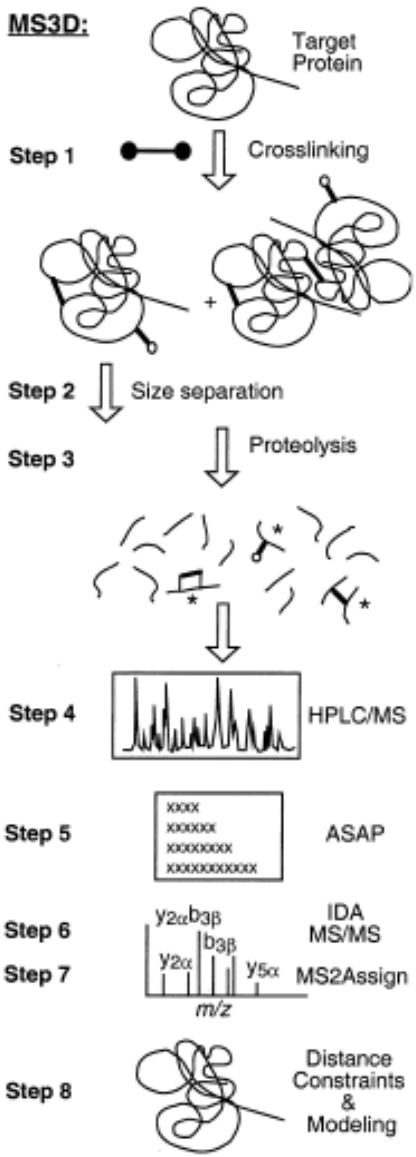
α - 50 LGFPINFLTLYVTVQH**KK**LRTPLNYILLNLAVADLFM 86

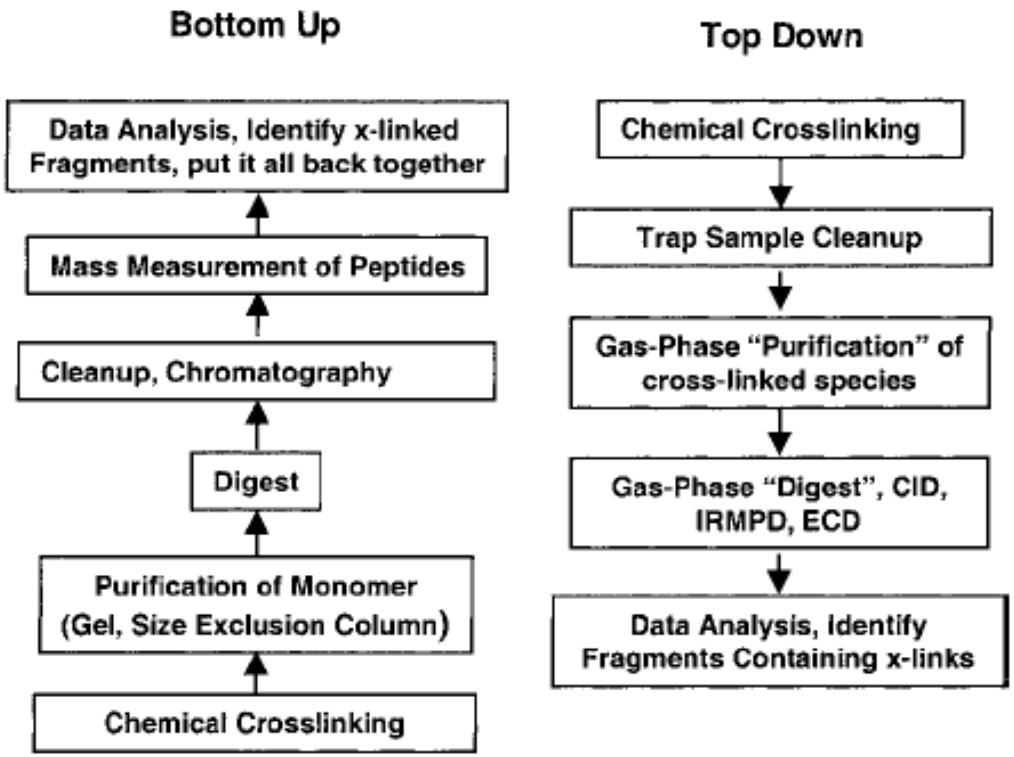
β2 - 318 VTTLCCG**K**NPLGDDEASTTVS**K**TETSQVAPA 348

