

---

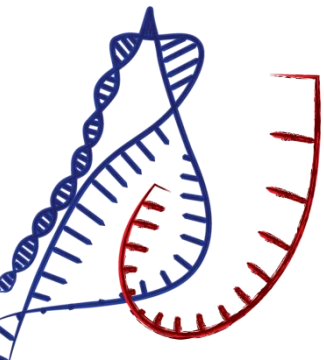
# Interpretation of PTM Data

---

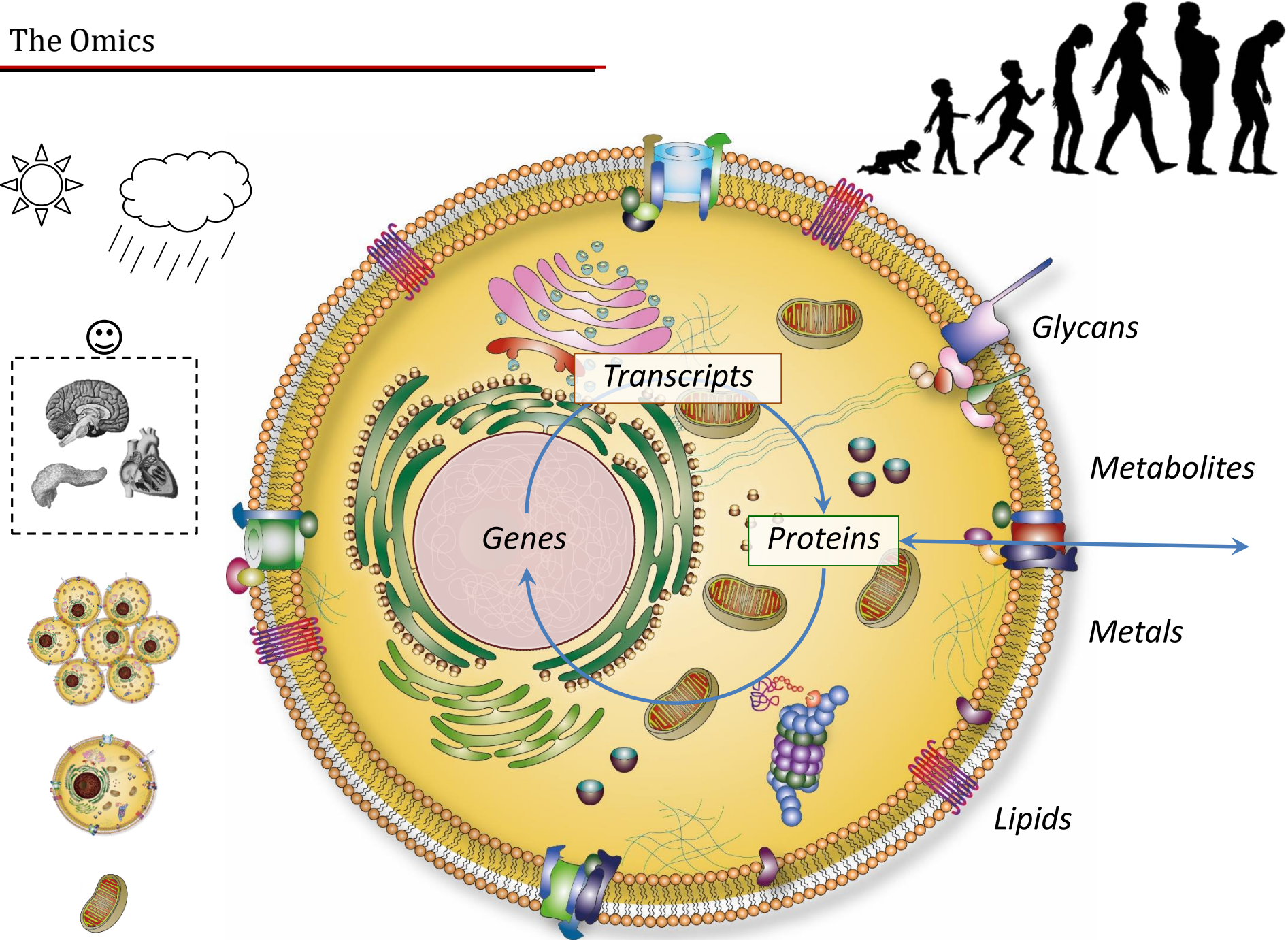
Marc Vaudel

*Center for Medical Genetics and Molecular Medicine,  
Haukeland University Hospital, Bergen, Norway*

*KG Jebsen Center for Diabetes Research, Department of Clinical Science,  
University of Bergen, Norway*



# The Omics



## Some PTMs and associated functions

---

### PTM

1- Glycosylation

2- Phosphorylation

3- Ubiquitination

4- Sumoylation

5- Acetylation

6- Oxydation

7- Methylation

### Function

A- Signalling

B- Tagging

C- Receptors

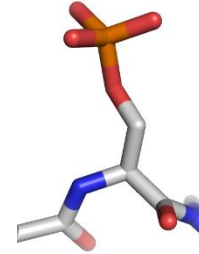
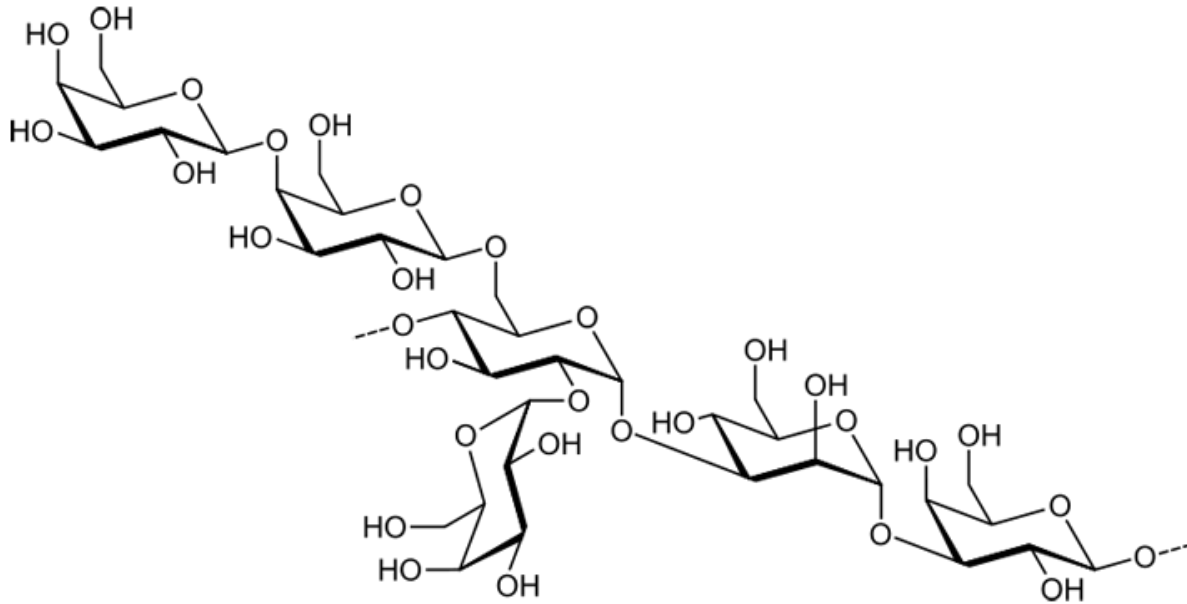
D- Adhesion/Contact

