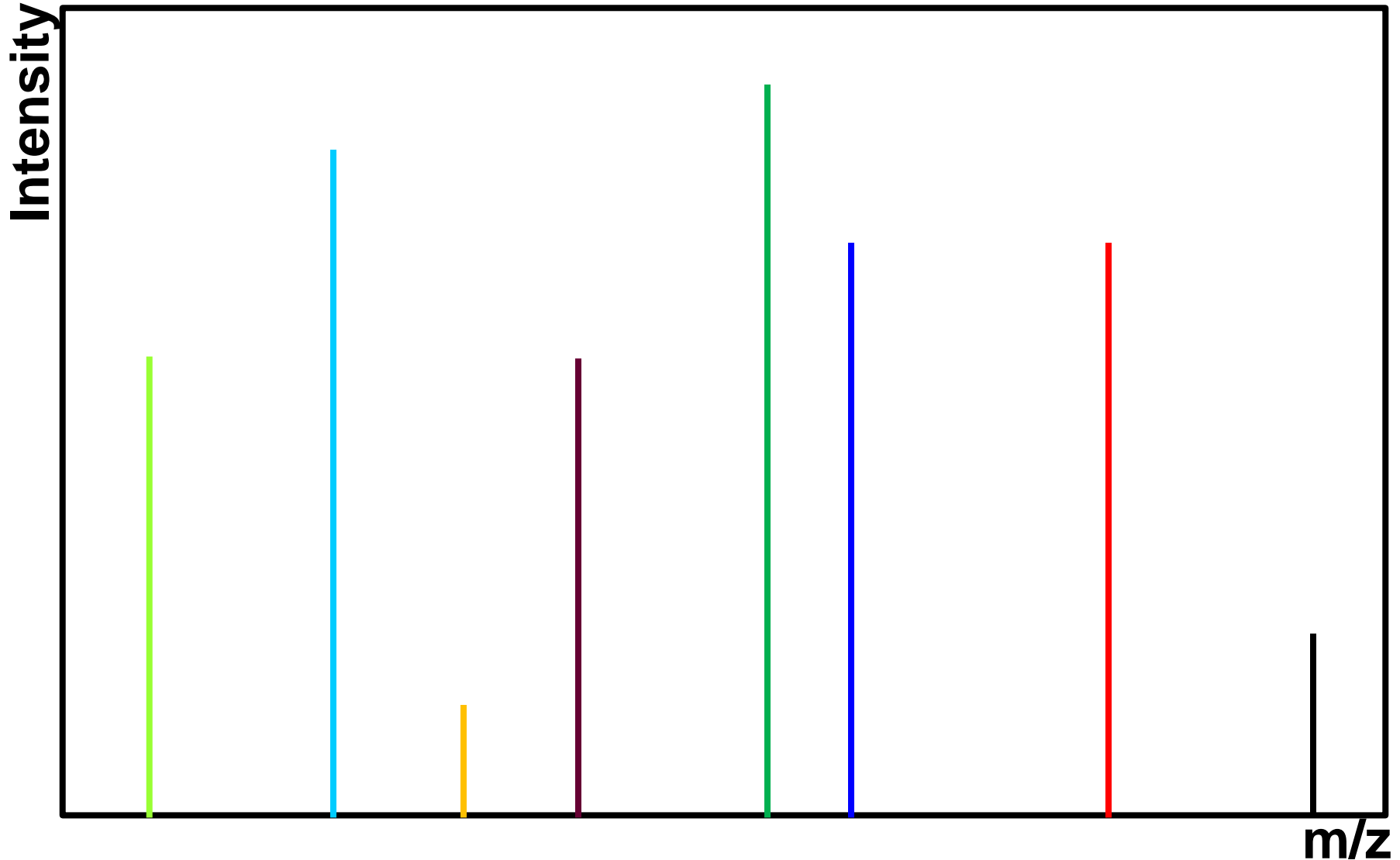


# Identifikace jednoho proteinu

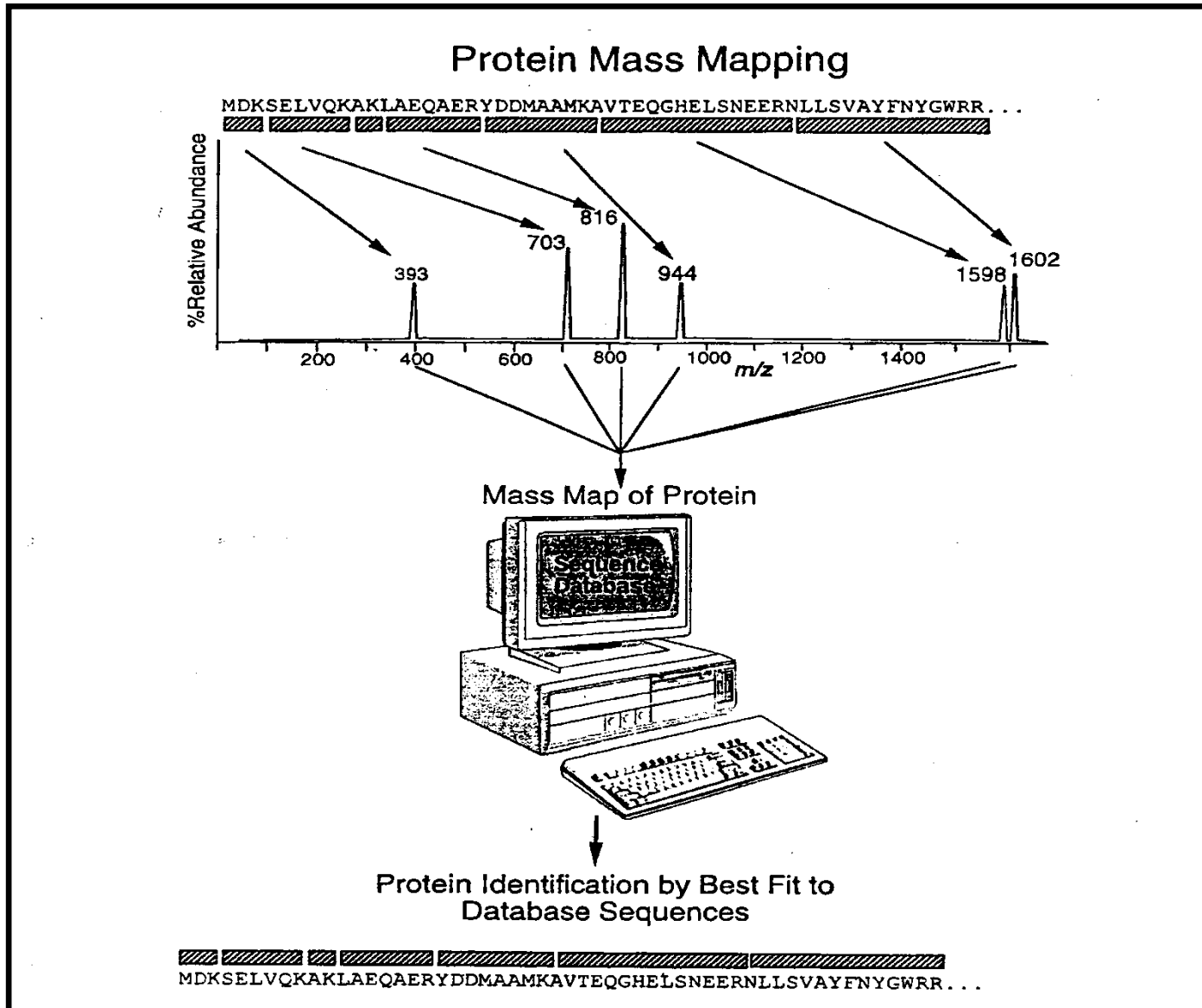
ADSMK EETSLINVC PWK VNMGGHDSIYTLR EEIYTLK  
MLIILLR TYSHEEDK EWQIDSLAEIR IQPLPMNVSA



# Proteinové štěpení

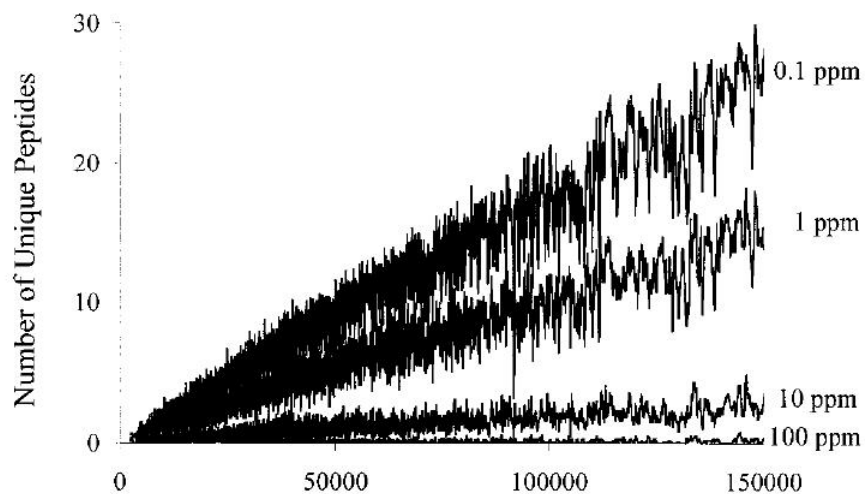
- Trypsin K/R- \-P
- TrypN -K/R
- AspN -D
- GluC E-, E/D- \-P
- ArgC R- \-P
- LysC K- \-P
  
- CNBr M-

# Identifikace proteinu - princip



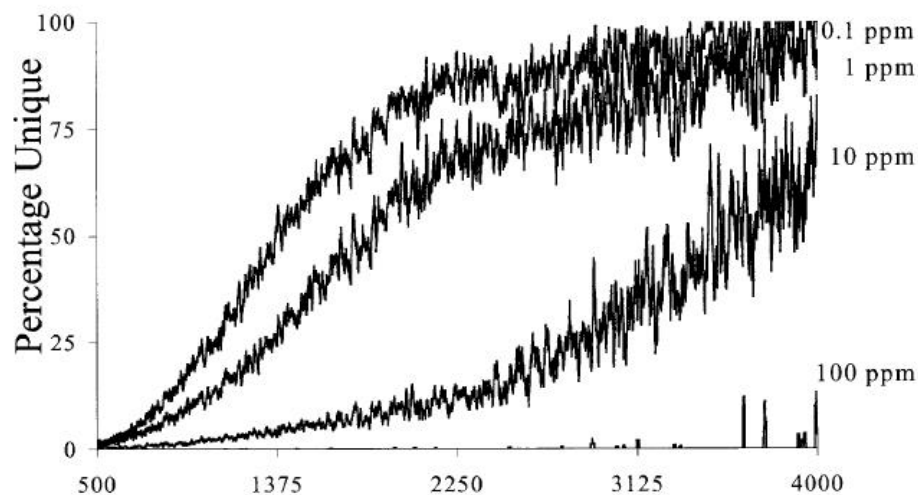
# Důležitá je přesnost...

Čím vyšší přesnost měření, tím méně peptidů potřebujeme pro bezchybnou identifikaci



**Počet unikátních peptidů  
jako funkce velikosti  
proteinu**

**Unikátnost peptidu jako  
funkce jeho MW**



# Peptidové mapování

**MASCOT** – Matrix Science - [http://www.matrixscience.com/cgi/search\\_form.pl?FORMVER=2&SEARCH=PMF](http://www.matrixscience.com/cgi/search_form.pl?FORMVER=2&SEARCH=PMF)

**MS-Fit – Protein Prospector, UCSF** - <http://prospector.ucsf.edu/prospector/mshome.htm>

**Pro-Found – Rockefeller University** <http://prowl.rockefeller.edu/prowl-cgi/profound.exe>  
<http://hs2.proteome.ca/prowl/profound/control.html>

# MASCOT Peptide Mass Fingerprint

Your name  Email

Search title

Database(s)   
 Plants\_EST   
 Prokaryotes\_EST   
 Rodents\_EST   
 Vertebrates\_EST   
 contaminants

Enzyme

Allow up to  missed cleavages

Taxonomy

Fixed modifications   
 --- none selected ---

Display all modifications

Variable modifications   
 --- none selected ---

Acetyl (K)   
 Acetyl (N-term)   
 Acetyl (Protein N-term)   
 Amidated (C-term)   
 Amidated (Protein C-term)   
 Ammonia-loss (N-term C)   
 Carbamidomethyl (C)   
 Carbamidomethyl (N-term)   
 Carbamyl (K)   
 Carbamyl (N-term)   
 Carboxymethyl (C)

Protein mass  kDa

Peptide tol.  $\pm$   Da

Mass values  MH<sup>+</sup>  M<sub>r</sub>  M-H<sup>-</sup>

Monoisotopic  Average

Data file  Soubor nevybrán.

Query

Data input

Decoy

Report top  hits

Start Search ...

Reset Form

# MS-Fit

<p><b>Database</b></p> <p>SwissProt.2016.9.6 SwissProt.2016.9.6.random SwissProt.2016.9.6.random.concat SwissProt.2017.11.01</p> <p><b>[+] User Protein Sequence</b></p> <p><b>DNA Frame Translation</b> 3 <b>N Term AA Limit</b> <input type="text"/></p> <p><b>Taxonomy</b></p> <p>All HUMAN MOUSE HUMAN RODENT MODEL PLANTS RODENT ROACH LOCUST BEETLE</p> <p><b>Output</b> HTML <input type="checkbox"/> <b>Hits to file</b> <input type="checkbox"/> <b>Name</b> lastres</p>	<p><b>Digest</b> Trypsin <b>Max. Missed Cleavages</b> 1</p> <p><b>Constant Mods</b></p> <p>Biotin (N-term) Biotin-HPDP (C) bisANS (K) bisANS-sulfonates (K) Butyryl (K) Carbamidomethyl (C)</p>
<p><b>[+] Pre-Search Parameters</b></p>	
<p><b>Start Search</b></p>	<p><b>Sample ID (comment)</b> <input type="text"/></p> <p><b>Display Graph</b> <input type="checkbox"/></p>
<p><b>Maximum Reported Hits</b> 5</p> <p><b>Sort By</b> Score Sort</p> <p><b>Report Homologous Proteins</b> Interesting</p> <p><b>Min. # peptides required to match</b> 4</p> <p><b>Report MOWSE Scores</b> <input checked="" type="checkbox"/> <b>Pfactor</b> 0.4</p> <p><b>Masses are</b> monoisotopic</p> <p><b>Tol</b> 20 <b>ppm</b> <b>Sys Err</b> 0</p> <p><b>Contaminant Masses</b></p> <p><input type="text"/></p>	<p><b>Possible Modifications</b></p> <p>Peptide N-terminal Gln to pyroGlu Oxidation of M Protein N-terminus Acetylated Acrylamide Modified Cys</p> <p><b>User Def Mod 1</b> Acetyl (K) <b>User Def Mod 2</b> Acetyl (K) <b>User Def Mod 3</b> Acetyl (K) <b>User Def Mod 4</b> Acetyl (K)</p> <p><b>OR</b></p> <p><b>Unknown Amino Acid</b> <input type="checkbox"/> <b>Single Base Change</b> <input type="checkbox"/> <b>Homology</b> <input type="checkbox"/></p> <p><b>Max Mods</b> 1 <b>Min. # match with NO AA subs</b> 1</p>
<p><b>Instrument</b> ESI-Q-TOF <b>Data Format</b> PP M/Z Charge</p>	
<p><b>Data Paste Area</b></p> <p>842.5100 856.5220 864.4733 870.5317 940.4754 943.4885 959.4934 970.4308 975.4785</p>	

## PROWL

- [ProFound](#)
- [ProteinInfo](#)
- [PeptideMap](#)
- [X! Tandem](#)
- [X! Hunter](#)
- [GPMDB](#)
  
- [PROWL Home](#)
- [Chait Lab](#)



The Rockefeller University  
1230 York Avenue,  
New York, NY 10021  
(212) 327-8000



**National Center for  
Research Resources**

National Resource  
for the Mass Spectrometric  
Analysis of Biological  
Macromolecules

## PROFOUND

### General

Sample ID

Database

Taxonomy

Protein Mass  -  kDa

Protein pI  -

Expect

Z  show  candidates

### Masses

Average Masses:

Mass tolerance (average): +/-

Tolerance unit:  Da  %  ppm

### Digestion

Allow  missed cleavages  
maximum

Enzyme

For user-defined cleavage, click [here](#).

### Modifications

Complete Modification(s)

4-vinyl-pyridine (Cys)

Acrylamide (Cys)

Iodoacetamide (Cys)

Iodoacetic acid (Cys)

Partial  Methionine  
Modification oxidation

For more partial modifications, click  
[here](#).

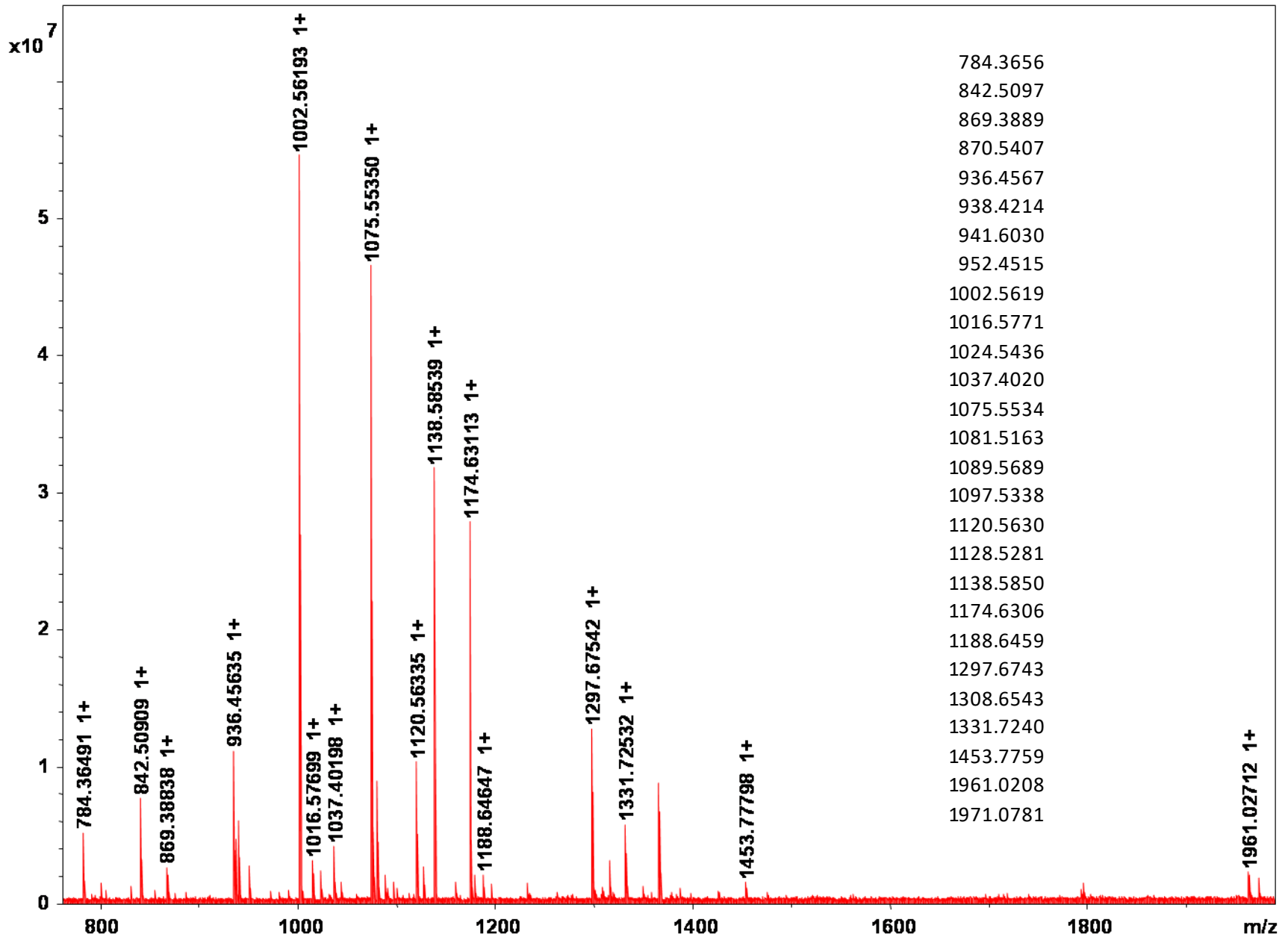
Monoisotopic Masses:

Mass tolerance (monoisotopic): +/-

Charge state:  M  MH+



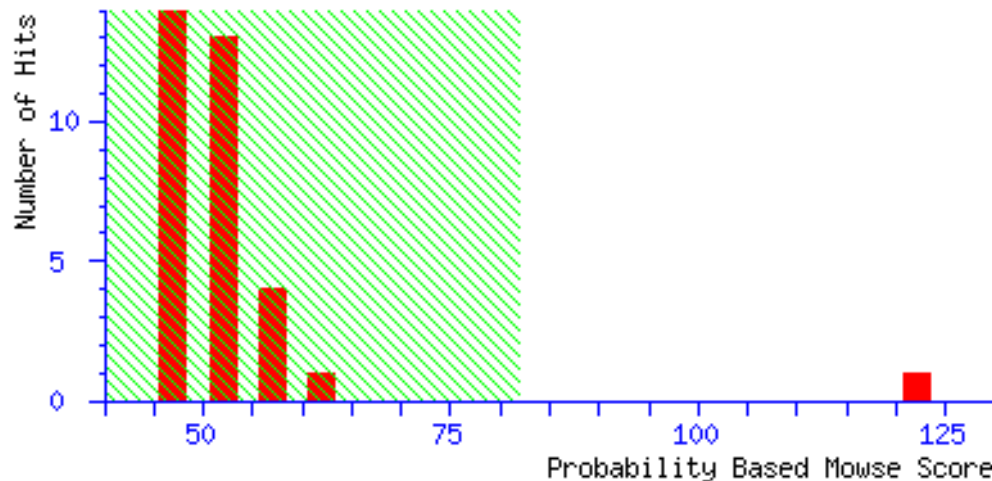
# Peptidové mapování – vliv parametrů



# Peptidové mapování – vliv parametrů

## Probability Based Mowse Score

Protein score is  $-10 \cdot \log(P)$ , where  $P$  is the probability that the observed match is a random event.  
Protein scores greater than 82 are significant ( $p < 0.05$ ).