The Cytoplasmic Face of Rhodopsin

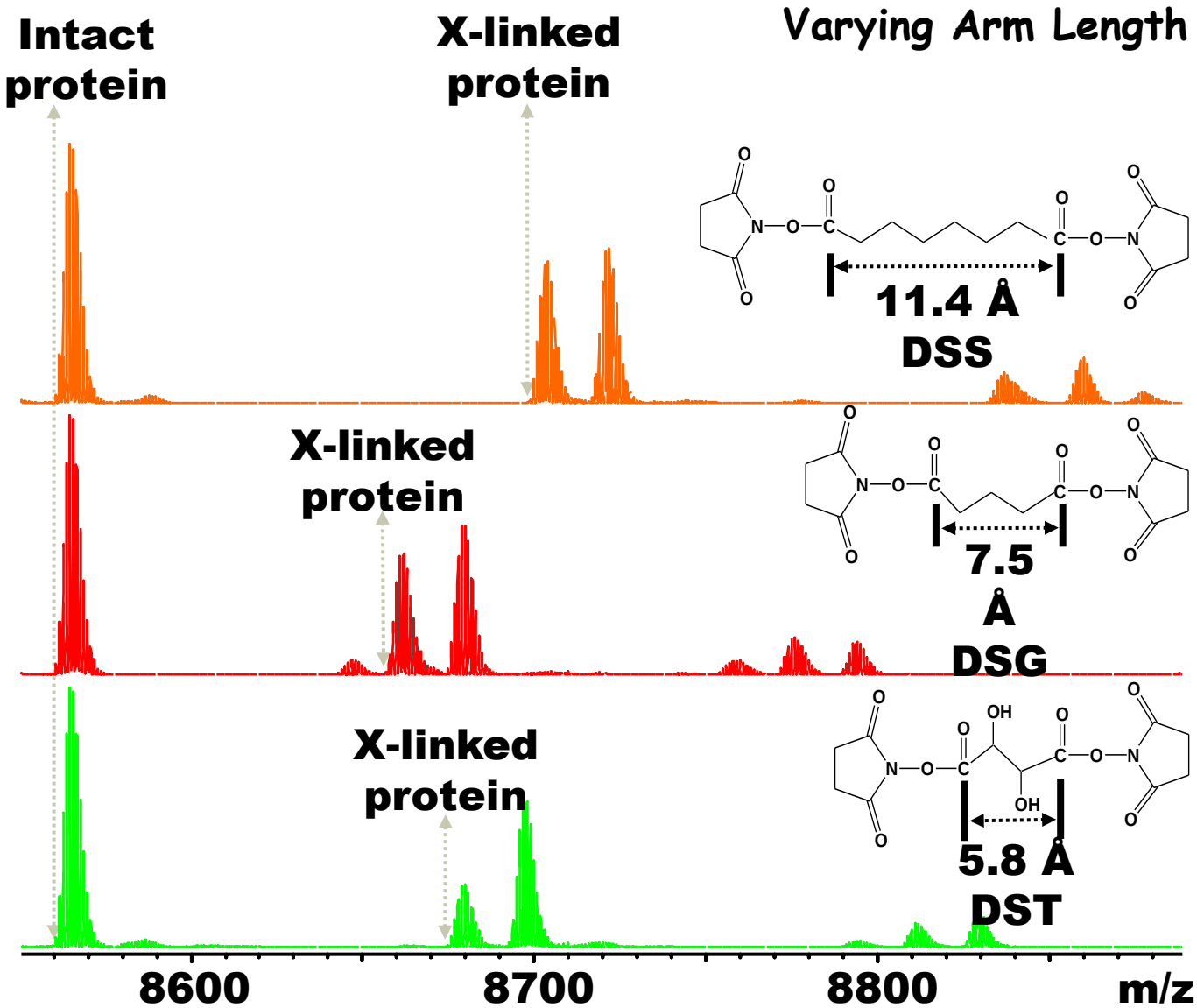


CXMS experiment: Top Down approach

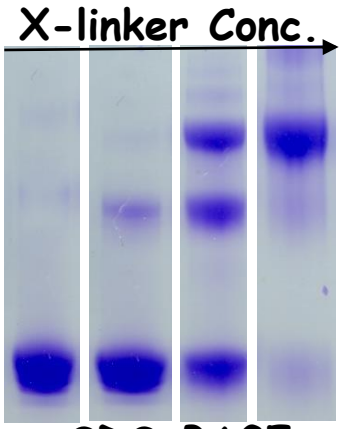


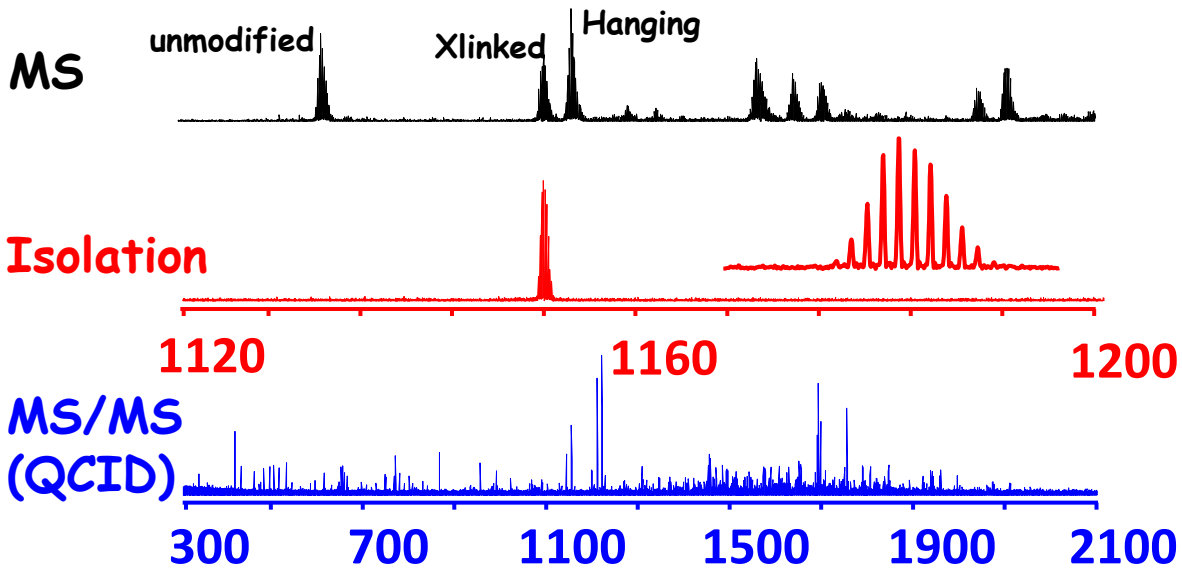


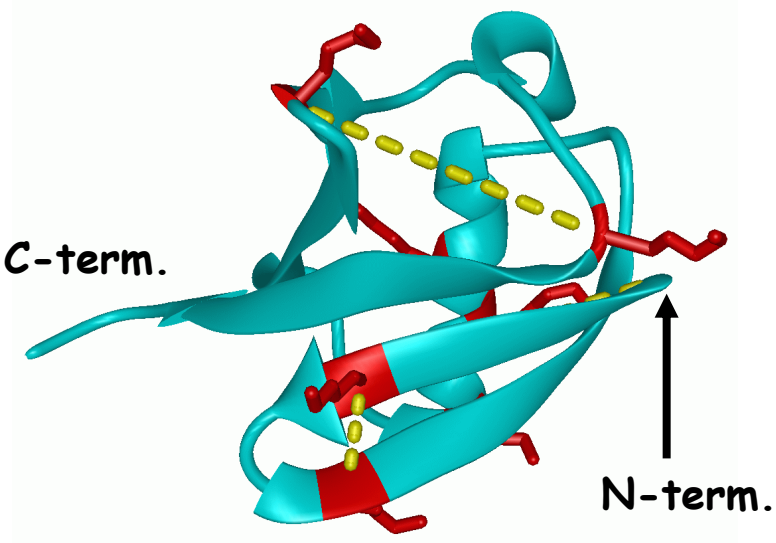
Top down: Cross-linked Ubiquitin with a Series of Cross-linkers



Chemical cross-linking: the identification of cross-link



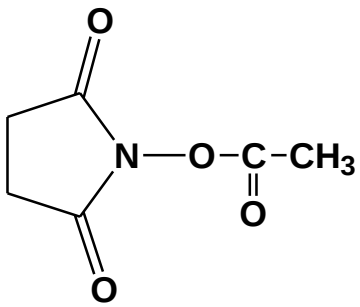


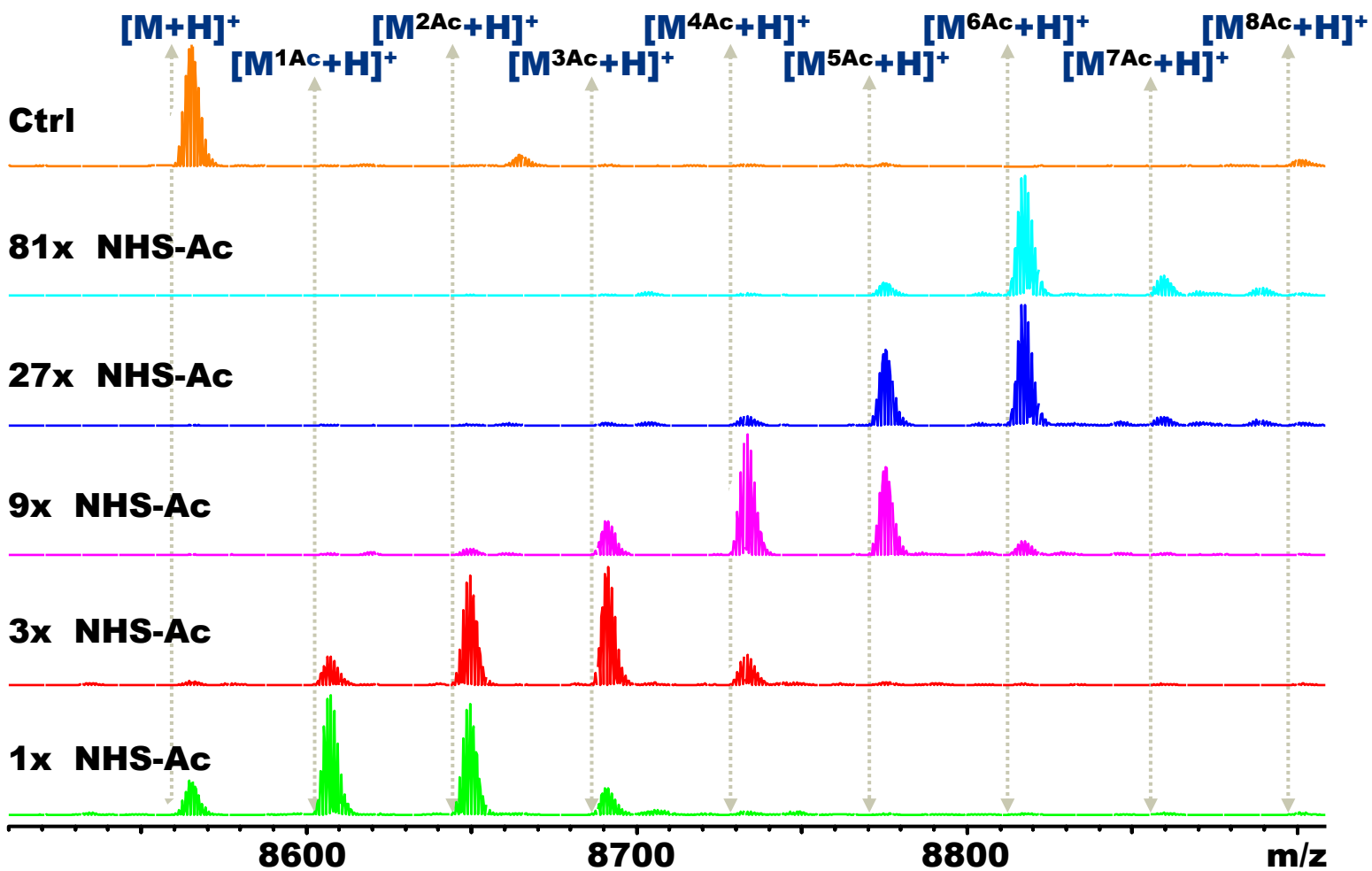


Cross-Link	DSS	DSG	DST	Constraint (N_e-N_e)
M1 – K6	Yes	Yes	Yes	Distance < 5.8 Å
K6-K11	Yes	Yes	Yes	Distance < 5.8 Å
K48-K63	Yes	Yes	No	5.8 Å < Distance < 7.5 Å

Novak et. al. *Eur. J. Mass Spectrom. 2003, 9, 623*

Protein covalent labeling: why only few cross-links?





Protein covalent labeling: a reactivity of lysine in an issue

^{1}MQIFVKTLTG ^{11}KTITLEVEPS ^{21}DTIENVKAKI ^{31}QDKEGIPPDQ
^{41}QRLIFAGKQL ^{51}EDGRTLSDYN ^{61}IQKESTLHLV ^{71}LRLRGG

(^{1}M~K6~K48~K63) > K33 > K11 > (K27,K29)

- In agreement with NMR data, which shows.
  - K11 interacts with E34; K29 interacts with D21
- Crystal structure indicates K27 H-bonds to D52.
- More reactive lysines don't H-bond (K63) or H-bond to backbone carbonyls (K48, K33).
- K48 and K63 participate in formation of polyubiquitin.
- Agrees with Cross-linking Results.
  - Cross-links observed only between the most reactive lysines and nearby reactive lysines !!!

Novak et. al. *J. Mass Spectrom. 2004, 39, 322*

Zero-Length and Carboxy-Carboxy Cross-linking

"ZERO-LENGTH" CROSS-LINKING

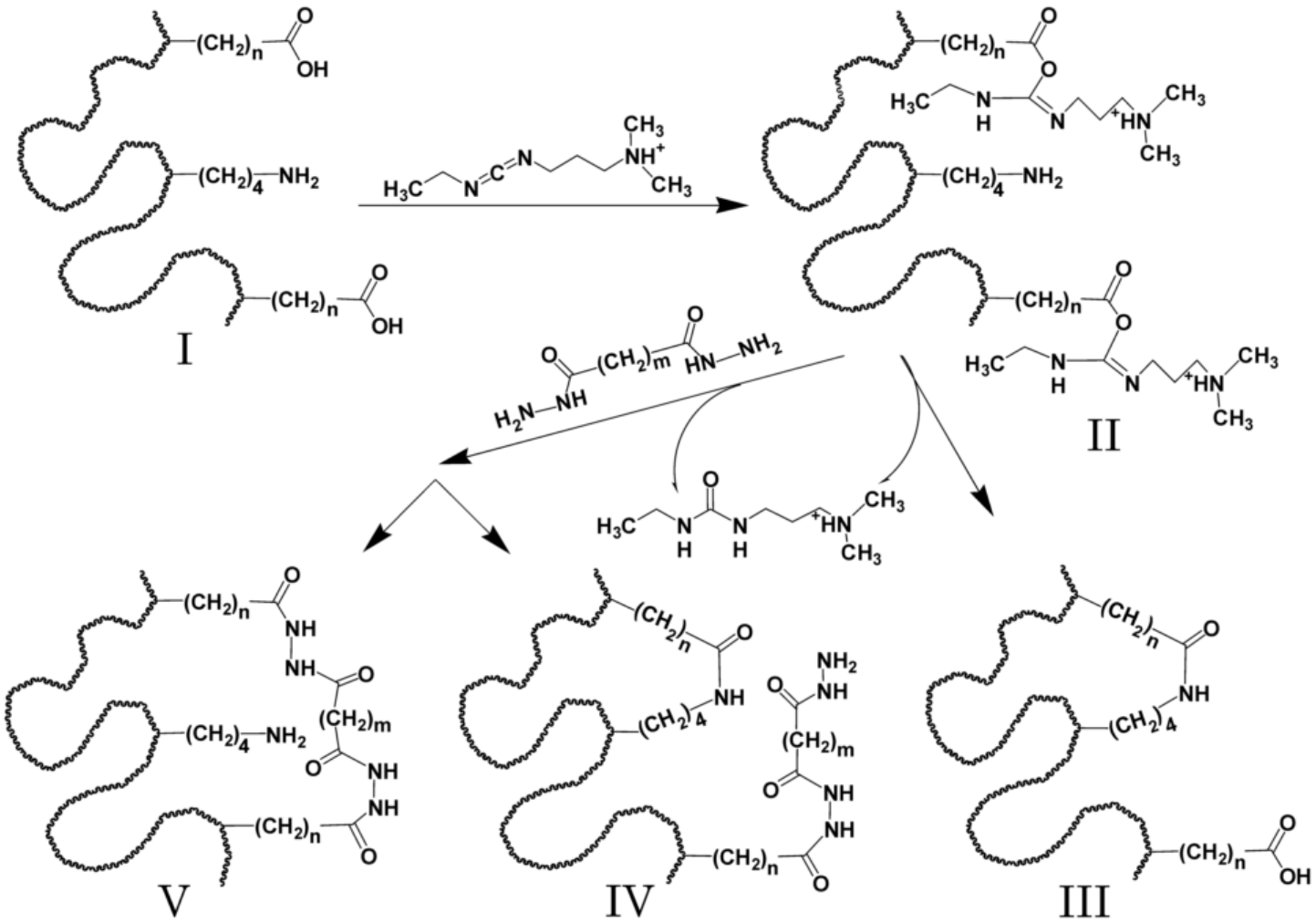
- No cross-linker used.
- Activate carboxylic acid groups with EDC.
- Activated acid side-chains react with primary amine side-chains (DEO-XK).
- Cross-link formed via new amide linkage.

