

Proteomické dáta

2D gélová elektroforéza

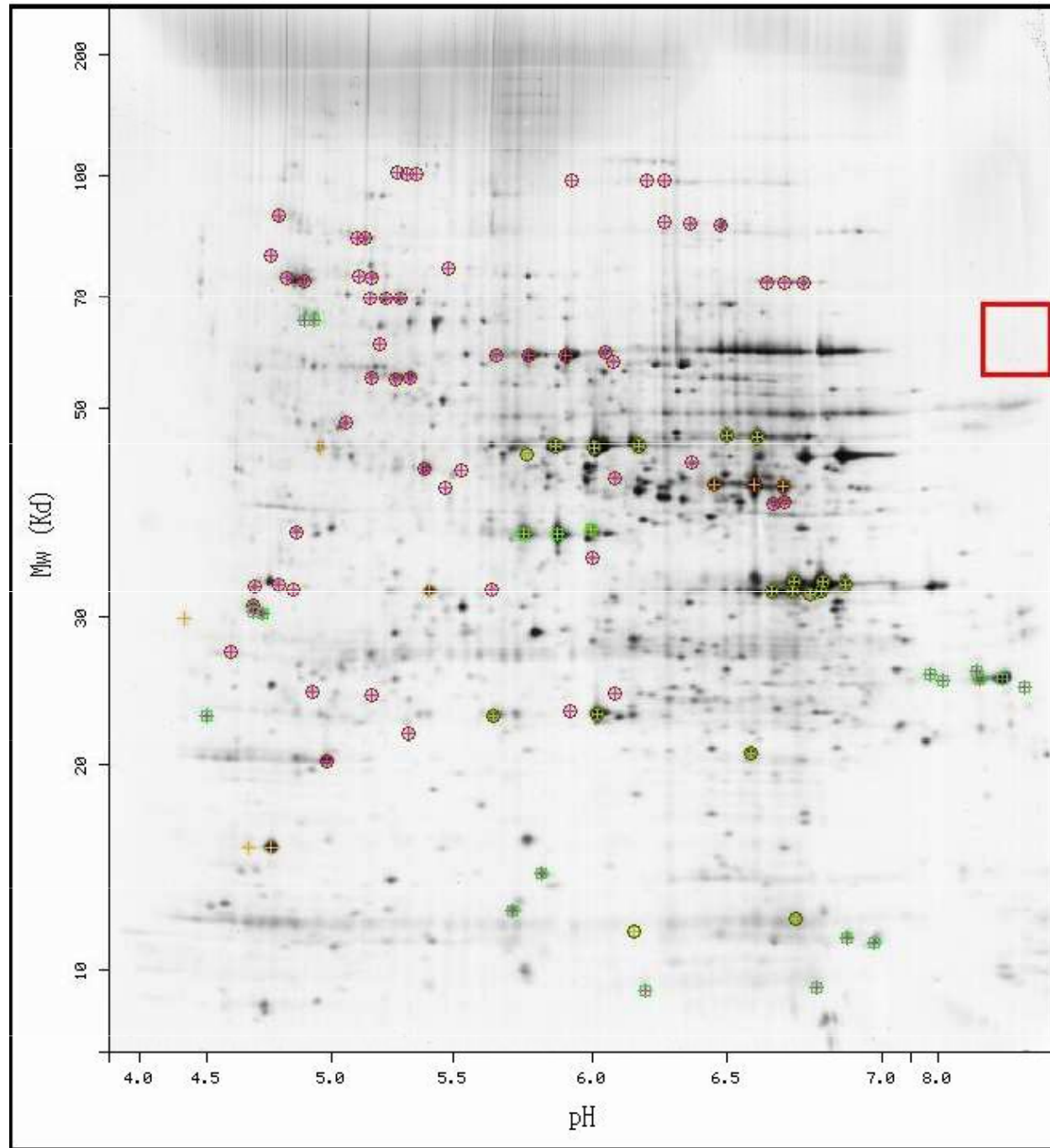
Princíp

- Proteíny sú separované na géle v dvoch dimenziách – na základe hmotnosti a na základe pH