Database	NCBI
Taxonomy	All taxa
Missed Cleavages	0
Fixed Mods	0
Variable Mods	0
Peptide tol	1000

## Concise Protein Summary Report

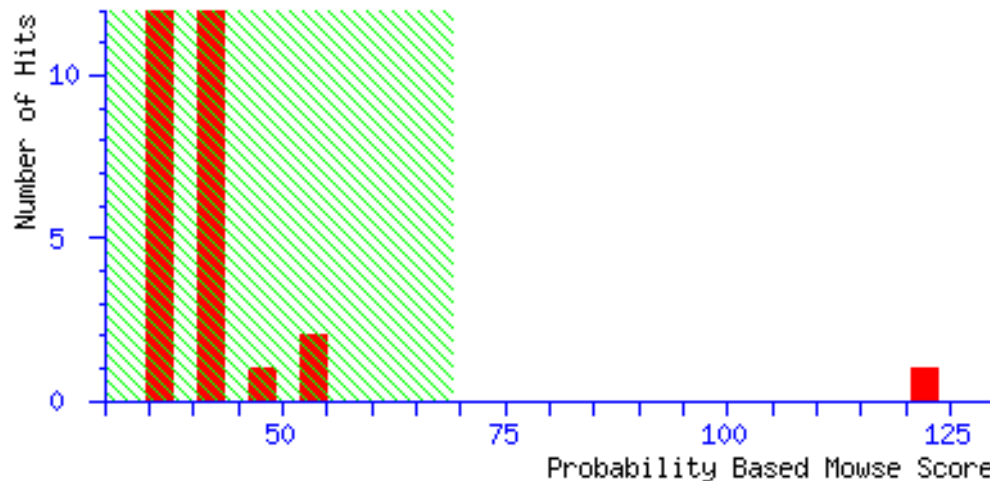
Format As	Concise Protein Summary	<a href="#">Help</a>	
Significance threshold $p <$	0.05	Max. number of hits	AUTO
Re-Search All	Search Unmatched		

- [gi|15900172](#) **Mass:** 34186 **Score:** 122 **Expect:** 6e-06 **Queries matched:** 12  
DNA-directed RNA polymerase subunit alpha [Streptococcus pneumoniae TIGR4]

# Peptidové mapování – vliv parametrů

## Probability Based Mowse Score

Protein score is  $-10 \cdot \log(P)$ , where  $P$  is the probability that the observed match is a random event. Protein scores greater than **69** are significant ( $p < 0.05$ ).



Database	SwissProt
Taxonomy	All taxa
Missed Cleavages	0
Fixed Mods	0
Variable Mods	0
Peptide tol	1000

## Concise Protein Summary Report

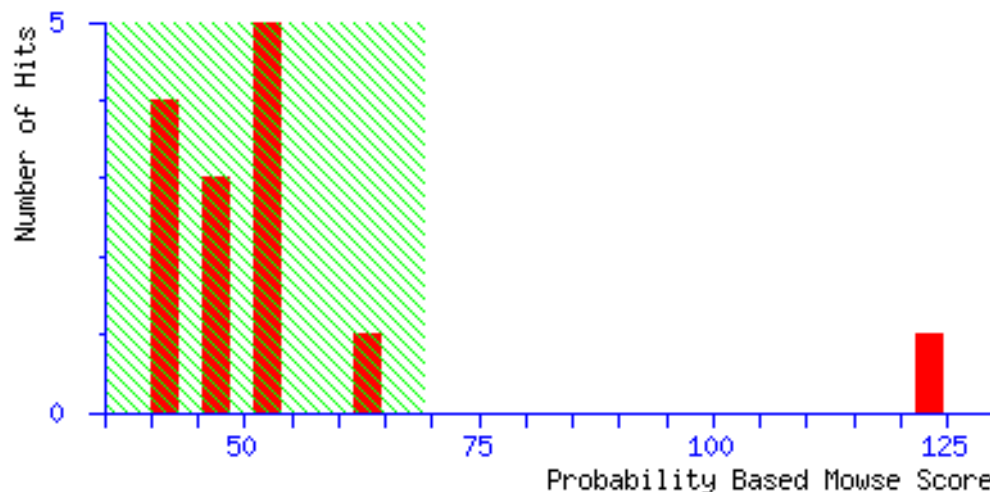
Format As	Concise Protein Summary	<a href="#">Help</a>	
Significance threshold $p <$	0.05	Max. number of hits	AUTO
Re-Search All	Search Unmatched		

- [RPOA STRP2](#)    **Mass:** 34186    **Score:** 122    **Expect:** 3.1e-07    **Queries matched:** 12  
DNA-directed RNA polymerase subunit alpha OS=Streptococcus pneumoniae serotype 2

# Peptidové mapování – vliv parametrů

## Probability Based Mowse Score

Protein score is  $-10 \cdot \log(P)$ , where  $P$  is the probability that the observed match is a random event.  
Protein scores greater than 69 are significant ( $p < 0.05$ ).



Database	SwissProt
Taxonomy	All taxa
Missed Cleavages	1
Fixed Mods	0
Variable Mods	0
Peptide tol	1000

## Concise Protein Summary Report

Format As	Concise Protein Summary	<a href="#">Help</a>	
Significance threshold $p <$	0.05	Max. number of hits	AUTO
Re-Search All	Search Unmatched		

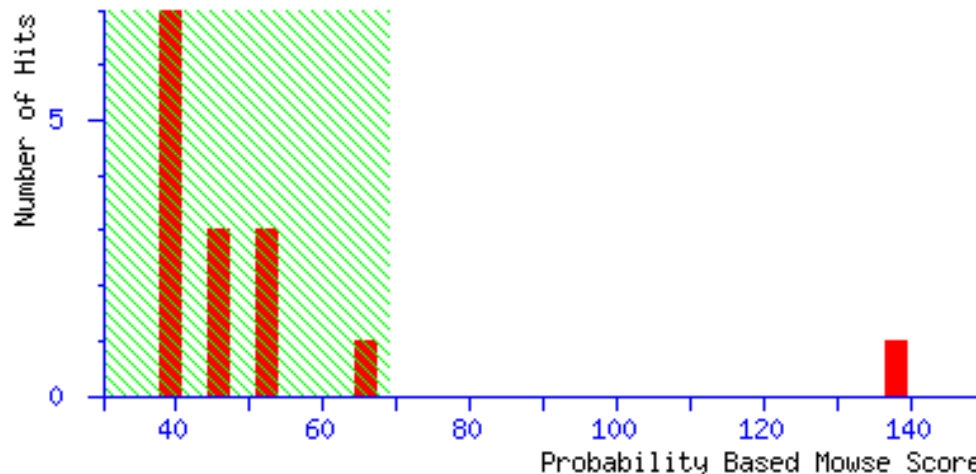
1. [RPOA\\_STRP2](#) **Mass:** 34186 **Score:** 123 **Expect:** 2.5e-07 **Queries matched:** 15  
DNA-directed RNA polymerase subunit alpha OS=Streptococcus pneumoniae serotype 2

# Peptidové mapování – vliv parametrů

## Probability Based Mowse Score

Protein score is  $-10 \cdot \log(P)$ , where  $P$  is the probability that the observed match is a random event.

Protein scores greater than 69 are significant ( $p < 0.05$ ).



Database	SwissProt
Taxonomy	All taxa
Missed Cleavages	1
Fixed Mods	C-Cam
Variable Mods	0
Peptide tol	1000

## Concise Protein Summary Report

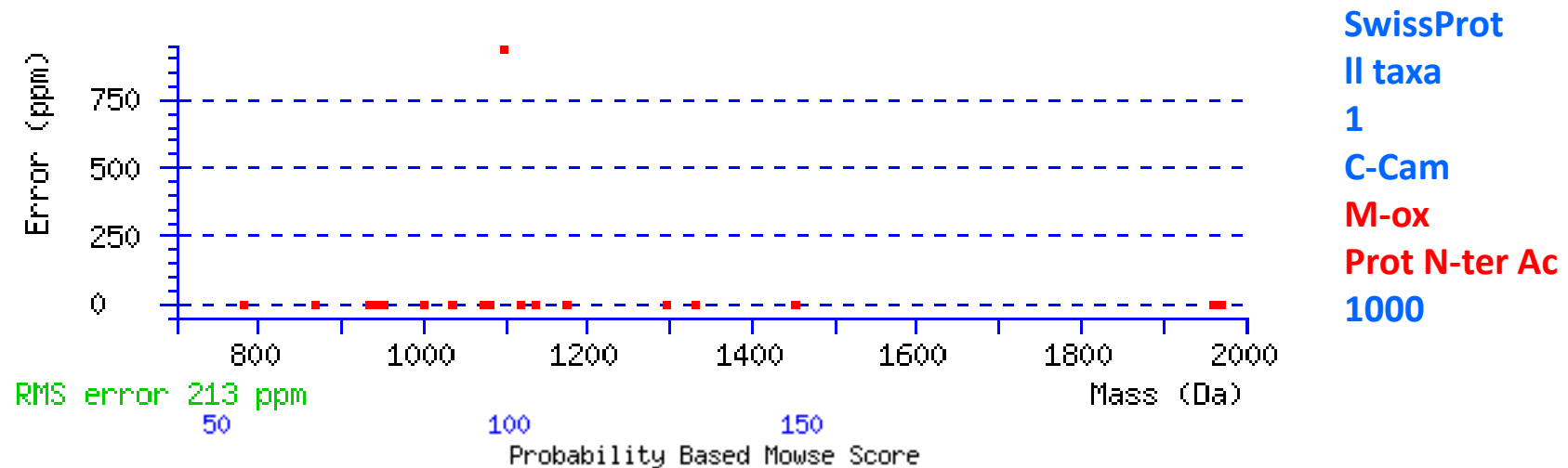
Format As	Concise Protein Summary	<a href="#">Help</a>	
Significance threshold $p <$	0.05	Max. number of hits	AUTO
Re-Search All	Search Unmatched		

- [RPOA STRP2](#) **Mass:** 34243 **Score:** 138 **Expect:** 7.9e-09 **Queries matched:** 16  
DNA-directed RNA polymerase subunit alpha OS=Streptococcus pneumoniae serotype 2

# Peptidové mapování – vliv parametrů

## Probability Based Mowse Score

Protein score is  $-10 \cdot \log(P)$ , where  $P$  is the probability that the observed match is a random event.  
Protein scores greater than 69 are significant ( $p < 0.05$ ).



## Concise Protein Summary Report

Format As  [Help](#)

Significance threshold  $p <$   Max. number of hits

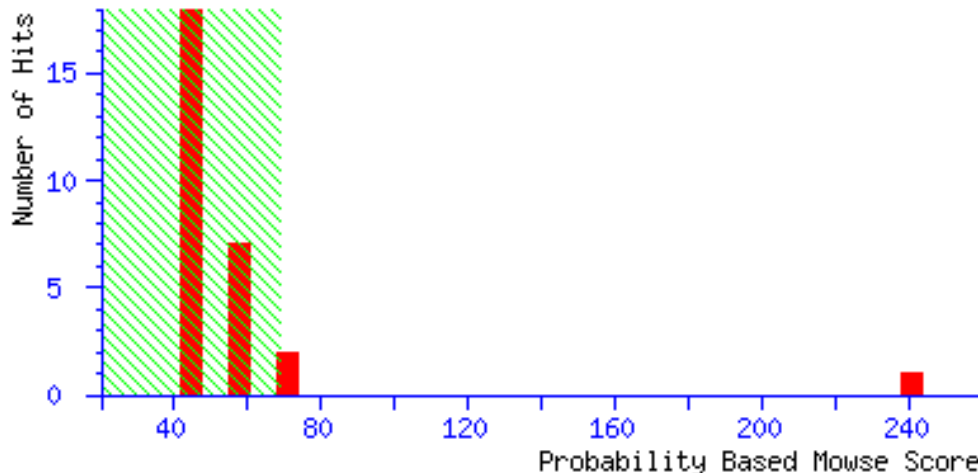
- [RPOA STRP2](#) **Mass:** 34243 **Score:** **156** **Expect:** 1.2e-10 **Queries matched:** **19**  
DNA-directed RNA polymerase subunit alpha OS=Streptococcus pneumoniae serotype 2

# Peptidové mapování – vliv parametrů

## Probability Based Mowse Score

Protein score is  $-10 \cdot \log(P)$ , where  $P$  is the probability that the observed match is a random event.

Protein scores greater than 69 are significant ( $p < 0.05$ ).



Database	SwissProt
Taxonomy	All taxa
Missed Cleavages	1
Fixed Mods	C-Cam
Variable Mods	M-ox
	Prot N-ter Ac
Peptide tol	100

## Concise Protein Summary Report

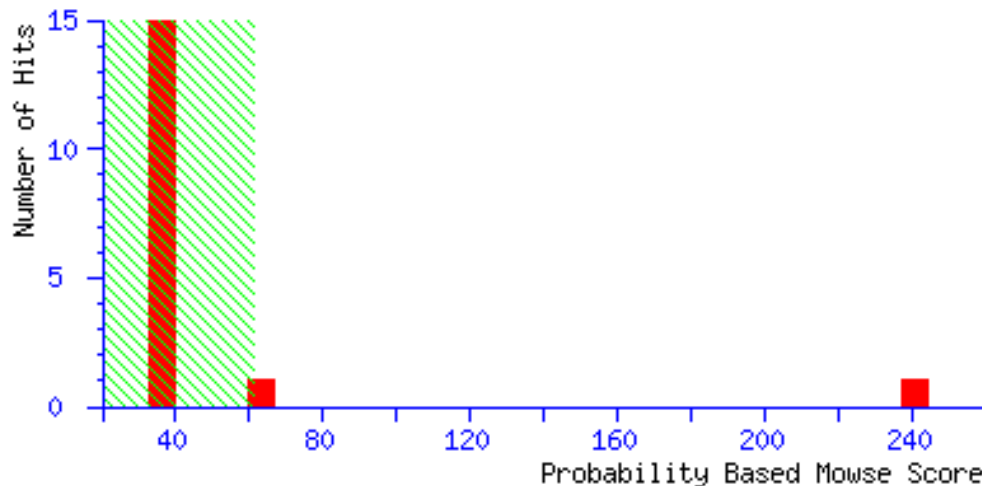
Format As	Concise Protein Summary	<a href="#">Help</a>	
Significance threshold $p <$	0.05	Max. number of hits	AUTO
Re-Search All	Search Unmatched		

1. [RPOA STRP2](#) **Mass:** 34243 **Score:** 241 **Expect:** 3.9e-19 **Queries matched:** 18  
DNA-directed RNA polymerase subunit alpha OS=Streptococcus pneumoniae serotype 2

# Peptidové mapování – vliv parametrů

## Probability Based Mowse Score

Protein score is  $-10 \cdot \log(P)$ , where P is the probability that the observed match is a random event.  
Protein scores greater than 61 are significant ( $p < 0.05$ ).



Database	SwissProt
Taxonomy	Firmicutes
Missed Cleavages	1
Fixed Mods	C-Cam
Variable Mods	M-ox
	Prot N-ter Ac
Peptide tol	100

## Concise Protein Summary Report

Format As	Concise Protein Summary	<a href="#">Help</a>	
Significance threshold p<	0.05	Max. number of hits	AUTO
Re-Search All	Search Unmatched		

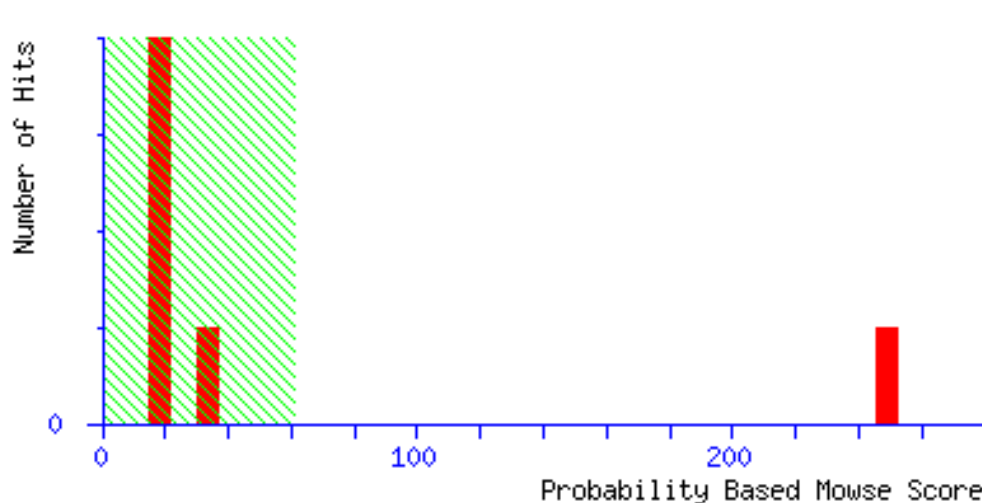
- [RPOA STRP2](#) **Mass:** 34243 **Score:** 241 **Expect:** 5.1e-20 **Queries matched:** 18  
DNA-directed RNA polymerase subunit alpha OS=Streptococcus pneumoniae serotype 2



# Peptidové mapování – vliv parametrů

## Probability Based Mowse Score

Protein score is  $-10 \cdot \log(P)$ , where P is the probability that the observed match is a random event.  
Protein scores greater than 61 are significant ( $p < 0.05$ ).