EDC ACTIVATION CAN ALSO BE USED TO CROSS-LINK ACIDIC SIDE-CHAINS TO EACH OTHER (DEO-DEO)

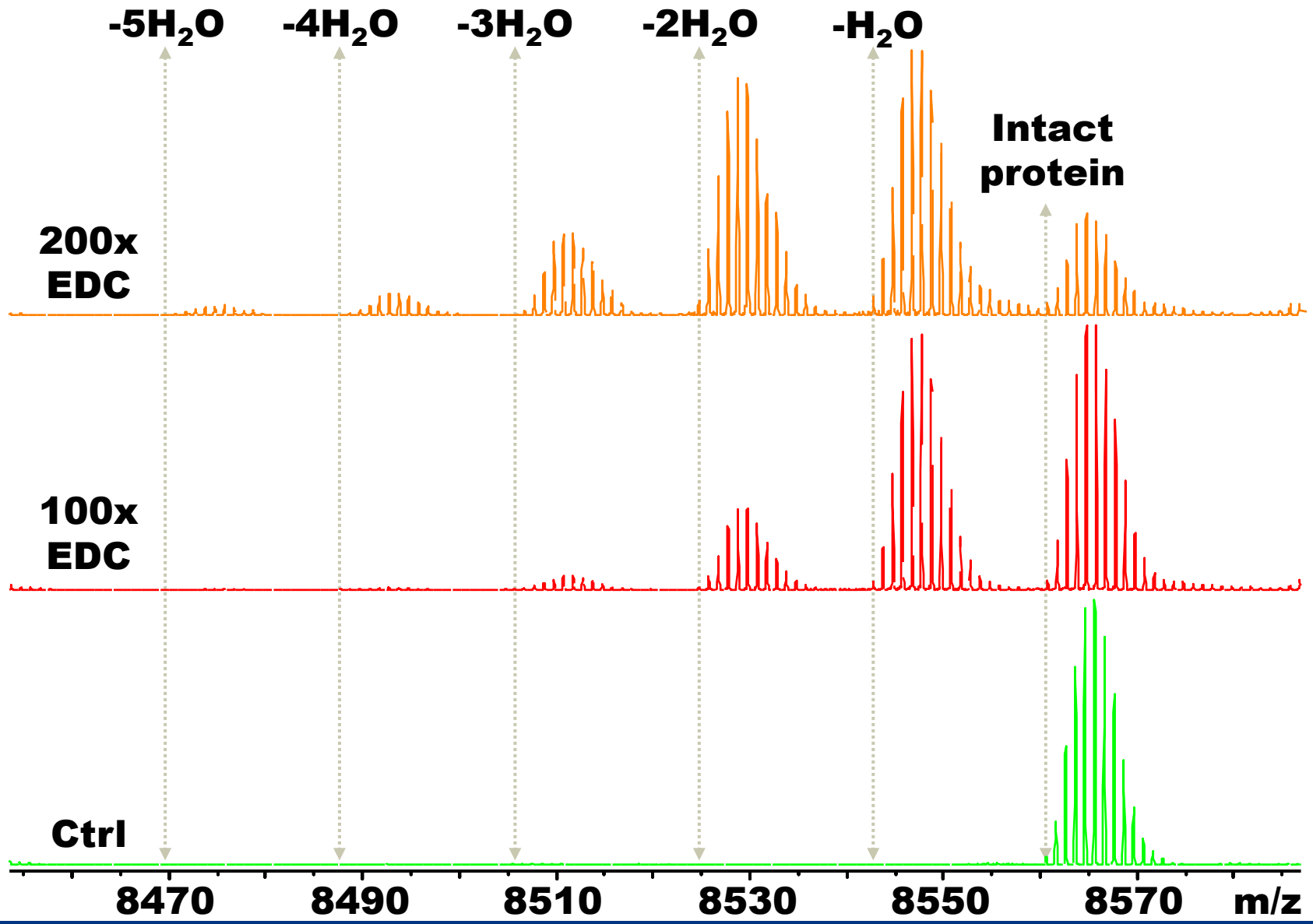
- Use dihydrazides as the cross-linking reagent.

