

Identifying Post-translational Modifications

Nuno Bandeira

Executive Director

NIH/NCRR Center for Computational Mass Spectrometry

Skaggs School of Pharmacy and Pharmaceutical Sciences

Dept. Computer Science and Engineering

University of California, San Diego



Center for
Computational
Mass
Spectrometry

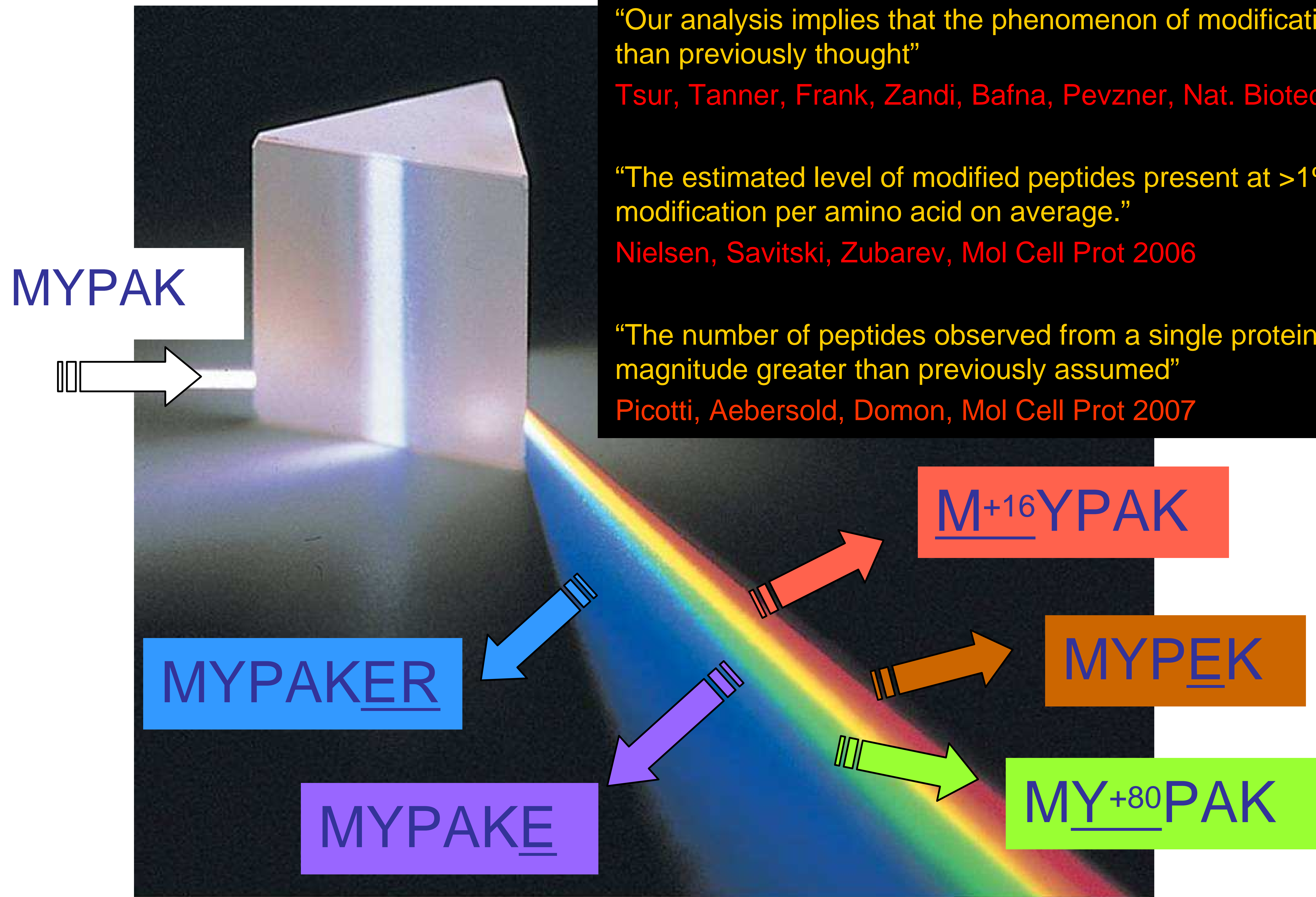


National Institutes of Health
The Nation's Medical Research Agency



**National Center for
Research Resources**

The dynamic proteome



“Our analysis implies that the phenomenon of modification is much more widespread than previously thought”

Tsur, Tanner, Frank, Zandi, Bafna, Pevzner, Nat. Biotech 2005

“The estimated level of modified peptides present at >1% level is approaching one modification per amino acid on average.”

Nielsen, Savitski, Zubarev, Mol Cell Prot 2006

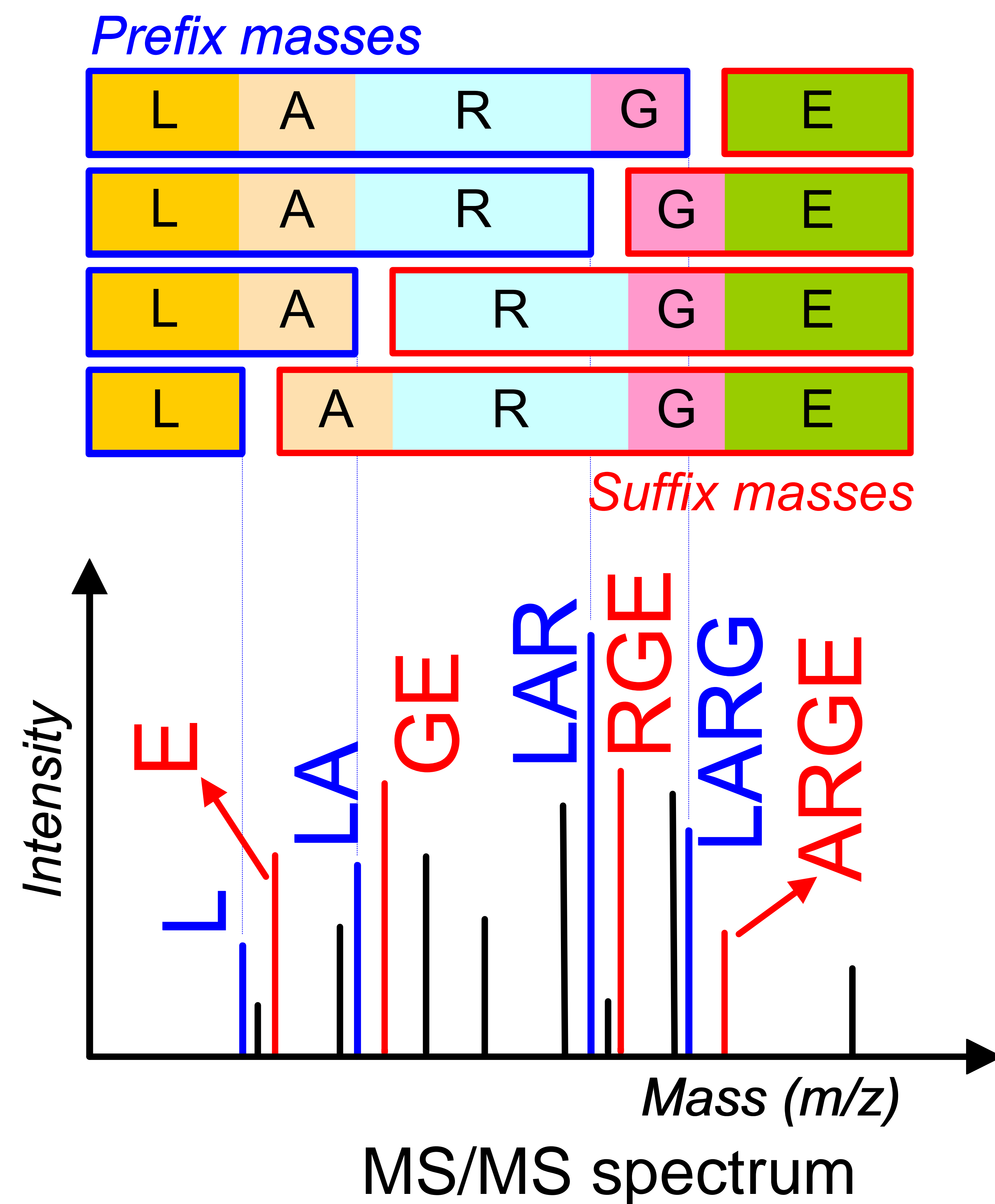
“The number of peptides observed from a single protein is at least one order of magnitude greater than previously assumed”

Picotti, Aebersold, Domon, Mol Cell Prot 2007

© 2007 Thomson Higher Education

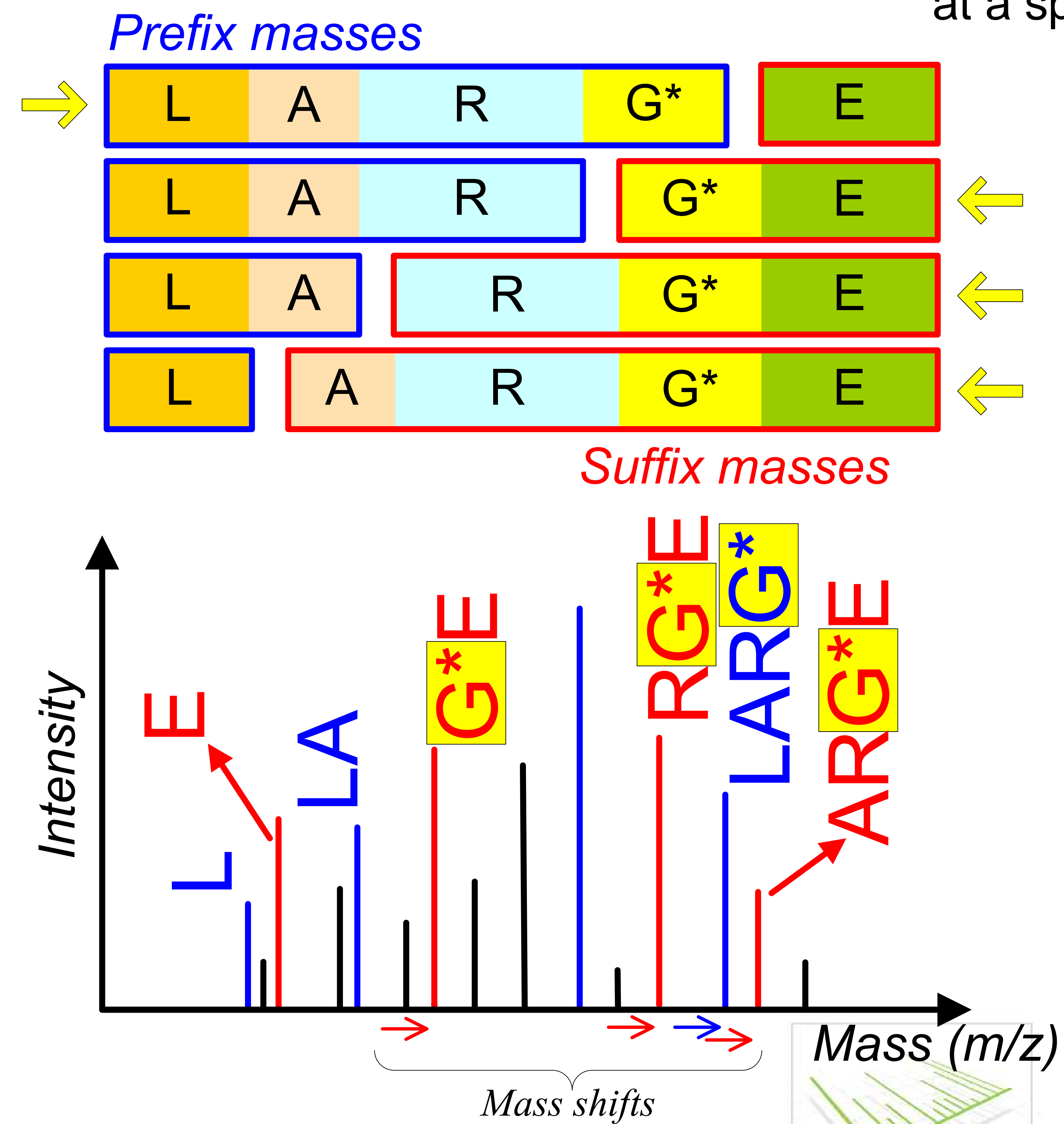
Tandem Mass Spectrometry (MS/MS)