E- Mobility

F- Packaging

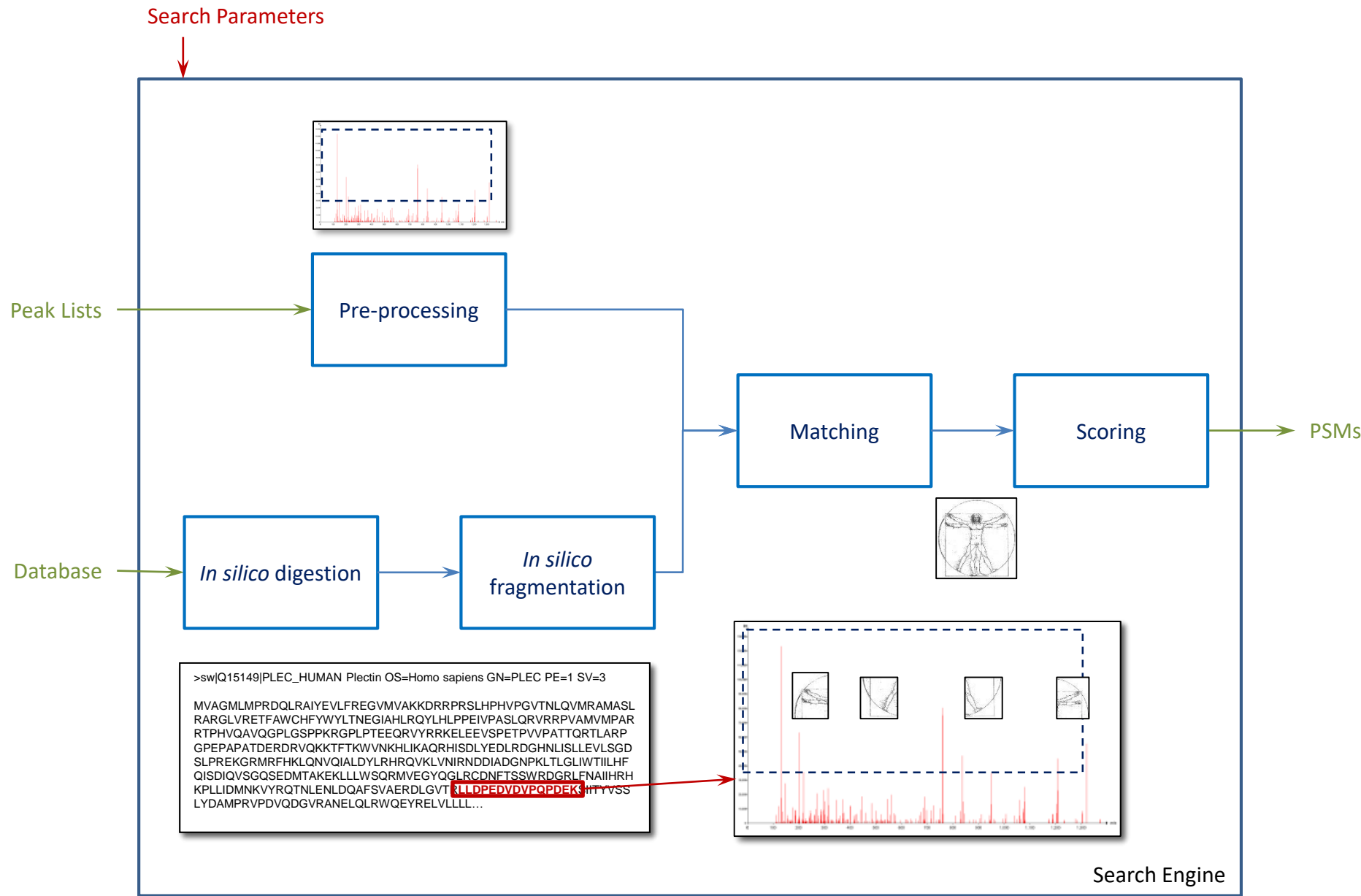
There is PTM and PTM

---

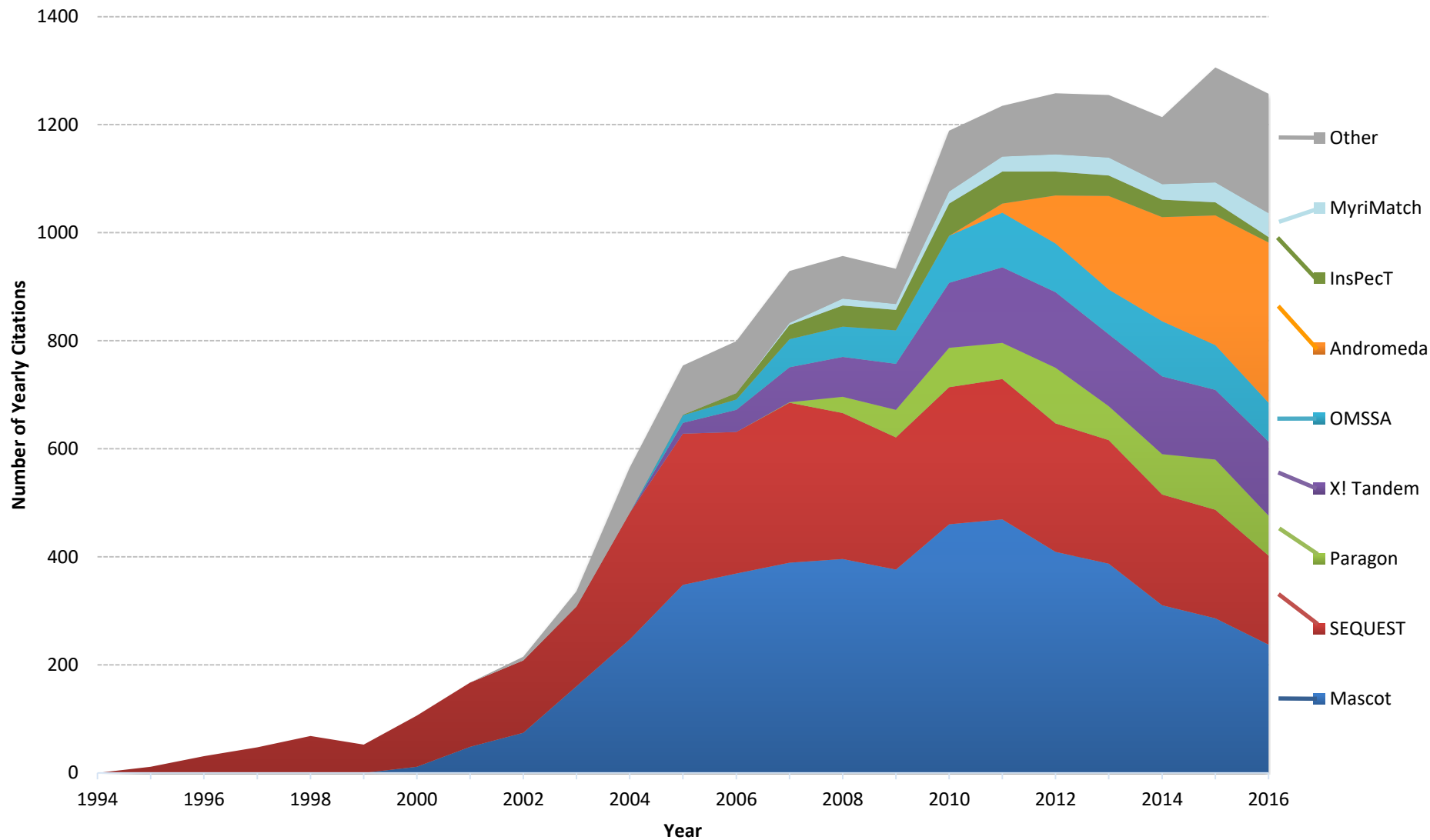


➡ ~~One pipeline fits all~~

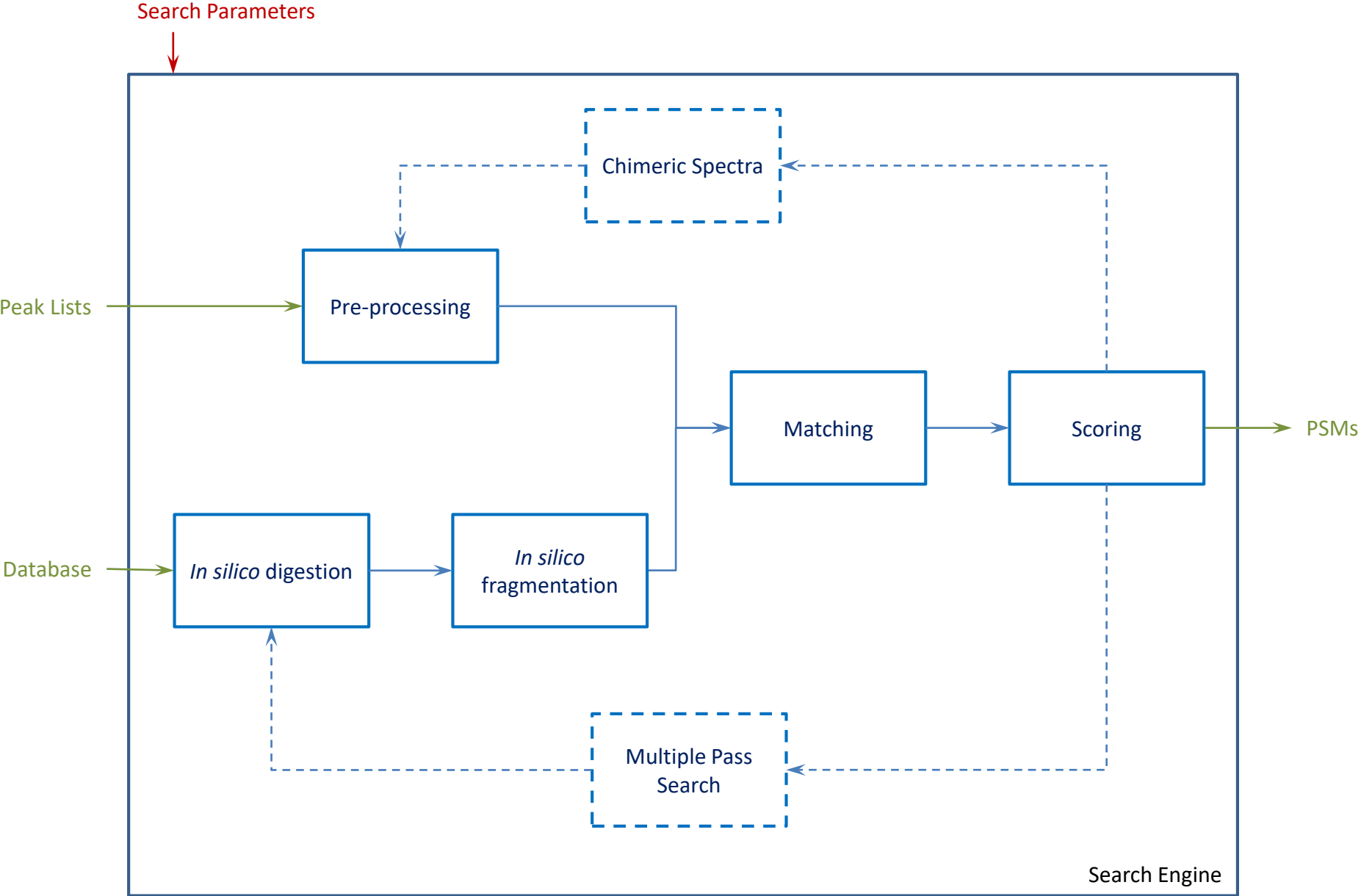
# Database Search Engines



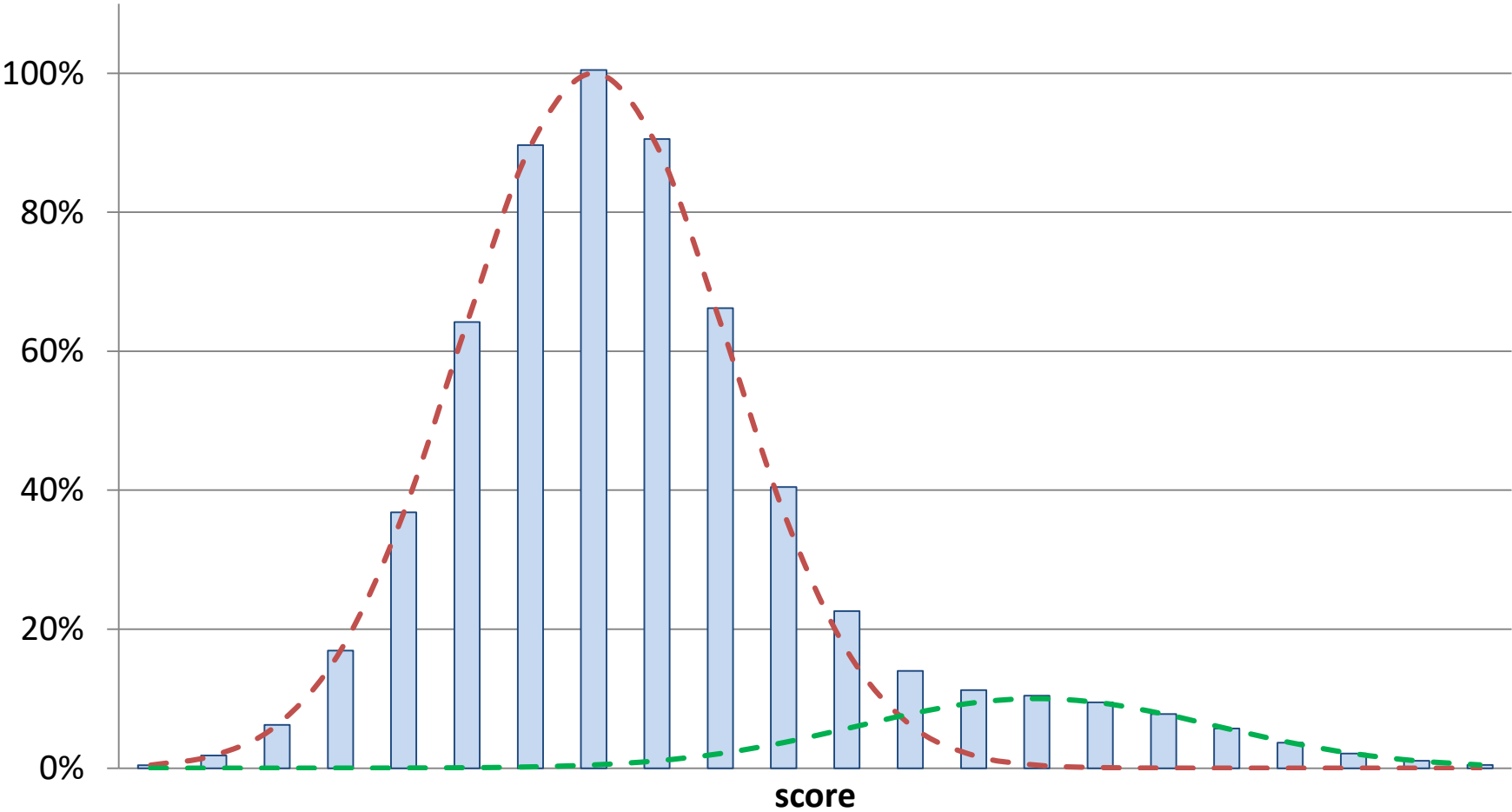