Database	SwissProt
Taxonomy	Firmicutes
Missed Cleavages	1
Fixed Mods	C-Cam
Variable Mods	M-ox
	Prot N-ter Ac
Peptide tol	10

## Concise Protein Summary Report

Format As  [Help](#)

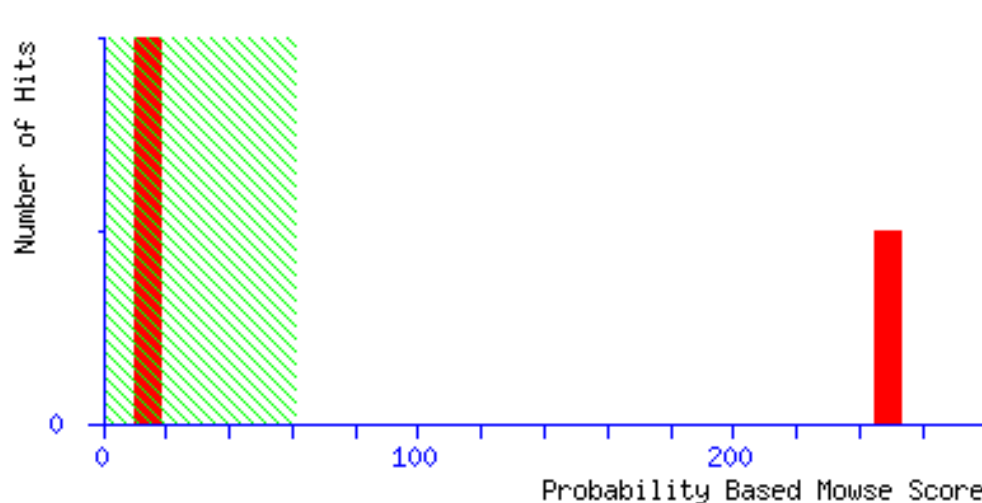
Significance threshold  $p <$   Max. number of hits

- [RPOA STRP2](#) **Mass:** 34243 **Score:** **249** **Expect:** 8e-21 **Queries matched:** 18  
DNA-directed RNA polymerase subunit alpha OS=Streptococcus pneumoniae serotype 2

# Peptidové mapování – vliv parametrů

## Probability Based Mowse Score

Protein score is  $-10 \cdot \log(P)$ , where  $P$  is the probability that the observed match is a random event.  
Protein scores greater than 61 are significant ( $p < 0.05$ ).



Database	SwissProt
Taxonomy	Firmicutes
Missed Cleavages	1
Fixed Mods	C-Cam
Variable Mods	M-ox
Peptide tol	Prot N-ter Ac <b>1</b>

## Concise Protein Summary Report

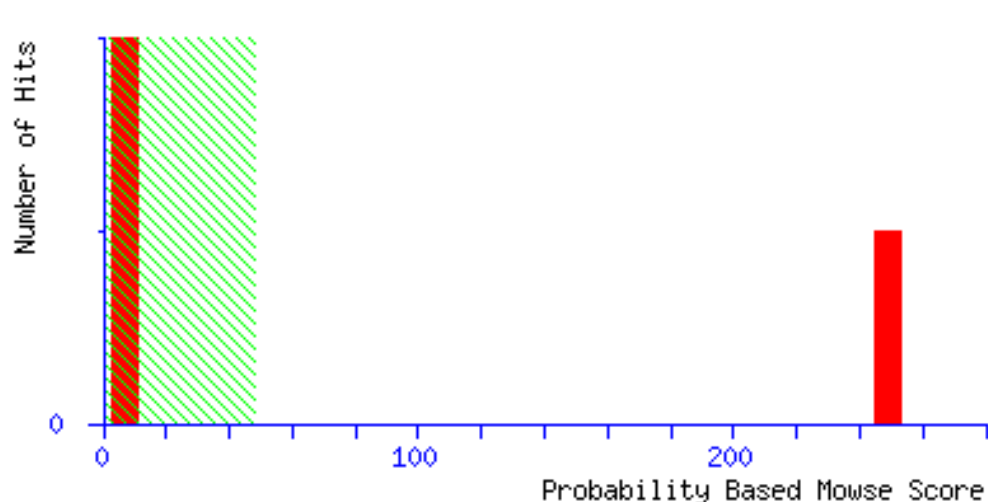
Format As	Concise Protein Summary	<a href="#">Help</a>	
Significance threshold $p <$	0.05	Max. number of hits	AUTO
Re-Search All	Search Unmatched		

- [RPOA STRP2](#) **Mass:** 34243 **Score:** **249** **Expect:**  $8e-21$  **Queries matched:** 18  
DNA-directed RNA polymerase subunit alpha OS=Streptococcus pneumoniae serotype 2

# Peptidové mapování – vliv parametrů

## Probability Based Mowse Score

Protein score is  $-10 \cdot \log(P)$ , where  $P$  is the probability that the observed match is a random event. Protein scores greater than 48 are significant ( $p < 0.05$ ).



Database	SwissProt
Taxonomy	<i>S. pneumoniae</i>
Missed Cleavages	1
Fixed Mods	C-Cam
Variable Mods	M-ox
Peptide tol	Prot N-ter Ac
	1

## Concise Protein Summary Report

Format As	Concise Protein Summary	<a href="#">Help</a>	
Significance threshold $p <$	0.05	Max. number of hits	AUTO
Re-Search All	Search Unmatched		

- [RPOA STRP2](#) **Mass:** 34243 **Score:** 249 **Expect:** 4.5e-22 **Queries matched:** 18  
DNA-directed RNA polymerase subunit alpha OS=Streptococcus pneumoniae serotype 2

# Peptidové mapování – vliv parametrů

## Protein View

Match to: **RPOA\_STRP2** Score: 249 Expect: 4.5e-22  
DNA-directed RNA polymerase subunit alpha 05=Streptococcus pneumoniae serotype 2 (strain D39 / NCTC 7466) GN=rpoA PE=3 SV=1

Nominal mass ( $M_r$ ): 34243; Calculated pI value: 4.64  
NCBI BLAST search of [RPOA\\_STRP2](#) against nr  
Unformatted [sequence string](#) for pasting into other applications

Taxonomy: [Streptococcus pneumoniae](#)

Fixed modifications: Carbamidomethyl (C)  
Variable modifications: Acetyl (Protein N-term), Oxidation (M)  
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P  
Number of mass values searched: 27  
Number of mass values matched: 18  
Sequence Coverage: 53%

Matched peptides shown in **Bold Red**

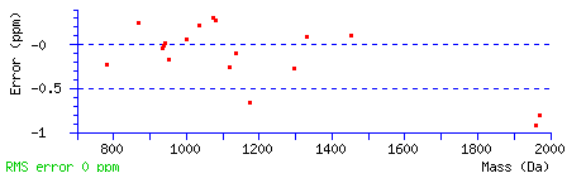
```
1 MIEFEKPNIT KIDENKDYGK FVIEPLERGY GTTLGNSLR VLLASLPGAA
51 VTSINIDGVL HEFDTVPQVGR EDVMQIILNI KGIAVKSYVE DEKIIELDVE
101 GPAEVTAGDI LTDSDEIIVN PDHYLFTIGE GSSLKATMTV NSGRGYVPAD
151 ENKKDNAPVG TLAVDSIYTP VTKVNYQVEP ARVGSNDGFD KLTLEILTNG
201 TIIPEDALGL SARILTEHLD LFTNLTEIAK STEVMREADT ESDDRILDRT
251 IEELDLSVRS YNCLKRAGIN TVHDLTEKSE AEMMKVRNLG RKSLEEVLK
301 LIDLGLGLKD K
```

Show predicted peptides also

Sort Peptides By  Residue Number  Increasing Mass  Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Sequence
260 - 265	784.3656	783.3583	783.3585	-0	0	R.SYNCLK.R
87 - 93	869.3889	868.3816	868.3814	0	0	K.SYVEDEK.I
136 - 144	936.4567	935.4494	935.4495	-0	0	K.ATMTVNSGR.G
183 - 191	938.4214	937.4141	937.4141	-0	0	R.VGSNDGFDK.L
301 - 309	941.6030	940.5957	940.5957	0	0	K.LIDLGLGLK.D
136 - 144	952.4515	951.4442	951.4444	-0	0	K.ATMTVNSGR.G Oxidation (M)
21 - 28	1002.5619	1001.5546	1001.5546	0	0	K.FVIEPLER.G
237 - 245	1037.4020	1036.3947	1036.3945	0	0	K.EADTESDDR.I
174 - 182	1075.5534	1074.5461	1074.5458	0	0	K.VNYQVEPAR.V
12 - 20	1081.5163	1080.5090	1080.5087	0	1	K.IDENKDYGK.F
145 - 154	1120.5630	1119.5557	1119.5560	-0	1	R.GYVPADENKK.D
29 - 39	1138.5850	1137.5777	1137.5778	-0	0	R.GYGTTLGNSLR.R
250 - 259	1174.6306	1173.6233	1173.6241	-1	0	R.TIEELDLSVR.S
267 - 278	1297.6743	1296.6670	1296.6674	-0	0	R.AGINTVHDLTEK.S
71 - 81	1331.7240	1330.7167	1330.7166	0	0	R.EDVMQIILNIK.G Oxidation (M)
266 - 278	1453.7759	1452.7686	1452.7685	0	1	K.RAGINTVHDLTEK.S
155 - 173	1961.0208	1960.0135	1960.0153	-1	0	K.DNAPVGTTLAVDSIYTPVK.V
214 - 230	1971.0781	1970.0708	1970.0724	-1	0	R.ILTEHLDLFTNLTEIAK.S

No match to: 842.5097, 870.5407, 1016.5771, 1024.5436, 1089.5689, 1097.5338, 1128.5281, 1188.6459, 1308.6543



Unmatched masses:

- 842.50 – Trypsin
- 870.54 – very tiny peak
- 1016.57 - methylation of 1002.56
- 1024.54 – sodium adduct of 1002.56
- 1089.57 - methylation of 1075.55
- 1097.53 - sodium adduct of 1075.55
- 1128.53 – very tiny peak
- 1188.65 – methylation of 1174.63
- 1308.65 - very tiny peak

# Peptidové mapování – prezentace výsledků

Nutno uvést

- instrument a jeho nastavení, typ kalibrace (interní, externí, post-kalibrace)
- program/algorithmus použitý pro označení píků, dekonvoluci, vyhlazení/korekce
- program a všechny parametry použité pro identifikaci
- Skóre (+limit), pokrytí sekvence, počet použitých/přiřazených píků

MALDI FT-ICR (Apollo II Dual source Apex Qe 9.4T)

externí kalibrace na Bruker Peptide Mix

DataAnalysis 4.0, unprocessed, SNAP v 2.0

MASCOT v 2.2 – Swiss-Prot (verze/datum), *S. pneumoniae*, Trypsin, m.c. - 1, f.m. - C-Cam,  
v.m. – Prot N-term acetylated + M-ox, p.t. 1ppm

Skóre 249 (48 limit), pokrytí sekvence 53%, 23/26 píků přiřazeno

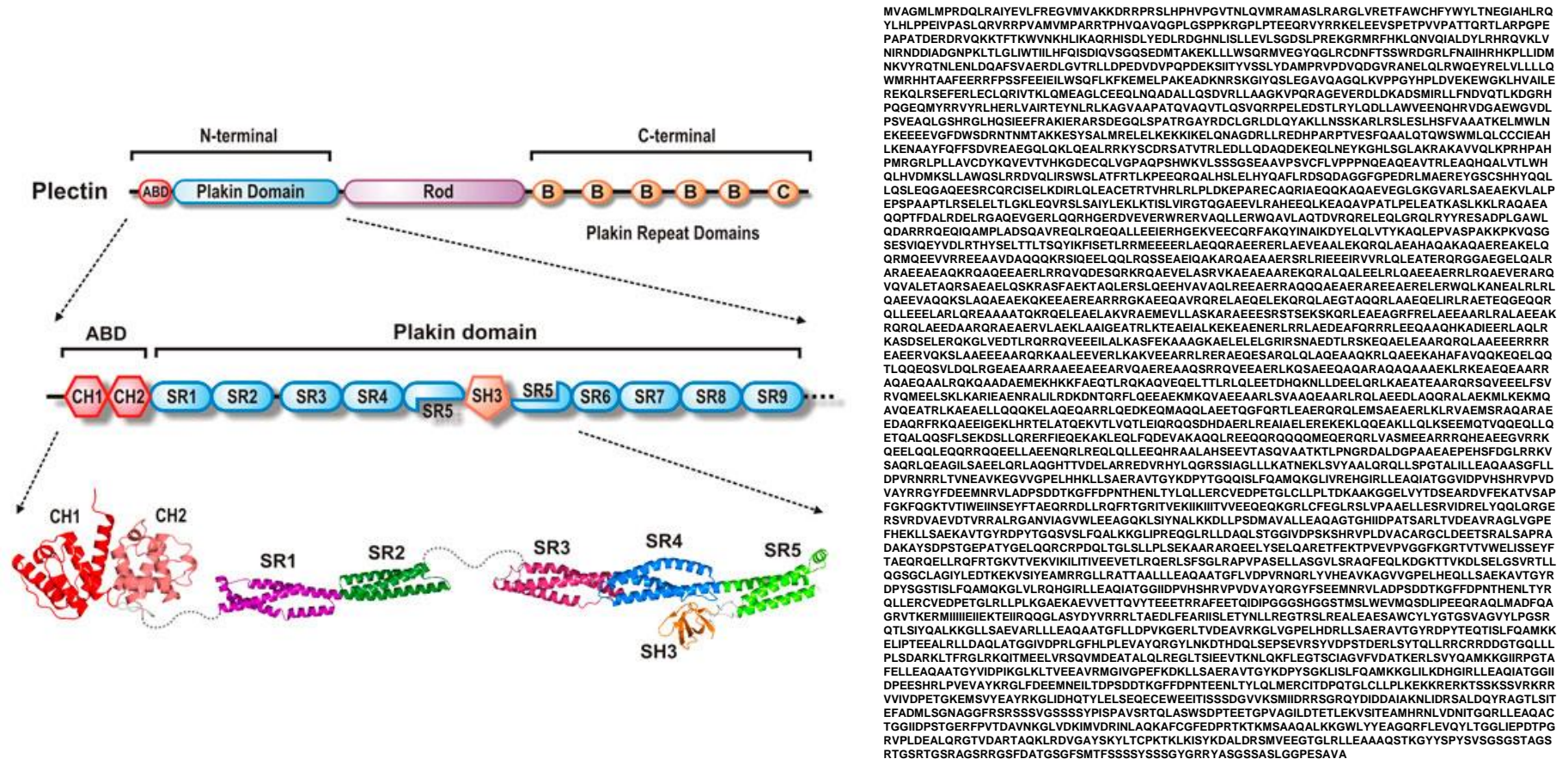
<https://hupo.org/Guidelines>

[http://www.mcponline.org/sites/default/files/assets/pdfs/Guidelines/Proteomic\\_Checklist.pdf](http://www.mcponline.org/sites/default/files/assets/pdfs/Guidelines/Proteomic_Checklist.pdf)

# Výhody vysokého rozlišení

Lepší a rychlejší identifikace proteinů – peptidové mapování

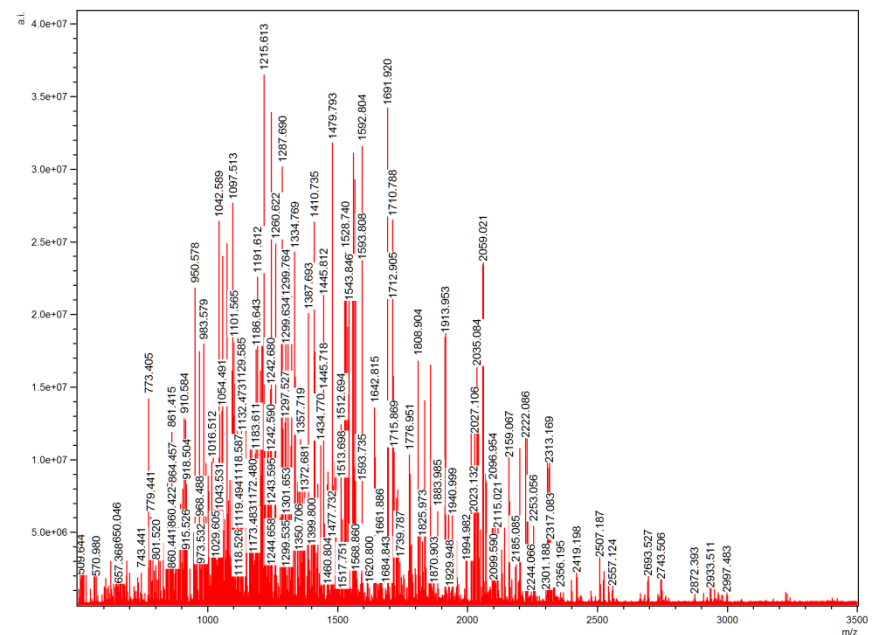
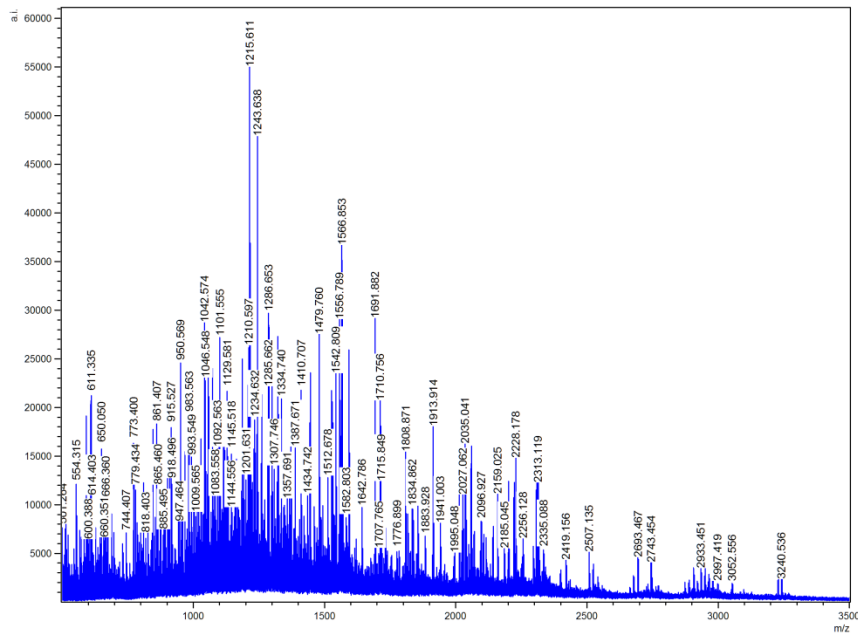
*Plectin* – obrovský protein z cytoskeletu (530kDa, 4900 AK, několik sestřihových variant)