Peptide LARGE

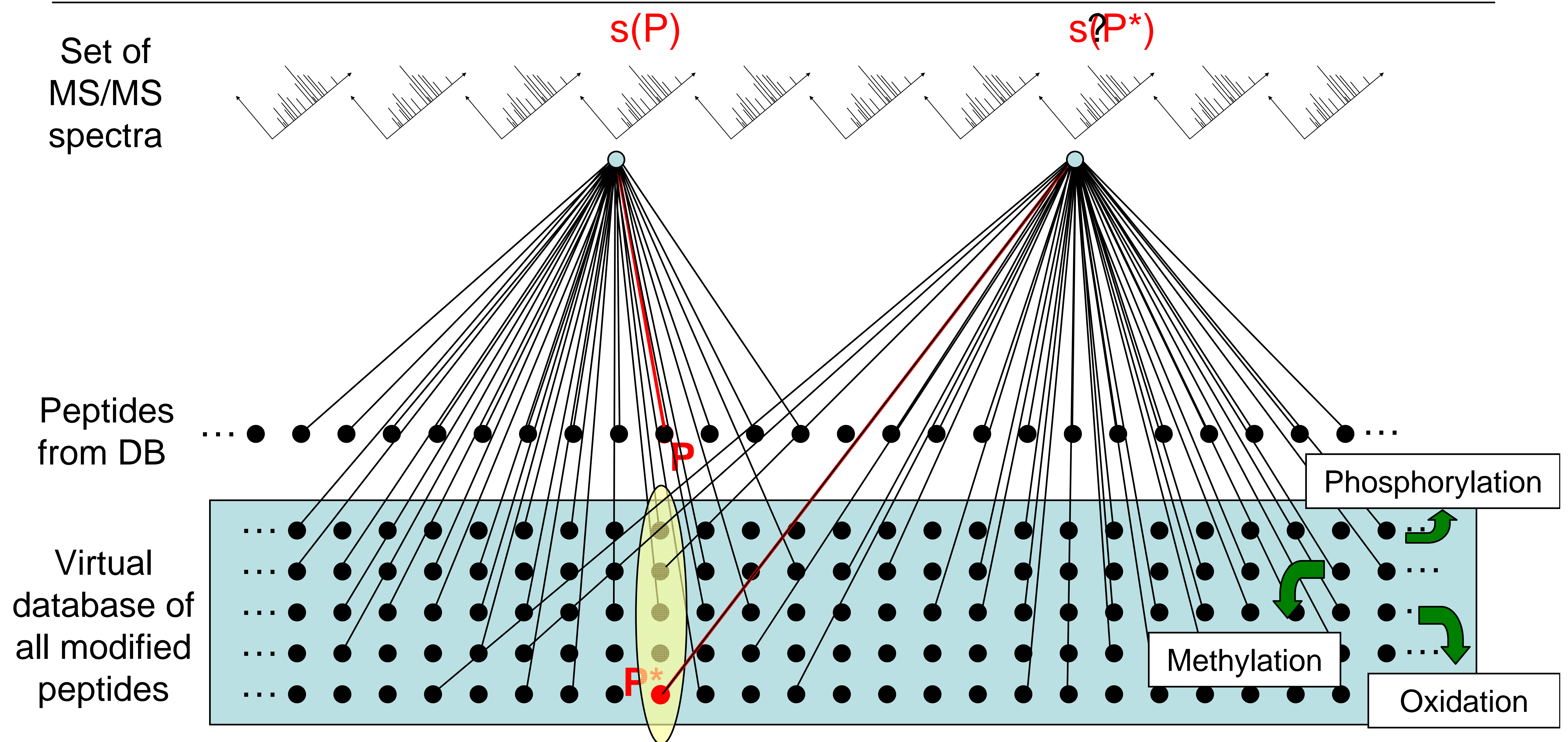


Modified peptide
LARG*E

Modification: any event that changes the mass at a specific site.



MS/MS spectrum identification



PTM database search:

- ⇒ Virtual database size restricts the allowed number of modifications
- ⇒ Becomes computationally heavy (i.e., slow)
- ⇒ Stricter thresholds for same False Discovery Rate

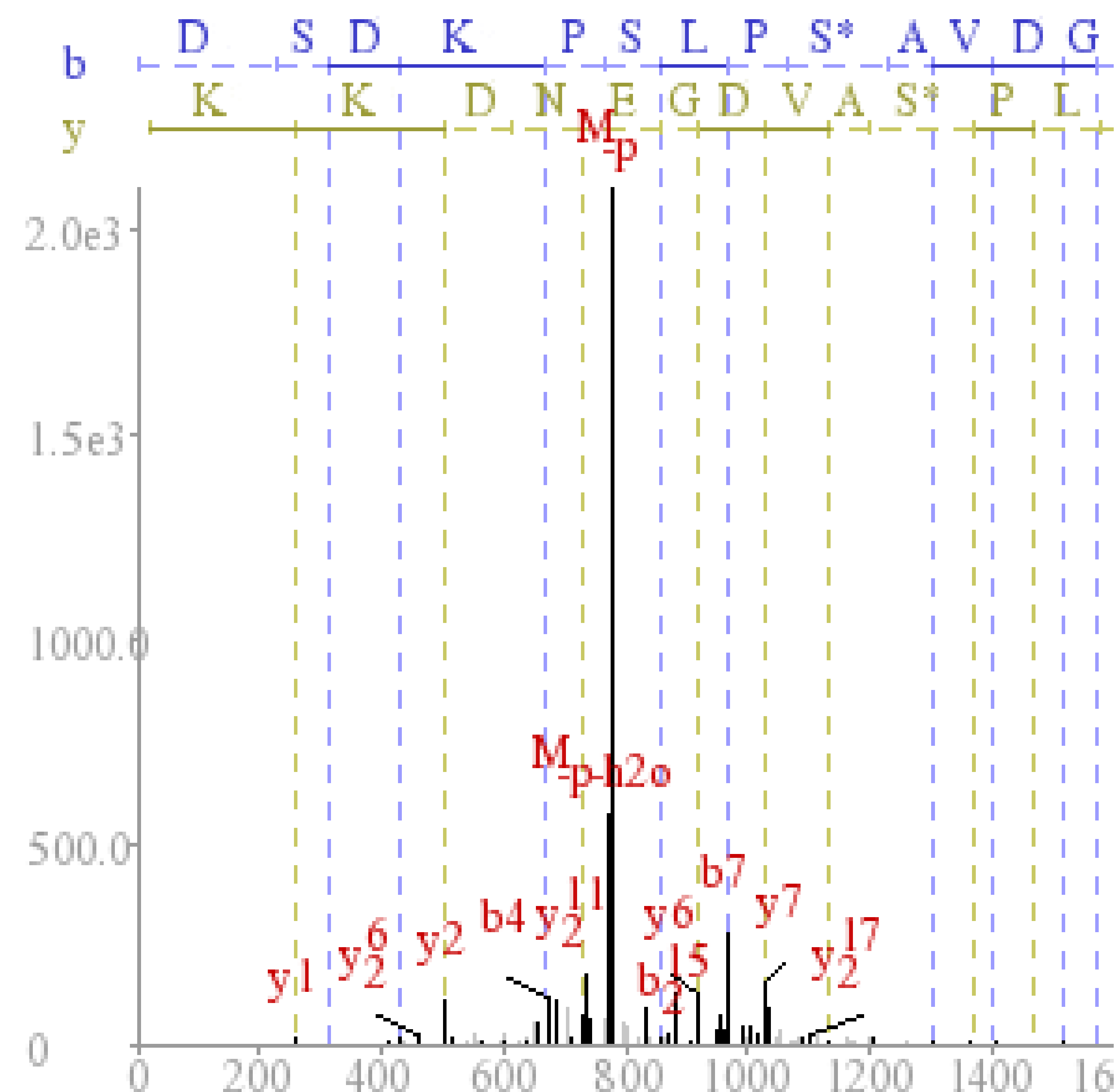
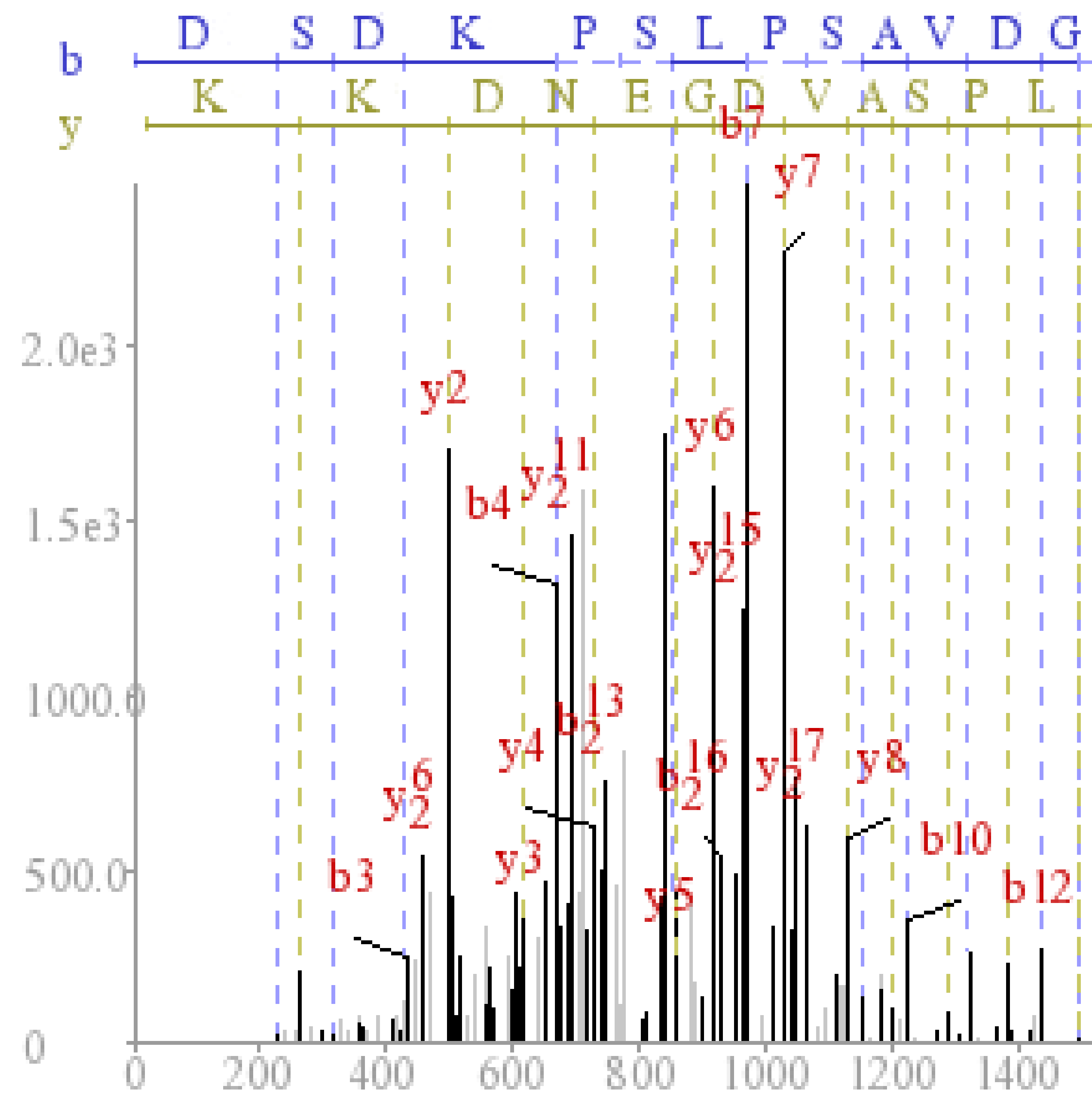
Computational strategies