# Database Search Engines



# Database Search Engines

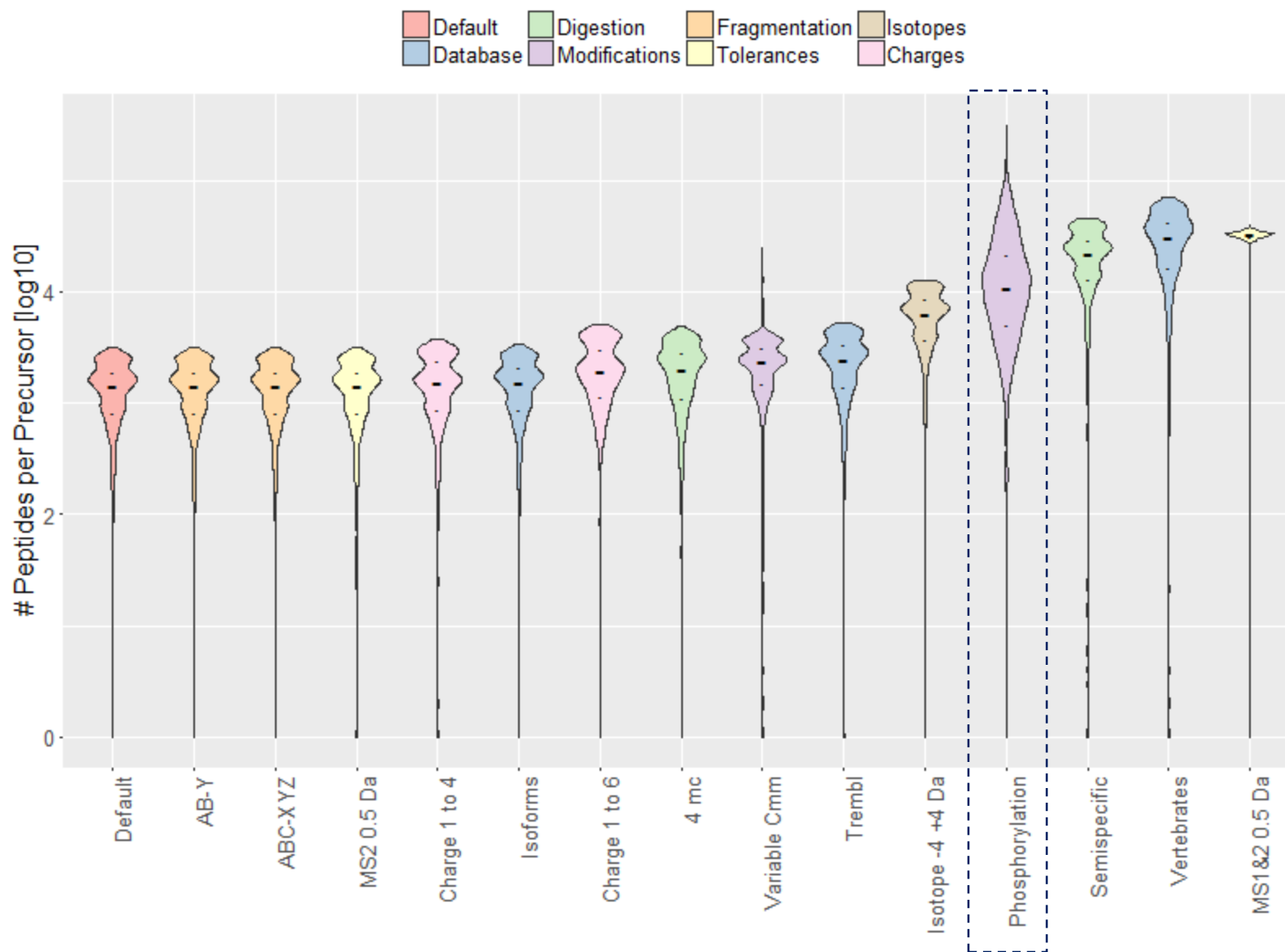


# Peptide score distribution

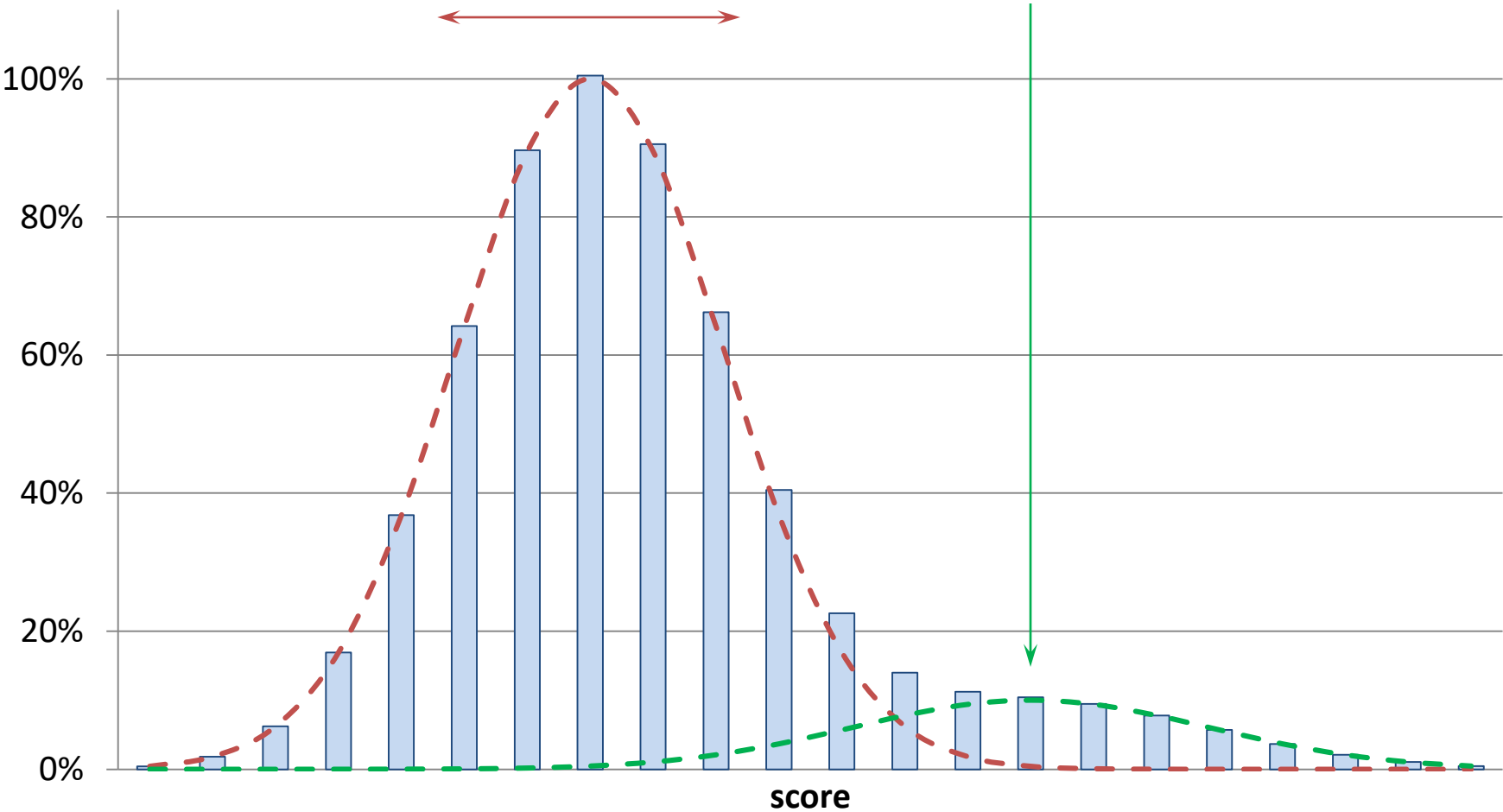




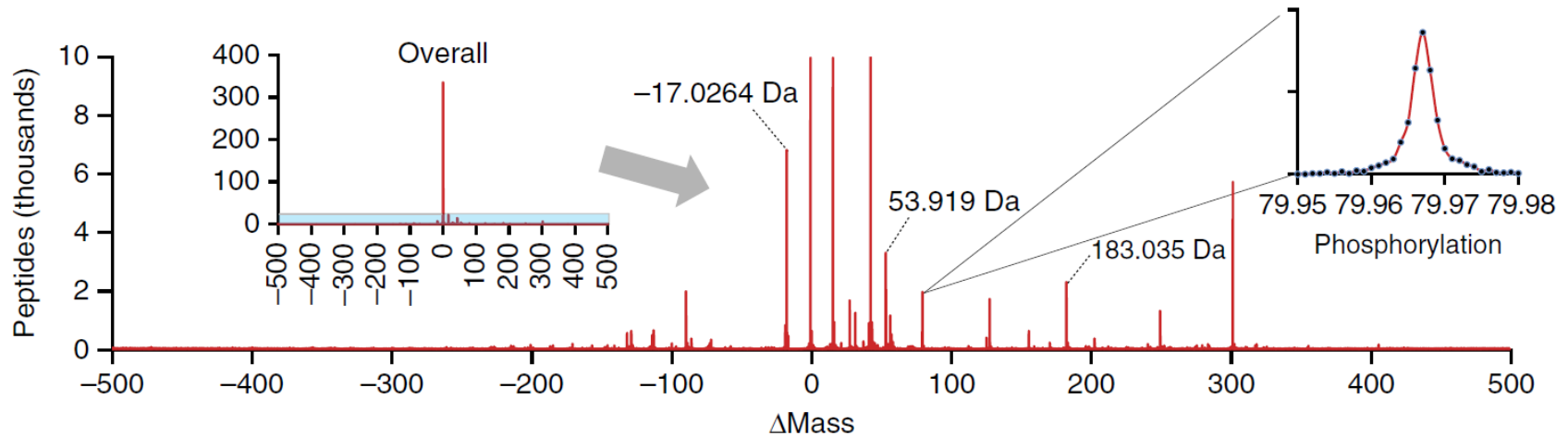
# The search space problem



Peptide score distribution