# Výhody vysokého rozlišení

## Lepší a rychlejší identifikace proteinů – peptidové mapování

### MALDI-TOF vs MALDI-FT-ICR

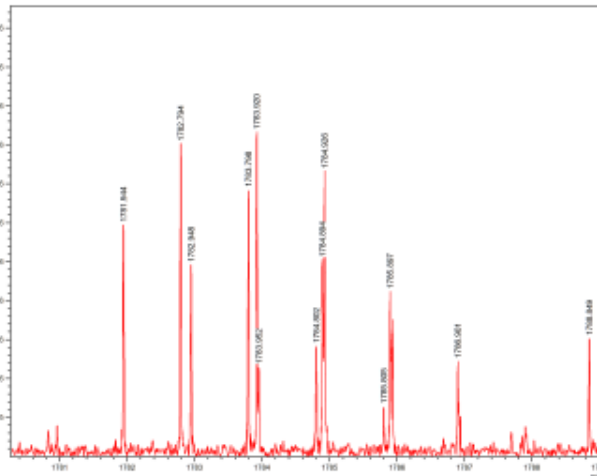
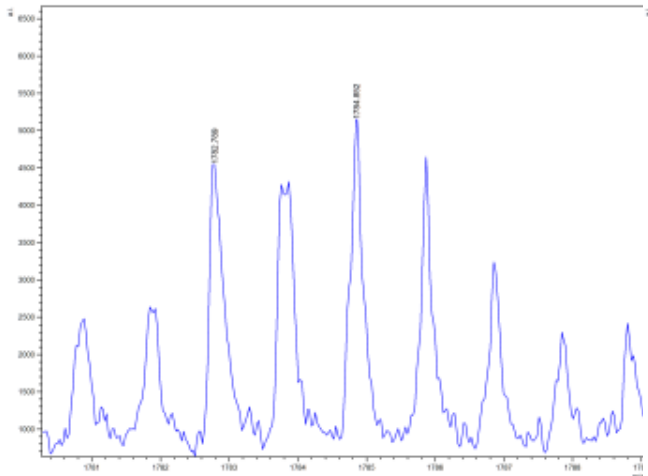


# Výhody vysokého rozlišení

TOF

ICR

Lenší a rychlejší identifikace proteinů - peptidové mapování

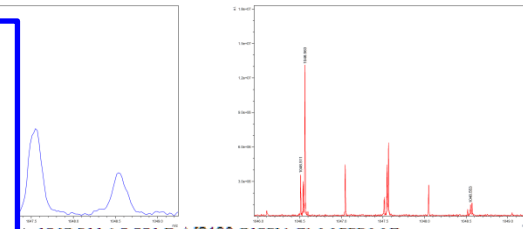


- ..... **h<sup>i</sup>** 1781.942 (-0.001 Da) [3244-3259] r.DPYTGQSVSLFQALKK.g
- ..... **h<sup>i</sup>** 1782.792 (-0.001 Da) [4610-4627] k.GYYSPYSVSGSGTAGSR.t
- ..... **h<sup>i</sup>** 1783.919 (0.0 Da) [2744-2761] r.AALAHSEEVTASQVAATK.t
- ..... **h<sup>i</sup>** 1784.892 (-0.004 Da) [1367-1382] r.QEQIQAMPLADSQAVR.e
- ..... **h<sup>i</sup>** 1788.848 (0.0 Da) [3575-3590] r.DPYSGSTISLFAQAMQK.g [1xOxidation]

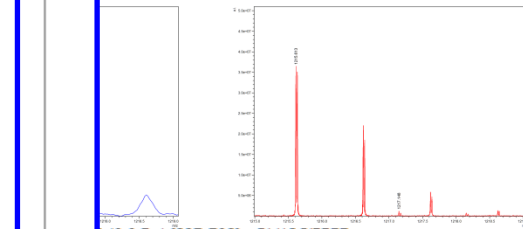
- ..... **h<sup>i</sup>** 1046.511 (-0.001 Da) [2130-2139] k.SLAEEEEAAAR.q
- ..... **h<sup>i</sup>** 1046.541 (-0.0 Da) [1534-1541] r.MQEEVVRR.e
- ..... **h<sup>i</sup>** 1046.562 (-0.0 Da) [3620-3628] r.VPVDVAYQR.g

```

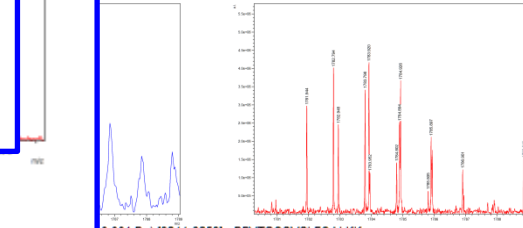
3901 MLSEETSEI REAEAEERW CYLITQVQ SVLPEEQQI LSTQAKRK
3951 LLAARVALL LEAQATQFL LEPVRSLLT VEGAVKGLV GPELDELLE
3961 LEEAVTQDR PYTQITSLR QMRRELPTT EEALLLIQA LATGVQVPR
3981 LKPLPLDVA VQKILSDIT IQGAKPKVQ KAVYKPTTE PLSTGELSE
4001 CRDDTQDL LPLLDARSL TFRLEKQIT MEELVKQVM DETALQLE
4051 GLIIEPTV IIGQFLDPT CIAGWQNT KERLVQAN RQGIKRPFL
4101 RELLQADAT QVYVITQIL VLVEAVRW GQVREKRW LSAARATIS
4151 YQVYSQELI SLQAWQGLI ILICRDLAL EAKIATGQII DPESHLPY
4201 KVAIQRGLD RENNELTDP SEQTRPFPF RTENALYIQ LMRKLTQD
4251 TQGLPLQLE IQKQKQVQV SPQKQVAV VQVETQRM VEGVQGLI
4301 DQITLLESE QEQREKITT SSSQVYVM IIRBNSQVQ DIDALAKIL
4351 IIRALAZDR AGTLSTERA DMLKRAQR RENNVSQVS SVYFSPAVS
4401 LKPLSDQSF IELISFQAL IICELIENV TQVAKRWV DNTQGLLE
4451 AQACTGQII PSTRGKPTT DAKSSQLCE LQVCLDAQ KAPQKQDRR
4501 TIKRNSAQA LQKWLATRA QGFLQVQIL TQGLIEPTT QVPLREALQ
4551 RQVTCARQI RLRGKQAKW YLQVQVCEI IIRKALERS IQVETQILC
4601 LKAACTQIC VYVYVQVQ GQVQKQVCS RQKQKQRA QEQATQSP
4651 INTFSSEIV SSQVQKVAS QSSASLQSE SAVA
    
```



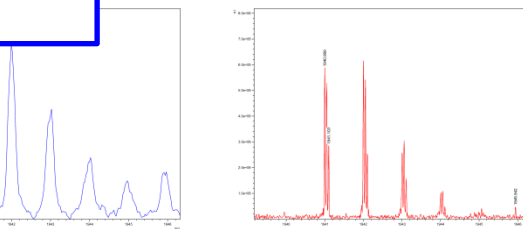
- ..... **h<sup>i</sup>** 1046.511 (-0.001 Da) [2130-2139] k.SLAEEEEAAAR.q
- ..... **h<sup>i</sup>** 1046.541 (-0.0 Da) [1534-1541] r.MQEEVVRR.e
- ..... **h<sup>i</sup>** 1046.562 (-0.0 Da) [3620-3628] r.VPVDVAYQR.g



- ..... **h<sup>i</sup>** 697.706 (0.0 Da) [697-706] r.GLHQSEEFRR.a
- ..... **h<sup>i</sup>** 651.660 (-0.0 Da) [651-660] r.RPELEDSTLR.y



- ..... **h<sup>i</sup>** 1781.942 (-0.001 Da) [3244-3259] r.DPYTGQSVSLFQALKK.g
- ..... **h<sup>i</sup>** 1782.792 (-0.001 Da) [4610-4627] k.GYYSPYSVSGSGTAGSR.t
- ..... **h<sup>i</sup>** 1783.919 (0.0 Da) [2744-2761] r.AALAHSEEVTASQVAATK.t
- ..... **h<sup>i</sup>** 1784.892 (-0.004 Da) [1367-1382] r.QEQIQAMPLADSQAVR.e
- ..... **h<sup>i</sup>** 1788.848 (0.0 Da) [3575-3590] r.DPYSGSTISLFAQAMQK.g [1xOxidation]



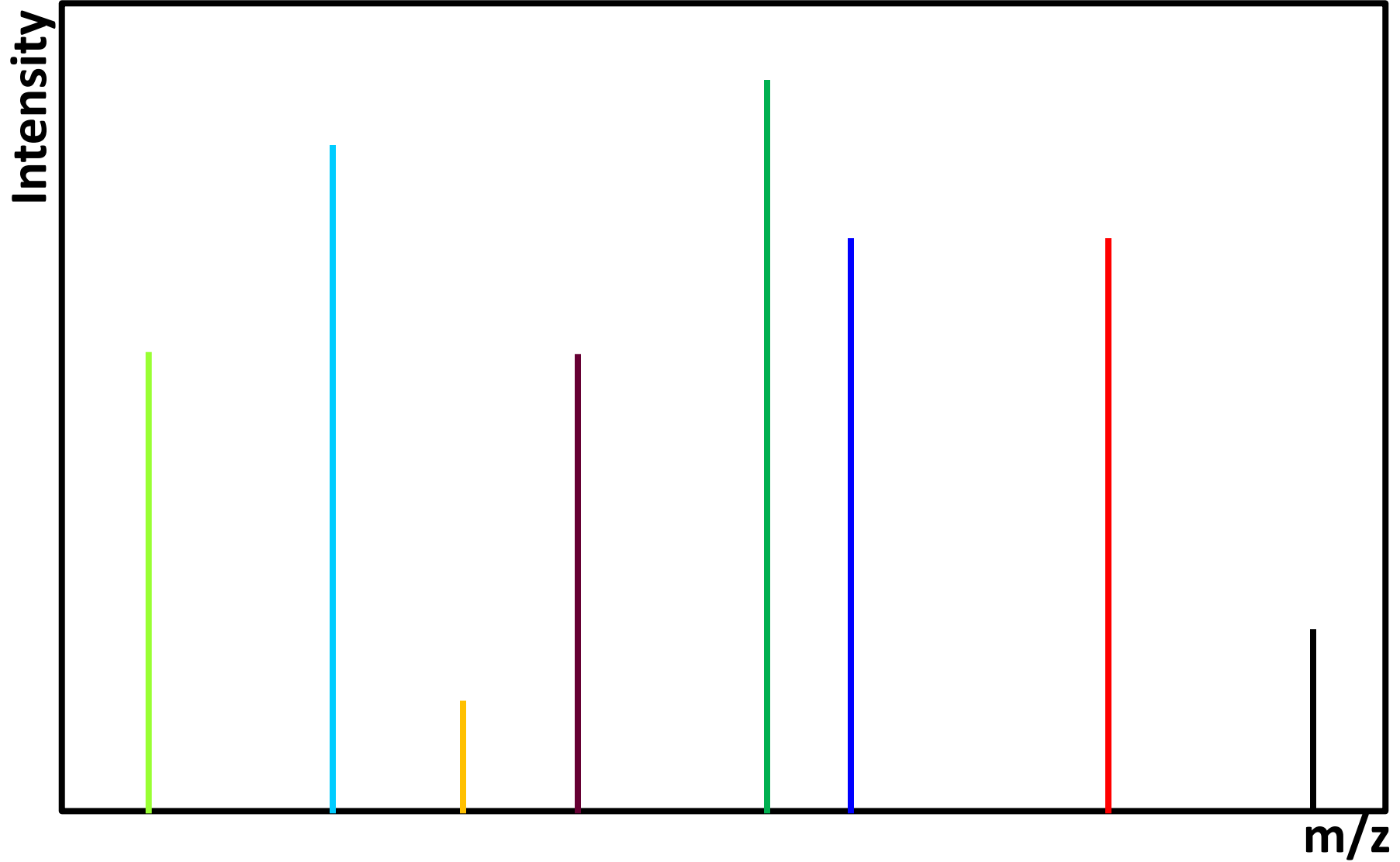
- ..... **h<sup>i</sup>** 1940.997 (-0.0 Da) [1366-1382] r.QEQIQAMPLADSQAVR.e
- ..... **h<sup>i</sup>** 1941.101 (-0.001 Da) [3424-3439] k.LITVVEETLQRER.I

Přiřazeno 384 peptidů  
Pokrytí 67%



# Identifikace proteinů ve směsi

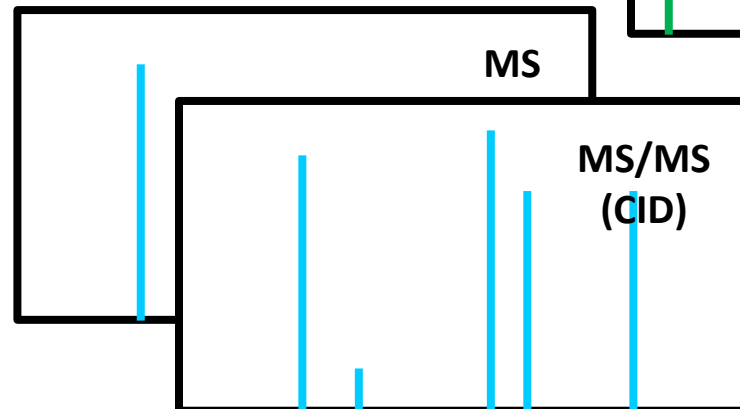
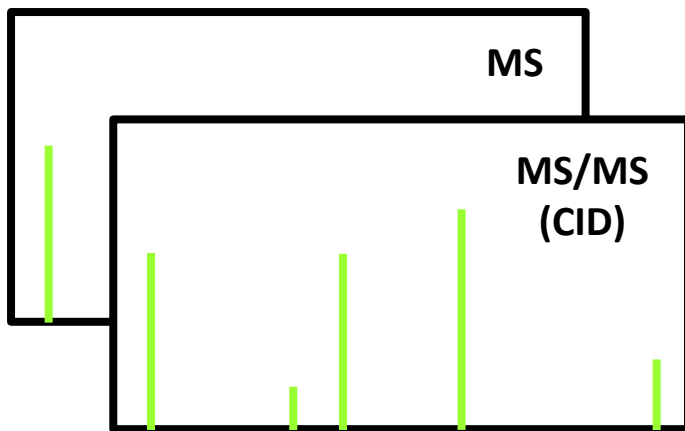
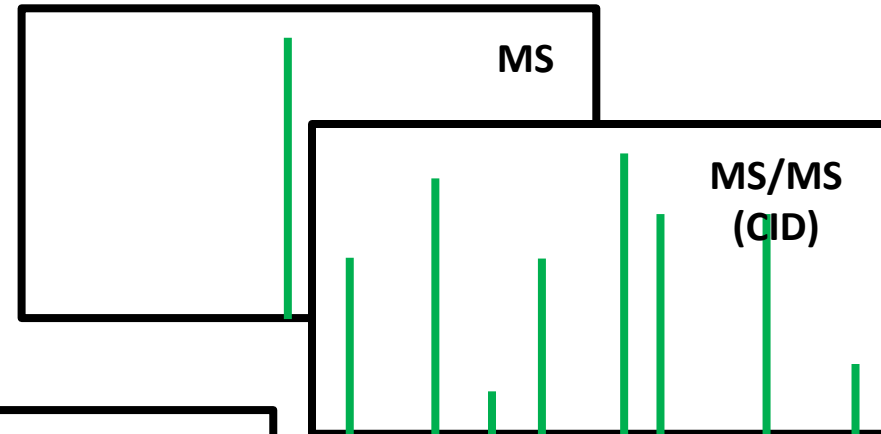
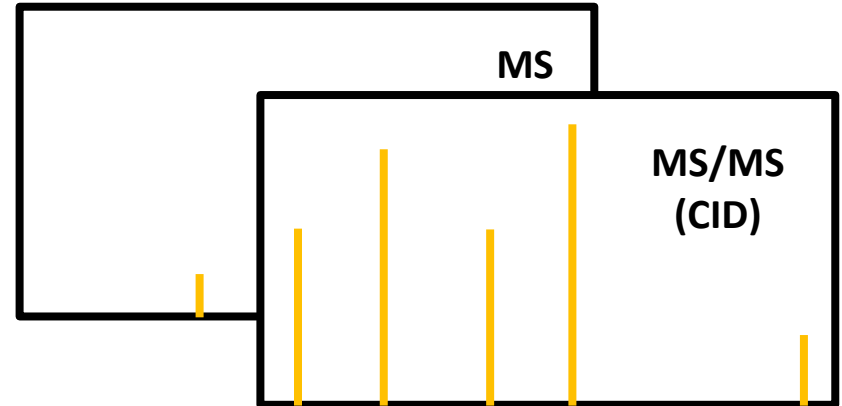
ADSMK EETSLINVCPWK VNMGGHDSIYTLR EEIYTLK MLIILLR  
TYSHEEDK EWQIDSLAEIR IQPLPMNVSA



# Identifikace proteinů ve směsi

## - separace a MS/MS

Separace – LC+schopnosti MS



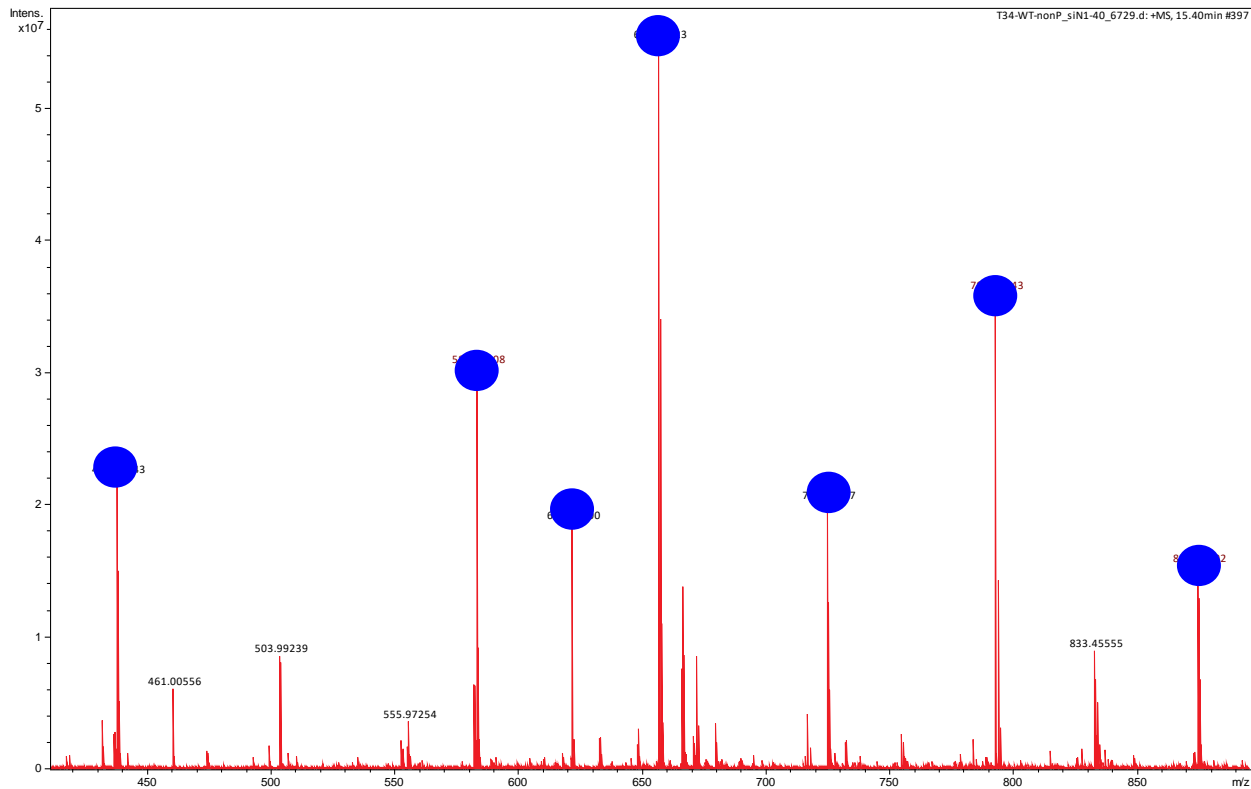
# Shotgun proteomics – Analýza proteinových směsí