Chemical cross-linking: an alternative chemistry

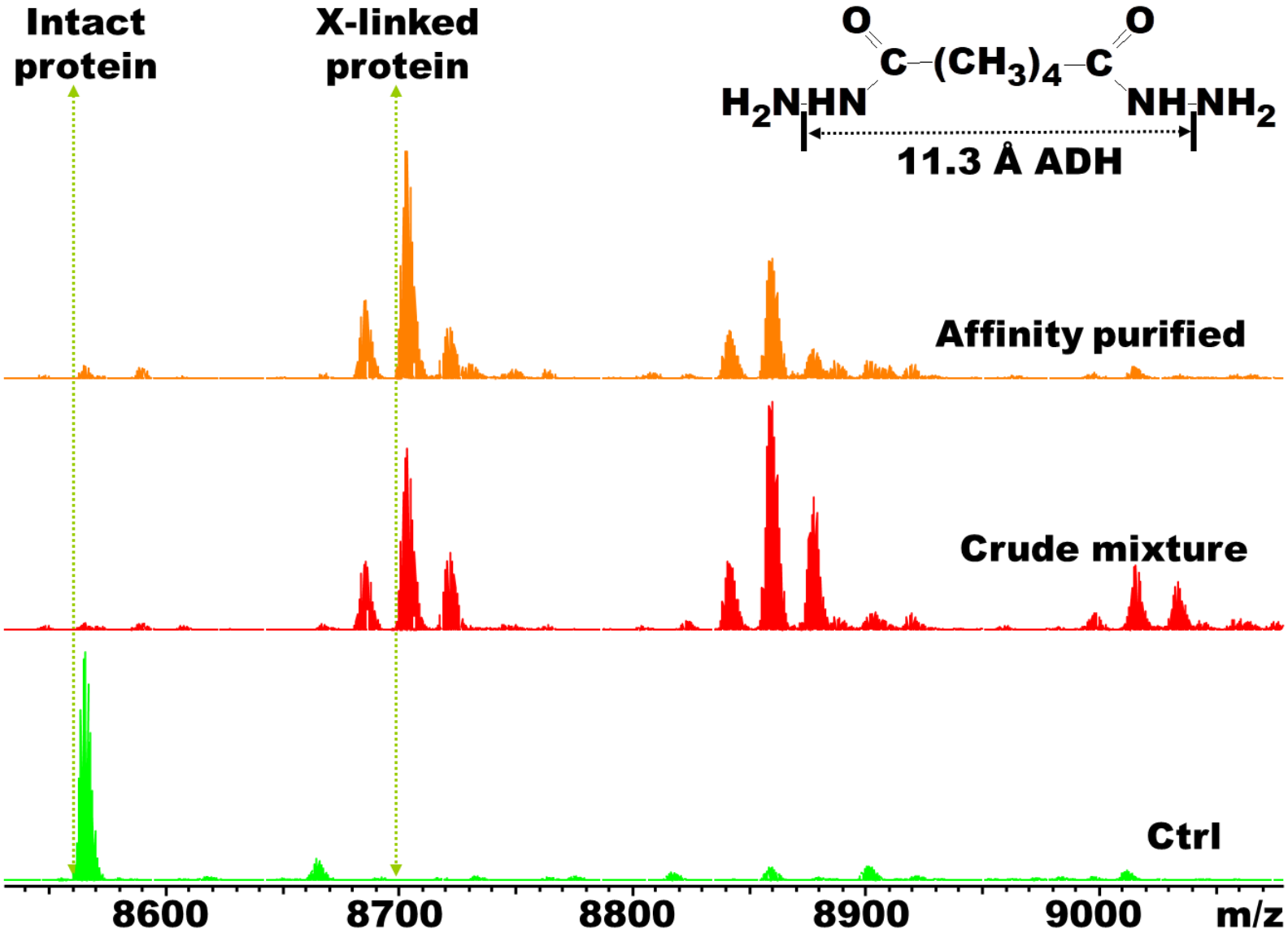
Carboxylic group reactive cross-linkers



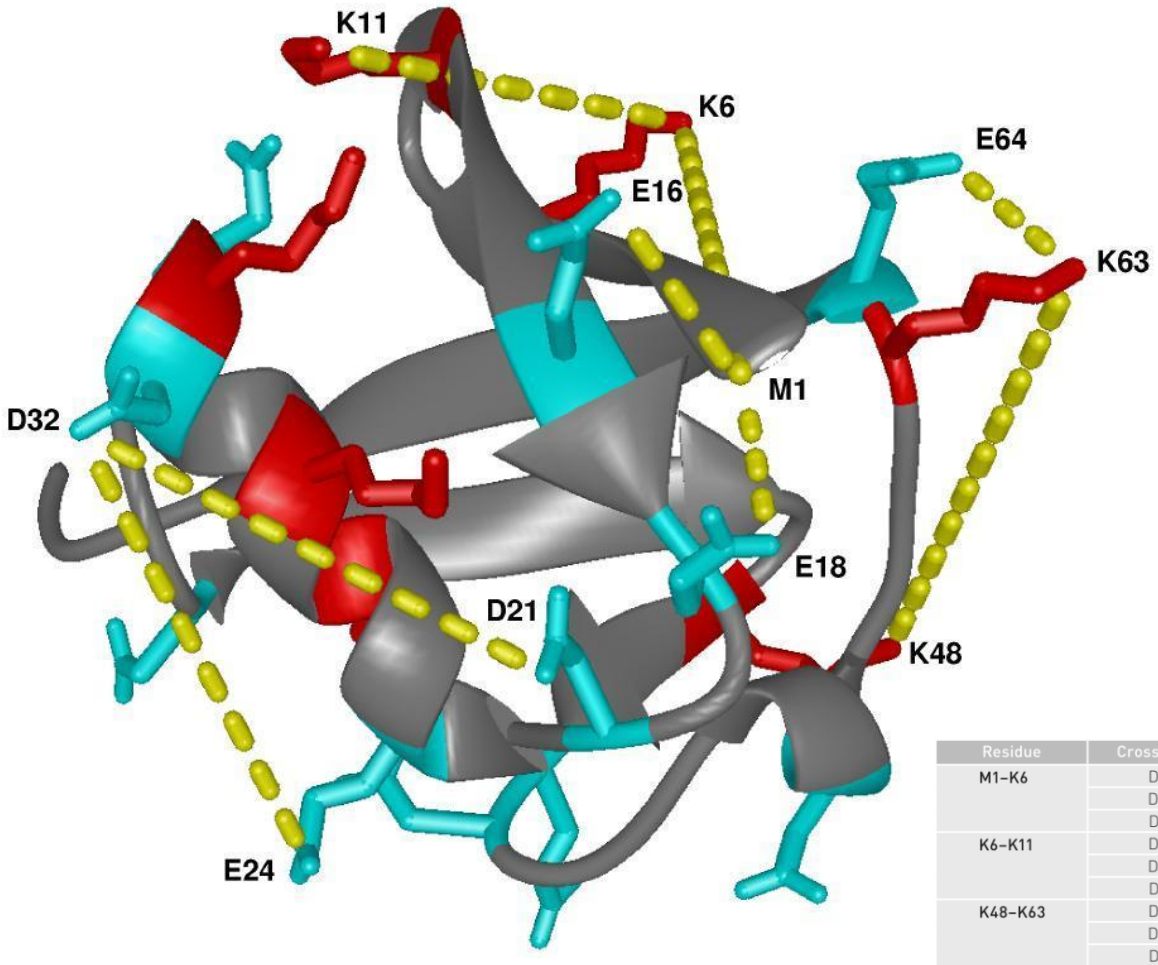
Zero-Length Cross-linking



Carboxy-Carboxy Cross-linking



Chemical cross-linking: an alternative chemistry



Residue	Cross-linker	Cross-link	Constraint (X-ray constraint)	Type
M1–K6	DSS	Yes	5.8Å < Distance < 7.5Å (20.0Å)	N_α–N_ε
	DSG	Yes		
	DST	No		
K6–K11	DSS	Yes	Distance < 5.8Å (14.0Å)	N_ε–N_ε
	DSG	Yes		
	DST	Yes		
K48–K63	DSS	Yes	5.8Å < Distance < 7.5Å (19.8Å)	N_ε–N_ε
	DSG	Yes		
	DST	No		
M1–E16	EDC	Yes	Distance < 1.5Å (6.2Å)	N_α–C_δ
M1–E18	EDC	Yes	Distance < 1.5Å (4.4Å)	N_α–C_δ
K63–E64	EDC	Yes	Distance < 1.5Å (4.8Å)	N_ε–C_δ
D21–D32	ADH	Yes	5.8Å < Distance < 7.5Å (12.9Å)	C_γ–C_γ
	SDH	No		
E24–D32	ADH	Yes	5.8Å < Distance < 7.5Å (14.0Å)	C_δ–C_γ
	SDH	No		

Novak et. al. *Eur. J. Mass Spectrom. 2008, 14, 355*

Bottom Up/Top Down

TOP DOWN MS3D WORKS WELL FOR:

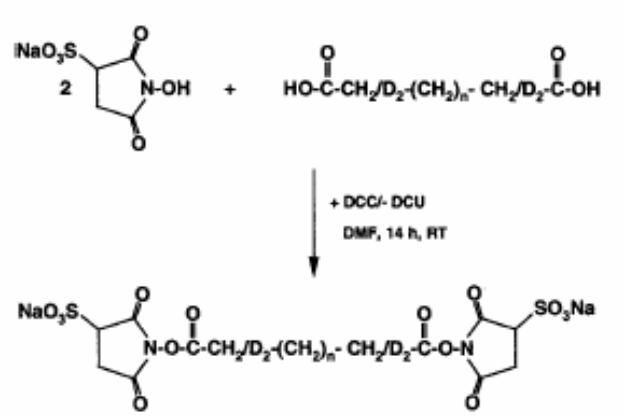
- Characterizing new cross-linking chemistry on small model proteins (< 20 kDa).
- Small proteins with unknown or partially known structures.
- Reactivity studies.

BOTTOM UP APPROACH WITH DIGEST FOR LARGER PROTEINS

- Localizing cross-links with digests alone is difficult
- Combined bottom up/top down approach
  - Digest protein with a single protease
  - Analyze the large cross-linked fragments with top down methods

Introduction of isotopically labeled probes

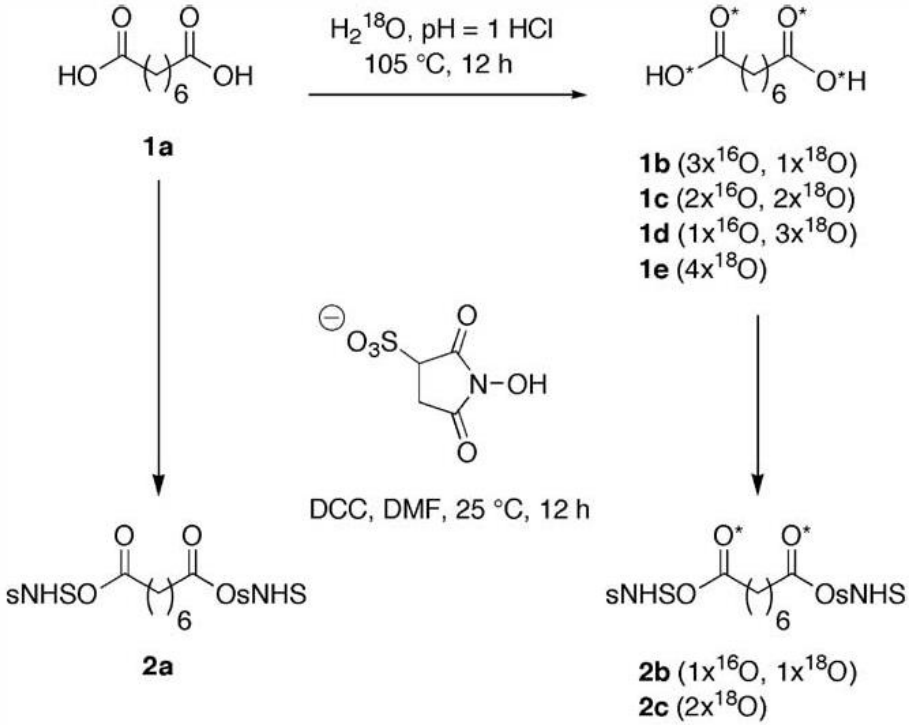
Incorporation of stable isotopes (deuterium) to the linker