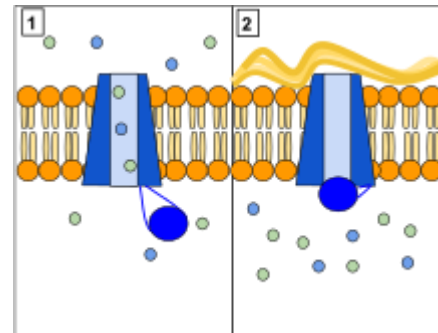
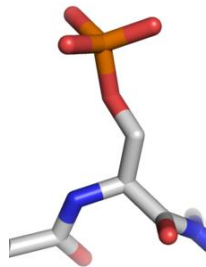
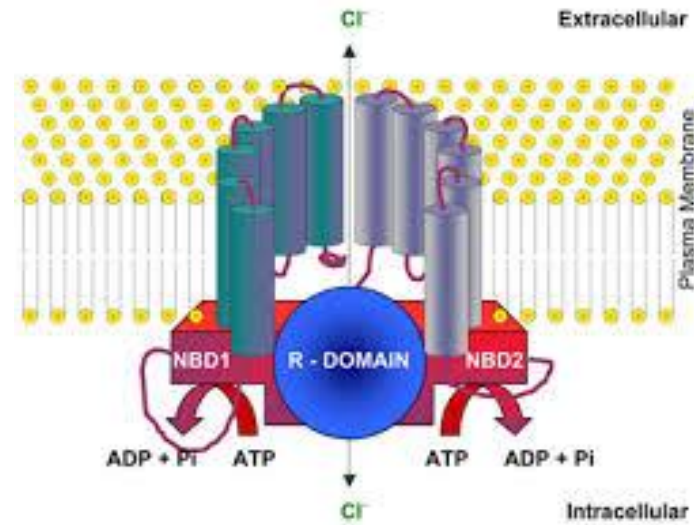


## Which PTMs include in the search?

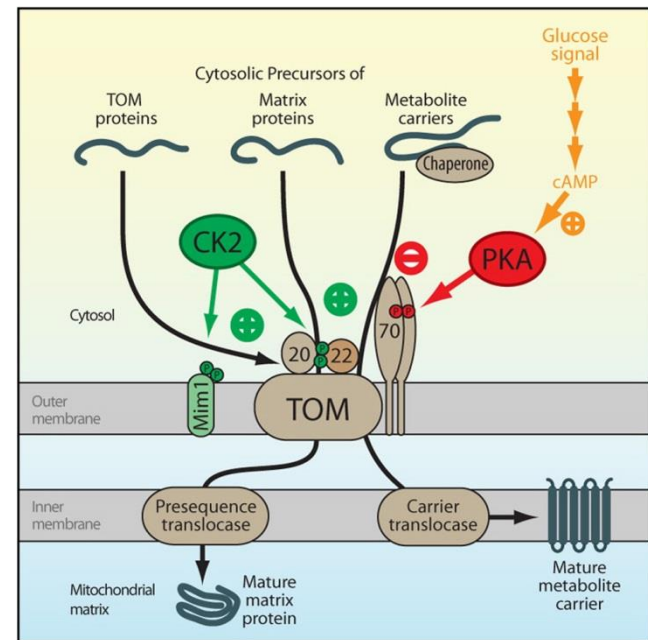
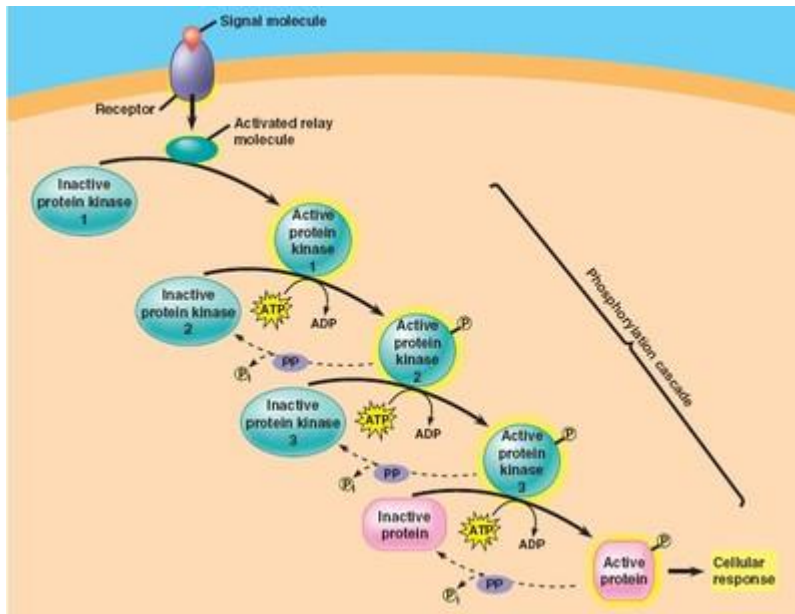