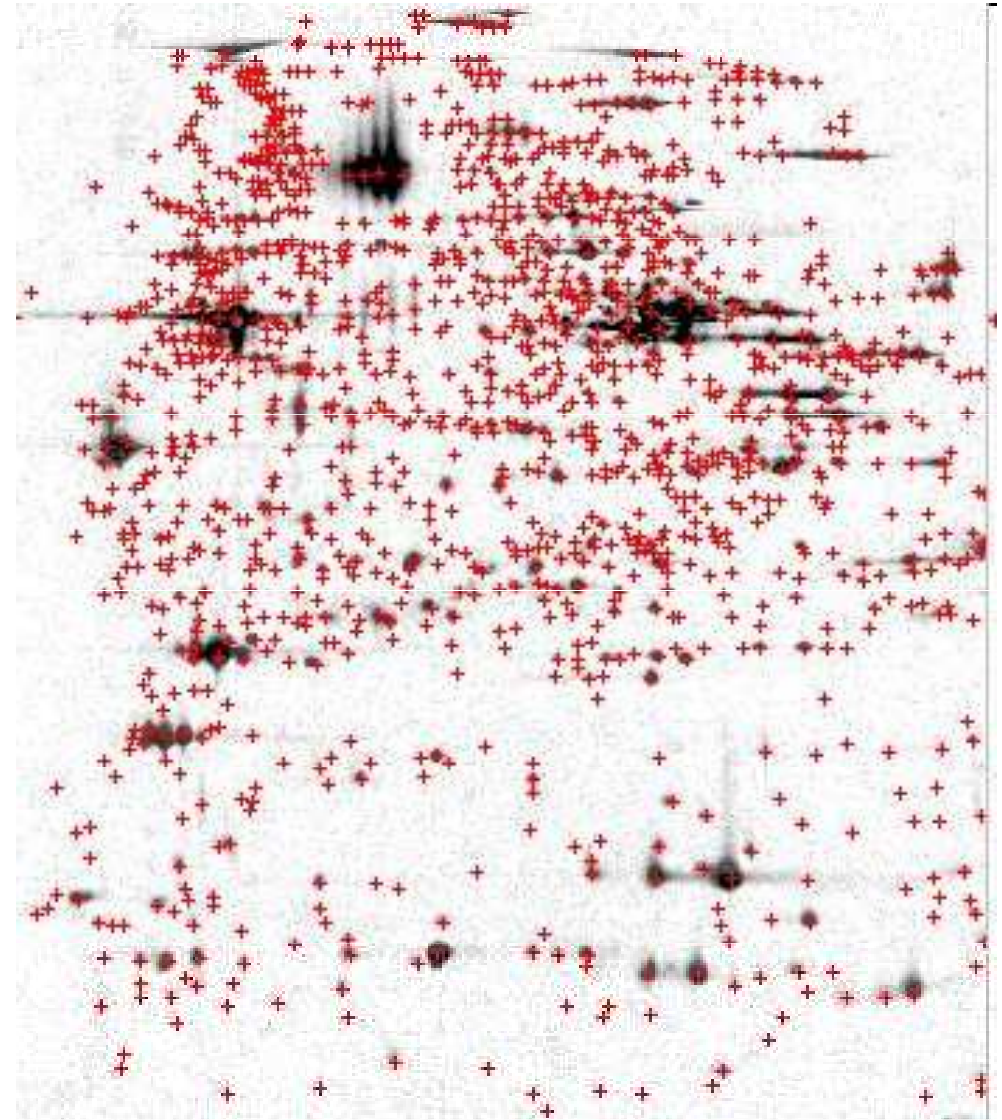
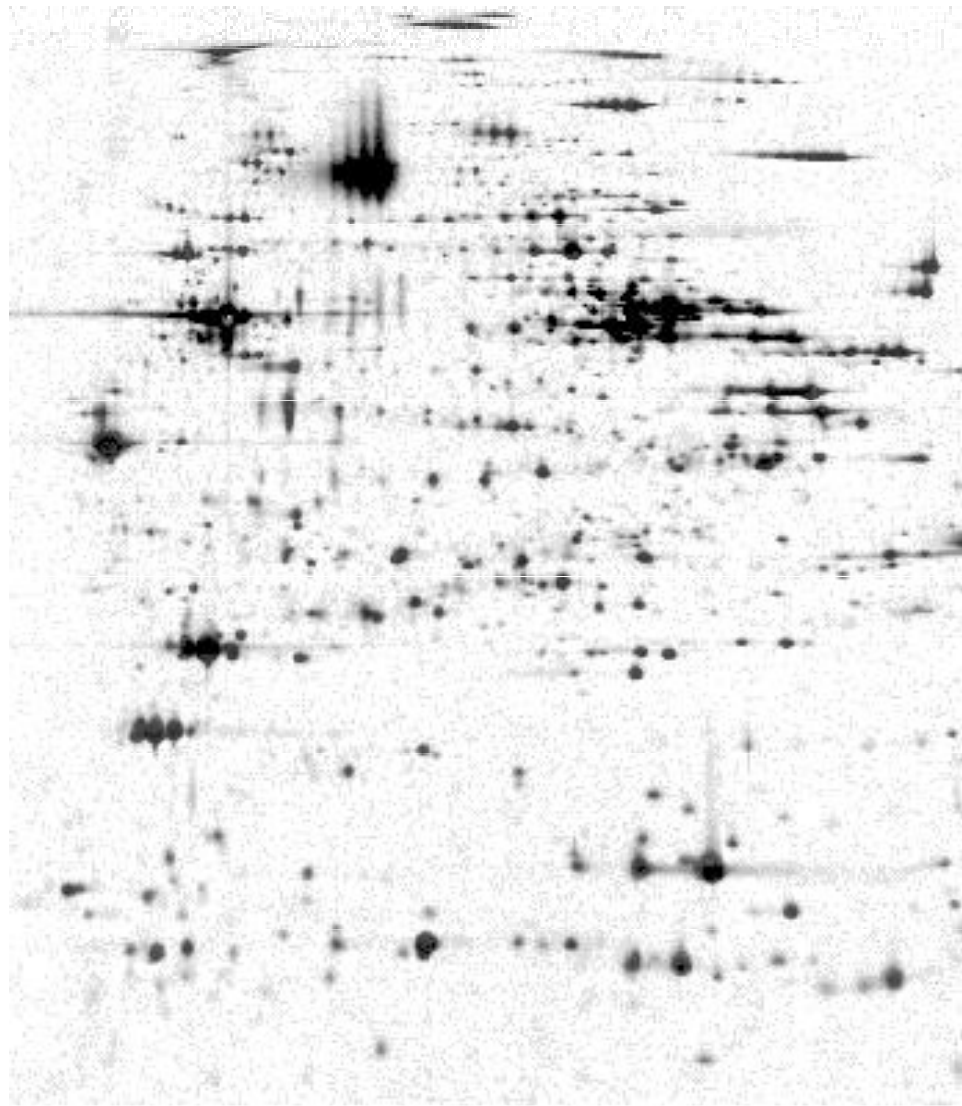
Princíp

1. **Proteíny sú extrahované** zo vzorky
2. **Vzorka sa umiestni na gél** a proteíny migrujú až kým dosiahnú izoelektronický bod (kedy ich náboj je nula)
 - Dôležitý je **výber gélu**, ktorý musí byť dostatočne pórovitý, aby umožnil proteínom pohyb (agaróza alebo polyacrylamidový gél)
3. Takto sa **proteíny oddeľia** vzhľadom k svojej hmotnosti a náboju.
4. Následne proteíny necháme pohybovať sa na základe **pH** v druhej dimenzii
5. Nakoniec je gél **zafarbený** aby sa detekoval jednotlivé spoty
6. Keď sú spoty **zafarbené**, gél sa môže odfoťiť
7. Intenzita pixelov koreluje **s množstvom proteínu**, používa sa špeciálny SW pre analýzu spotov