^{13}C available as well

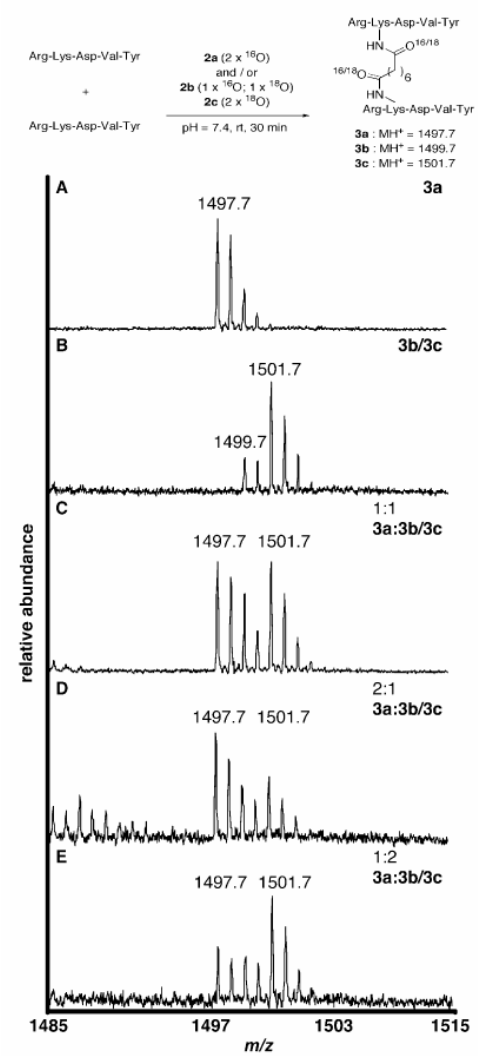
Muller DR. et al. Anal. Chem. 2001

Incorporation of stable isotopes (oxygen, ^{18}O) to the linker



Collins CJ. et al. Bioorg. Med. Chem. Lett. 2003

Introduction of isotopically labeled probes



Simplifies data analysis

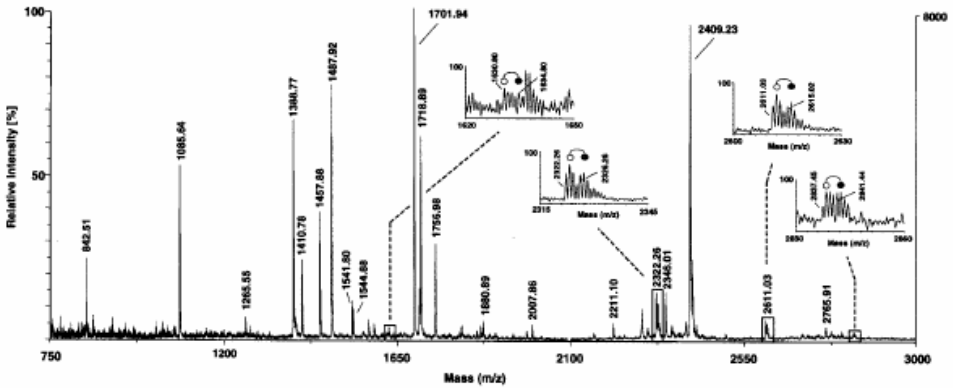
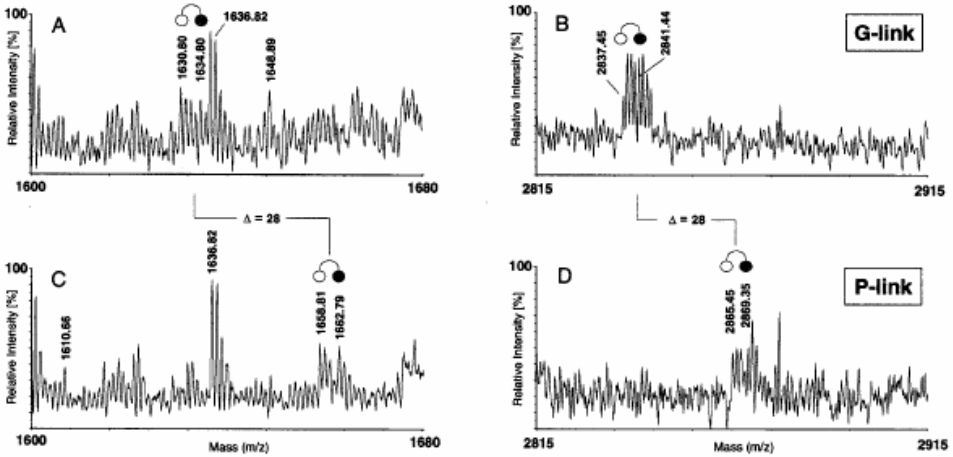


Figure 4. MALDI mass map obtained from G-linked Op18–tubulin complexes with doublet regions expanded in insets.

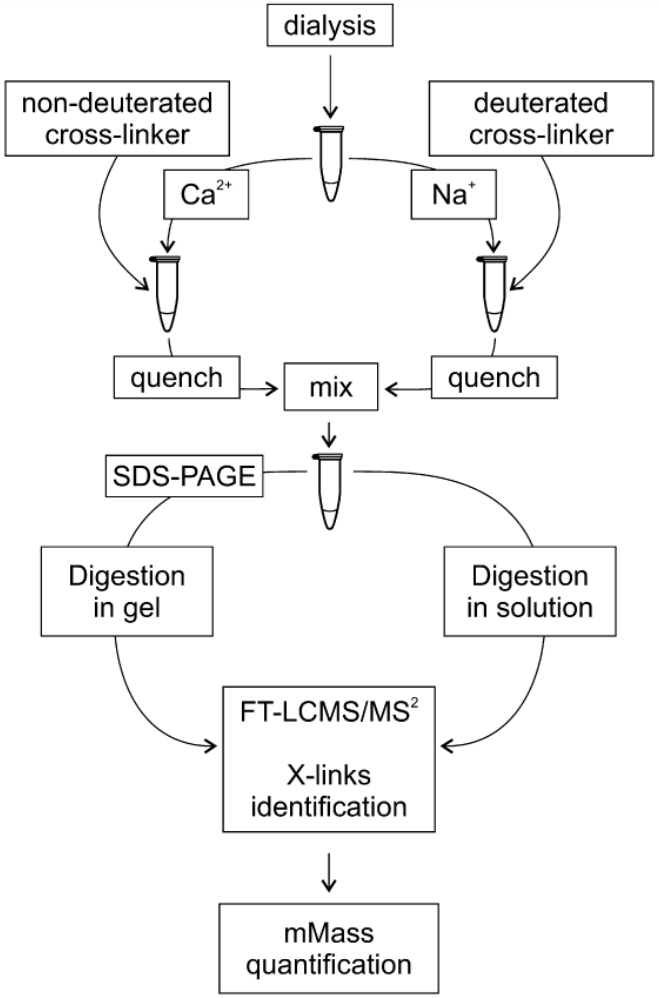


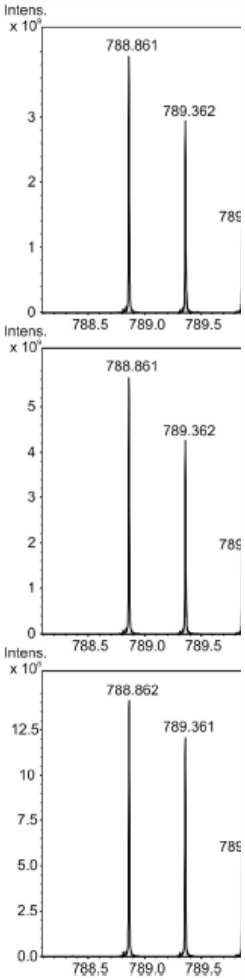
Collins CJ. et al. Bioorg. Med. Chem. Lett. 2003