Rychlost skenování, nastavení MS-MS/MS cyklu, dynamická  
exkluze, paralelní analýza a detekce

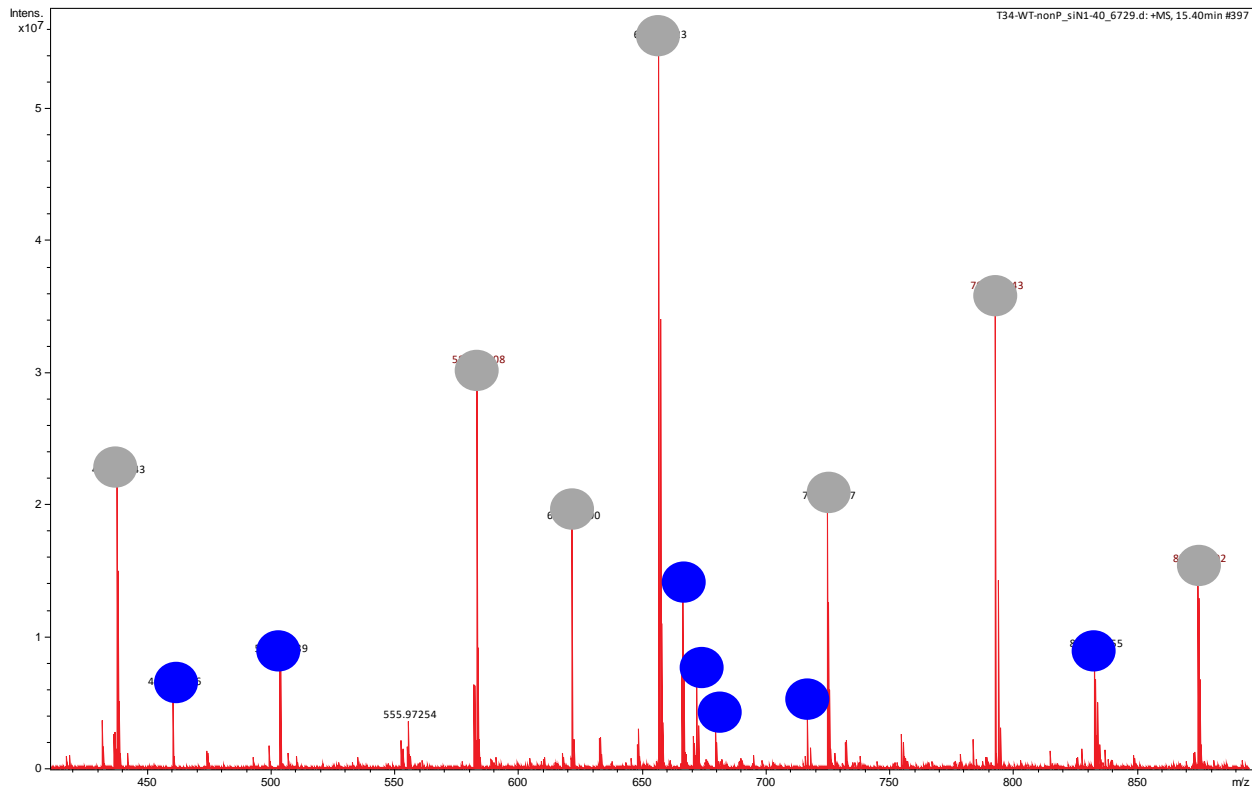
Jeden MS sken následovaný daným počtem MS/MS nebo variabilním počtem  
(s limitem na celý duty cycle).

Např. za 1 sec lze stihnout MS a 40MS/MS (skenovací rychlost 40Hz).  
Vybrány jsou aktuálně nejintenzivnější píky

# Shotgun proteomics – Analýza proteinových směsí

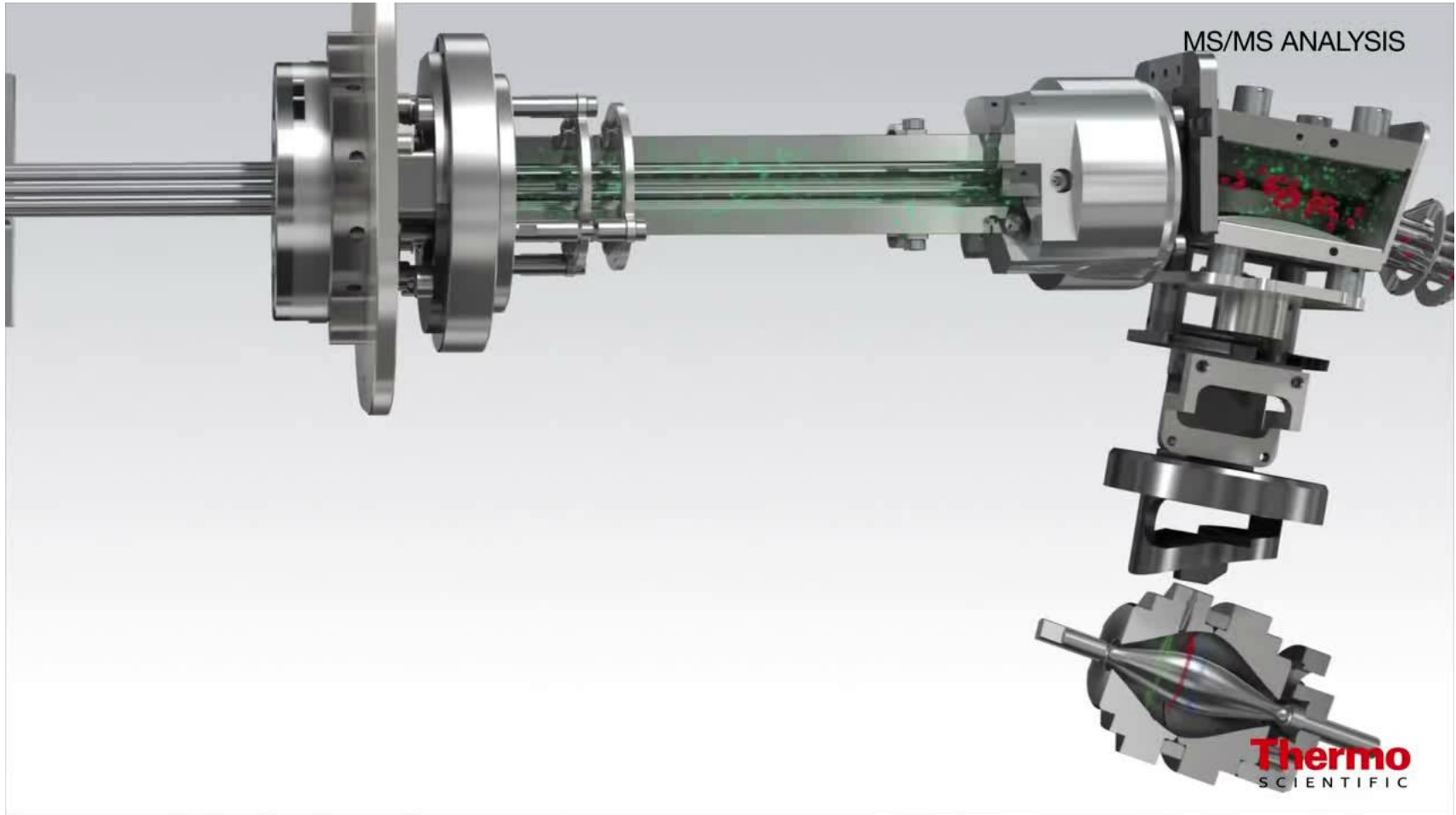


# Shotgun proteomics – Analýza proteinových směsí

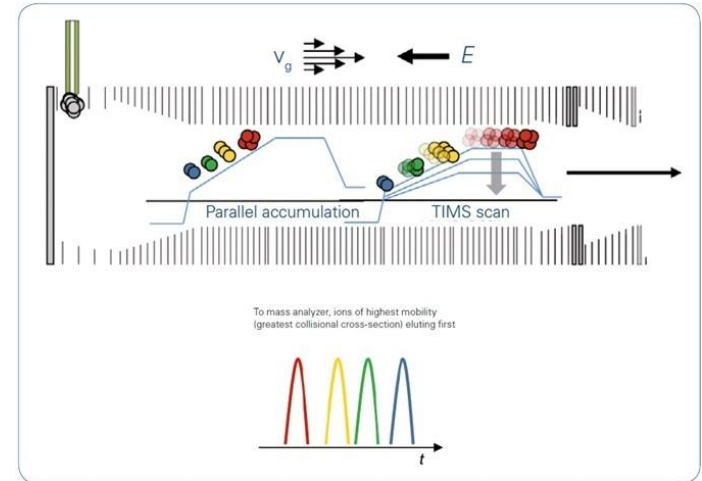
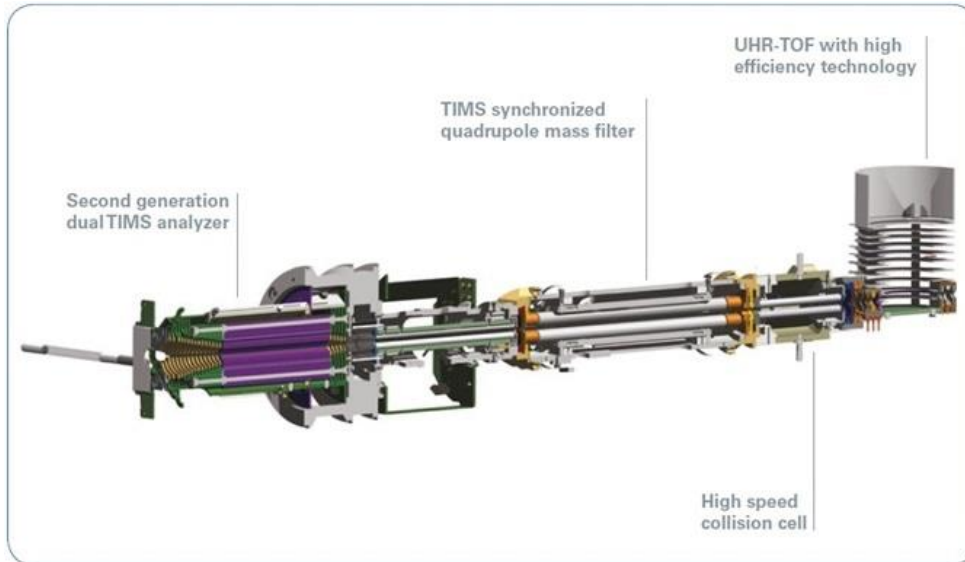


# Hybridní MS – výhoda současné analýzy na více místech.

Vysoce přesná M prekurzoru (intaktního peptidu) a méně rozlišená a přesná MS/MS spektra



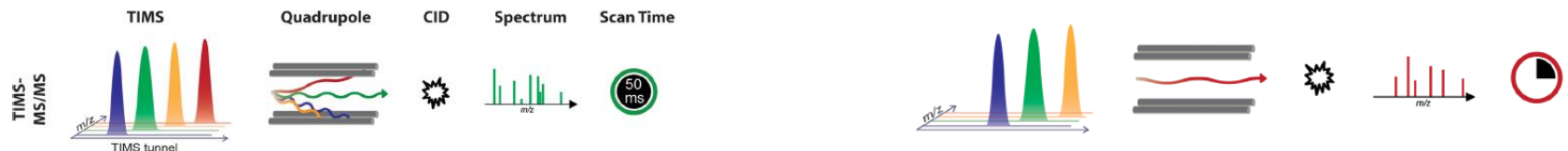
# TIMS + qTOF



**TIMS – trapped ion mobility mass spectrometry** – separace iontovou mobilitou „dle tvaru/velikosti“. Fokusace iontů = lepší S/N, větší citlivost.  
Postupné uvolňování

**Paralelní akumulace iontů/iontová mobilita a následná fragmentace – PASEF** – za 1sec = 10x TIMS separace, z každé až 12 MS/MS = 120Hz!

Další dimenze pro identifikaci –  $R_t$  z LC,  $m/z$  z qTOF a  $t_D$



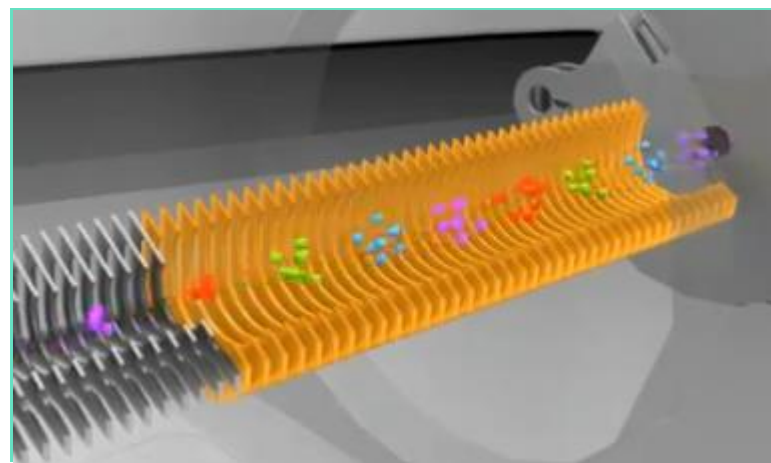
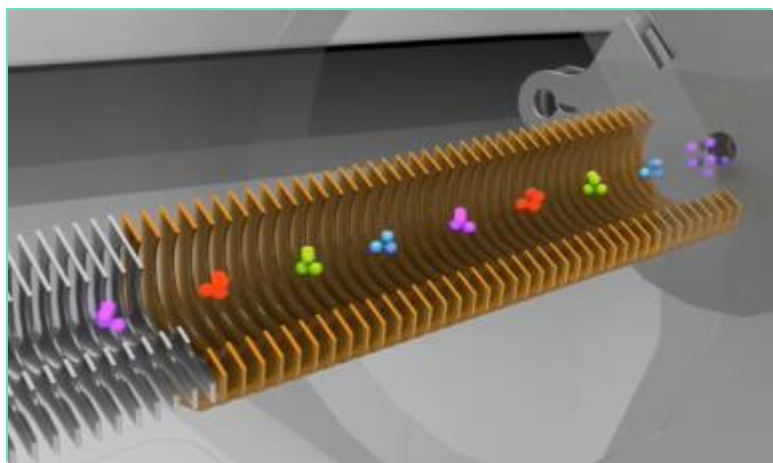
# Nové přístupy = MS<sup>e</sup>

## DIA – Data Independent Acquisition

MSe – cyklické přepínání mezi normálním MS skenem a MSMS skenem kde je fragmentováno vše najednou – potřebujeme vysokou přesnost, rozlišení a velmi dobrou chromatografii (UPLC). MSMS spektrům nerozumíme - data jsou dodatečně rozklíčována – fragmentové ionty přiřazeny prekurzorům.

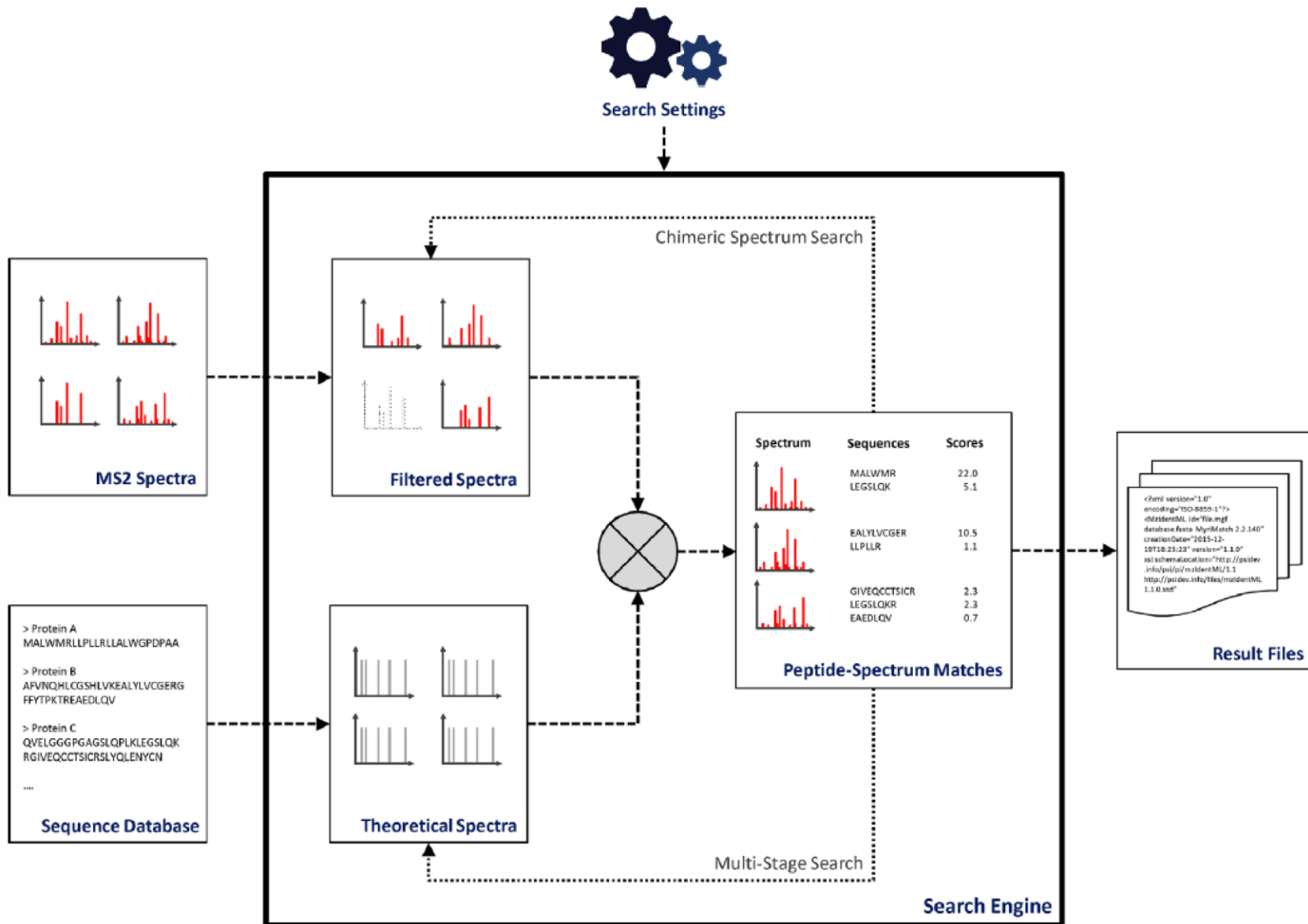
Rychlejší, větší dynamický rozsah, větší citlivost (odpadá ztráta při izolaci)

*Waters*





# Peptidové sekvenování / identifikace na základě MS/MS spekter



# Programy pro interpretaci MS/MS spekter

**MASCOT** (*Matrix Science*) – *error tolerant search*

[http://www.matrixscience.com/cgi/search\\_form.pl?FORMVER=2&SEARCH=MIS](http://www.matrixscience.com/cgi/search_form.pl?FORMVER=2&SEARCH=MIS)