Ako vyzerá obrázok



2-D gelová elektroforéza



Ako vyzerajú dáta

← Vzorky →

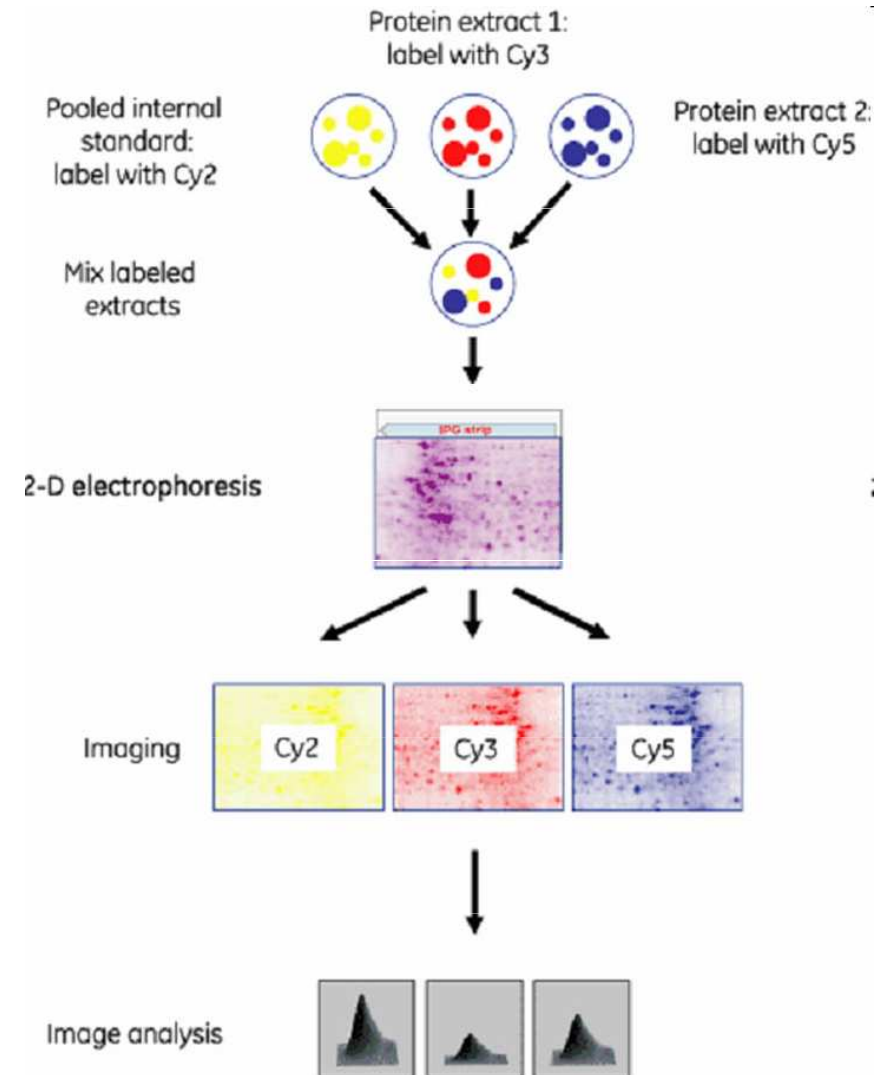
SSP	wt_A	wt_A	wt_A	wt_S
101	2338.84	2078.42	2625.1	2550.54
102	118.92	68.65	125.8	109.66
103	221.89	55.32	NA	NA
104	215.3	189.02	220.28	NA
105	106.56	NA	238.36	NA
202	328.32	226.46	522.52	1281.75
203	259.8	228.13	340.37	NA
205	1439.72	1213.28	1187.43	1353.14
206	1094.33	754.83	1291.89	1240.82
208	97.78	41.51	164.49	33.25
209	NA	NA	NA	22.42
301	212.63	92.12	307.19	317.67
302	1491.34	1703.79	1830.19	1976.66
304	71.25	72.72	127.87	199.31

Peptidy

DIGE

- Špeciálny typ 2-DE je 2-D Fluorescence Difference Gel Electrophoresis (DIGE).

- Proteíny sa najprv zafarbia fluorescenčným farbivom
- Každé farbivo sa skenuje pod iným filtrom
- Takto sa môže porovnávať viac vzoriek