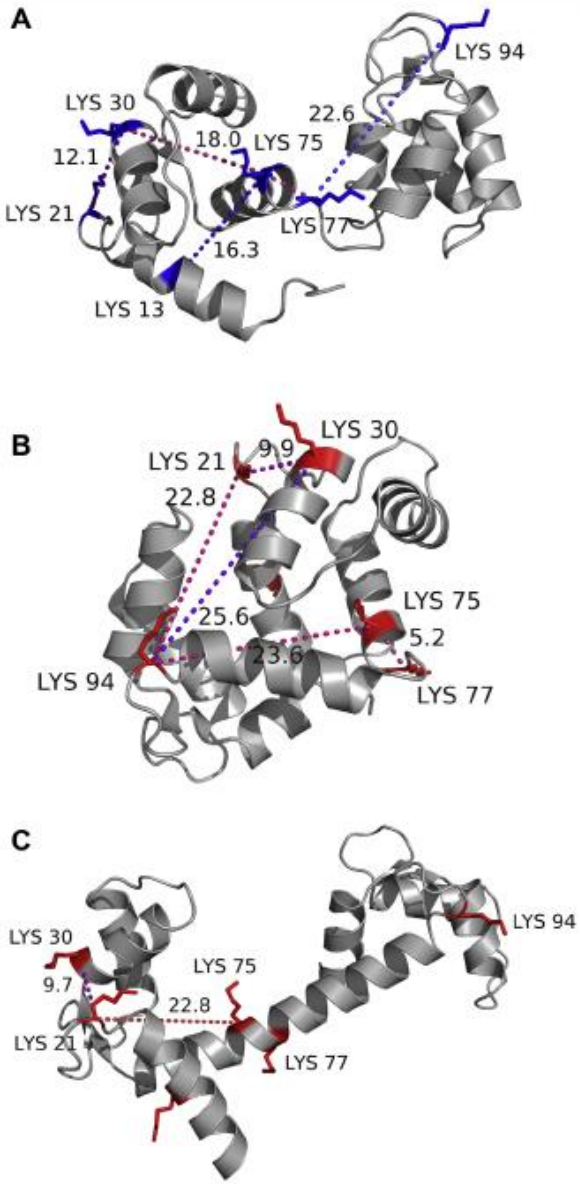
Muller DR. et al. Anal. Chem. 2001

Quantitative chemical cross-linking

Protein dynamics and conformat







er

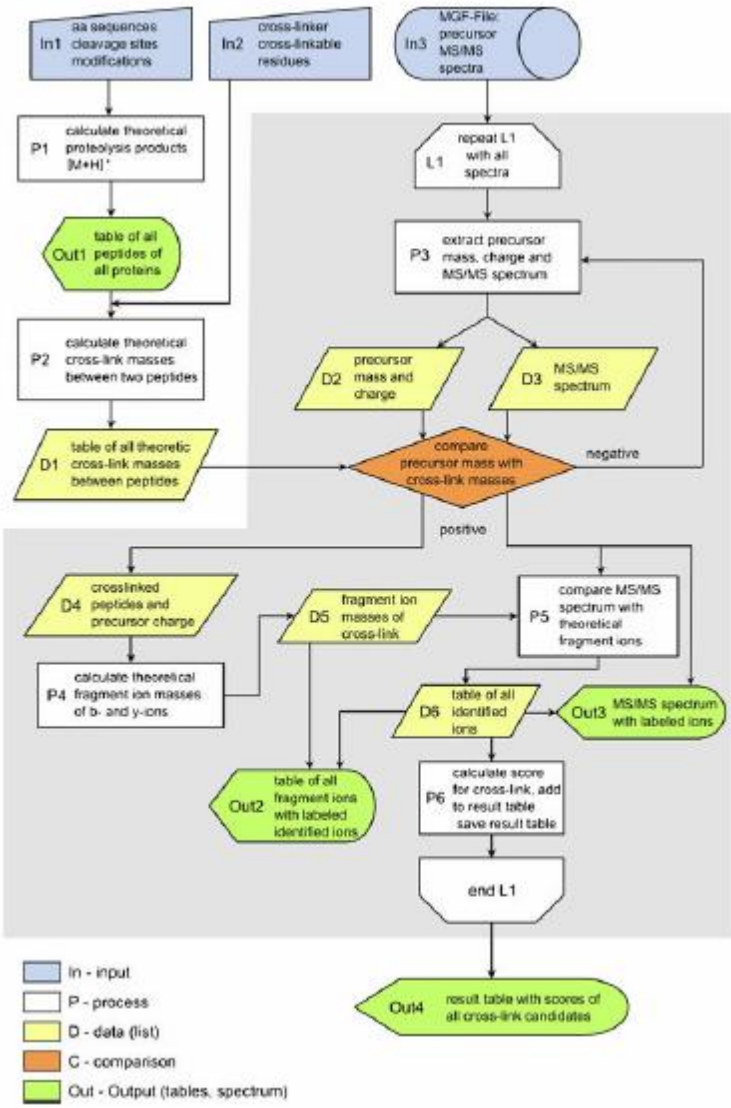
Na$^+$

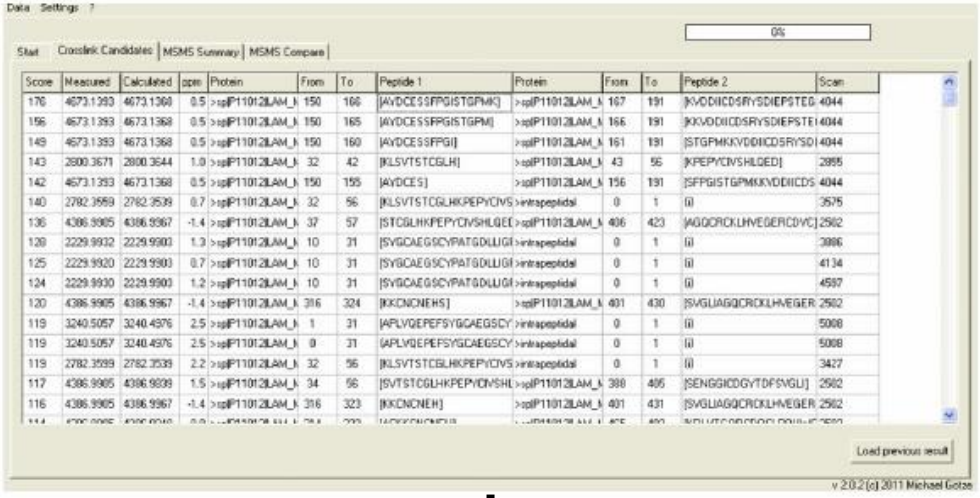
ers

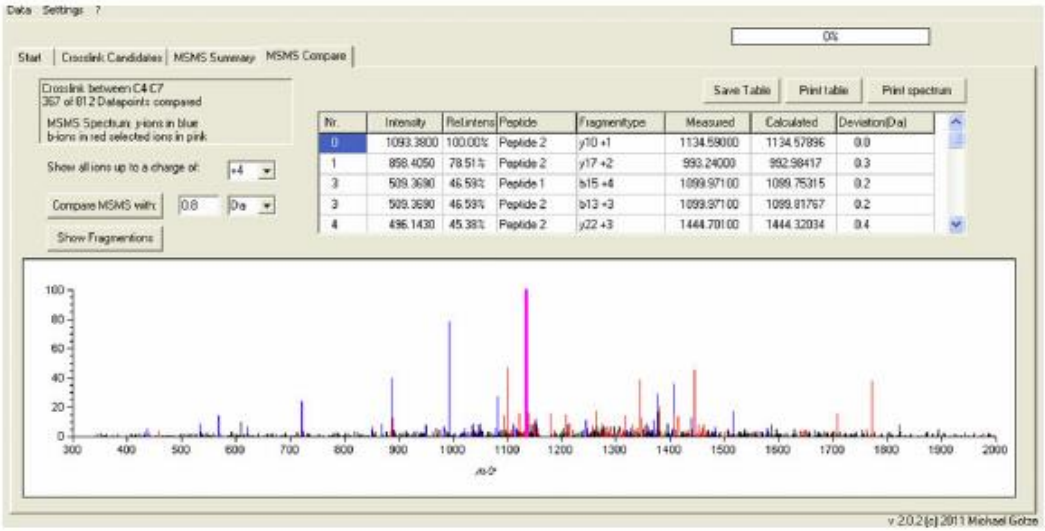
er

Kukacka Z. et al. Methods 2015

Data analysis: STAVROX



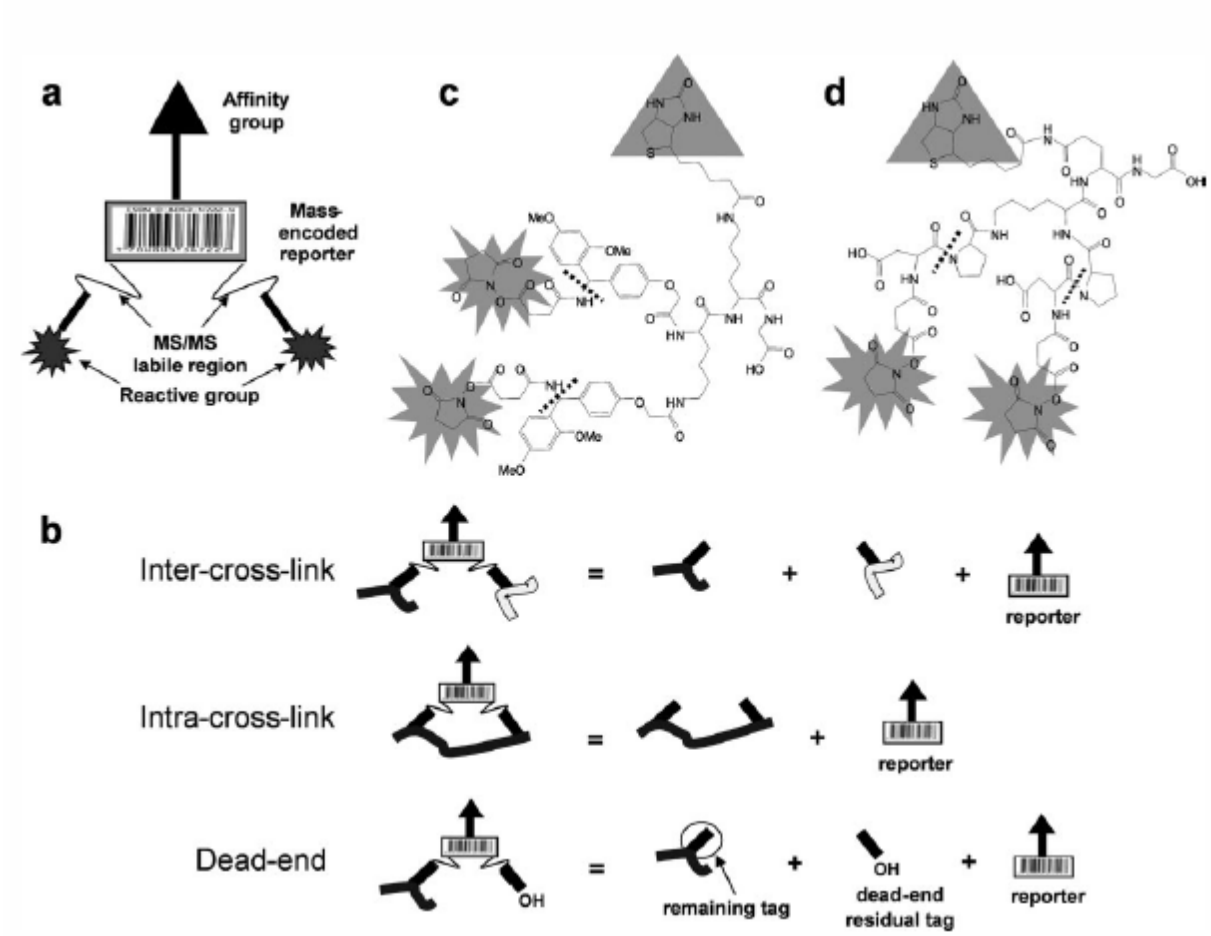




Gotze M. et al. JASMS 2012

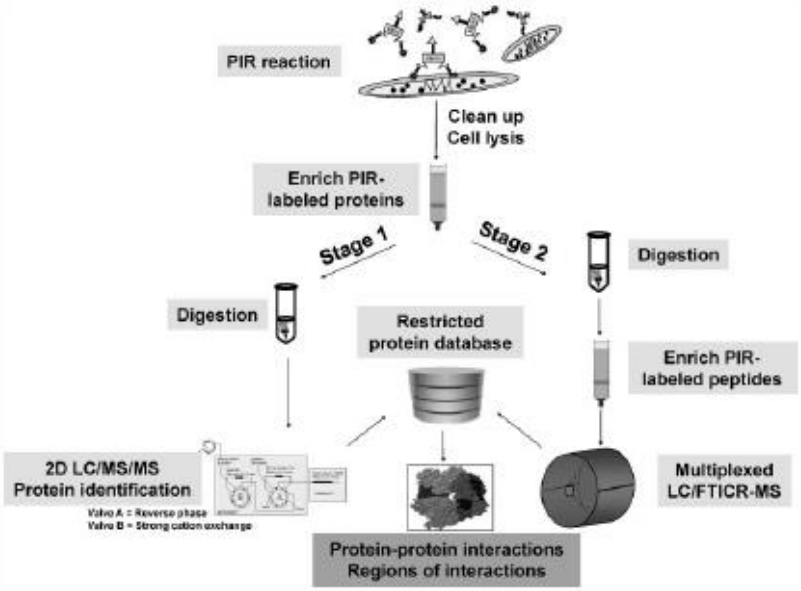
Novel cross-linker design

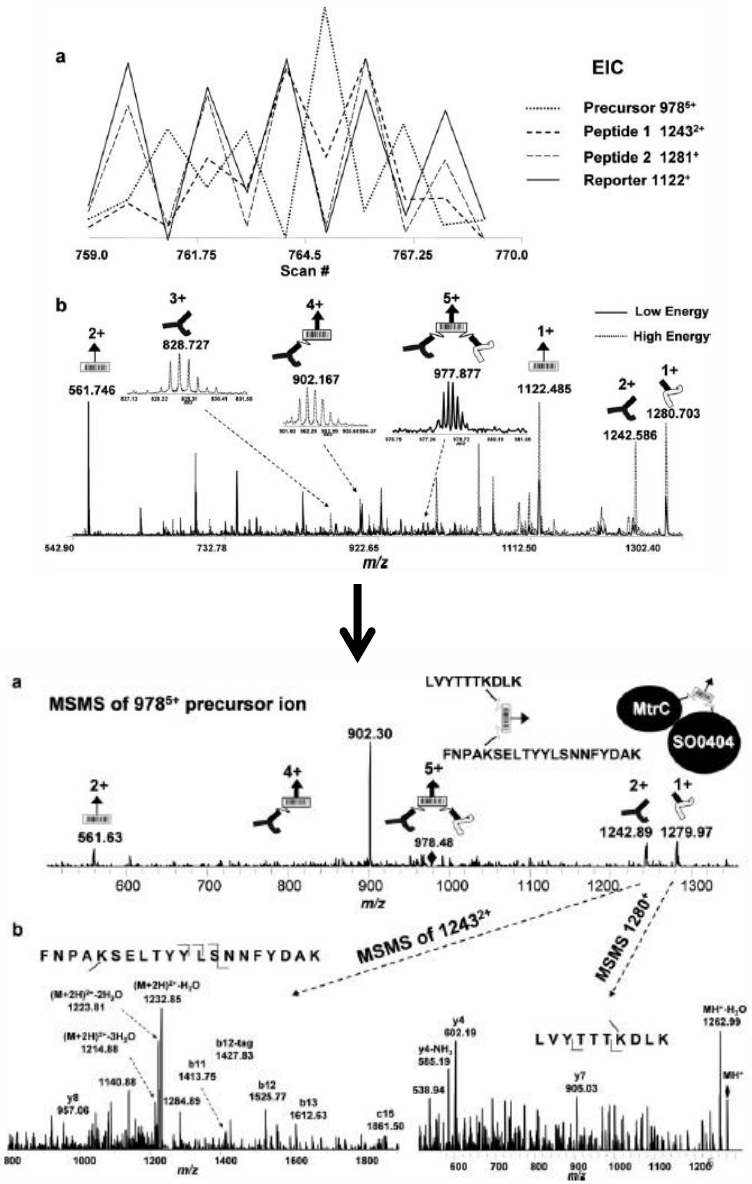
- Affinity tag for purification
- Gas phase cleavable linker for easy data analysis



Tang X. et al. Anal. Chem. 2005