**Peaks Studio** (*Bioinformatics Solutions*) – *de novo sequencing*

<http://www.bioinfor.com/peaks-studio/>

**!XTANDEM** – *open source*

<http://www.thegpm.org/tandem/>

**MaxQuant (Andromeda)** - *freely available*

<http://www.coxdocs.org/doku.php?id=:maxquant:start>

**MS-GF+** - *freely available*

<https://omics.pnl.gov/software/ms-gf>

**OMSSA** - *freely available*

**SEQUEST** – Scripps, ThermoFisher Scientific

**Protein Pilot** – ABI Sciex

**MassLynx** – Waters

**Free-access combined under GALAXY (ELIXIR)**

<https://usegalaxy.eu/>

<https://galaxy-dev.metacentrum.cz/galaxy/>

# MASCOT MS/MS Ions Search

## MASCOT – MS/MS

<b>Your name</b>	<input type="text"/>	<b>Email</b>	<input type="text"/>
<b>Search title</b>	<input type="text"/>		
<b>Database(s)</b>	contaminants (AA) NCBIProt (AA)	> <	Amino acid (AA) cRAP SwissProt Nucleic acid (NA) Environmental_EST Fungi_EST Human_EST Invertebrates_EST Mammals_EST Mus_EST
<b>Taxonomy</b>	..... Mammalia (mammals) ▾		
<b>Enzyme</b>	Trypsin ▾	<b>Allow up to</b>	1 ▾ missed cleavages
<b>Quantitation</b>	None ▾		
<b>Fixed modifications</b>	Carbamidomethyl (N-term)	> <	Acetyl (K) Acetyl (N-term) Amidated (C-term) Amidated (Protein C-term) Ammonia-loss (N-term C) Carbamidomethyl (C) Carbamyl (K) Carbamyl (N-term) Carboxymethyl (C) Cation:Na (C-term) Cation:Na (DE)
	<input type="checkbox"/> Display all modifications		
<b>Variable modifications</b>	Acetyl (Protein N-term) Oxidation (M)	> <	
<b>Peptide tol. ±</b>	10 <input type="text"/> ppm ▾	<b># <sup>13</sup>C</b>	0 ▾
<b>MS/MS tol. ±</b>	10 <input type="text"/> ppm ▾	<b>Monoisotopic</b>	<input checked="" type="radio"/> Average <input type="radio"/>
<b>Peptide charge</b>	2+ ▾		
<b>Data file</b>	Procházet... Soubor nevybrán.		
<b>Data format</b>	Mascot generic ▾		
<b>Instrument</b>	ESI-FTICR ▾		
<b>Decoy</b>	<input checked="" type="checkbox"/>		
<b>Precursor</b>	<input type="text"/> m/z		
<b>Error tolerant</b>	<input type="checkbox"/>		
<b>Report top</b>	AUTO ▾ hits		
<b>Start Search ...</b>		<b>Reset Form</b>	

# MASCOT – MS/MS

## MATRIX SCIENCE MASCOT Search Results

**User** : guest  
**E-mail** : guest@localhost  
**MS data file** : transporters-tryp.mgf  
**Database** : SwissProt\_ID 2017\_09 (555,594 sequences; 199,016,217 residues)  
**Taxonomy** : Bacteria (Eubacteria) (333,223 sequences)  
**Timestamp** : 23 Oct 2018 at 20:47:55 GMT

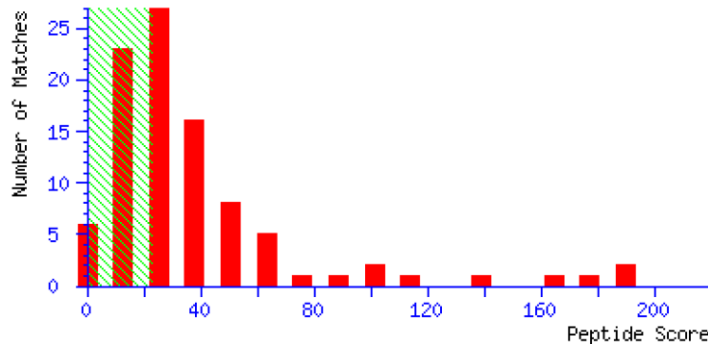
Re-search  All  Non-significant  Unassigned [\[help\]](#)

Export As

Not what you expected? Try [the peptide summary](#).

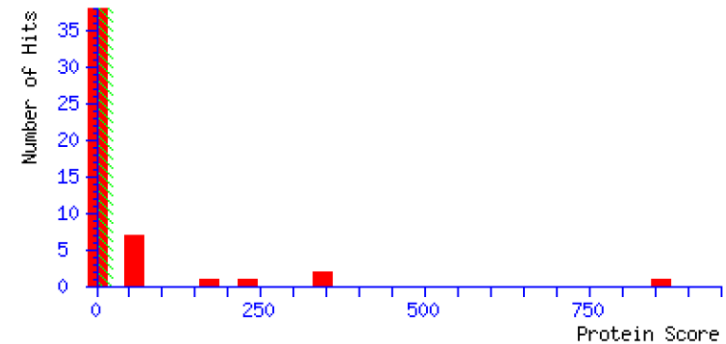
### ► Search parameters

### ▼ Score distribution



**Peptide score distribution.** Ions score is  $-10\log(P)$ , where  $P$  is the probability that the observed match is a random event.

There are **58** peptide matches above identity threshold and **65** matches above homology threshold for **799** queries. Histogram score range is (0, 190). On average, individual ions scores **> 23** (beyond green shading) indicate **identity or extensive homology** ( $p < 0.05$ ).



**[Deprecated]** Protein score distribution. Score distribution for family members in the first 50 proteins. Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein families.

	SwissProt_ID	<a href="#">Decoy</a>	False discovery rate
Peptide matches above identity threshold	58	1	1.72 %
Peptide matches above homology or identity threshold	65	1	1.54 %

# MASCOT – MS/MS

## Protein Family Summary

Filter	Significance threshold p<	<input type="text" value="0.05"/>	Max. number of families	<input type="text" value="AUTO"/>	<a href="#">[help]</a>
	Ions score or expect cut-off	<input type="text" value="0"/>	Dendrograms cut at	<input type="text" value="0"/>	
	Show Percolator scores	<input type="checkbox"/>			
	Preferred taxonomy	<input type="text" value="All entries"/>			

## ▶ Decoy search summary (reversed protein sequences)

Proteins (16)

[Report Builder](#)

[Unassigned \(711\)](#)

## Protein families 1–10 (out of 16)

per page 1 [2](#) [Next](#)

▶ 1	<b>SLYD_ECOLI</b>	860	FKBP-type peptidyl-prolyl cis-trans isomerase SlyD OS=Escherichia coli (strain K12) GN=slyD PE=1 SV=1
▶ 2	<b>CYOA_ECOLI</b>	368	Cytochrome bo(3) ubiquinol oxidase subunit 2 OS=Escherichia coli (strain K12) GN=cyoA PE=1 SV=1
▶ 3	<b>ARNA_ECOBW</b>	351	Bifunctional polymyxin resistance protein ArnA OS=Escherichia coli (strain K12 / MC4100 / BW2952) GN=arna PE=3 SV=1
▶ 4	<b>CYOB_ECOLI</b>	219	Cytochrome bo(3) ubiquinol oxidase subunit 1 OS=Escherichia coli (strain K12) GN=cyoB PE=1 SV=1
▶ 5	<b>AHPF_ECOLI</b>	181	Alkyl hydroperoxide reductase subunit F OS=Escherichia coli (strain K12) GN=ahpF PE=1 SV=2
▶ 6	<b>ARNC_ECOBW</b>	73	Undecaprenyl-phosphate 4-deoxy-4-formamido-L-arabinose transferase OS=Escherichia coli (strain K12 / MC4100 / BW2952) GN=arnC PE=3 SV=1
▶ 7	<b>MPL_ECOLI</b>	58	UDP-N-acetylmuramate--L-alanyl-gamma-D-glutamyl-meso-2,6-diaminoheptandioate ligase OS=Escherichia coli (strain K12) GN=mpl PE=1 SV=3
▶ 8	<b>YFBU_ECOBW</b>	56	UPF0304 protein YfbU OS=Escherichia coli (strain K12 / MC4100 / BW2952) GN=yfbU PE=3 SV=1
▶ 9	<b>GATD_ECOLI</b>	43	Galactitol 1-phosphate 5-dehydrogenase OS=Escherichia coli (strain K12) GN=gatD PE=1 SV=1
▶ 10	<b>YAJC_ECOLI</b>	42	Sec translocon accessory complex subunit YajC OS=Escherichia coli (strain K12) GN=yajC PE=1 SV=1

per page 1 [2](#) [Next](#)

# MASCOT – MS/MS

▼1 **SLYD\_ECOLI** 860 FKBP-type peptidyl-prolyl cis-trans isomerase SlyD OS=Escherichia coli (strain K12) GN=slyD PE=1 SV=1

2.1 [SLYD\\_ECOLI](#) **Score** 860 **Mass** 20840 **Matches** 15 (11) **Sequences** 6 (6) **empPAI** 1.06 FKBP-type peptidyl-prolyl cis-trans isomerase SlyD OS=Escherichia coli (strain K12) GN=slyD PE=1 SV=1  
 ▶3 samesets of SLYD\_ECOLI

▼15 peptide matches (6 non-duplicate, 9 duplicate)

Auto-fit to window

Query	Dupes	Observed	Mr (expt)	Mr (calc)	ppm	M	Score	Expect	Rank	U	Peptide
<a href="#">540</a>	▶1	523.8035	1045.5924	1045.5920	0.33	0	40	0.00018	▶1	U	K.FNVEVVAIR.E
<a href="#">599</a>	▶1	631.8589	1261.7033	1261.7030	0.19	0	64	1.9e-006	▶1	U	K.DLVVSLAYQVR.T
<a href="#">666</a>		798.3819	1594.7492	1594.7484	0.49	0	85	1e-008	▶1	U	K.DVFMGVDELQVGM.R.F
<a href="#">730</a>	▶1	782.0330	2343.0772	2343.0768	0.20	0	45	0.00016	▶1	U	K.FDVAVGANDAYGQYDENLVQR.V
<a href="#">781</a>	▶4	965.2306	3856.8934	3856.8942	-0.21	0	190	2.2e-019	▶1	U	R.FLAETDQGPVFEITAVEDDHVVVDGNHMLAGQNLK.F
<a href="#">796</a>	▶2	1095.5272	6567.1195	6567.1266	-1.10	1	16	0.026	▶1	U	R.TEDGVLVDESFPVSAPLDYLHGHSLSISGLETALEGHEVGDKFDVAVGANDAYGQYDENLVQR.V

▶1 subset or intersection (1 subset protein in total)

▼2 **CYOA\_ECOLI** 368 Cytochrome bo(3) ubiquinol oxidase subunit 2 OS=Escherichia coli (strain K12) GN=cyoA PE=1 SV=1

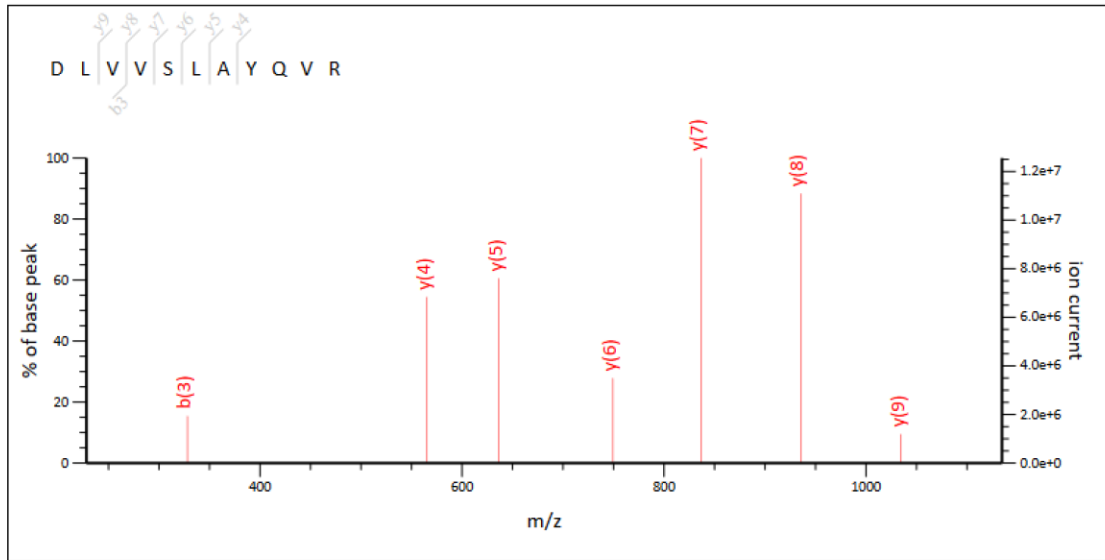
2.1 [CYOA\\_ECOLI](#) **Score** 368 **Mass** 34889 **Matches** 9 (8) **Sequences** 5 (5) **empPAI** 0.44 Cytochrome bo(3) ubiquinol oxidase subunit 2 OS=Escherichia coli (strain K12) GN=cyoA PE=1 SV=1  
 ▶1 sameset of CYOA\_ECOLI

▼9 peptide matches (7 non-duplicate, 2 duplicate)

Auto-fit to window

Query	Dupes	Observed	Mr (expt)	Mr (calc)	ppm	M	Score	Expect	Rank	U	Peptide
<a href="#">658</a>		763.8765	1525.7384	1525.7381	0.18	0	62	1.7e-006	▶1	U	R.LGSQIYAMAGMQTR.L
<a href="#">659</a>		509.5871	1525.7393	1525.7381	0.82	0	10	1.1	▶1	U	R.LGSQIYAMAGMQTR.L
<a href="#">668</a>		799.9036	1597.7926	1597.7923	0.21	0	99	5.3e-010	▶1	U	K.VTSNSVMNSFFIPR.L
<a href="#">740</a>	▶1	923.7779	2768.3119	2768.3116	0.11	0	62	1.7e-006	▶1	U	R.LHLIANEPGTVDGISASYSGGPGFSGMK.F
<a href="#">758</a>		1024.5150	3070.5230	3070.5287	-1.86	1	133	2.7e-013	▶1	U	K.LAAPSEYNQVEYFSNVKPDLEADVINK.F
<a href="#">762</a>	▶1	632.0428	3155.1775	3155.1783	-0.27	0	29	0.0012	▶1	U	K.SMDMTQPEGEHSAHEGMEGMDMSHAESA.H.-
<a href="#">763</a>		789.8017	3155.1777	3155.1783	-0.20	0	20	0.01	▶1	U	K.SMDMTQPEGEHSAHEGMEGMDMSHAESA.H.-

# MASCOT – MS/MS



Label all possible matches  Label matches used for scoring

Monoisotopic mass of neutral peptide Mr(calc): 1261.7030

Ions Score: 64 Expect: 1.9e-006

Matches : 7/92 fragment ions using 7 most intense peaks [\(help\)](#)

#	b	b <sup>++</sup>	b <sup>*</sup>	b <sup>+++</sup>	b <sup>0</sup>	b <sup>0++</sup>	Seq.	y	y <sup>++</sup>	y <sup>*</sup>	y <sup>+++</sup>	y <sup>0</sup>	y <sup>0++</sup>	#
1	116.0342	58.5207			98.0237	49.5155	D							11
2	229.1183	115.0628			211.1077	106.0575	L	1147.6834	574.3453	1130.6568	565.8320	1129.6728	565.3400	10
3	<b>328.1867</b>	164.5970			310.1761	155.5917	V	<b>1034.5993</b>	517.8033	1017.5728	509.2900	1016.5887	508.7980	9
4	427.2551	214.1312			409.2445	205.1259	V	<b>935.5309</b>	468.2691	918.5043	459.7558	917.5203	459.2638	8
5	514.2871	257.6472			496.2766	248.6419	S	<b>836.4625</b>	418.7349	819.4359	410.2216	818.4519	409.7296	7
6	627.3712	314.1892			609.3606	305.1840	L	<b>749.4305</b>	375.2189	732.4039	366.7056			6
7	698.4083	349.7078			680.3978	340.7025	A	<b>636.3464</b>	318.6768	619.3198	310.1636			5
8	861.4716	431.2395			843.4611	422.2342	Y	<b>565.3093</b>	283.1583	548.2827	274.6450			4
9	989.5302	495.2688	972.5037	486.7555	971.5197	486.2635	Q	402.2459	201.6266	385.2194	193.1133			3
10	1088.5986	544.8030	1071.5721	536.2897	1070.5881	535.7977	V	274.1874	137.5973	257.1608	129.0840			2
11							R	175.1190	88.0631	158.0924	79.5498			1

# Validace identifikací z LC-MS/MS ?

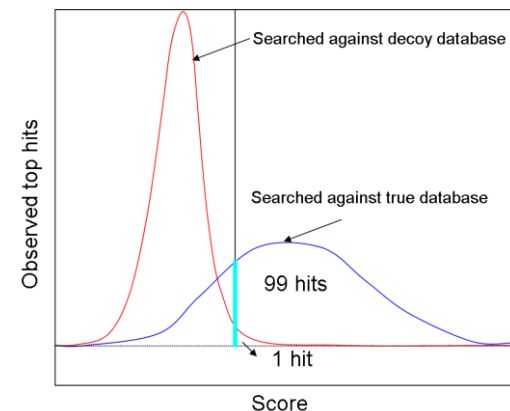
Obrovské objemy dat, nemožnost manuální validace

Statistické přístupy

Decoy databáze – prohledání proti nesmyslné datbázi (invertovaná nebo zpřeházené pořadí AK,...)