- ➡ Open search: MSFragger, PMID 28394336
- ➡ Parameters optimization: Param-Medic, PMID 28263070
- ➡ Limit PTMs per peptide

# Phosphorylation



# Phosphorylation



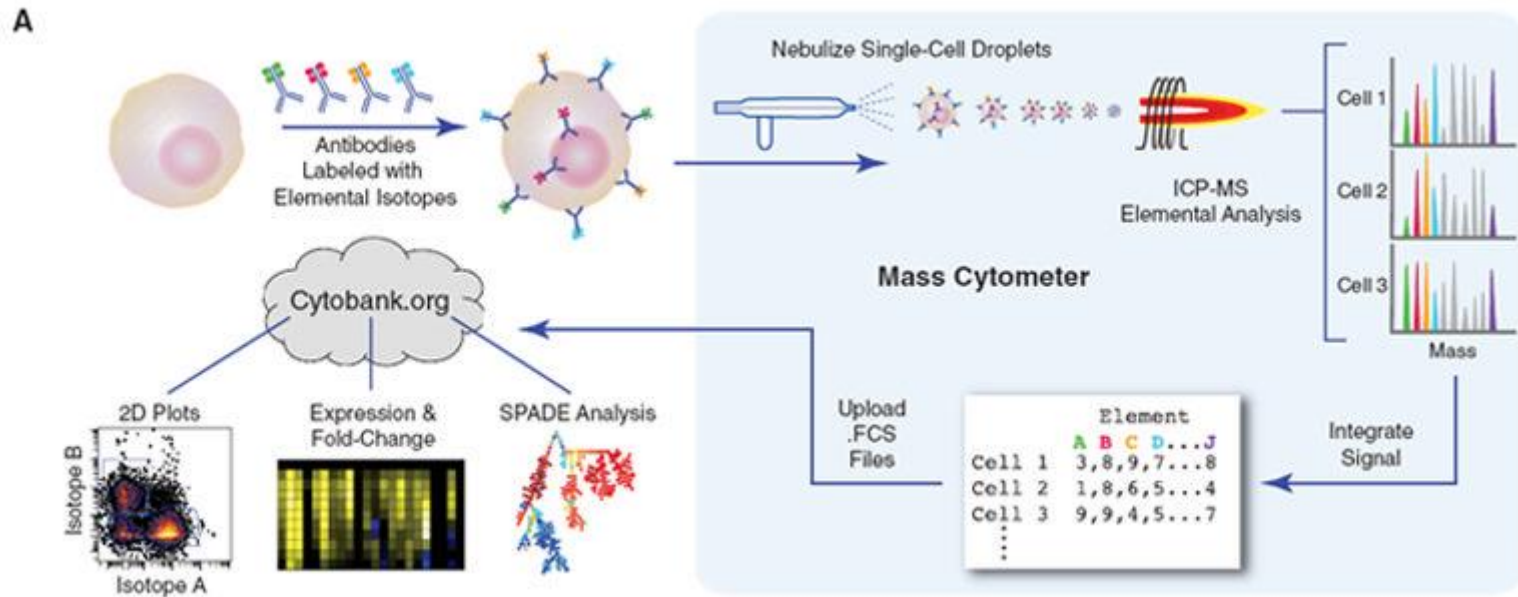
# Analyzing phosphoproteins

Antibody

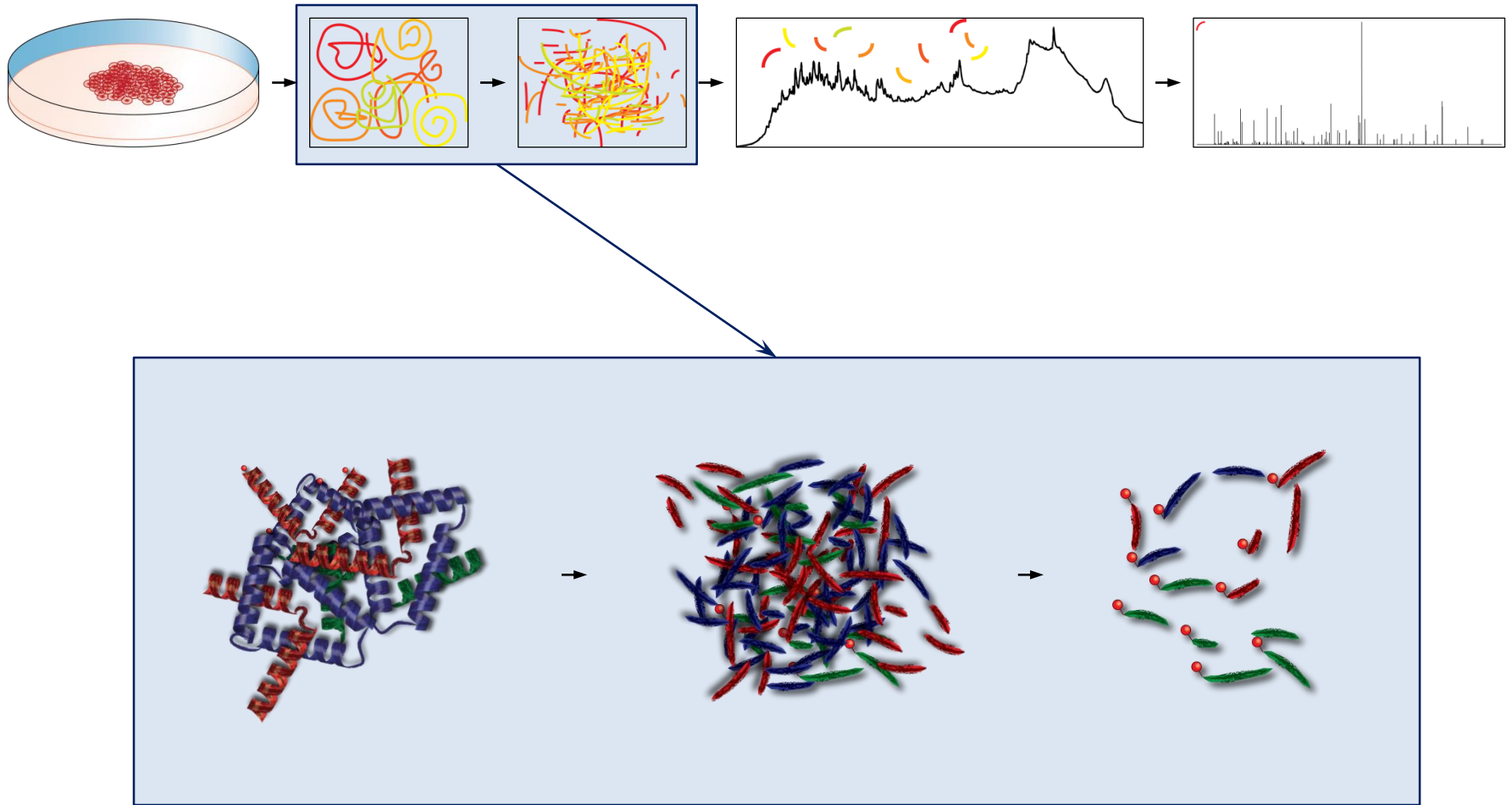


Mass Spectrometry

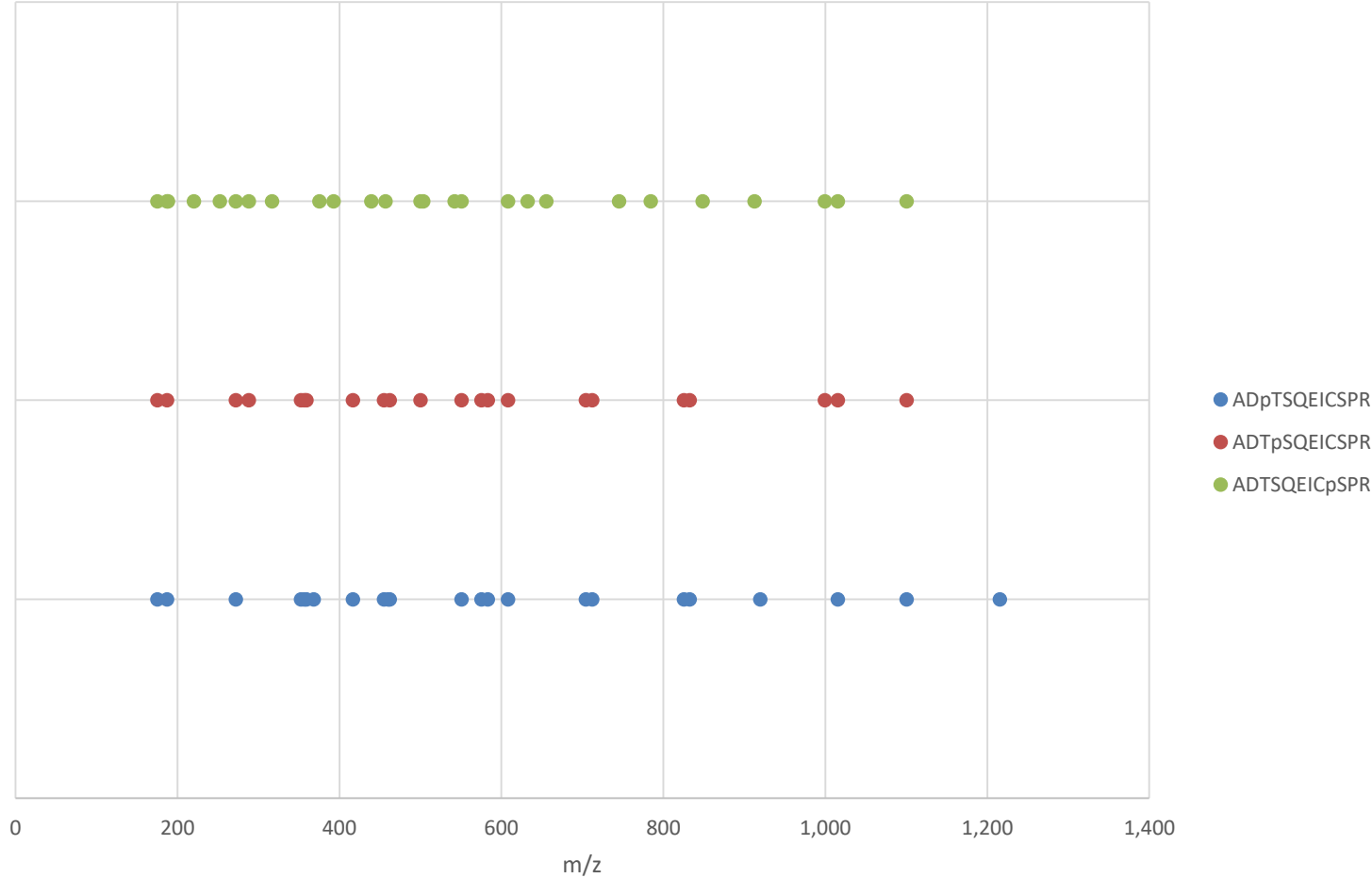
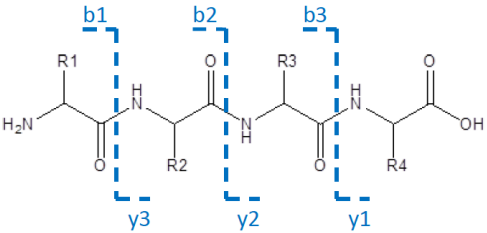
Cy-TOF