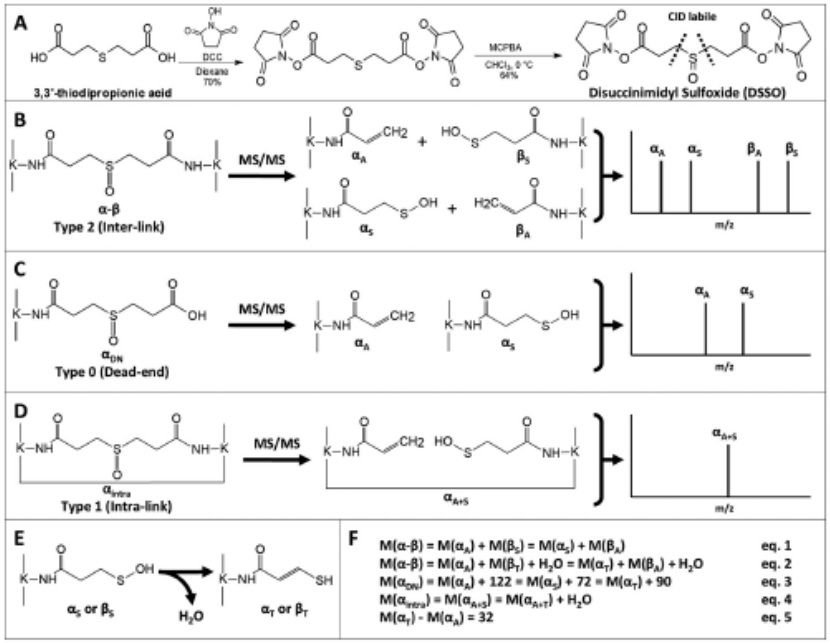
Smart design enables in vivo cross-linking





New generation of cleavable cross-linkers

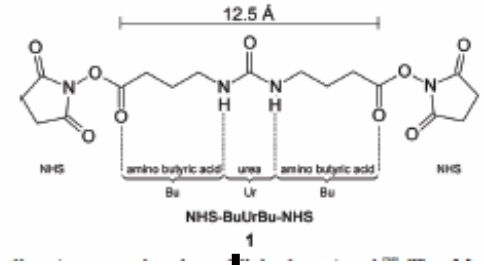
DSSO



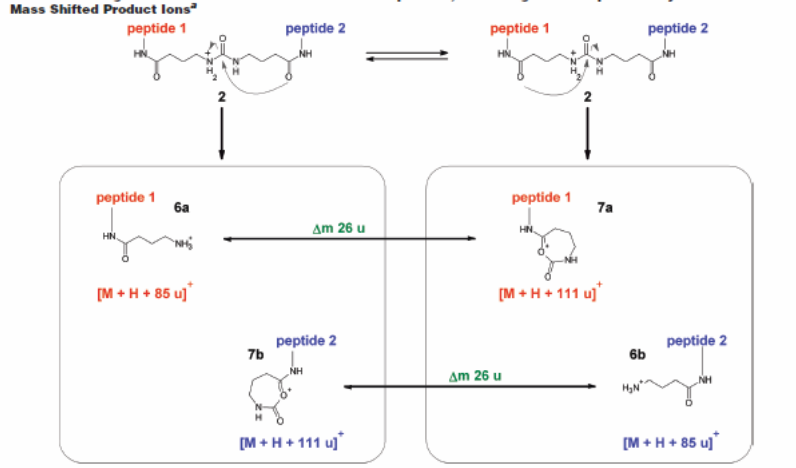
Kao MQ. et al. Mol. Cell Prot. 2010

DSSBu

Scheme 1. Structure of the Symmetric NHS-BuUrBu-NHS Compound (1) for Chemical Cross-Linking

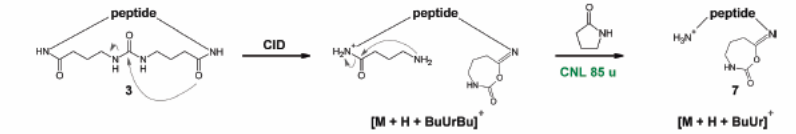


Scheme 3. Fragmentation Mechanism of Protonated 2 upon CID, Delivering Two Complementary Doublets of 26 u Mass Shifted Product Ions[a]



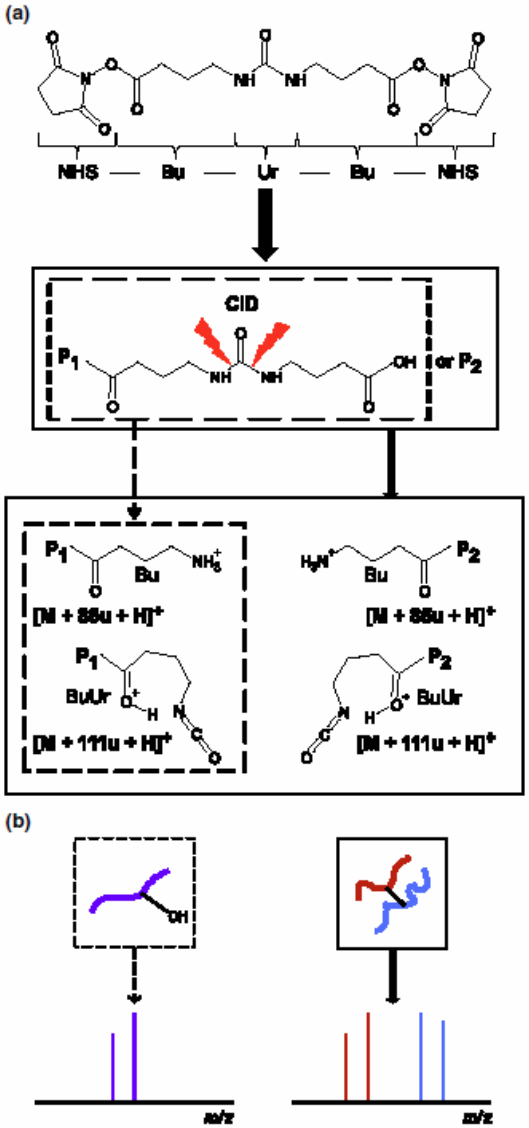
[a] Product ions of peptide 1 are **6a** and **7a**, and product ions of peptide 2 are **6b** and **7b**.