- **InsPecT: tag-based search**
 - Derives amino acid sequence tags from each modified spectrum and only considers DB peptides containing one of the reconstructed tags
 - Pro: filtered virtual database reduces FDR; much faster than standard approaches
 - Con: misses identifications if spectrum has no correct sequence tag (typically 95%+ sensitivity for top 50 tags)
- **Alternative approach: two-pass search**
 - First identify proteins using spectra from unmodified peptides then search for modifications only on proteins from the first pass
 - Pro: speedup inversely proportional to complexity of the sample
 - Con: misses modified proteins with no unmodified peptides, difficulties estimating FDRs (small Decoy databases, should not re-search spectra identified in first pass)



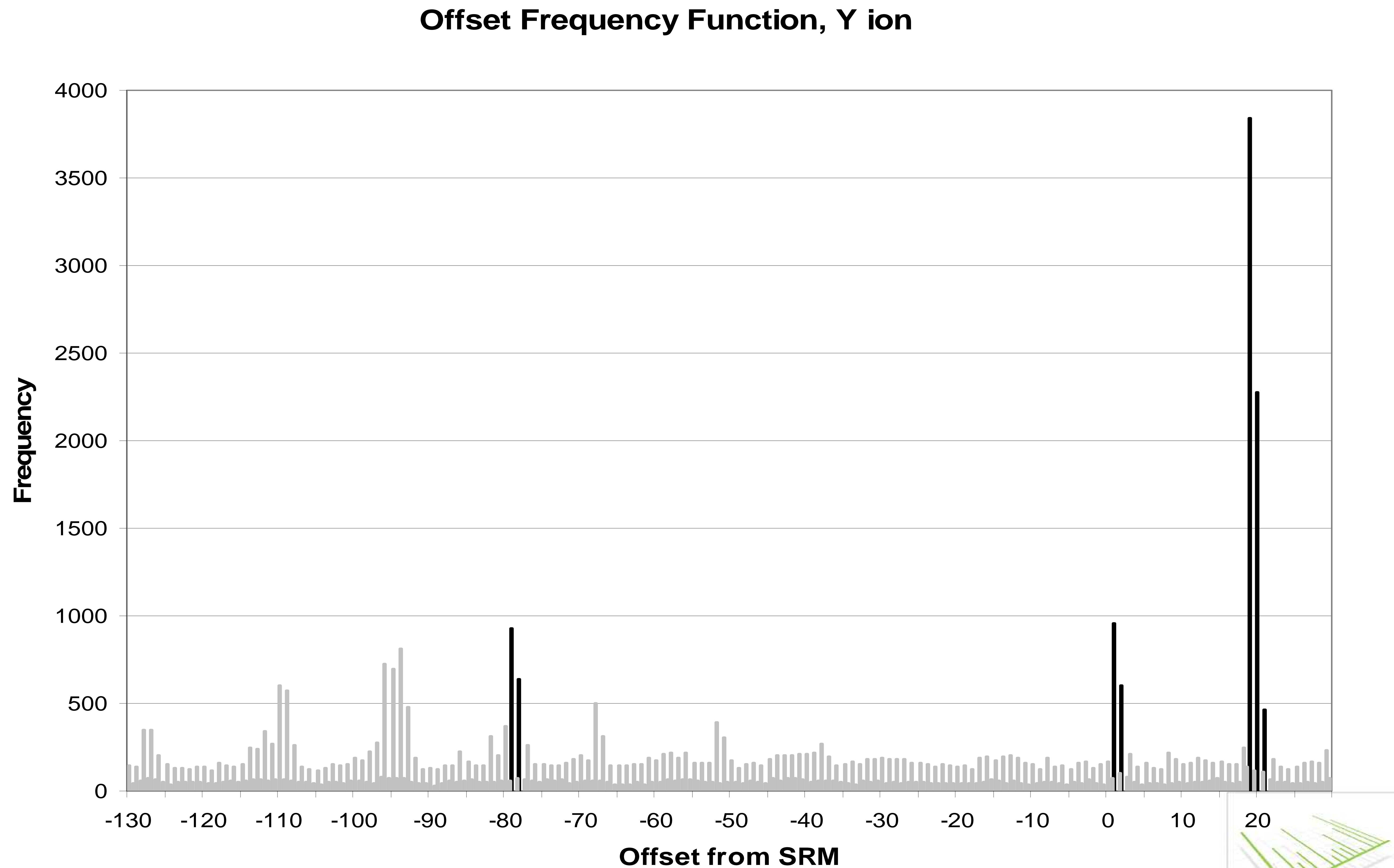
PTMs may change fragmentation

Phosphorylation: weak signal in b and y ions due to phosphate loss



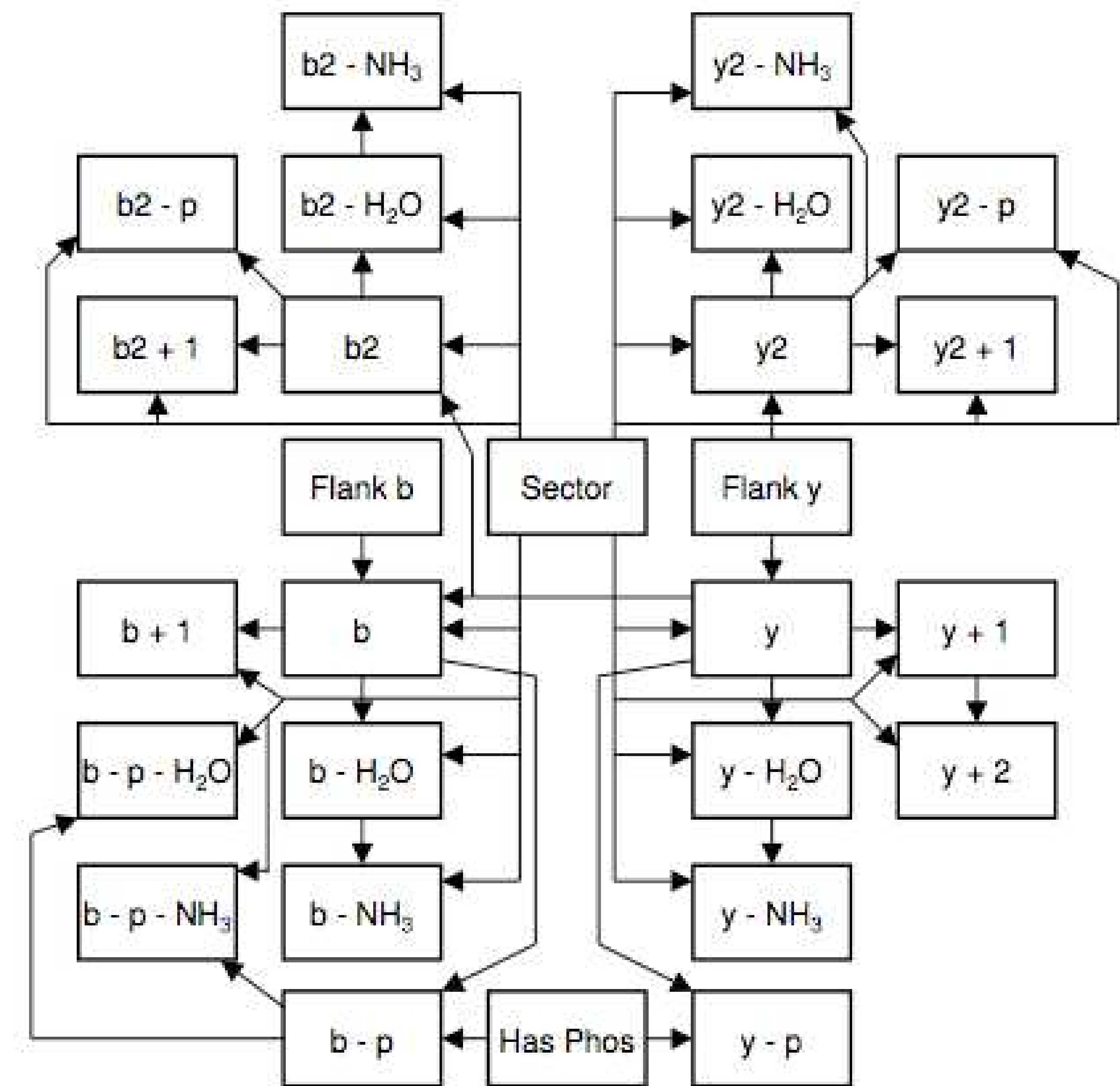
Modification Changes Fragmentation

- New ion observed, fragment neutral loss



InsPecT Scoring Paradigm