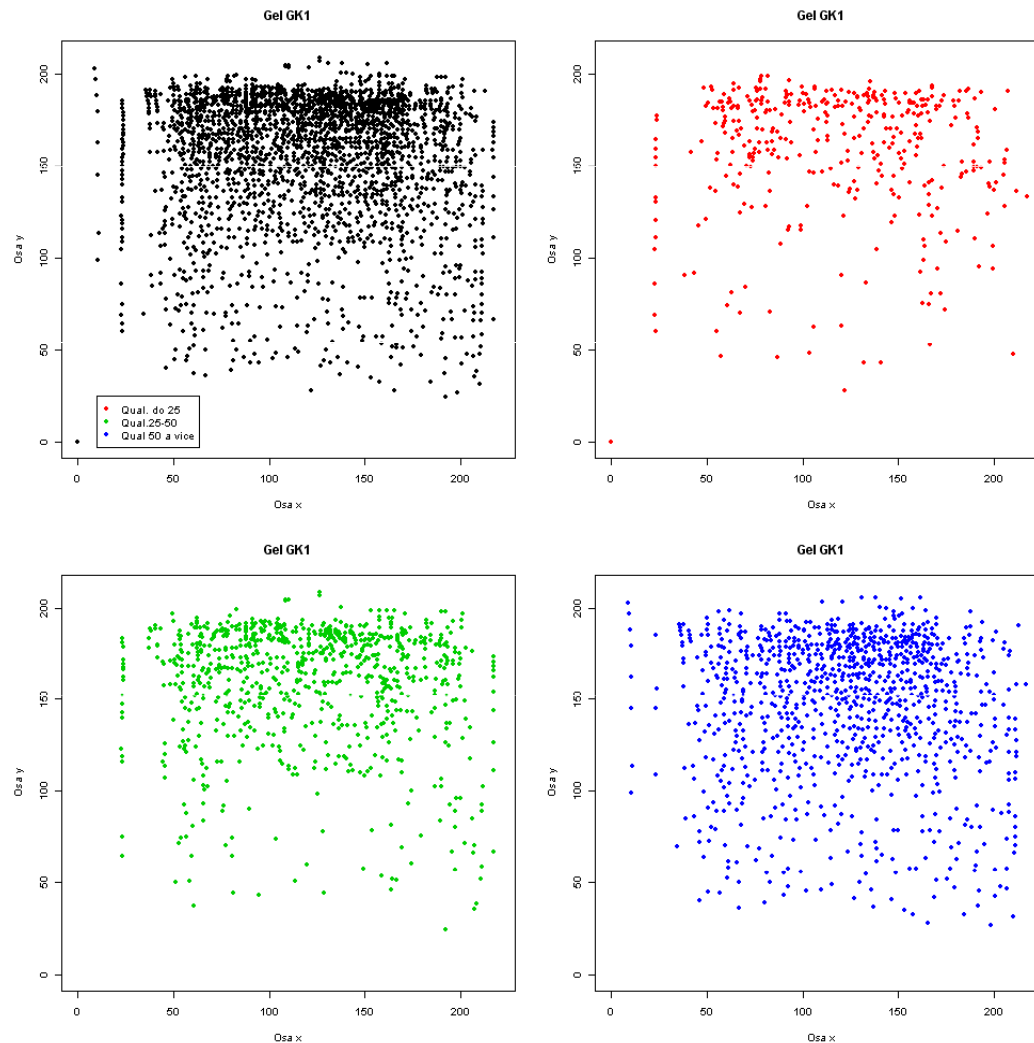


Nutnosť úpravy dát

- Tak ako mikročipový experiment i 2-DE je vystavená experimentálnym chybám ktoré sú zdrojom šumu
- Je nutná úprava a normalizácia dát
- Neexistuje tu ale taká automatická kvantifikácia spotov tak ako u microarrays, pretože spoty nie sú fixne dané predom!
 - tá čo existuje, vyžaduje manuálnu úpravu
 - premenné kvality spotov
- Dáta z 2DE nie sú normálne rozložené – je nutná transformácia (log)