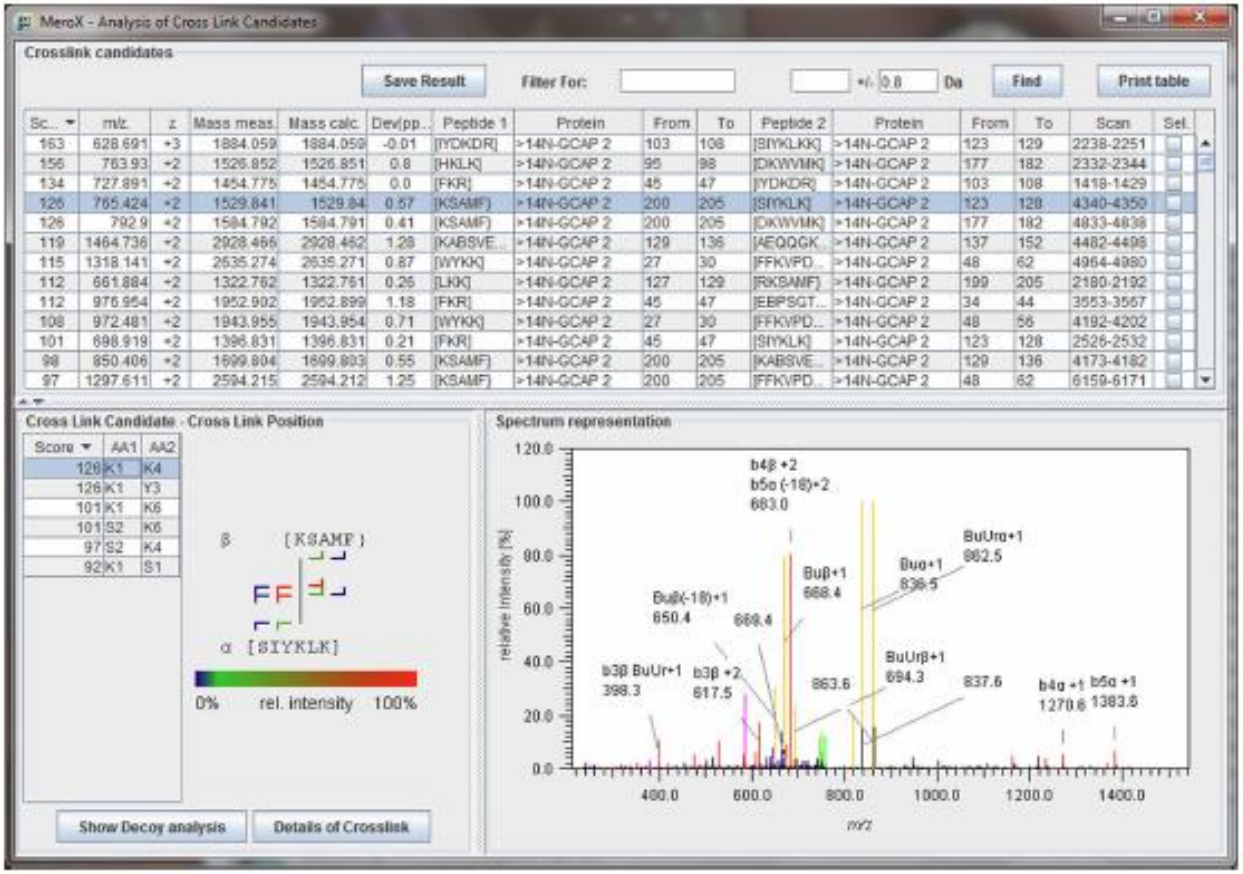
Scheme 4. Fragmentation Mechanism of a Protonated Type 1 Modified Peptide (3) upon CID, Delivering a Product Ion That Is Modified with BuUr [M + H + BuUr]+ (7) by a CNL of Pyrolidinone (85 u)



Muller A. et al. Anal. Chem. 2011

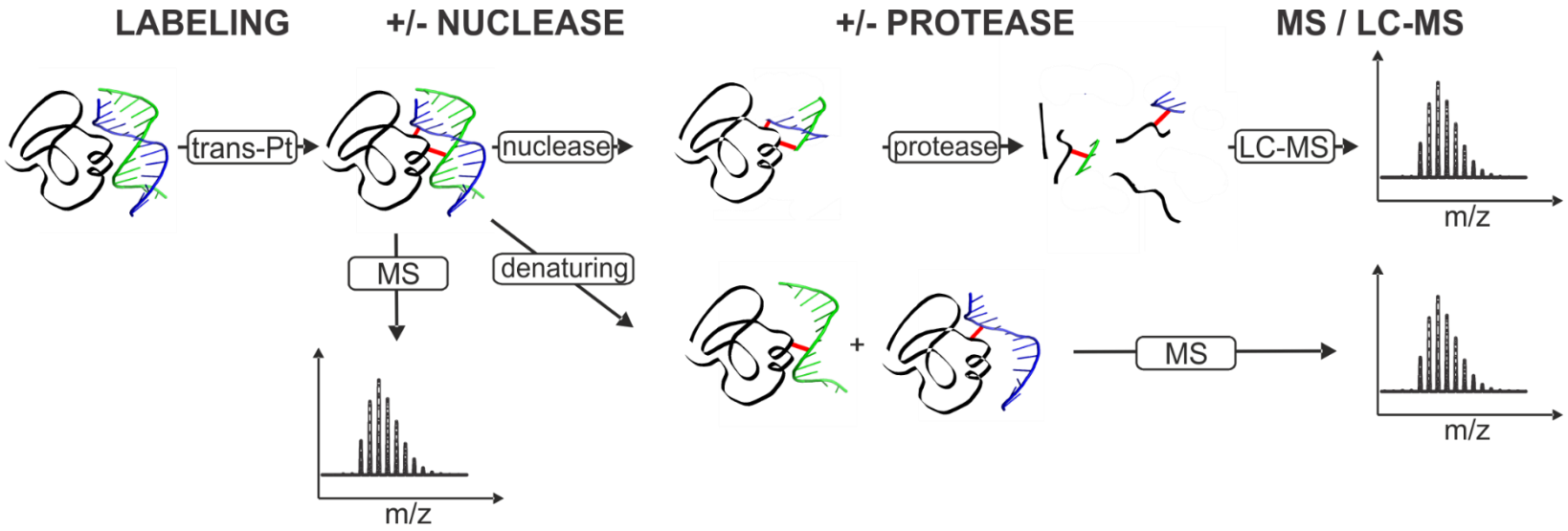
Data analysis: MEROX

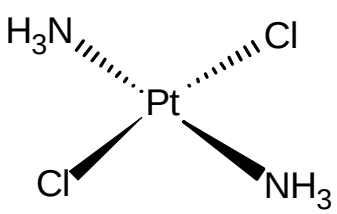




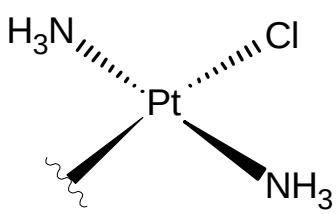
Gotze M. et al. JASMS 2015

Protein-nucleic acid cross-linking

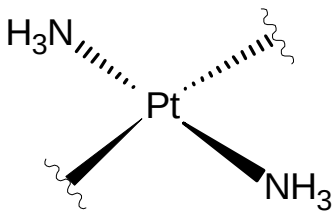




Trans-platinum

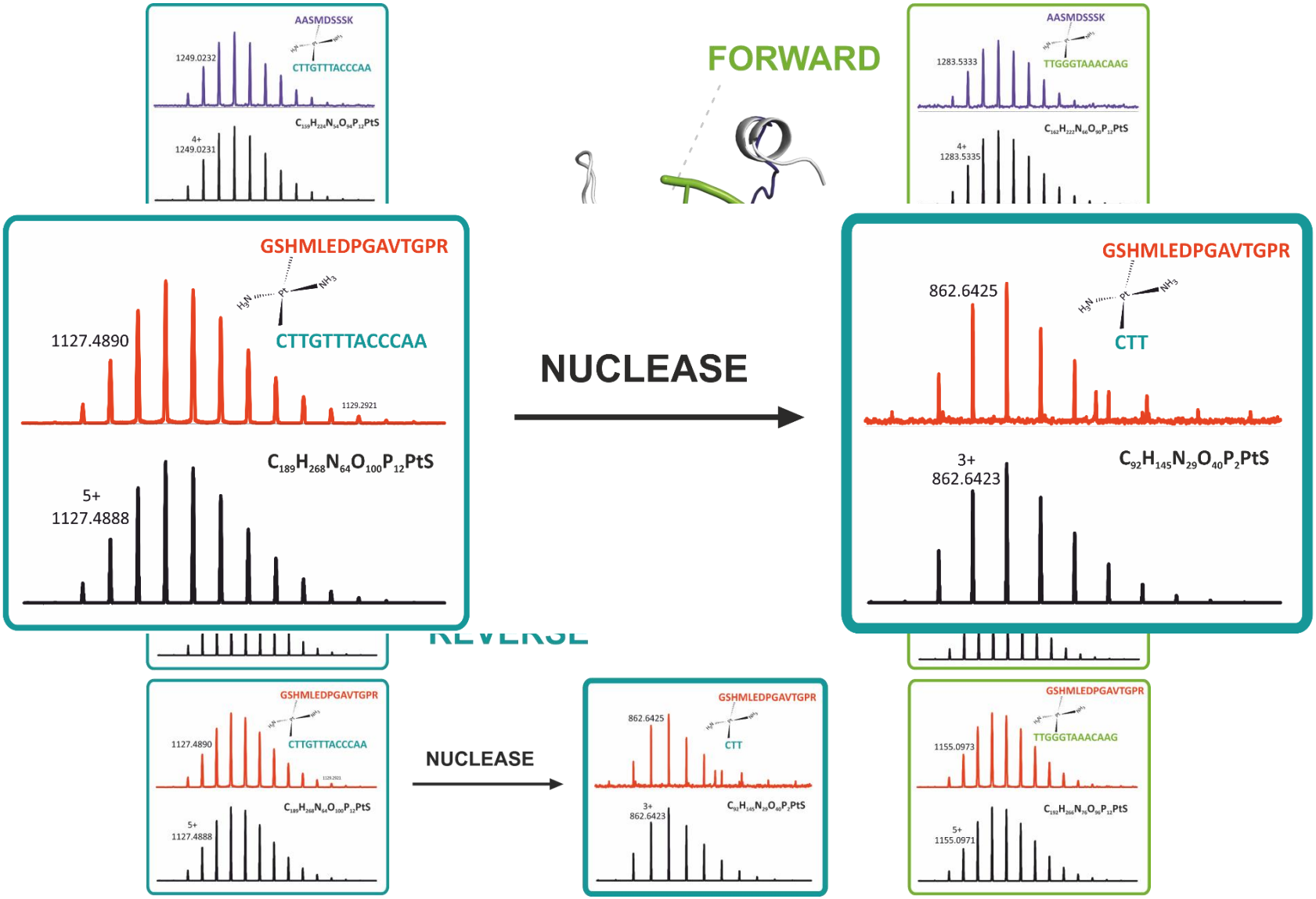


Trans-platinum (monofunctional)



Trans-platinum (bifunctional)

Protein – DNA cross-links



Footprinting

Assay examining higher structures of biomacromolecules by monitoring solvent accessibility of their regions

- **Single molecule conformation / Conformational changes**
- **Ligand binding / biomacromolecular interactions**

Different techniques

- **Enzymatic / chemical cleavage**
- **Stable covalent labeling**

Stable covalent labeling

- **Chemical footprinting**
- **Radical footprinting ($^{\bullet}OH$, $^{\bullet}I$, $^{\bullet}CF3$)**