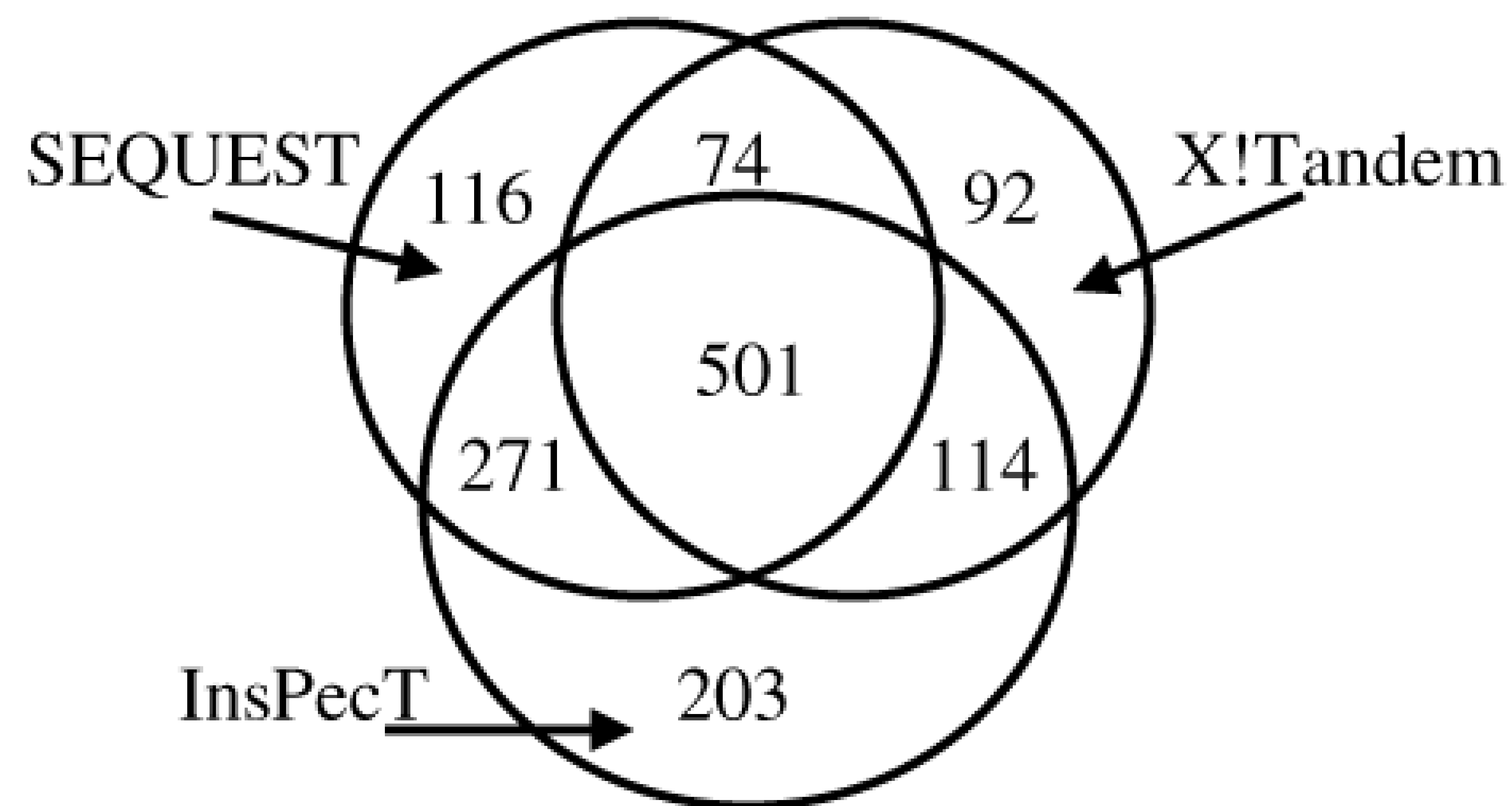
- Ions generated by fragmentation are not independent
- Peak intensities taken into account
- Model the probability of observing in CID with a Bayesian network.



$$P_{CID}(\vec{I} = [I_0, I_1, I_2, \dots] | P_j, S) \approx \prod_i P_{CID}(I_i | P_j, I_{\pi(i)}, S)$$

InsPecT results

- Benchmark with SEQUEST and X!Tandem
 - 6410 LTQ MS/MS, IMAC, *S. cerevisiae*
 - Up to 2 phosphorylations (+80 on S,T,Y) per peptide
 - 1% FDR



Run Time

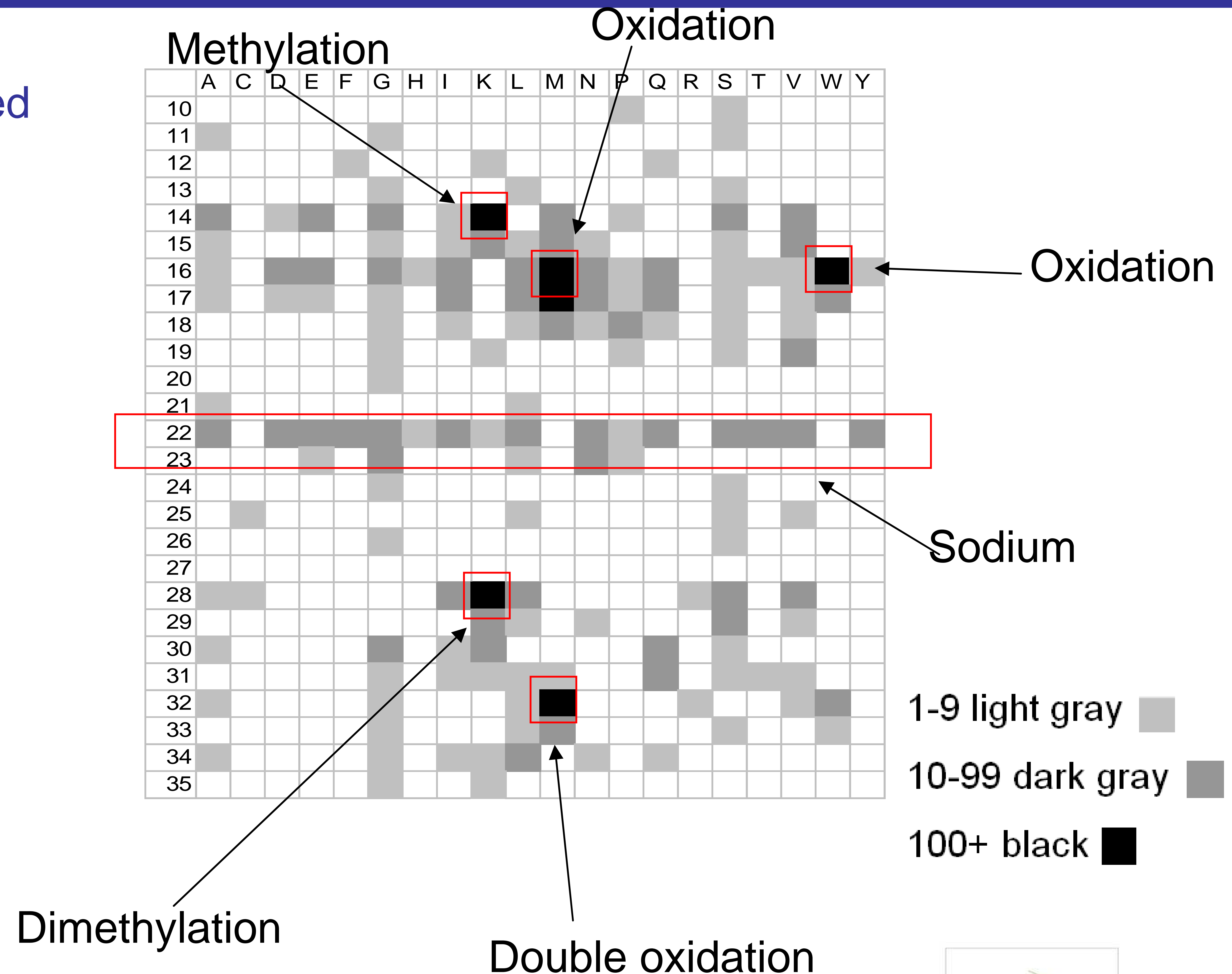
Inspect: 30 min

X!Tandem: 6 hours

SEQUEST: 36 hours

PTM Frequency Matrix: strength in numbers

- Over-represented mass-shifts represent the ubiquitous modifications.
- Can we reliably detect the lower abundance modifications?



Center for
Computational
Mass
Spectrometry

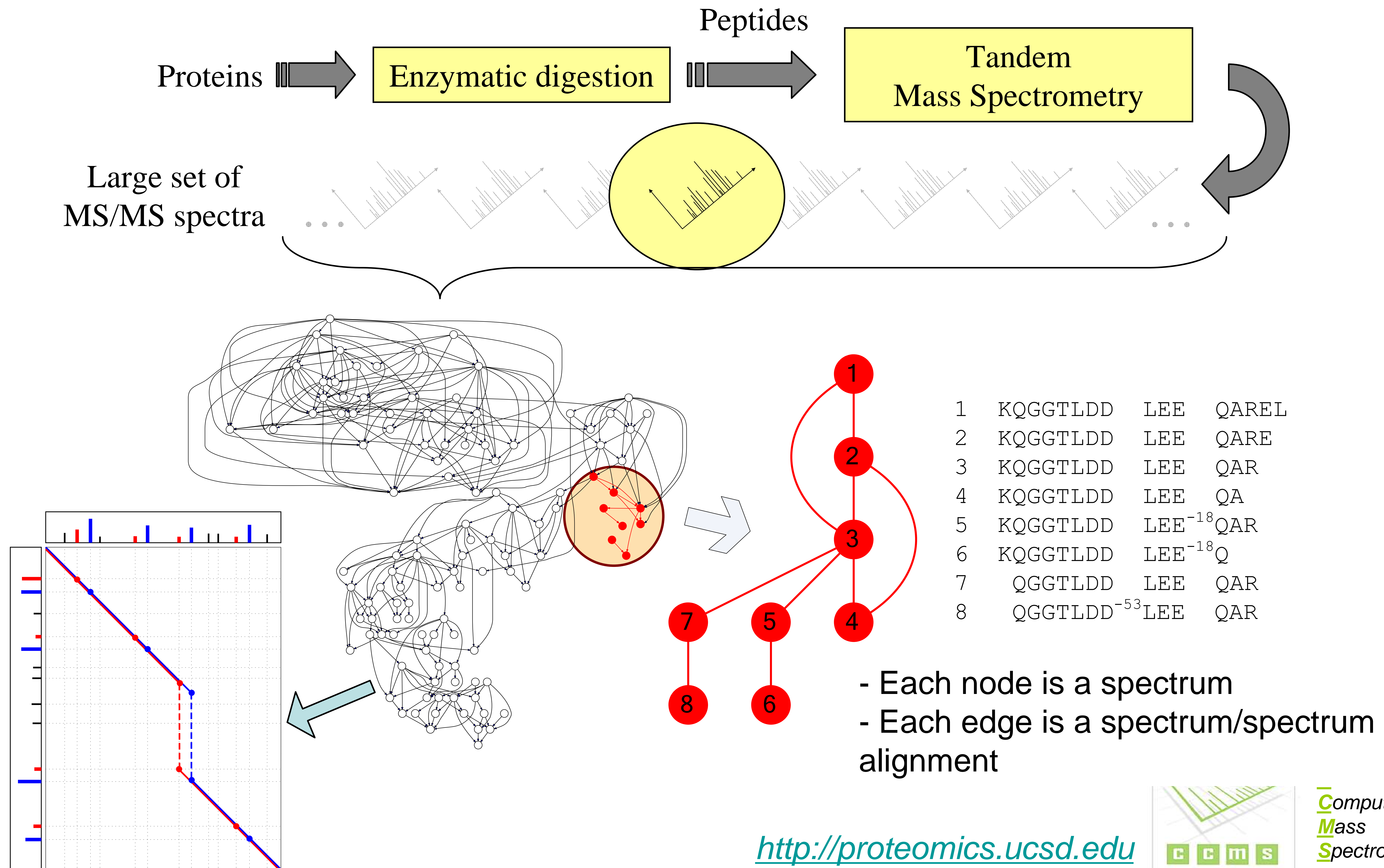
PTMFinder

14 on K (methylation)		
K*LSSPATL	9	0
K*LSSPATLN	1	0
K*LSSPATLNS	36	0
K*LSSPATLNSR	8	0
IMLIK*LSSPATLNSR	1	0
TLDNDIM+16LIK*	4	11
IITHPNFNGNTLDNDIMLIK*	4	6
IITHPNFN+1GNTLDNDIMLIK*	2	2
IITHPNFNGNTLDNDIM+16LIK*	4	24