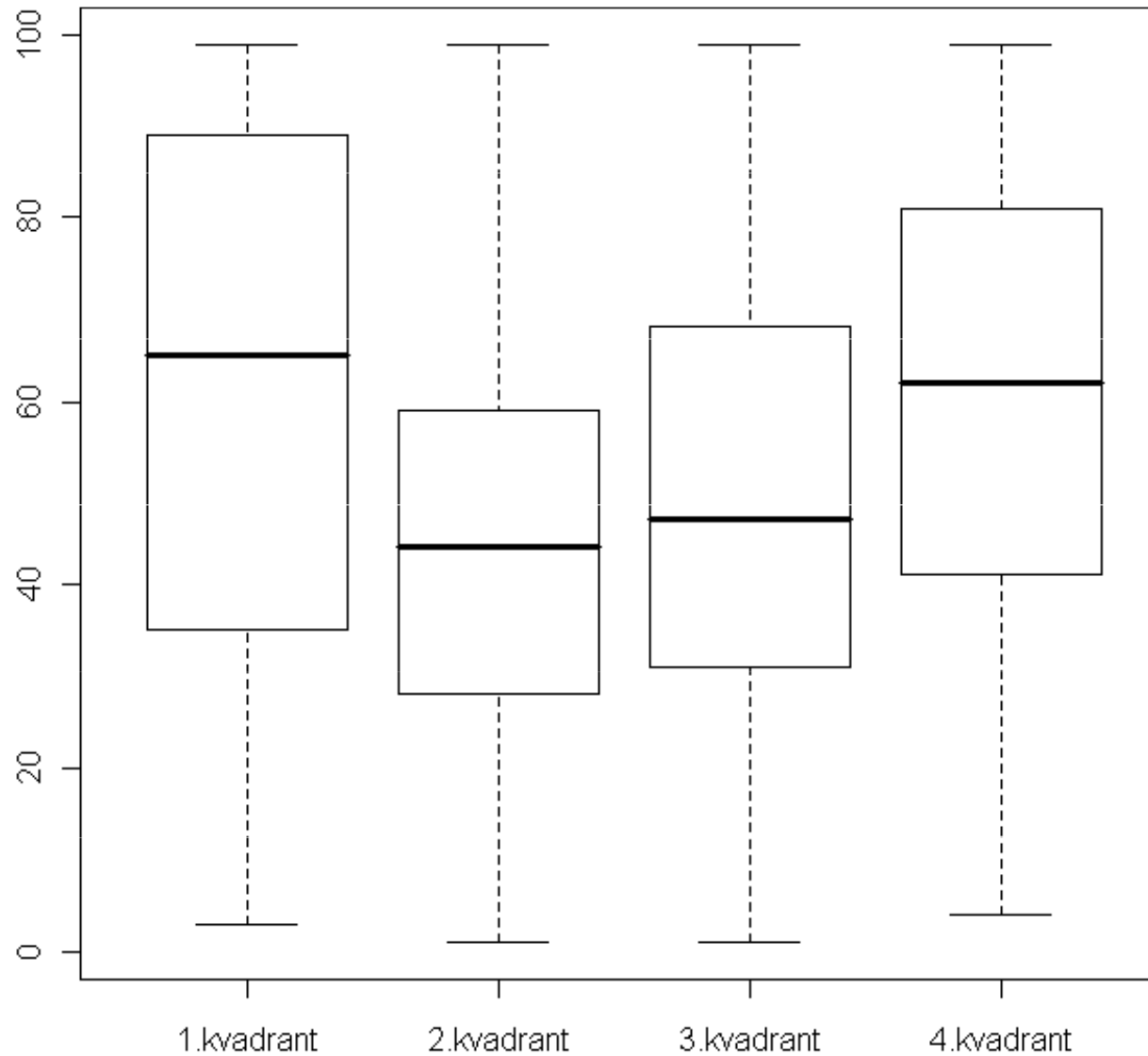
Normalizácia

- Dôležitým krokom v úprave dát je **kalibrácia všetkých expresných hodnôt a gélov navzájom**
- V tomto procese sa odstraňuje priestorový efekt, aj efekt farbiva
- Na každom géle sú kontrolné proteíny, podľa ktorých gél kalibruje



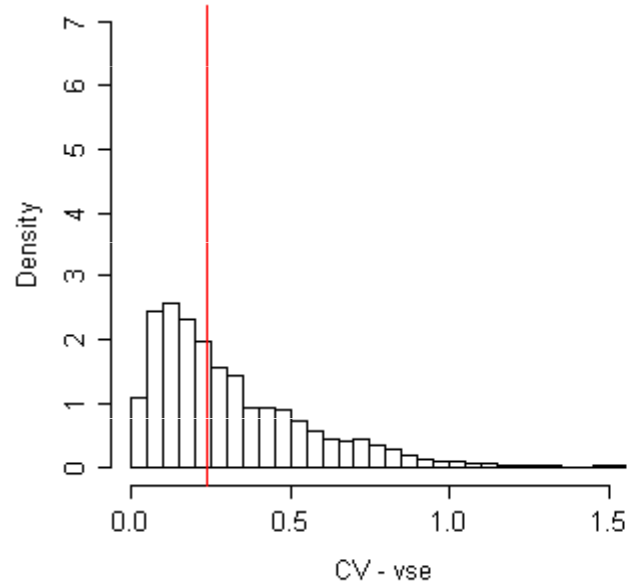
Kontrola kvality spotov

Spot quality (N1=53, N2=598, N3=1217, N4=105)

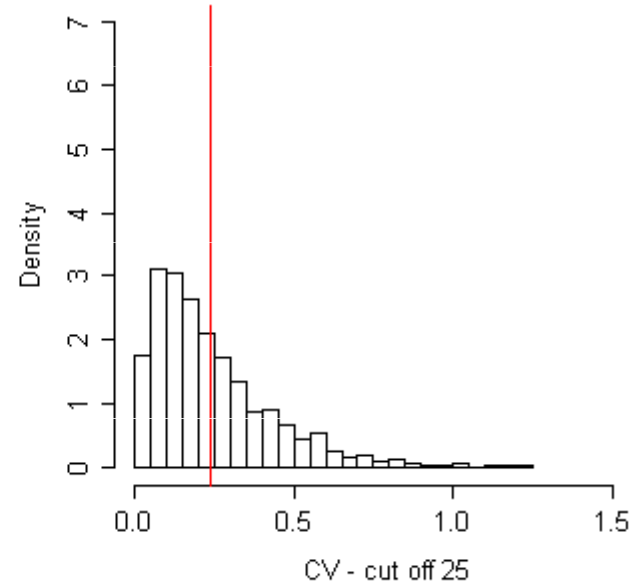


Kontrola kvality spotov 2.