# Phosphoproteomics



# Site determining ions





# Scoring PTM isoforms

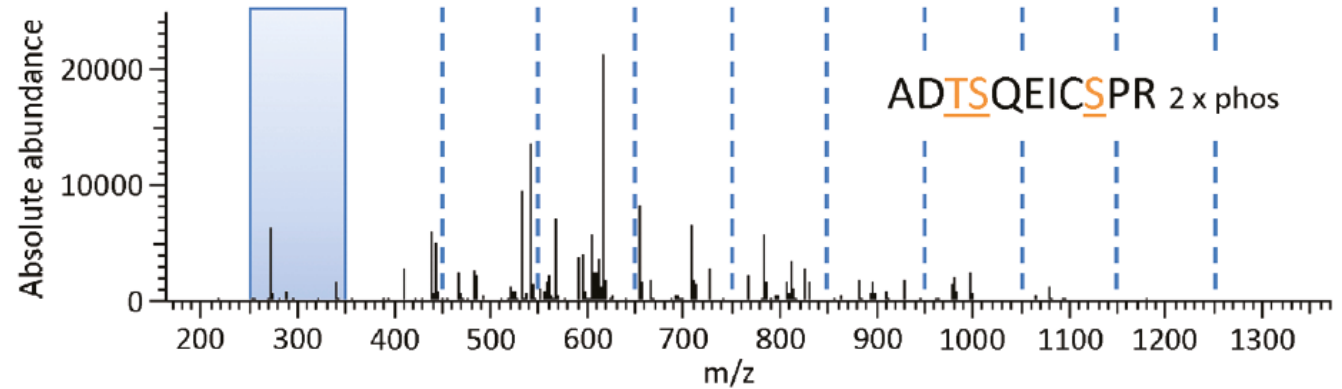
Noise level



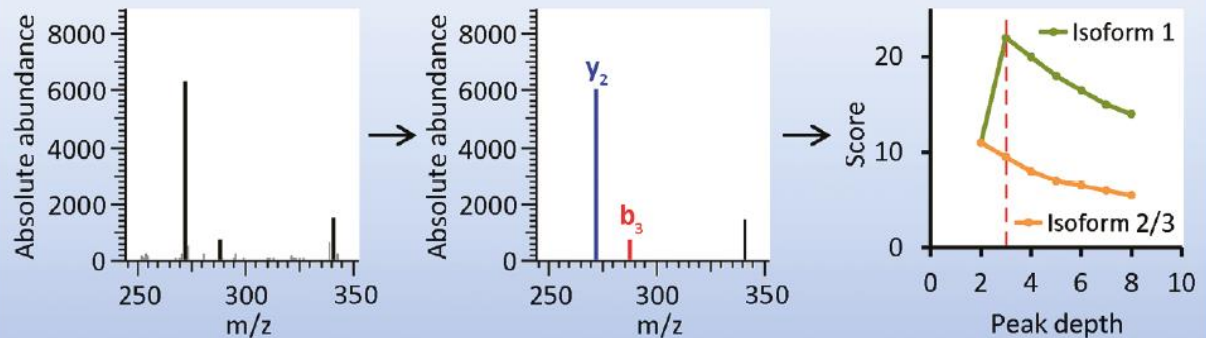
Isoform score



Significance



Determine optimal peak depth for current m/z window



## Scoring PTM isoforms

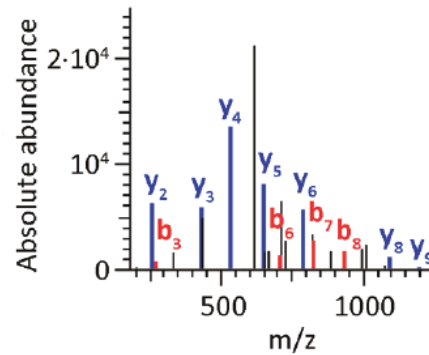
Noise level



Isoform score



Significance



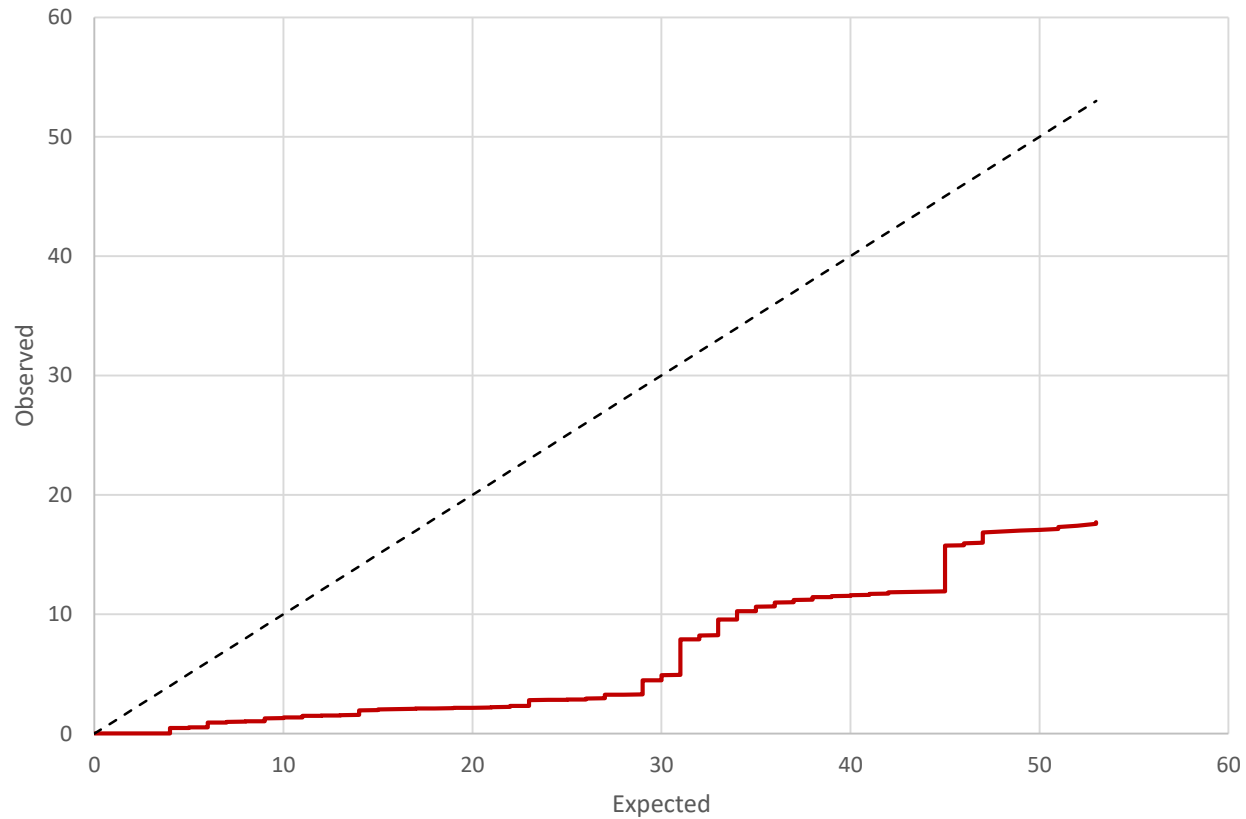
$$p = \frac{N_{\text{peaks}} \cdot d}{w}$$