28 on S (mutation to D)		
GPGTS*ILSTWIGGSTR	3	0
FGPGTS*ILSTWIGGSTR	1	0
DIFGPGTS*ILSTWIGGSTR	21	0
DIFGPGTS*ILSTWIGGSTRSISGT	2	0
DIFGPGTS*ILSTWIGGSTRSISGTSMATPHVAGLA	3	0

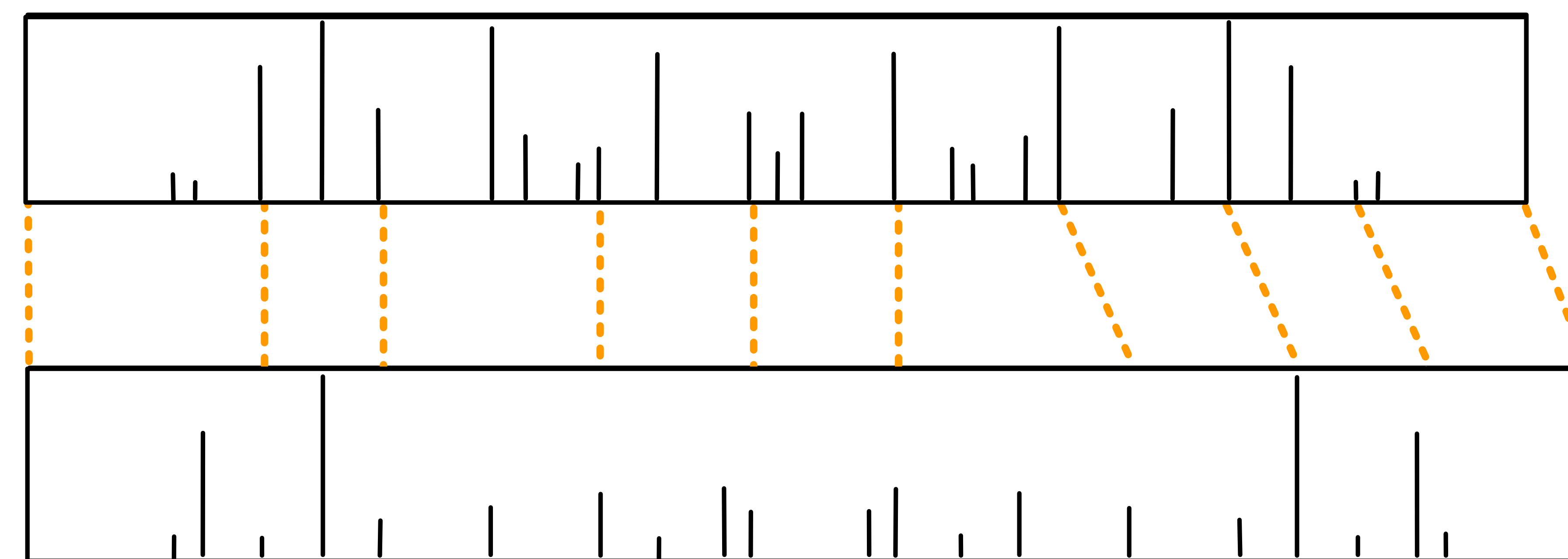
Overlapping peptides help confirm modifications

Spectral Networks



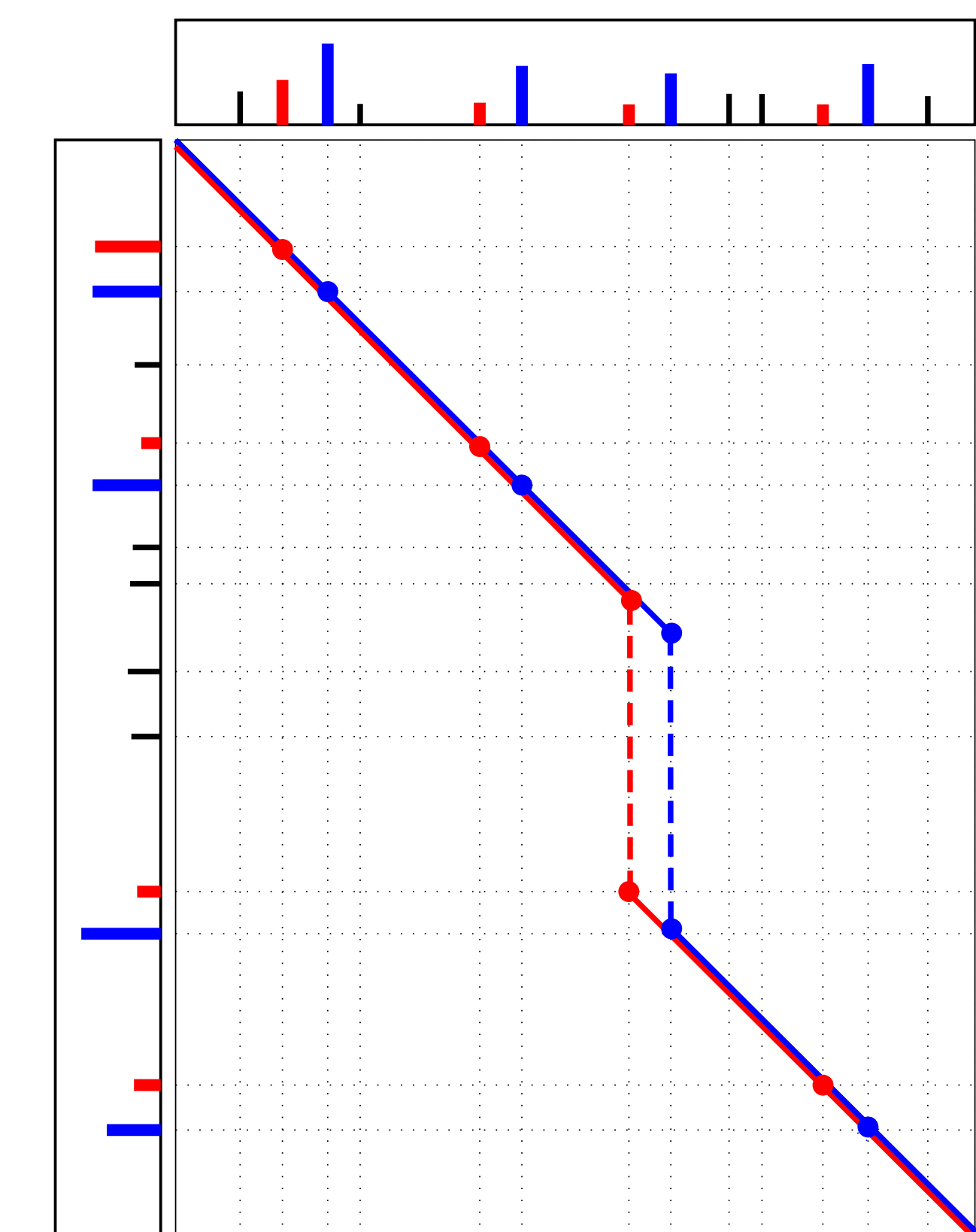
Spectral Alignment

Spectral alignment reveals the mass and location of post-translational modifications.



Modification site

Modification mass



Sample of cataractous lens from a 93-year old patient

- Collaboration with Larry David @ Oregon Health and Science University
- Lens proteins do not turnover and accumulate modifications over time
- Intensively studied in Searle et al.'04, Tsur et al.'05 and Wilmarth et al.'06
- Detected over 70,000 spectral alignments



Modifications on cataractous lens

Location	Modification Mass	Putative annotation
S,T	-18	dehydration
Q	-17	deamidation
W	-2	cross-linking
H	14	methylation
M,W	16	oxidation
S,H	28	double methylation
N-term	42	acetylation
N-term	43	carbamylation
K,non-terminal	43	carbamylation
W	44	carboxylation
R	55	unknown
K	58	carboxymethylation
K	72	carboxyethylation

Table 1: Rediscovered all modifications previously identified by blind database search.