Kolik procent dobrých identifikací získám v „nesmyslné“ databázi – FDR (False Discovery Rate)  $FDR = FP / (FP + TP)$  FP – false positive, TP – true postive

Informace že např. méně než 1% je špatně, ale ne co konkrétně je špatně





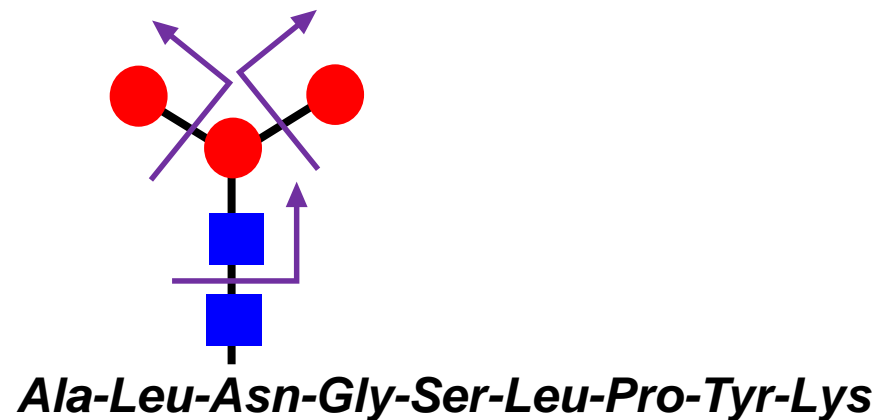
# Post-Translační Modifikace

**!!! změna MW !!!**

- **Fosforylace (+80 Da)**
- **Glykosylace (+162,203,146 Da)**
- **S-S můstky (-2 Da)**
- **Deamidace (+1 Da)**
- **Gln -> pyro-Glu (-17 Da)**
- **Oxidace (+16 Da)**
- **Acetylace (+42 Da)**
- **Acylace (+238 Da)**
- **Ztráta N-konc.Met (-131 Da)**
- **Zkrácení proteinu (??? Da)**

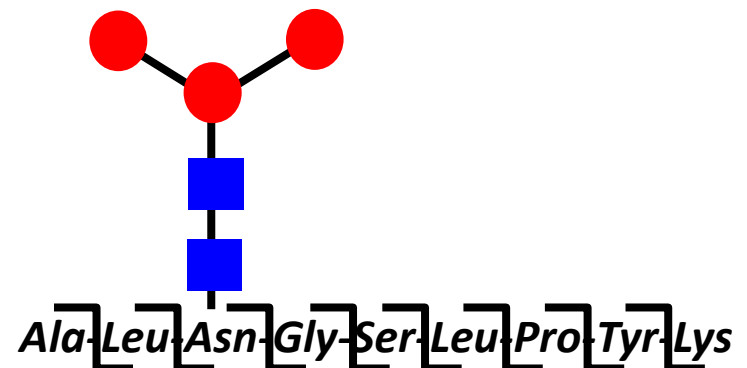
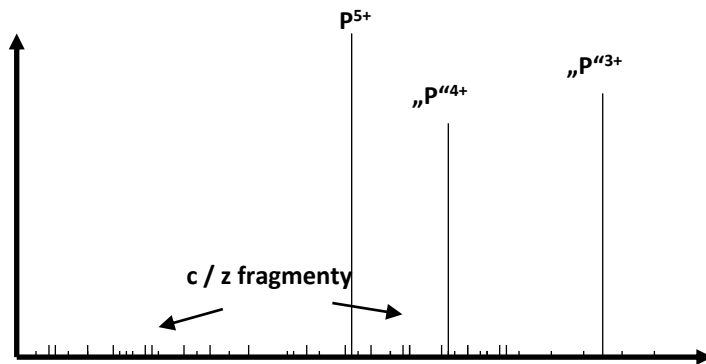
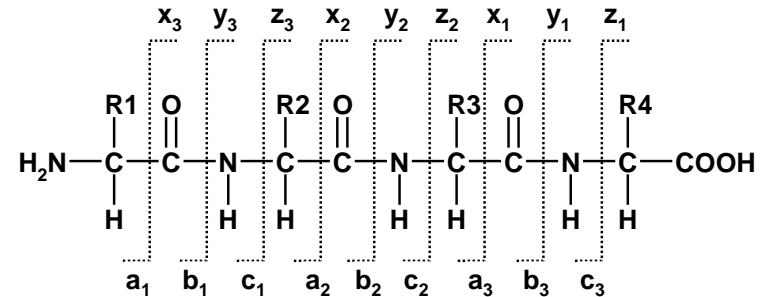
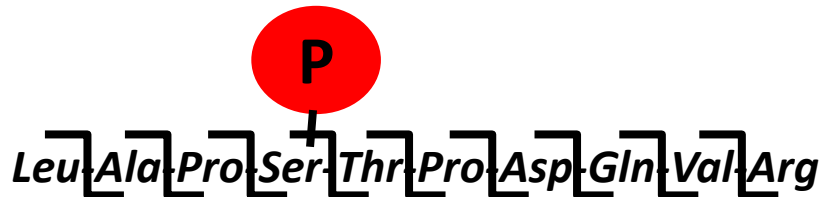
# CID/CAD, IRMPD...

... fragmentují peptid v peptidové vazbě, ALE – jsou-li přítomny post-translační modifikace, dojde především k jejich fragmentaci



# Electron Capture/Transfer Dissociation (ECD/ETD) UVPD

- Záchyt elektronu nebo jeho přenos, fragmentace UV laserem, c a z ionty
- Preferenční štěpení peptidové vazby, jsou zachovány labilní skupiny modifikací



# Fosforylace

Signalizace, regulace proteinové aktivity, lokalizace, interakce. . .

S (90%), T (10%), Y (0.05%),. . . .H, K, D, E, C

P skupina = 80 Da ( $\text{PO}_3$ )

Nízká stechiometrie

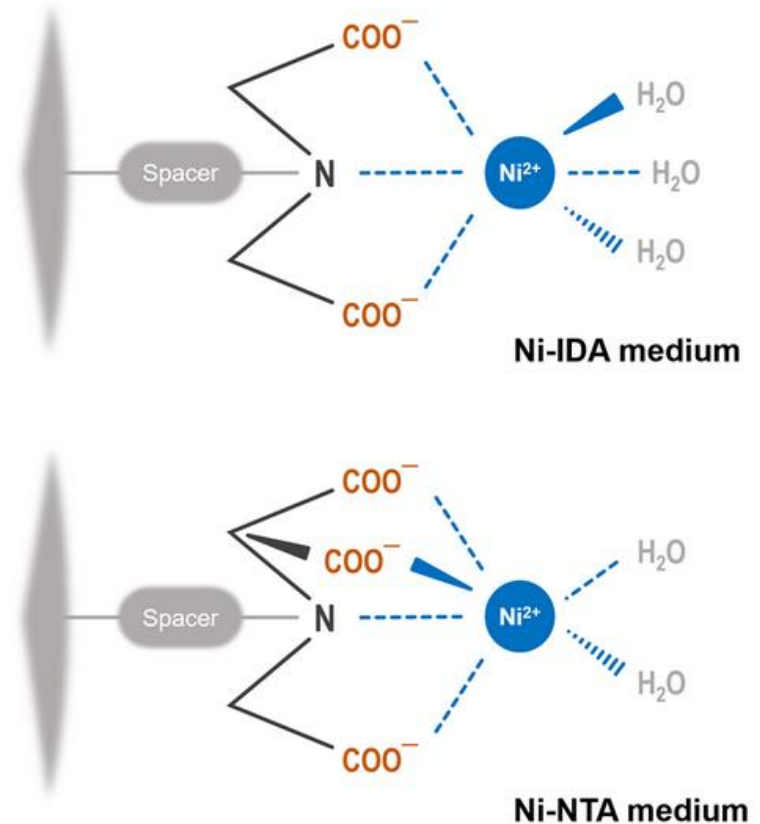
P-peptidy se hůře ionizují, ztrácejí P při MSMS (CID)

## Fosforylace – cíle...

1. Identifikace/nabohacení fosfopeptidů
2. Lokalizace fosforylace
3. Kvantifikace fosforylací

# Selektivní vychytání fosfopeptidů pomocí chromatografie IMAC – Immobilized Metal Ion Affinity Chromatography

- IDA (iminodiacetic acid)  
NTA (nitrilo-triacetic acid)
  - Ionty –  $\text{Ga}^{3+}$ ,  $\text{Fe}^{3+}$
- Eluce  $\text{NH}_4\text{OH}$  (nutná neutralizace), přídavek DHB, nebo využití kys. fosforečné a acetonitrilu
  - Odsolení na RP
  - Nespecifická vazba COOH esterifikace (potíže s deamidací...) nebo vazba za velmi nízkého pH (pKa Asp, Glu – 3.6, 4.2 vs. pKa fosfátu – 2.1)



# Obohacení fosfopeptidů pomocí $\text{TiO}_2$

Digeste v gelu

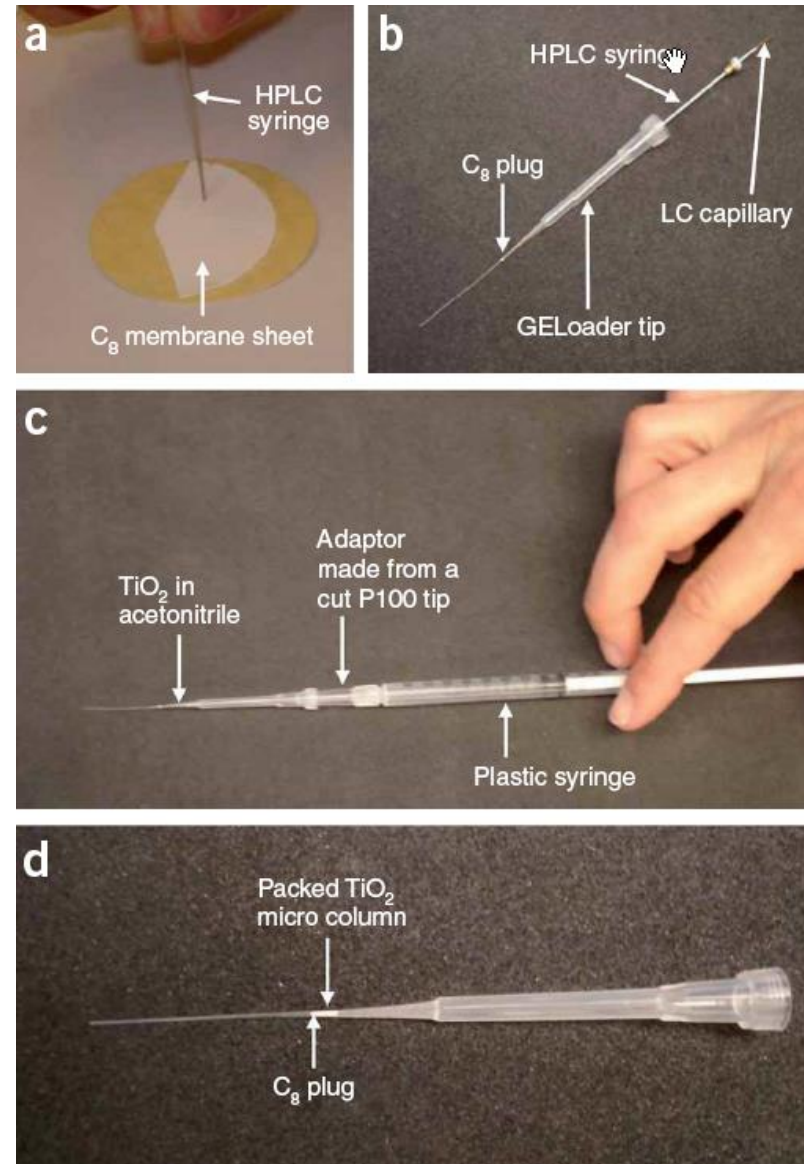
Zachycení P-peptidů na  $\text{TiO}_2$  mikrokolonce  
kyselé prostředí a přidavek DHB blokuje  
nespecifické interakce

Eluce P-peptidů (pH 10.5)

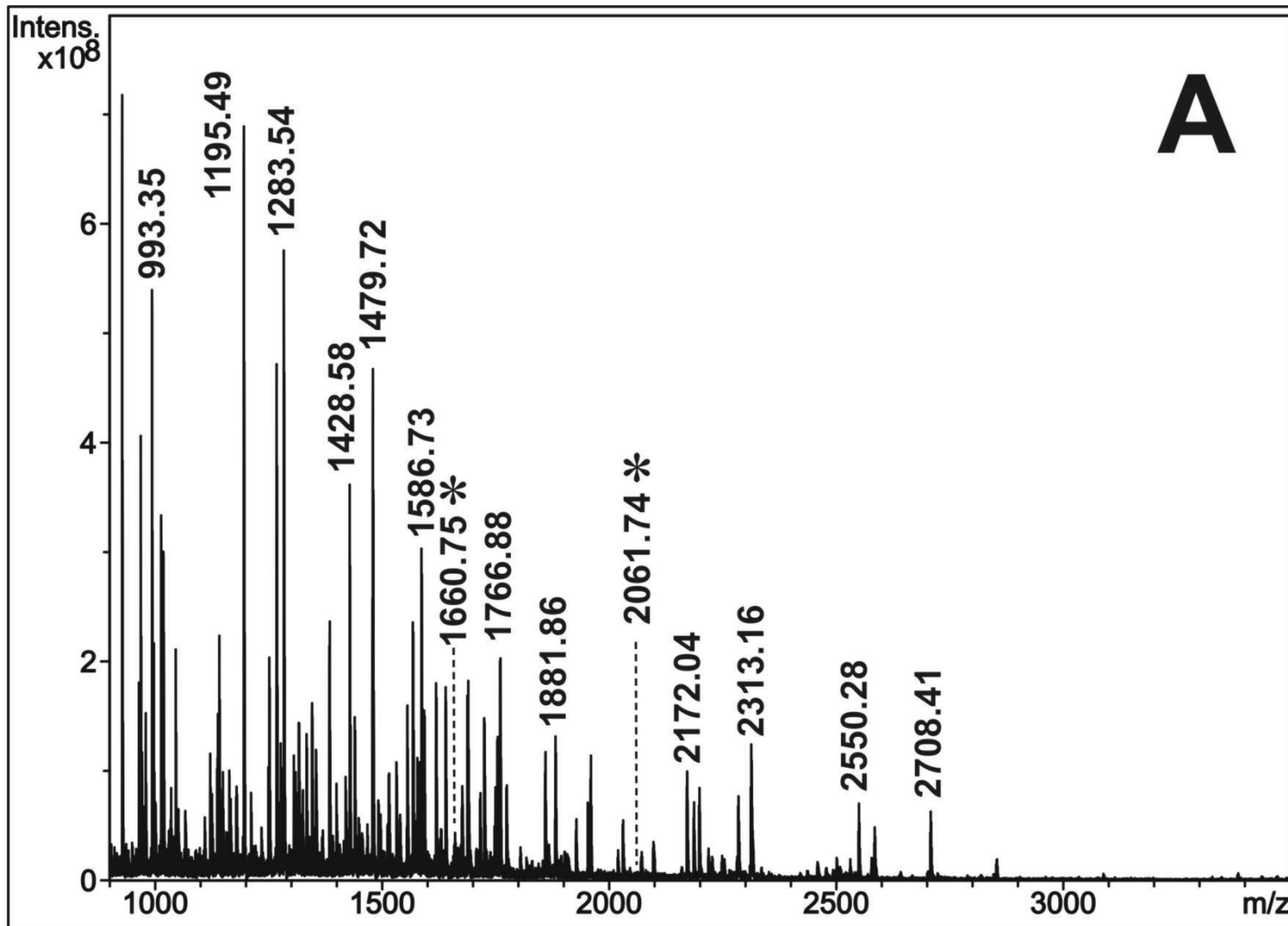
Okyselení a purifikace na RP

Eluce maticí DHB přímo na MALDI desku

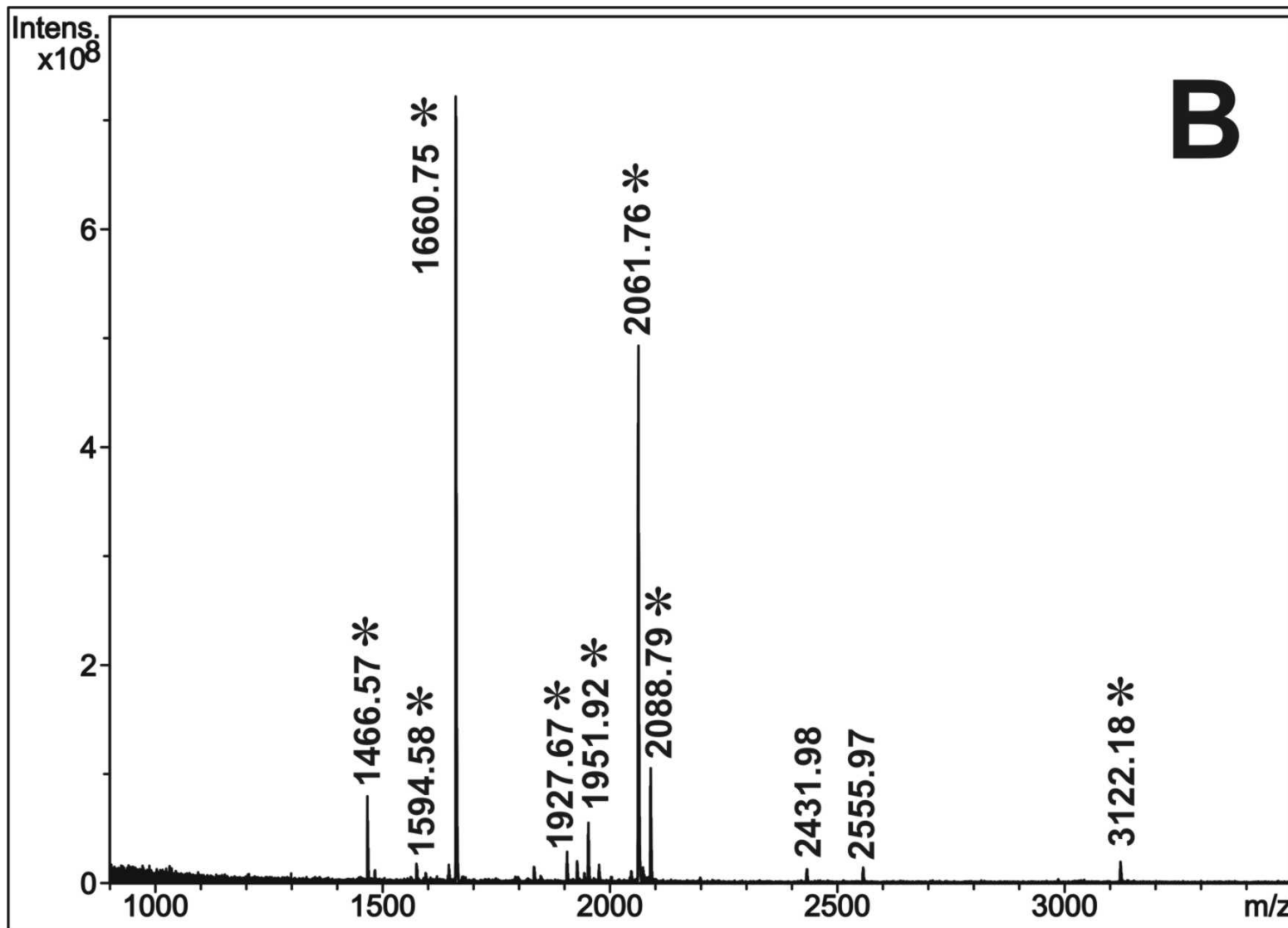
Může pevněji vázat vícenásobně fosforylované peptidy