$$P = \sum_k^n p^k \cdot (1 - p)^{n-k}$$

$$\text{phosphoRS Peptide Score} = -10 \cdot \log(P)$$

# Performance

---



- ➡ Simplistic fragmentation model
- ➡ Challenges in noise estimation

# Scoring PTM isoforms

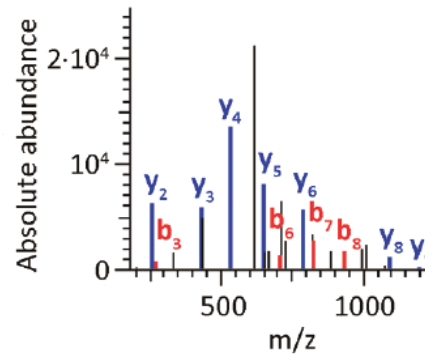
Noise level



Isoform score



Significance



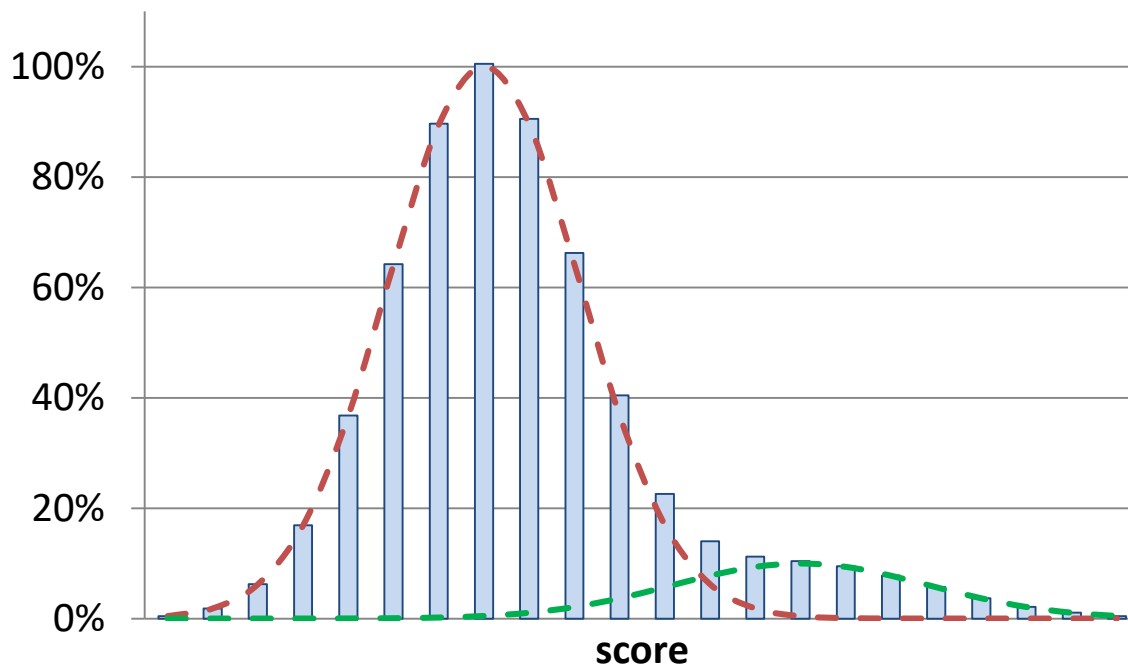
$$p = \frac{N_{Peaks} \cdot d}{w}$$

$$P = \sum_k^n p^k \cdot (1 - p)^{n-k}$$

$$\text{phosphoRS Peptide Score} = -10 \cdot \log(P)$$

phosphoRS Peptide Score	1/P value	phosphoRS Sequence Probability	phosphoRS Site Probability Calculation					
121.7	$1.48 \cdot 10^{12}$	99.8 %	AD	T	(pS)	QEIC	(pS)	PR
93.8	$2.40 \cdot 10^9$	0.2 %	AD	(pT)	S	QEIC	(pS)	PR
24.8	$3.02 \cdot 10^2$	0.0 %	AD	(pT)	(pS)	QEIC	S	PR
$\Sigma = 1.48 \cdot 10^{12}$		$\Sigma = 100.0 \%$	0.2 %	99.8 %	100.0 %			

## False Localization Rate Estimation



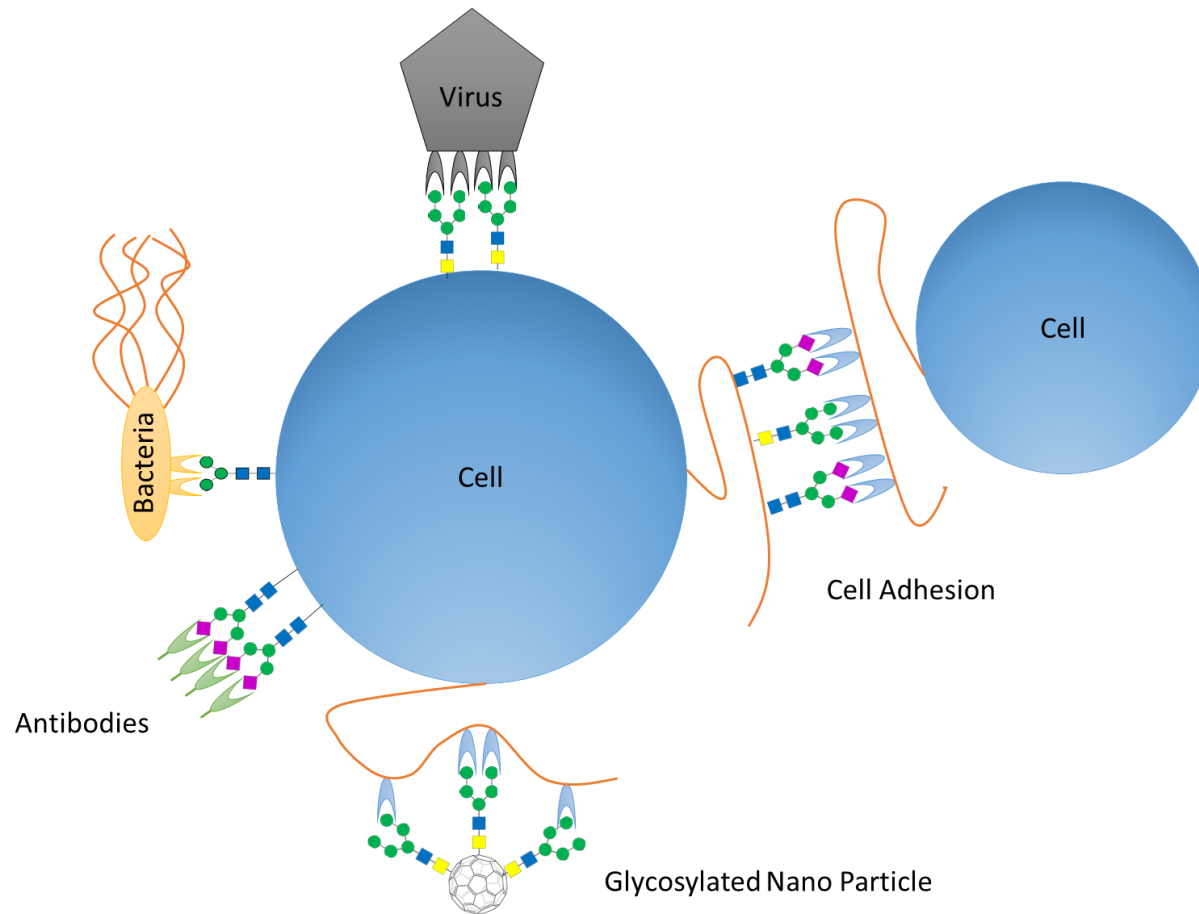
### *target permutations*

LQ <b>T</b> VHSIPLTINK	124.66
LQTVH <b>S</b> IPLTINK	84.35
LQTVHSIPL <b>T</b> INK	10.99

### *decoy permutations*

<b>L</b> QTVHSIPLTINK	51.09
<b>L</b> QTVHSIPLTINK	68.26
LQ <b>T</b> VHSIPLTINK	64.66
LQTV <b>H</b> SIPLTINK	49.83
LQTVHS <b>I</b> PLTINK	33.12
LQTVHSI <b>P</b> LTINK	13.55
LQTVHSIPL <b>L</b> TINK	14.01
LQTVHSIPLT <b>I</b> NK	9.90
LQTVHSIPLTINK <b>N</b>	6.24
LQTVHSIPLTINK <b>K</b>	4.46