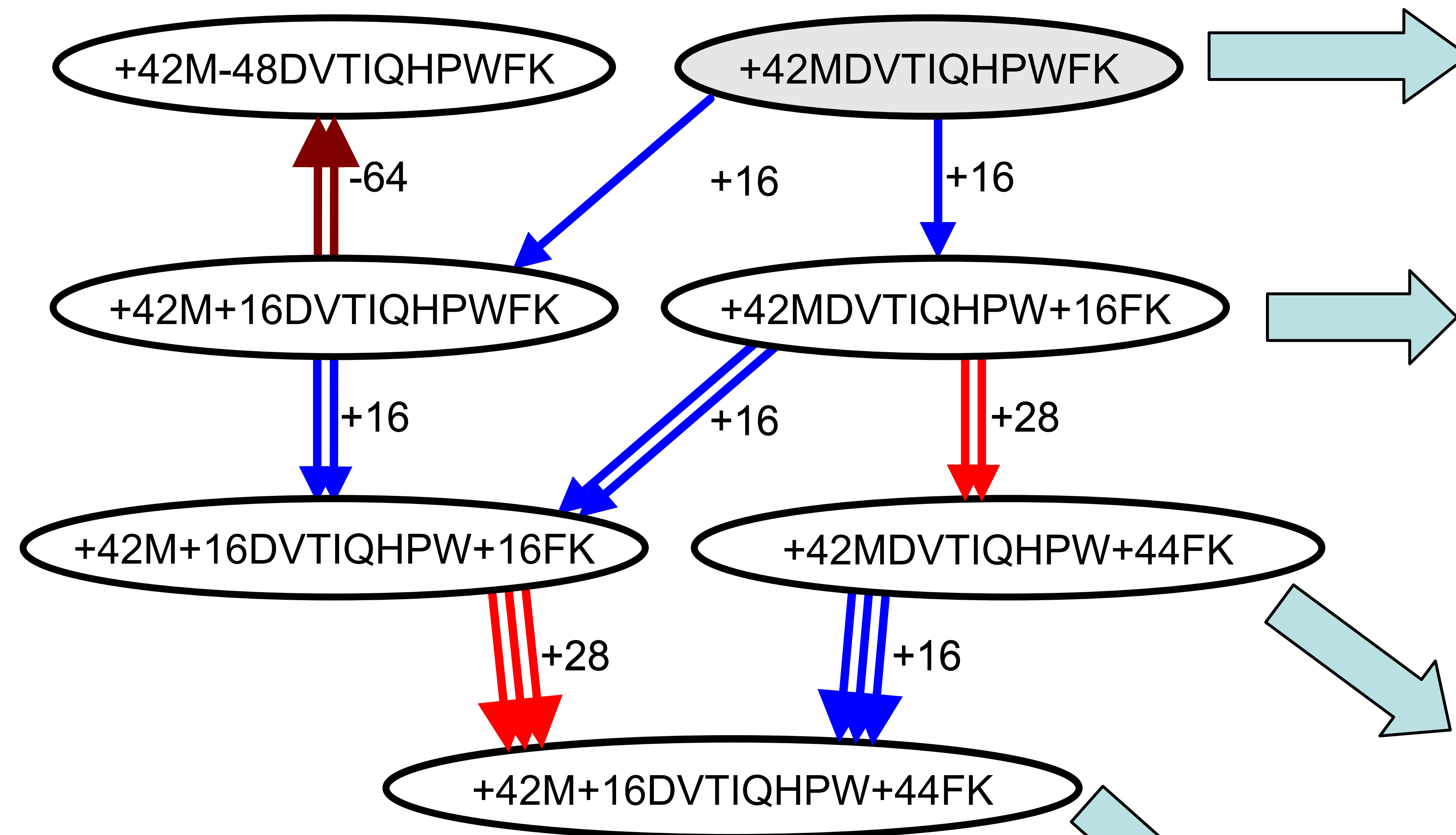
Table 2: Identified 6 new modification events

Location	Modification mass	Type	Putative annotation	Comment
M	-48	Chem. artifact	loss of methane sulfenic acid	<i>reported on same site</i>
W	4	PTM	kynurenine	<i>reported in cataractous lenses</i>
S	30/73	unknown	unknown	
W	32	PTM	formylkynurenine	<i>reported in cataractous lenses</i>
N-term	57	unknown	carboxyamidomethylation	<i>In-vivo N-term modification?</i>
N-term	271	unknown	unknown	



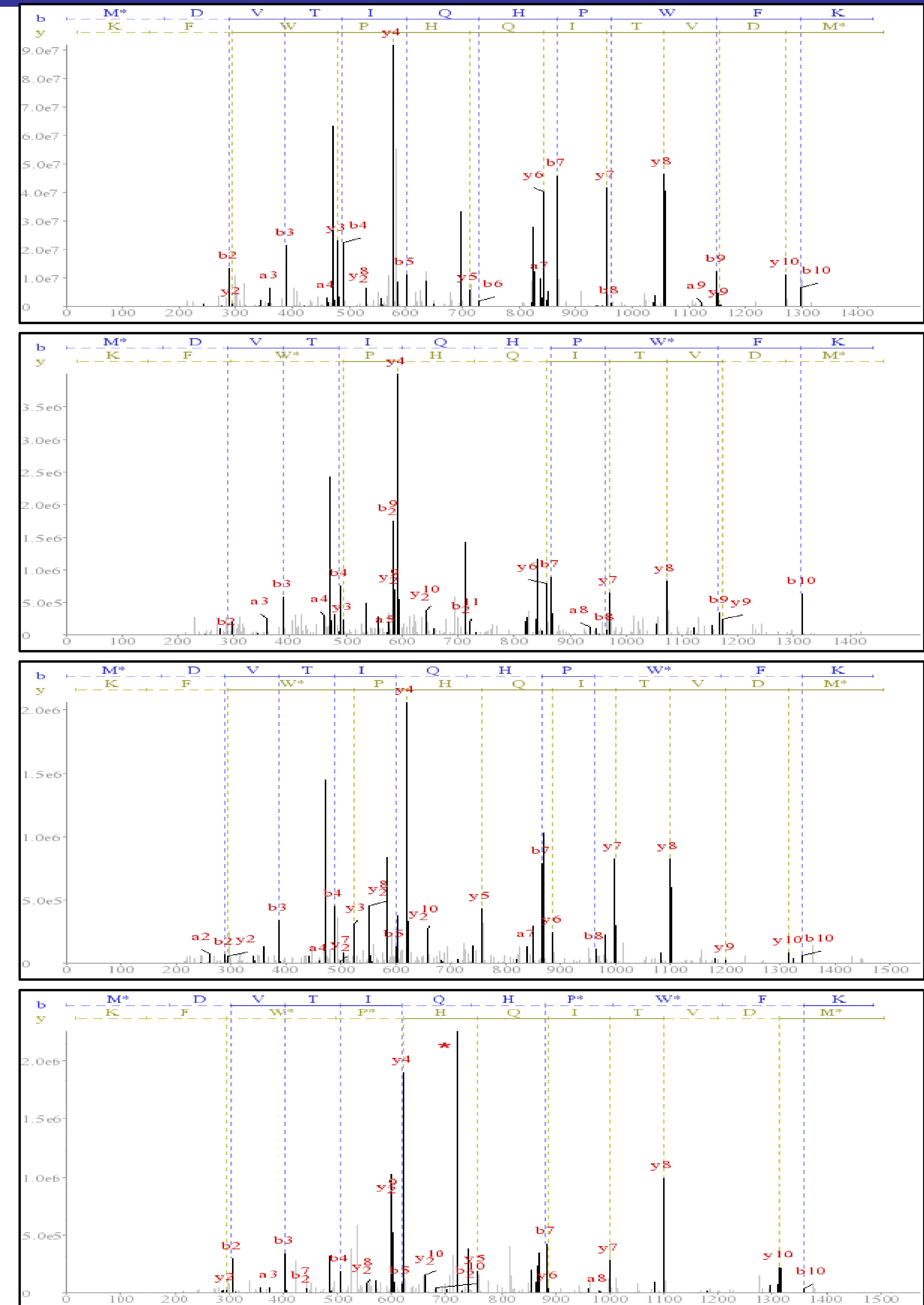
Spectral networks of modified variants

Focus on a single peptide: MDVTIQHPWFK

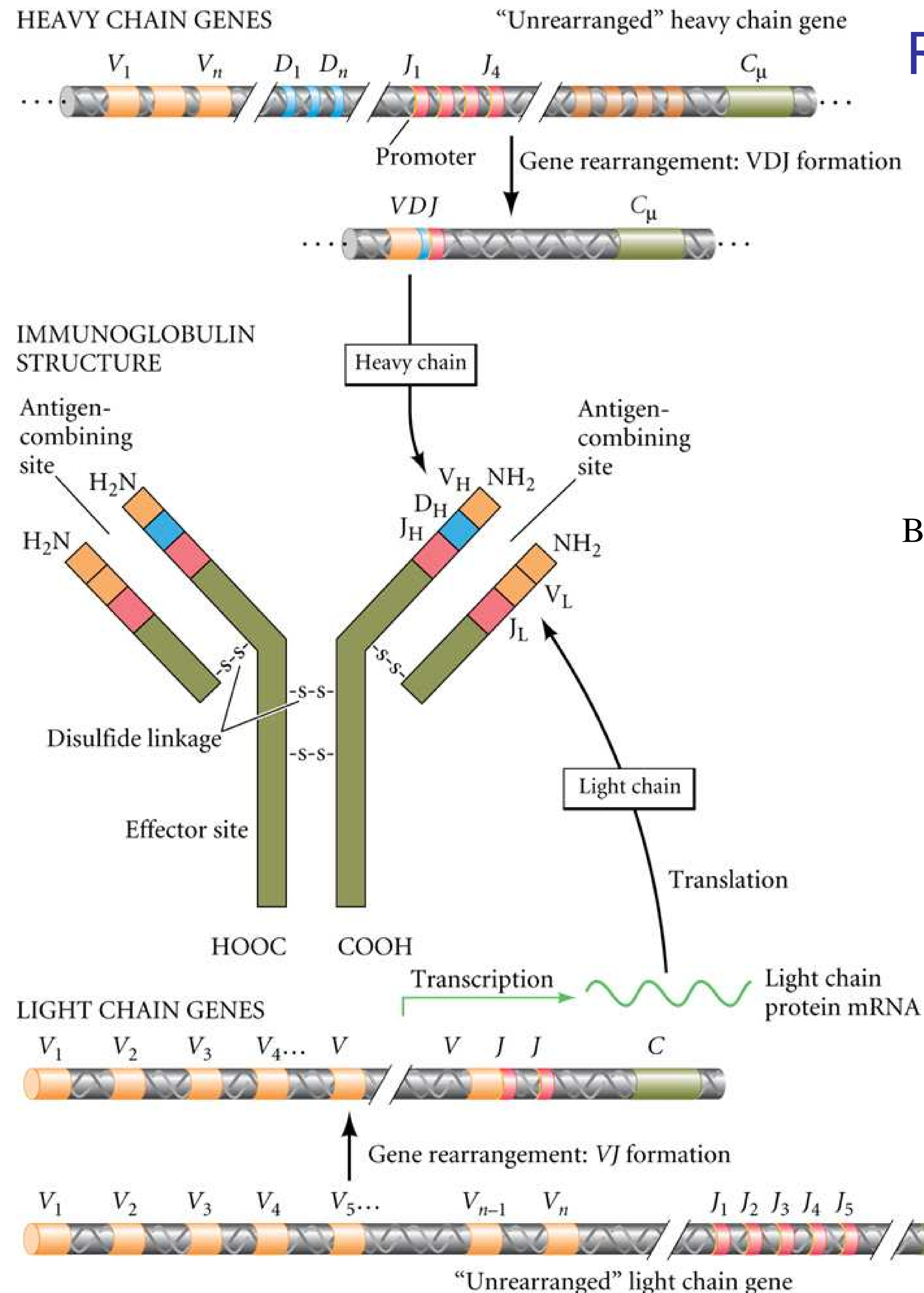


All modifications discovered on cataractous lens were supported by spectral networks.