# Obohacení fosfopeptidů pomocí TiO<sub>2</sub>



# Obohacení fosfopeptidů pomocí TiO<sub>2</sub>

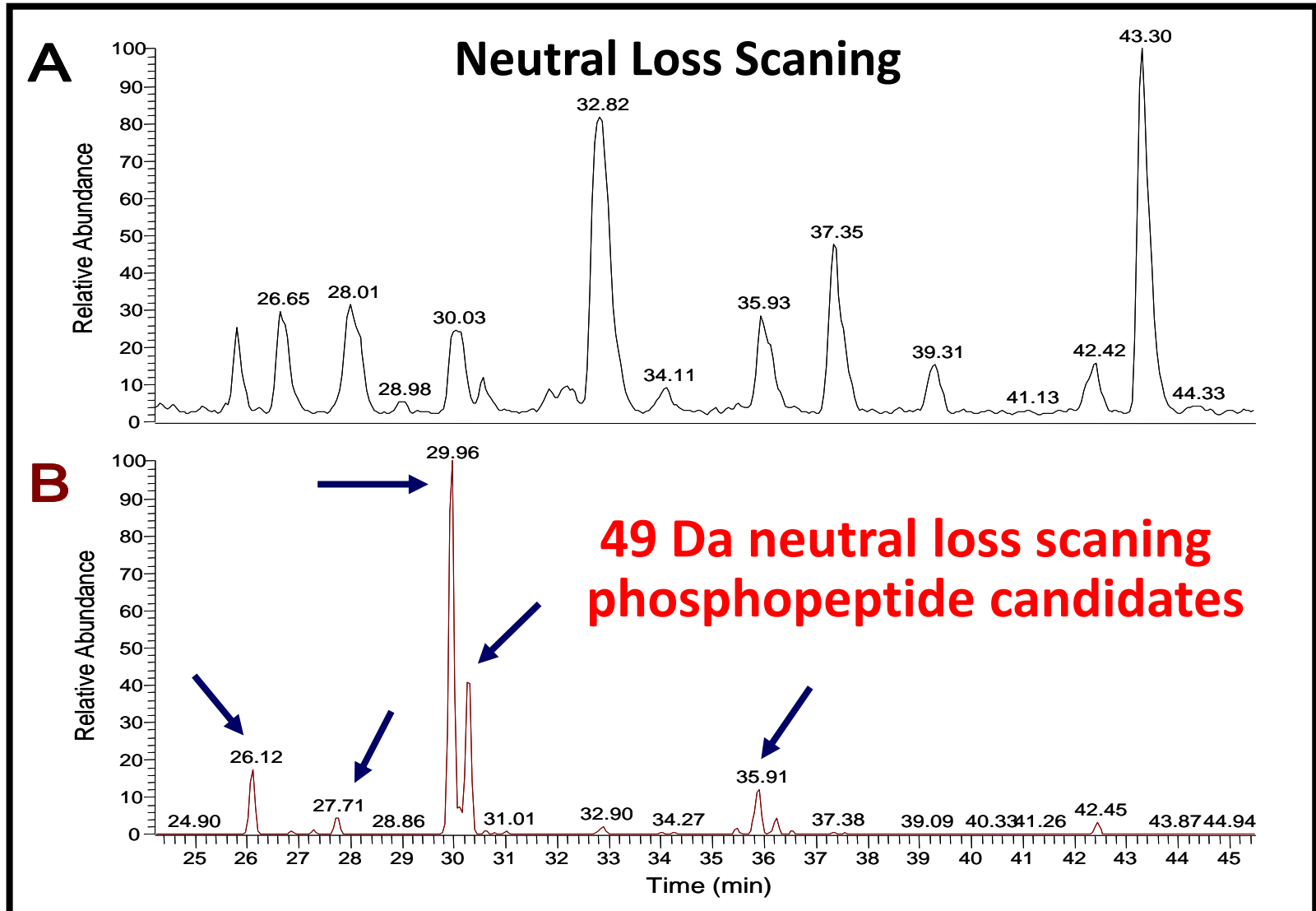




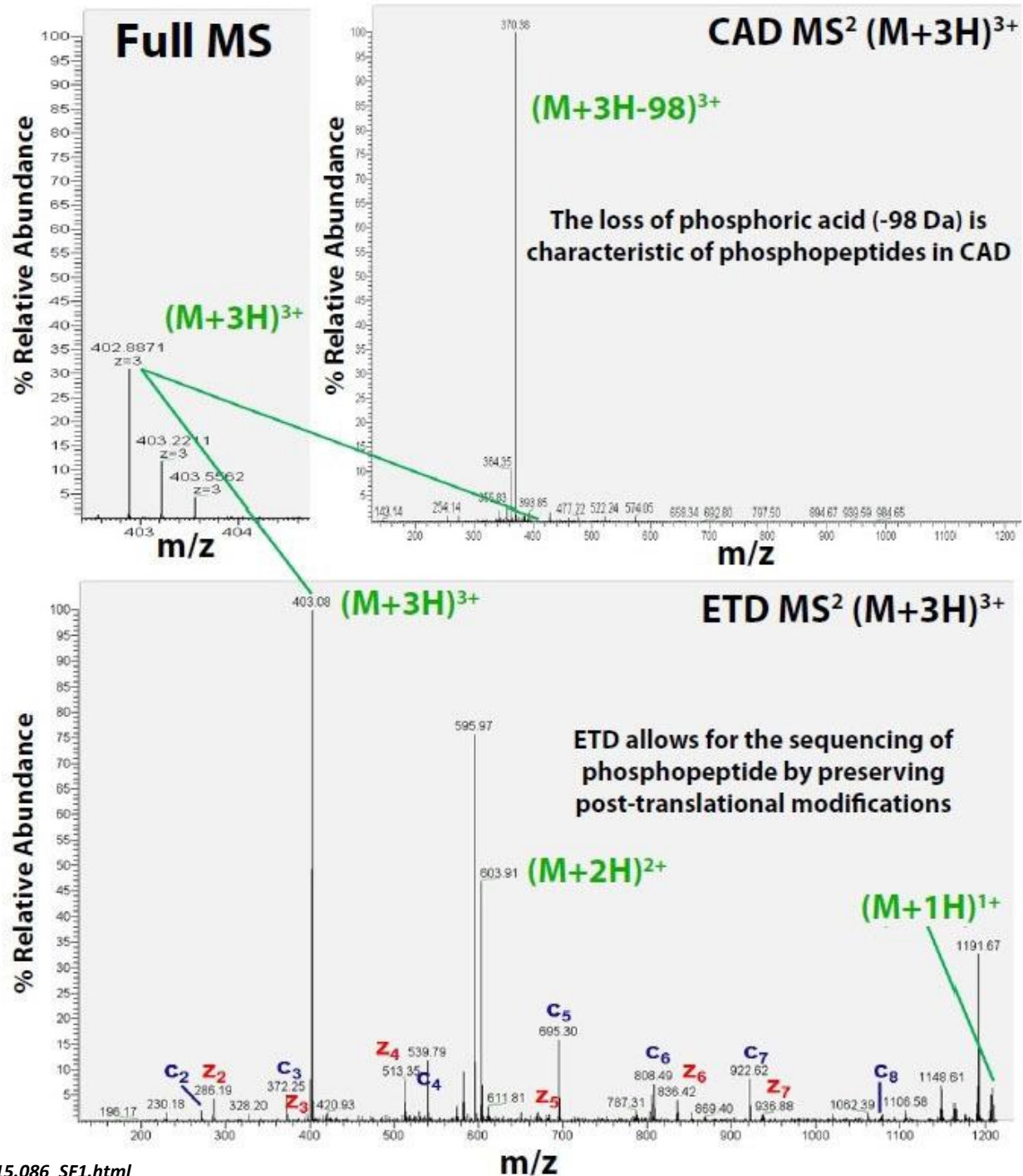
## Identifikace P-peptidů pomocí MS

- pY [MH-80]<sup>+</sup>, pS,pT [MH-98]<sup>+</sup>
- Neutrální ztráta (pS, pT: 98, 49, 33)
- P-peptidy spouštěné MS<sup>3</sup> (MS<sup>2</sup> – je-li ztráta P -> MS<sup>3</sup> na hlavním píku v MS<sup>2</sup>)
  - Prekurzorový sken +ní mód (pY m/z 216.043)
  - Prekurzorový sken -ní mód (pS, pT, pY: m/z 79)

# Identifikace P-peptidů pomocí MS

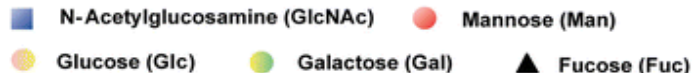
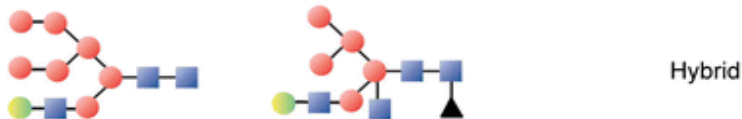
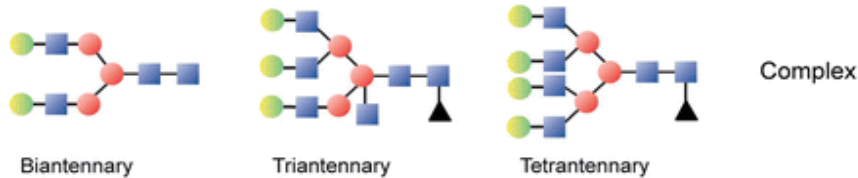
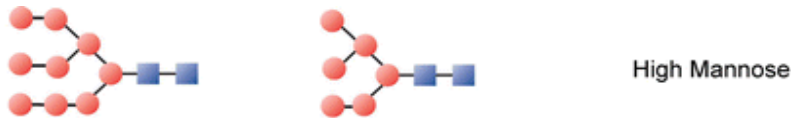


# Fragmentce – lokalizace fosforilace



# Glykosylace

- Mnoho důležitých funkcí: stabilizace struktury, dobré poskládání, ochrana před degradací, interakce ligand-receptor, targeting, modulace aktivity
- N-glykosylace  
na sekvenci: NXT/S/C – v b-strukturách nebo smyčkách sacharid: větvený, 3 typy: mannosový, komplexní, hybridní



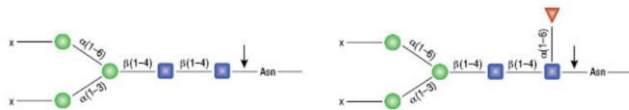
# N-glykosylace

1. Charakterizace sacharidových struktur+určení heterogenity
2. Nalezení míst glykosylace
3. Přiřazení struktur na jednotlivá glykosylační místa

Deglykosylace:

PNGasa F – **N->D** (využití  $H_2^{18}O$ ), manosové, hybridní i komplexní, ale ne 1-3 Fuc (rostliny) – PNGase A

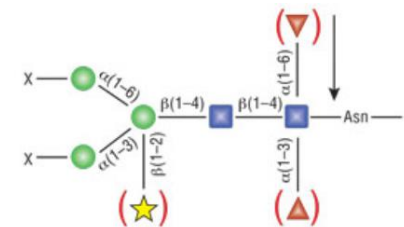
Endo H – (zůstává kotevní HexNAc – snadná modifikace pro vyhledávací program), manosové a hybridní oligosacharidy



A. PNGase F can cleave when an 1-6 Fucose is on the core GlcNAc

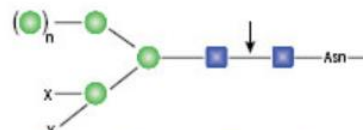


PNGase F



PNGase A

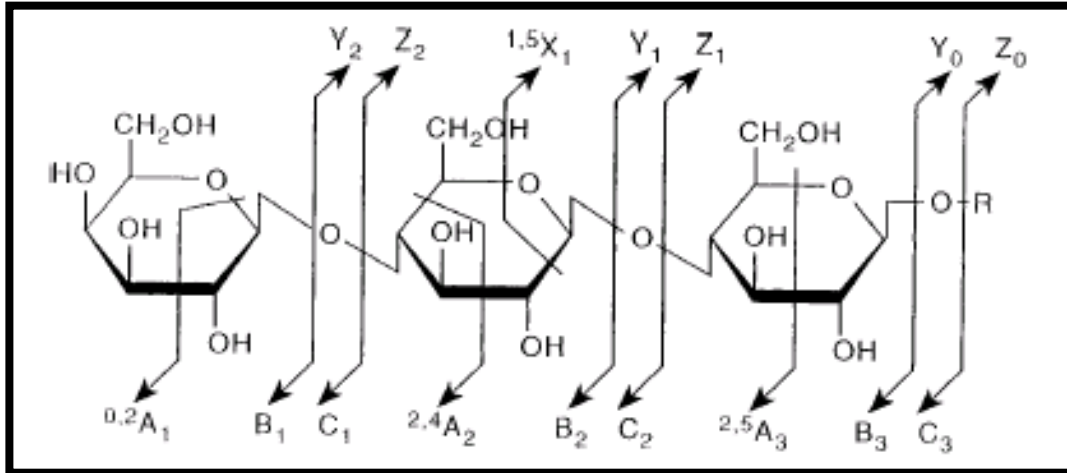
Endo H



Gal ● Glc ● Man ● GalNAc ● GlcNAc ● Fuc ▲ NeuAc ◆ R = any sugar

# Určení struktury glykanu

Štěpení proteinu na (glyko-)peptidy a přímá analýza



Nomenklatura  
*Domon-Costello (1998)*

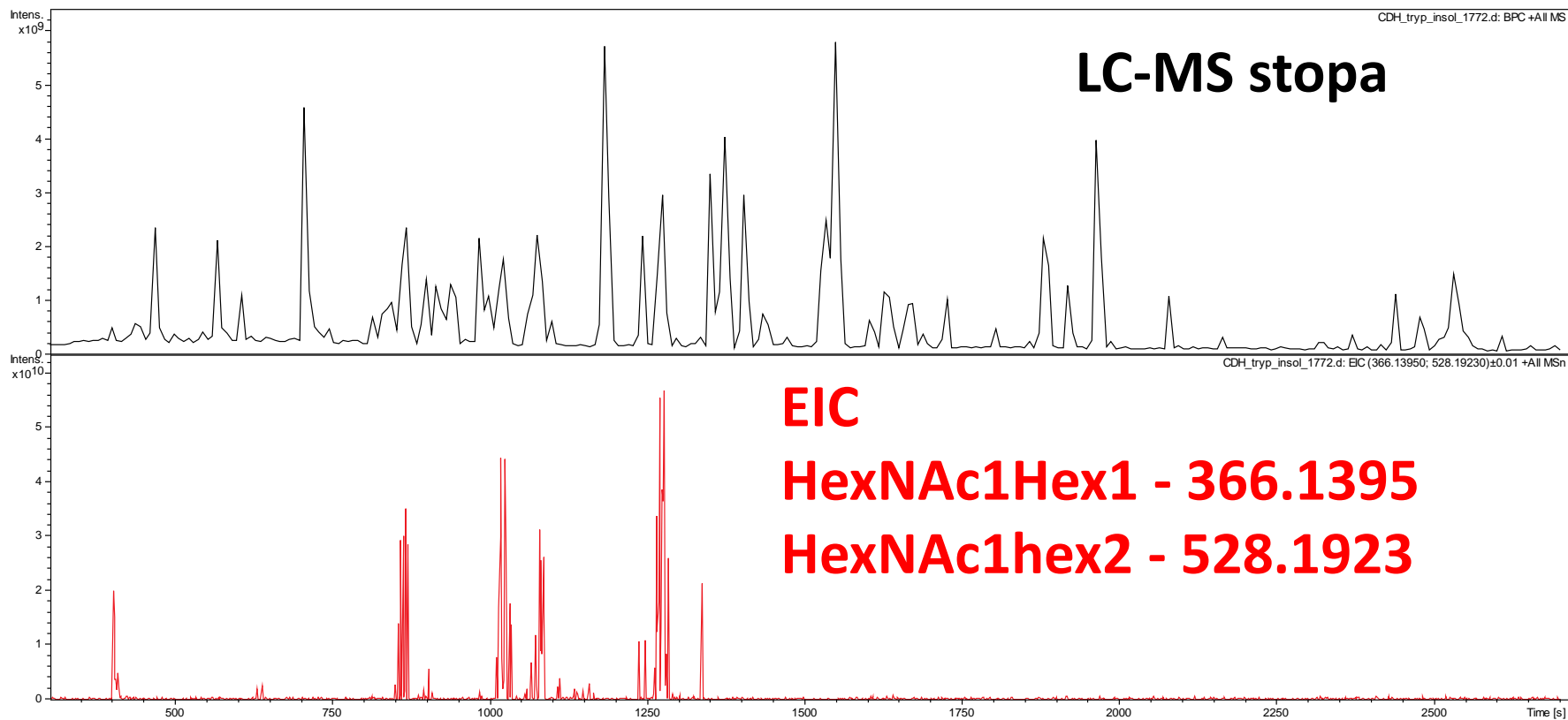
MS "vidí" hmotnosti – nerozliší např. Glc a Man - Hex (162 Da), dHex (146 Da), HexNAc (203 Da), Sia (291 Da),...

V MS stopě hledáme difference odpovídající sacharidům (např. 162, 81, 54 pro Hexosu)

V MS/MS hledáme oxoniové ionty – fragmenty glykanu (366, 528, 690)

# Analýza glykopeptidů

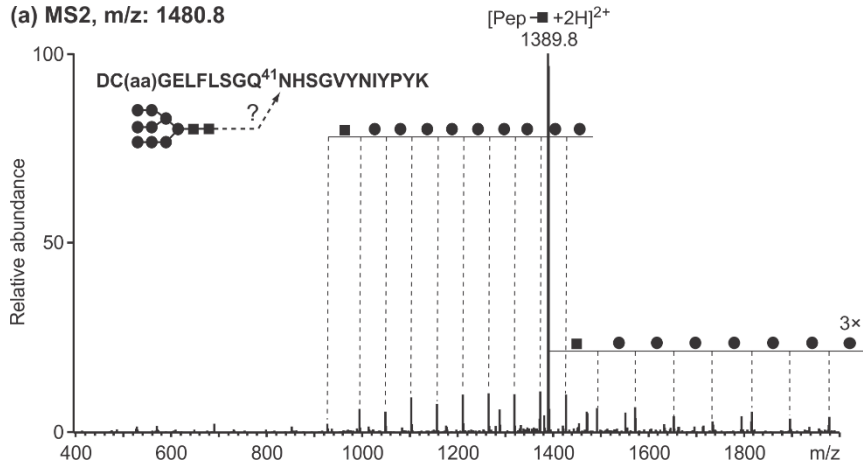
Selektivní stopa pro oxoniové ionty – fragmenty sacharidové části glykopeptidu



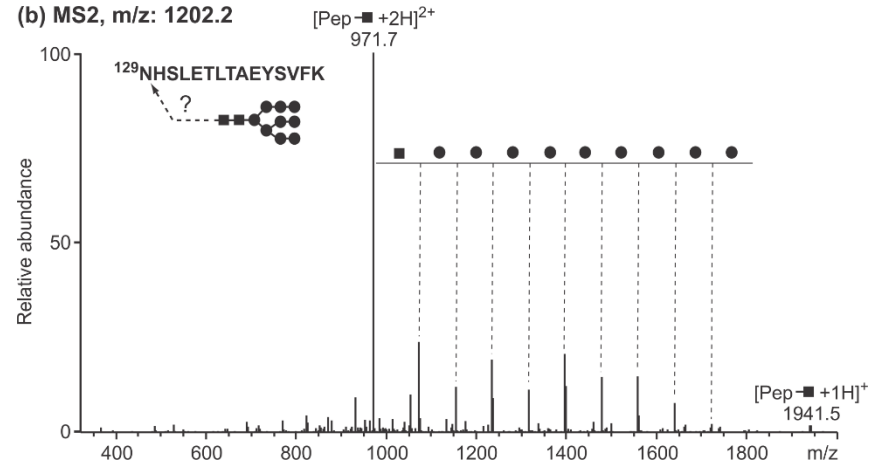
# MSMS (CID) glykopeptidu

Fragmentace glykopeptidu – rozpad sacharidové struktury  
Sekvence peptidu odvoditelná díky použití specifické proteasy

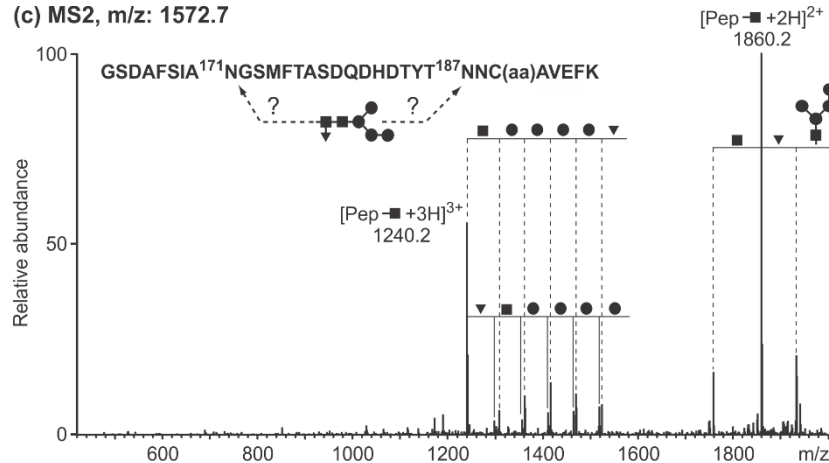
(a) MS2, m/z: 1480.8



(b) MS2, m/z: 1202.2



(c) MS2, m/z: 1572.7





# MSMS (CID vs ETD) glykopeptidu