# Glycosylation



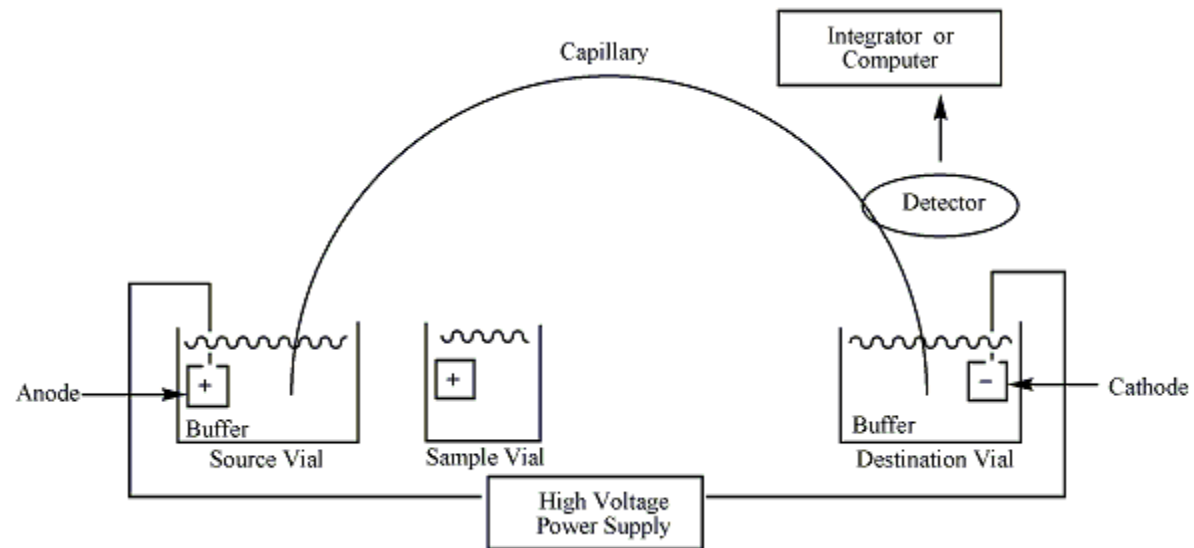
# Analyzing glycoproteins

Antibody

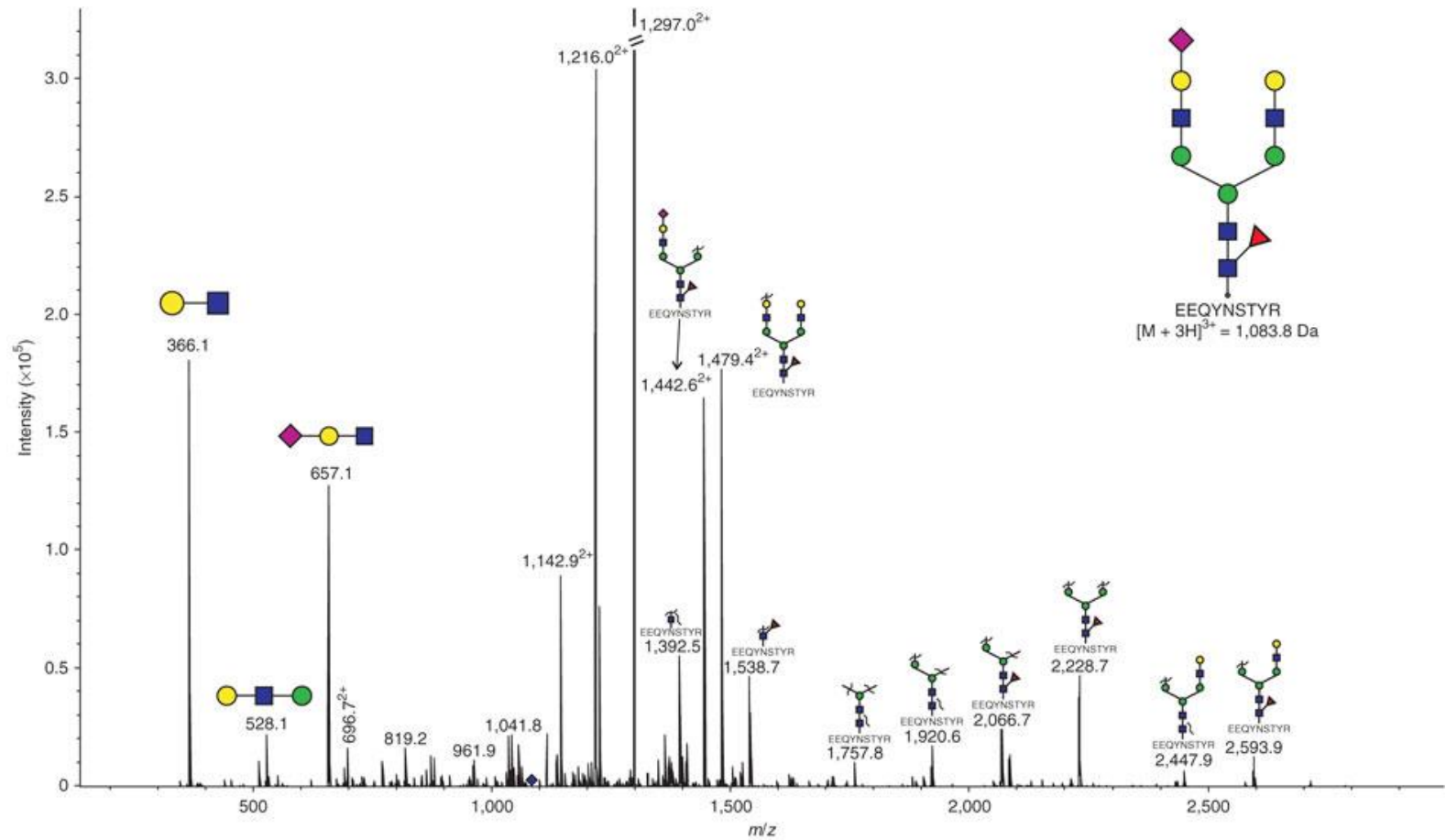


Mass Spectrometry

Capillary Electrophoresis



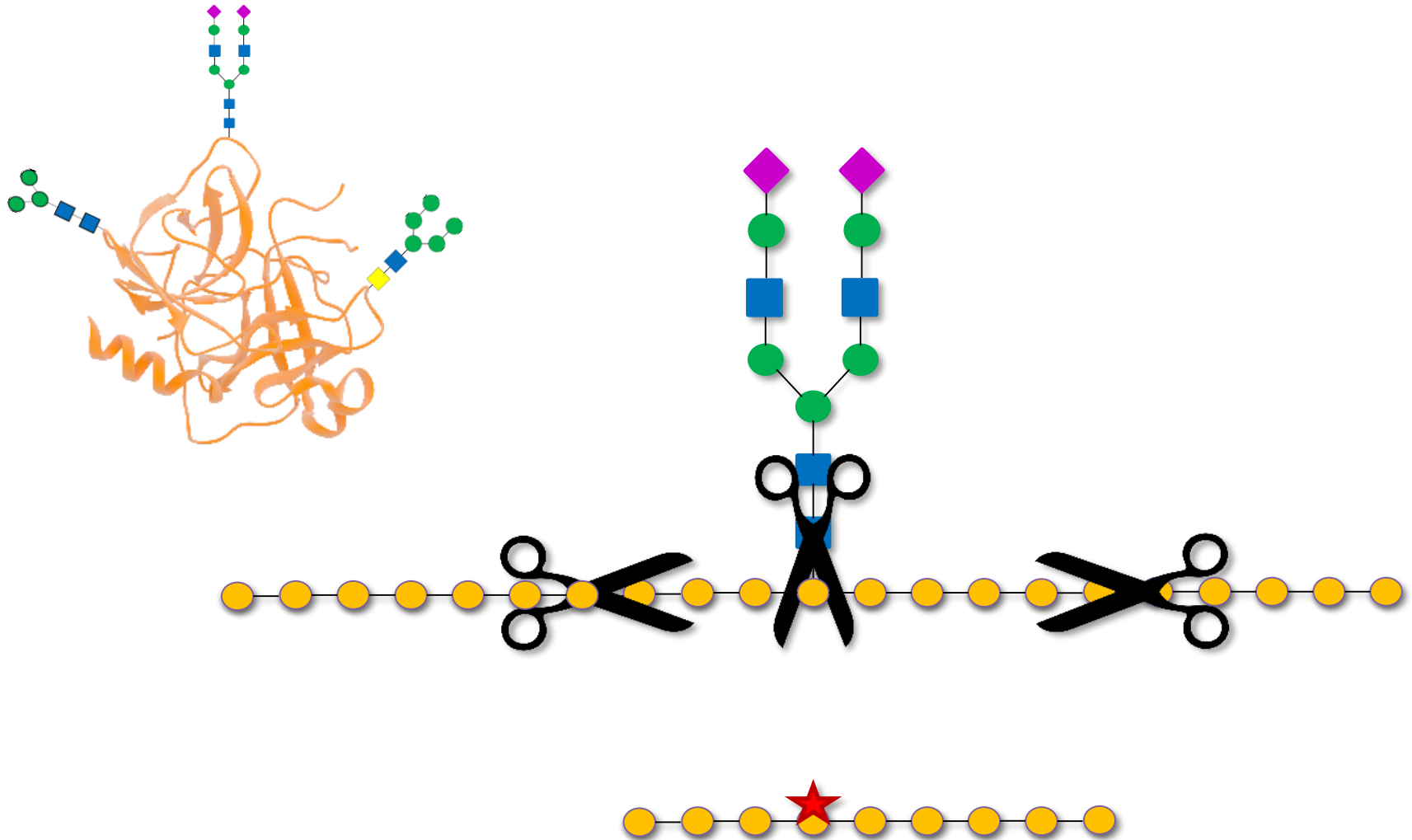
# No database and infinite combinations





## Simplify to standard problem

---



# De novo identification

*MS1*



*MS2*



*MS3*