- PTMs without a database
- Triply-modified peptide
- All unrestricted modifications



Characterizing monoclonal antibodies



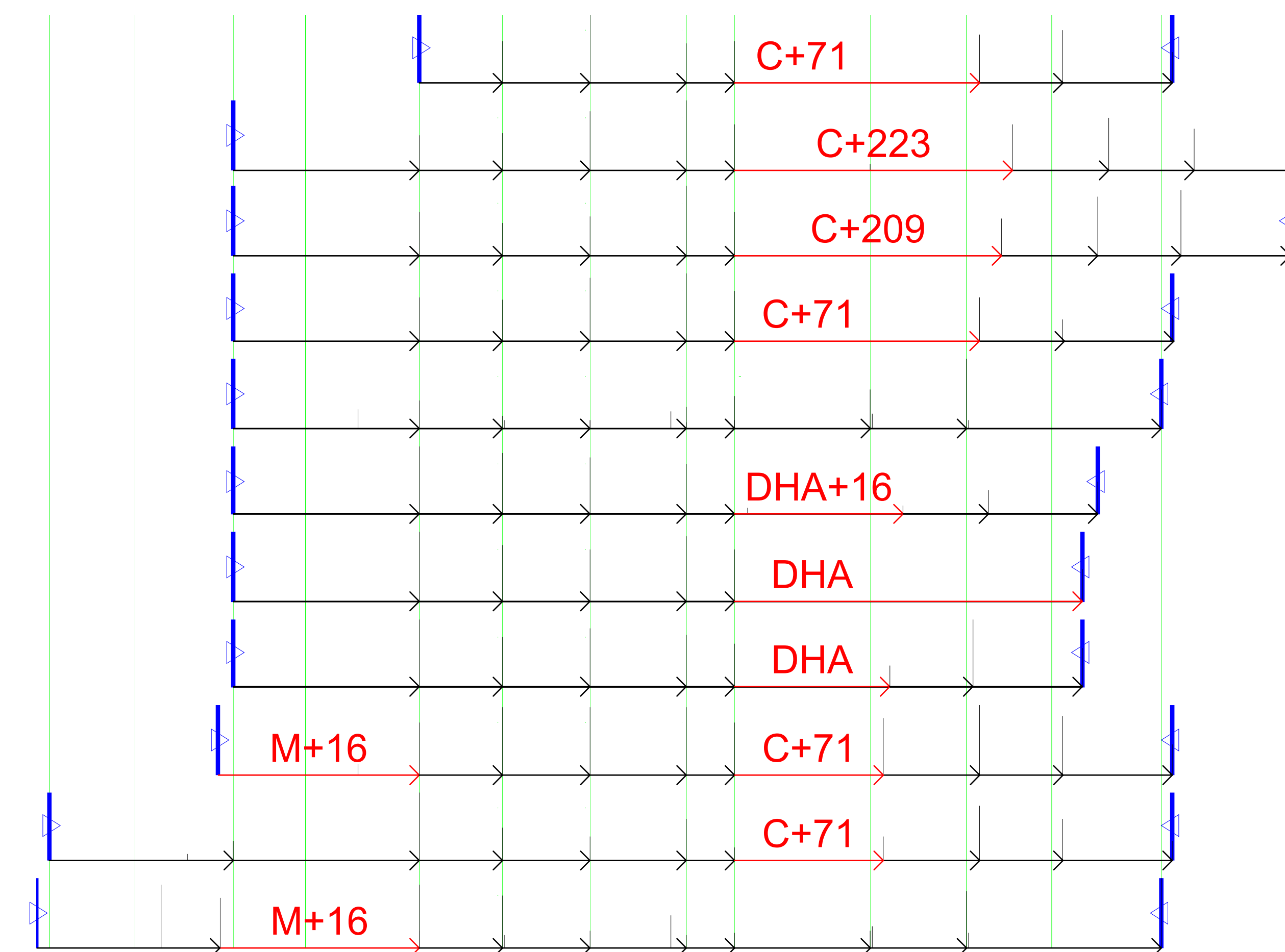
Recombinant sequences create a challenge for de novo sequencing

- Full characterization includes PTMs
- Mutations are a *"special kind of modification"*

(see PSGR, Tue 10:45am)

Bandeira et al, Nat Biotech 2008

Target sequence



— Markers of true amino acid masses

→ - Sequenced mass intervals

— Start/end of assembled spectra

→ - Sequenced mass intervals (with modifications)

<http://proteomics.ucsd.edu>

CCMS

Spectrometry



Genentech
IN BUSINESS FOR LIFE

er for
putational

Conclusions

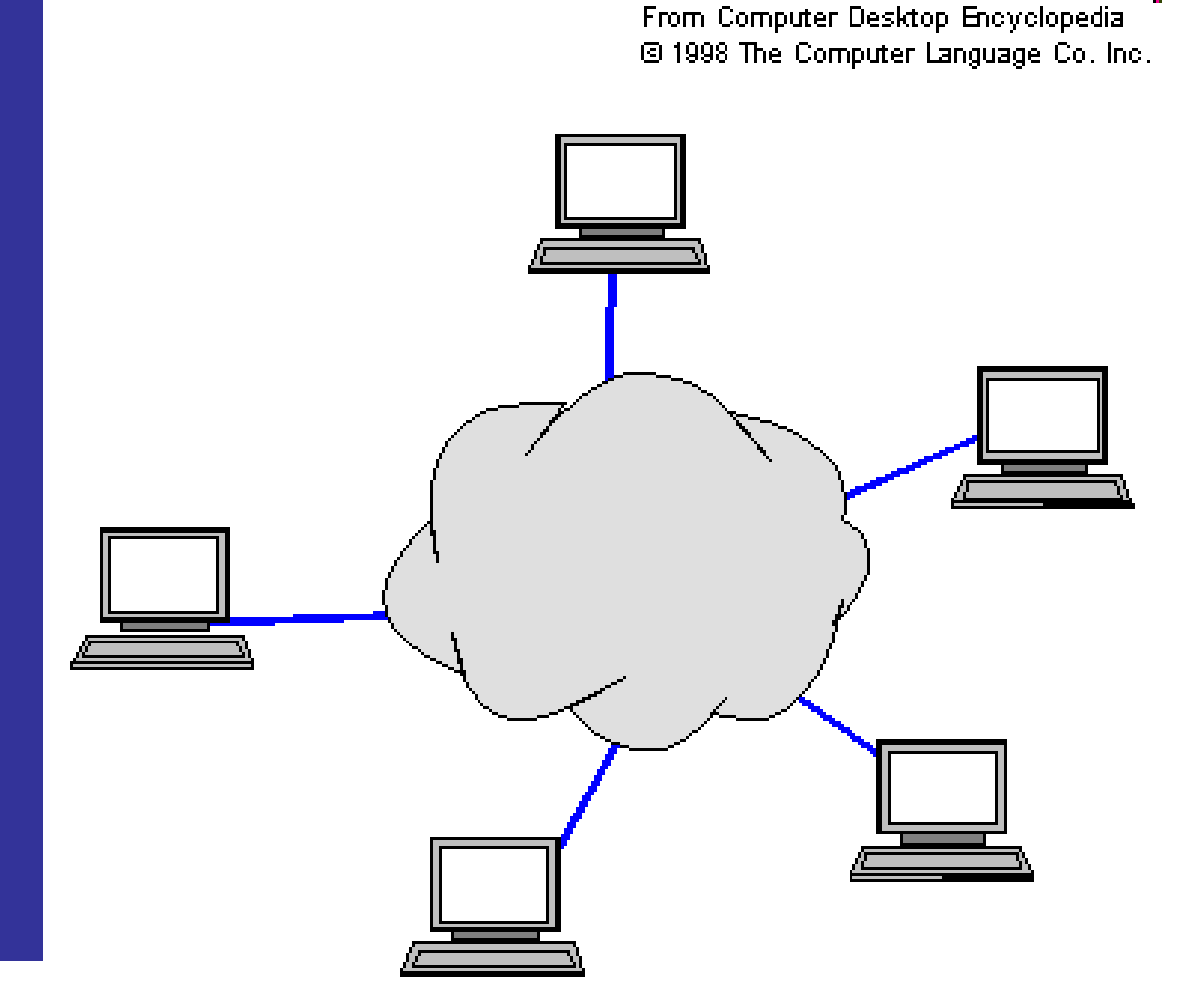
Possible strategies

- Known modifications: filtration, PTM-specific scoring
- Blind search: search any mass offsets, singly-modified peptides
- Spectral Networks: search spectra against spectra, consensus interpretation, highly modified peptides

Main considerations

- False Discovery Rate stringency depends on size of virtual database, strategies may not be Target/Decoy compliant
- PTM site assignments are often ambiguous
 - AScore, Phospho-Loc. Score (PLS)
- Charged PTMs are typically not considered (e.g., phosphopantetheinyl)
- Glycosylation, SUMOylation and Ubiquitination (chains) require special approaches

Center for Computational Mass Spectrometry



***Compute-intensive
discovery proteomics at
the click of a button***



Tool selection

Tool: ☒ InsPect ☐ MS-Alignment ☐ PepNovo

Spectrum file:

OR [SELECT SERVER-SIDE FILES](#)

Description:

Instrument:

Cysteine protecting group:

Protease:

Parent mass tolerance: Da between 0 and 2.5

Ion tolerance: Da between 0 and 1

by Professors
ng algorithms to
developed a
ine has it's own
ad, or are also

3 spectra from
5 (but inaccurate
o significant gains
e had a limited
ie Spectral
ous probabilistic
des, resulting in

Latest Releases

MS-GeneratingFunction

[2008.09.04](#)

PepNovo

[2008.07.08](#)

MS-Clustering

[2008.06.09](#)

Inspect, MS-Alignment

[2008.04.04](#)

MS-Dictionary

[2007.11.30](#)

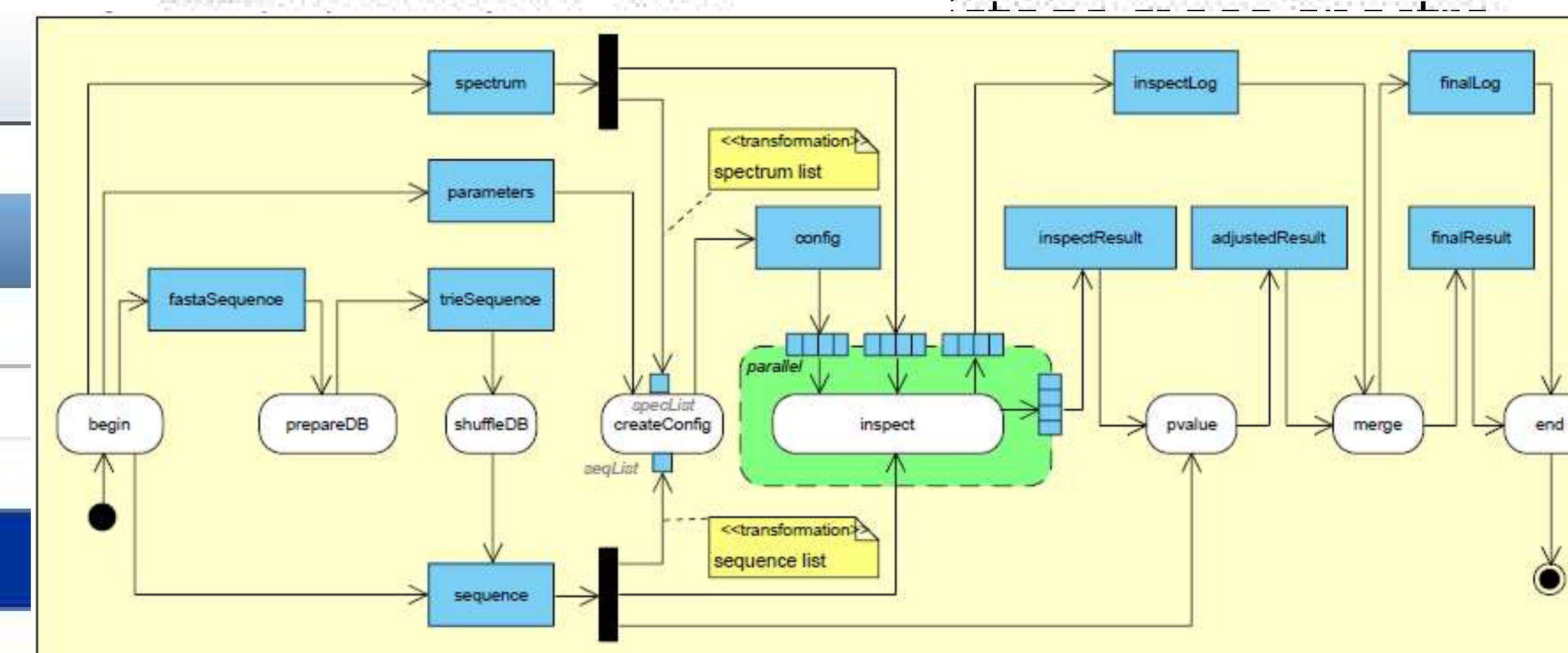
Spectral Networks

[Sept 2007](#)