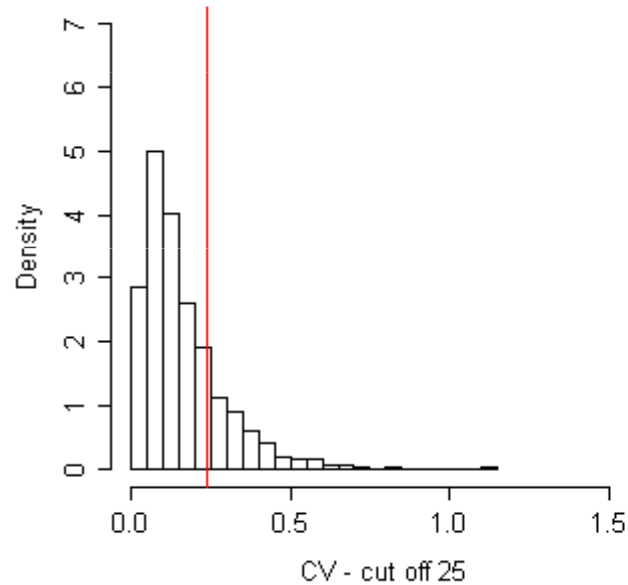
Histogram of CV.good



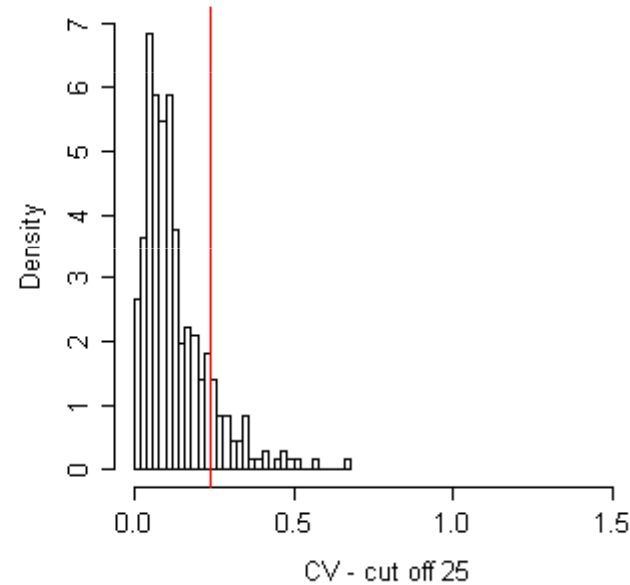
Histogram of CV.good.25



Histogram of CV.good.50



Histogram of CV.good.75

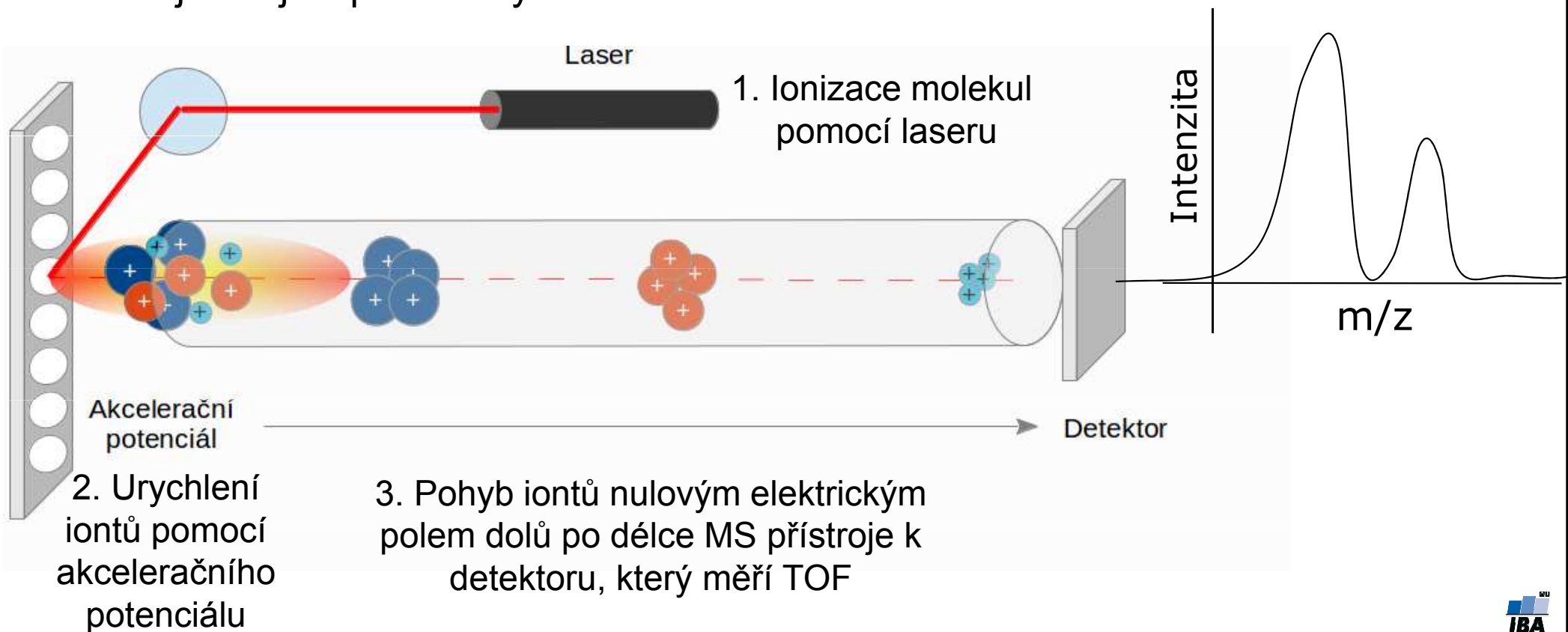


Kapitola IX

Hmotnostná spektrometria

Hmotnostná spektrometria

- Technika používaná pro charakterizaci proteomu v biologickém vzorku (plasma, sérum, . . .)
- Založena na rozdílné hmotnosti peptidů a proteinů
- hmotnostní spektrometr je separuje na základě poměru **hmotnosti k náboji** (anglicky *mass to charge ratio* – m/z , jednotka Dalton), který je specifický pro každou molekulu.
- Nejčastěji používaný - TOF



Hmotnostný spektrometer TOF - princíp

- TOF (time-of-flight) závisí na hmotnosti proteinů nebo přesněji na jejich m/z a představuje sumu těchto časů:

$$TOF = t_a + t_D + t_d$$

t_a je čas letu v akcelerační oblasti,

t_D je čas přeletu v oblasti s nulovým elektrickým polem

t_d je čas detekce

- TOF lze aproximovat pouze pomocí t_D

mass-to-charge ratio je vypočteno podle:

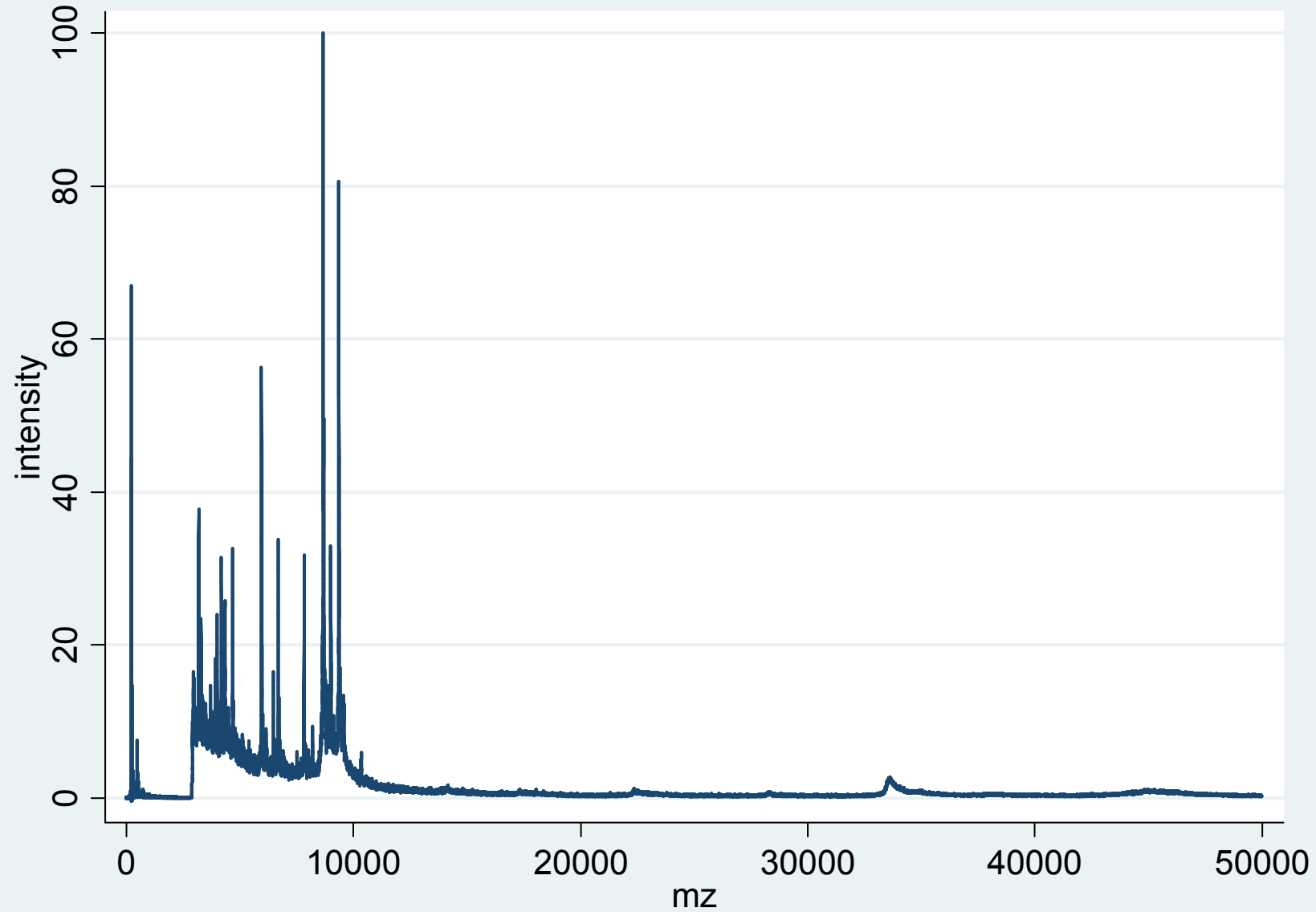
$$m / z = B (t_D - A)^2$$

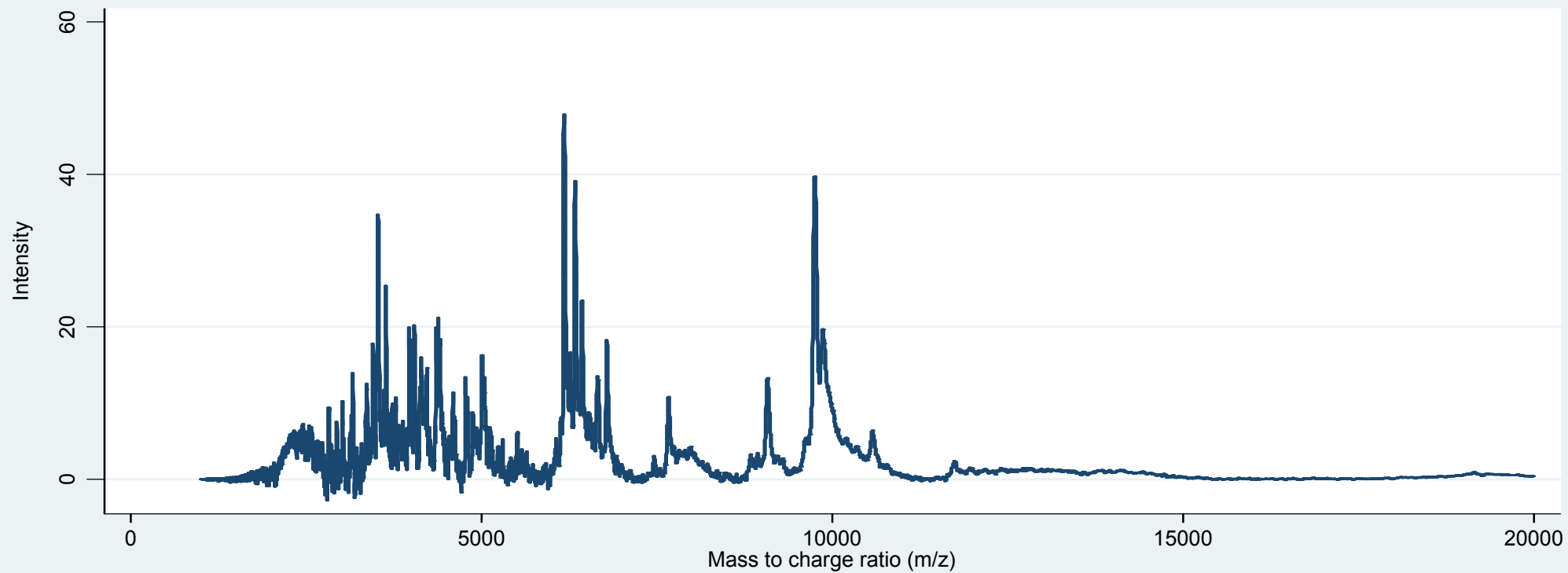
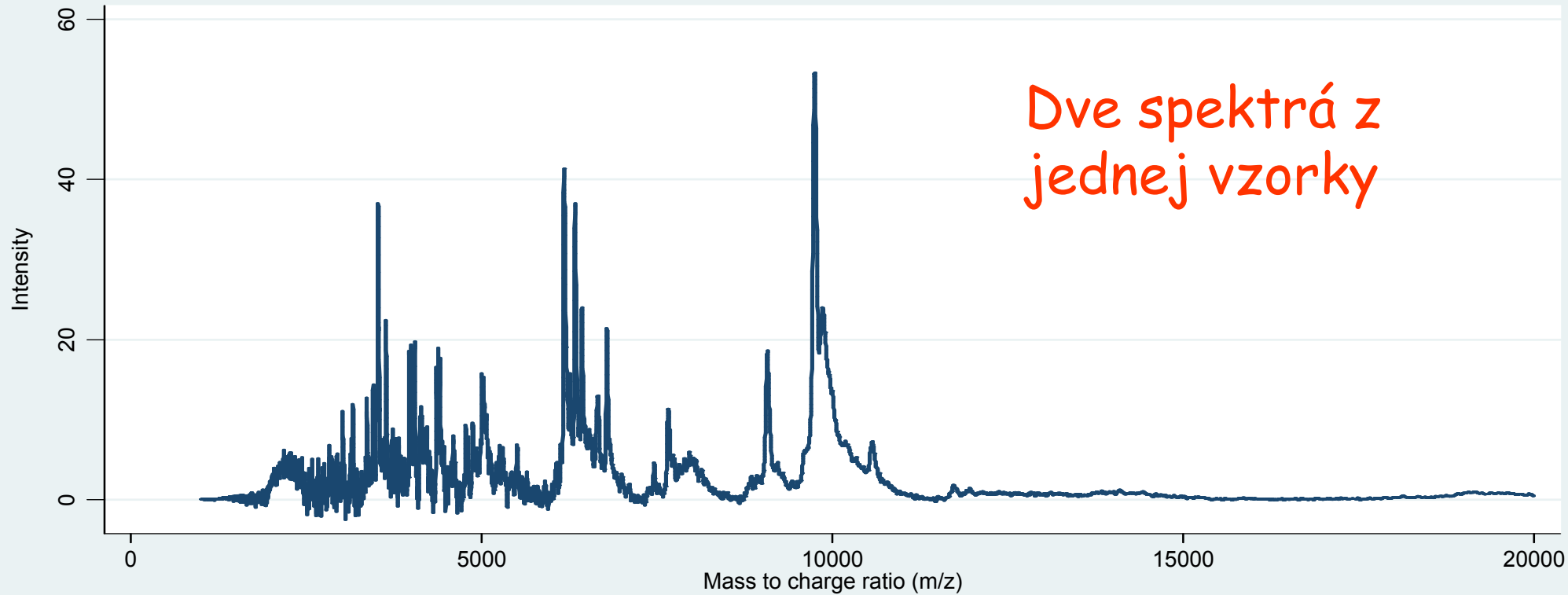
- A a B jsou stanoveny pomocí kalibrace