Allowed Post-Translational Modifications

Maximal number of PTMs permitted in a single peptide :

	Mass (Da)	Residues:
<input type="checkbox"/> Oxidation	15.9994	MMW
<input type="checkbox"/> Carbamidomethylation	44.0222	K



Media Coverage

A powerful tool for PTM
discovery (Jan 2008,
[Journal of Proteome
research, Vol 7, Issue 1](#))



**National Center for
Research Resources**

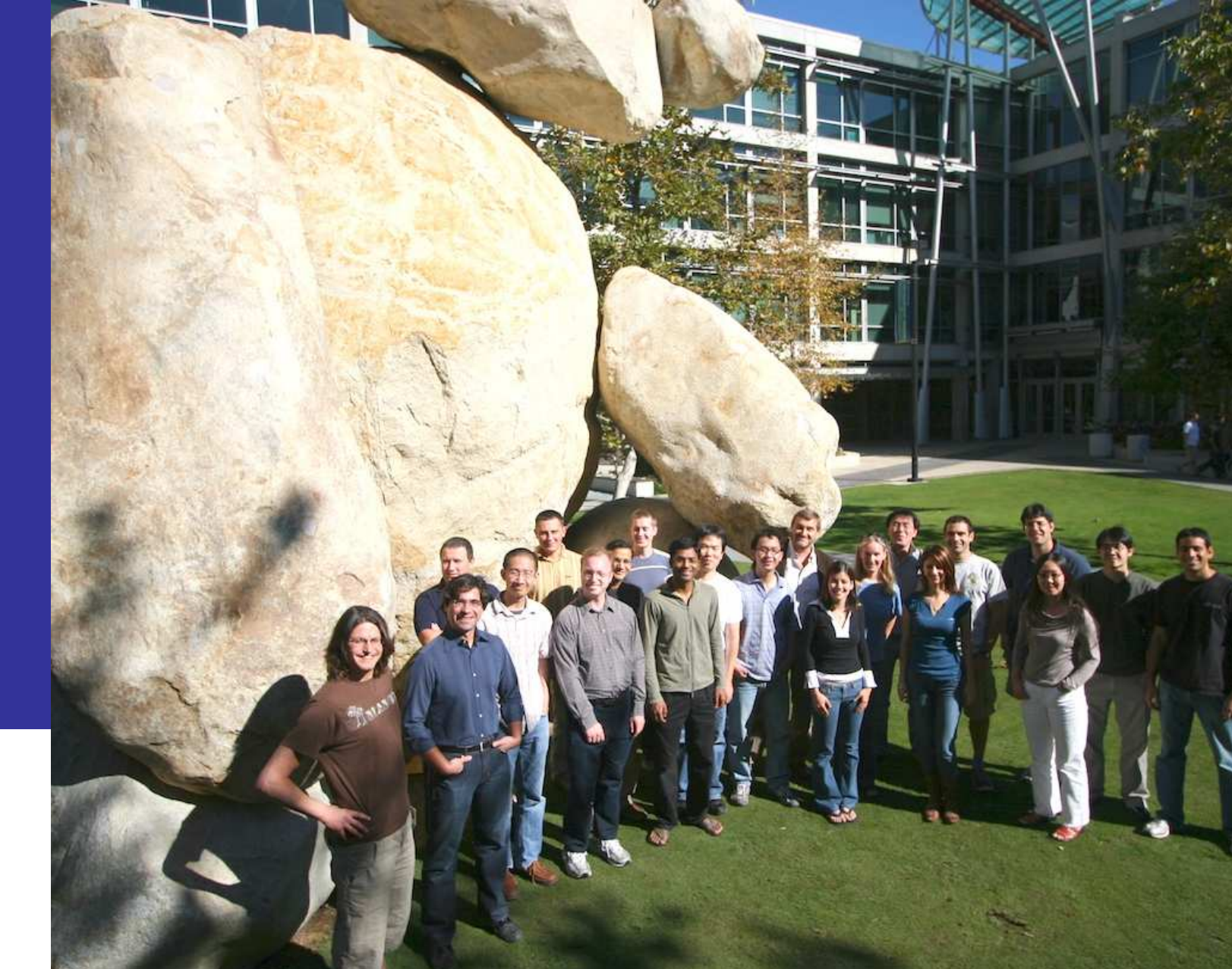
<http://proteomics.ucsd.edu>



**Center for
Computational
Mass
Spectrometry**



Acknowledgements



CCMS, UCSD

Pavel Pevzner

Vineet Bafna

Dumitru Brinza

Yuval Inbar

Ari Frank

Stephen Tanner

Banu Dost

Nitin Gupta

Sangtae Kim

Roy Liu

Julio Ng



UCSD

Steven Briggs

Pieter Dorrestein

Vivian Hook

Susan Taylor

Virgil Woods

Huilin Zhou

Steven Bark

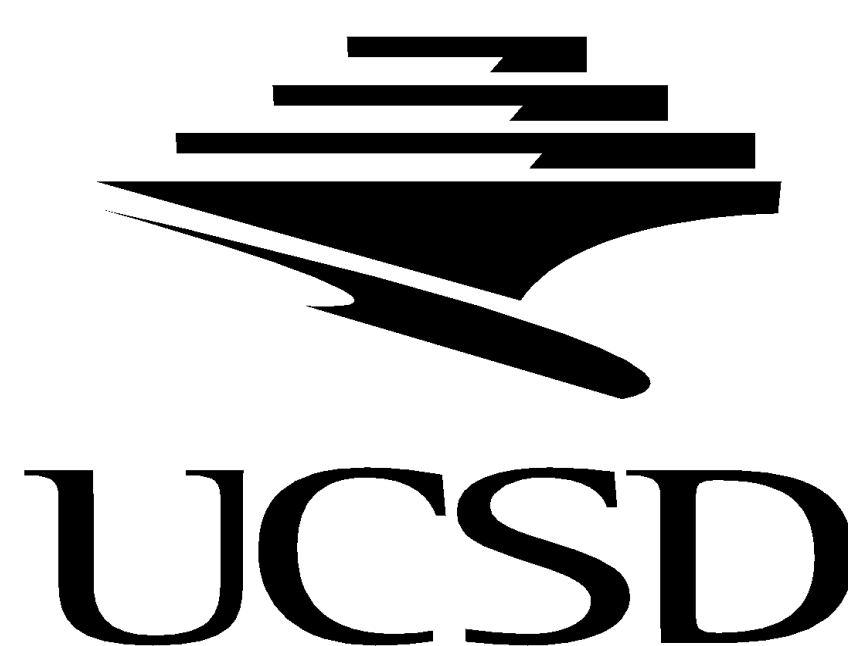
Zhouxin Shen

Wei-Ting Liu

Dario Meluzzi

Josué Pérez

Jian Wang



Broad Institute MIT/Harvard

Karl Clauser



Max Planck Institute, Germany

Matthias Mann

Jesper Olsen



Max Planck Institute
of Biochemistry

OHSU

Larry David

Phil Wilmarth



Pacific Northwest National Lab

Richard Smith

Joshua Adlkins

Gordon Anderson

Kim Hixson

Mary Lipton



Indiana University

Haixu Tang



Genentech

Jennie Lill

David Arnott

Victoria Pham

Wendy Sandoval



UCLA

Parag Mallick

Roland Luethy

NIH NIAID/NIMH

Jon Yewdell

Sanford Markey

Utrecht University, Netherlands

Albert Heck

U Seoul, Korea

Eunok Paek

Seungjin Na

U Victoria, Canada

Cristoph Borchers

Inst. Ricardo Jorge, Portugal

Deborah Penque

Nuno Charro

Merck

Arana Therapeutics

Agilent

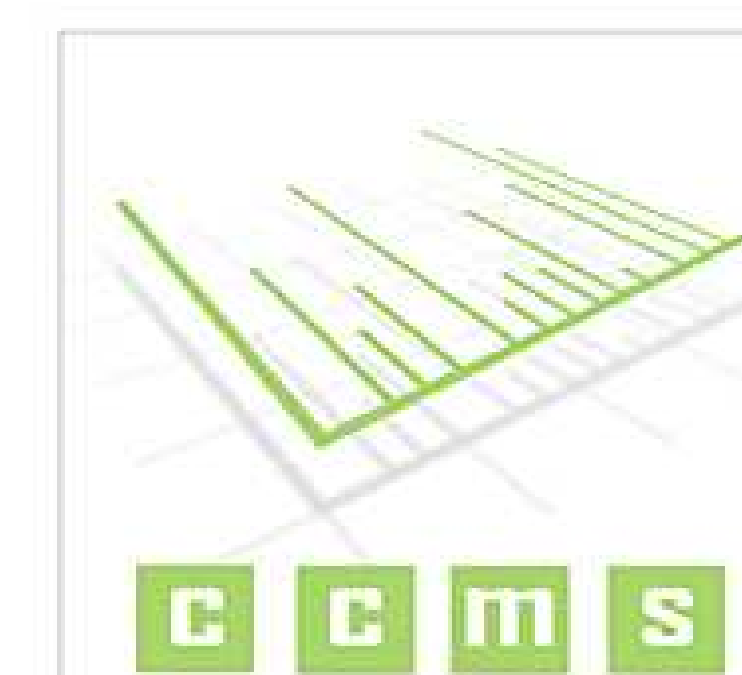


National Center for
Research Resources



Calit2

[//proteomics.ucsd.edu](http://proteomics.ucsd.edu)



Center for
Computational
Mass
Spectrometry