Hmotnostný spektrometer TOF - druhy

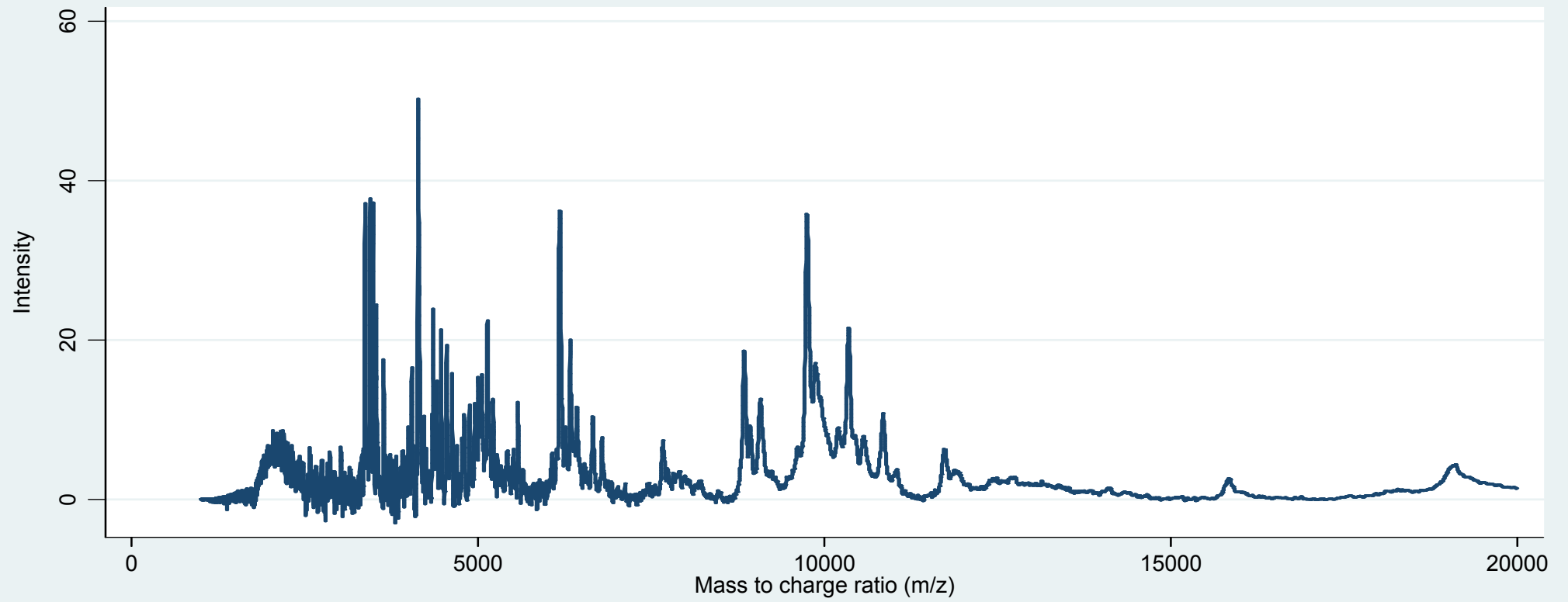
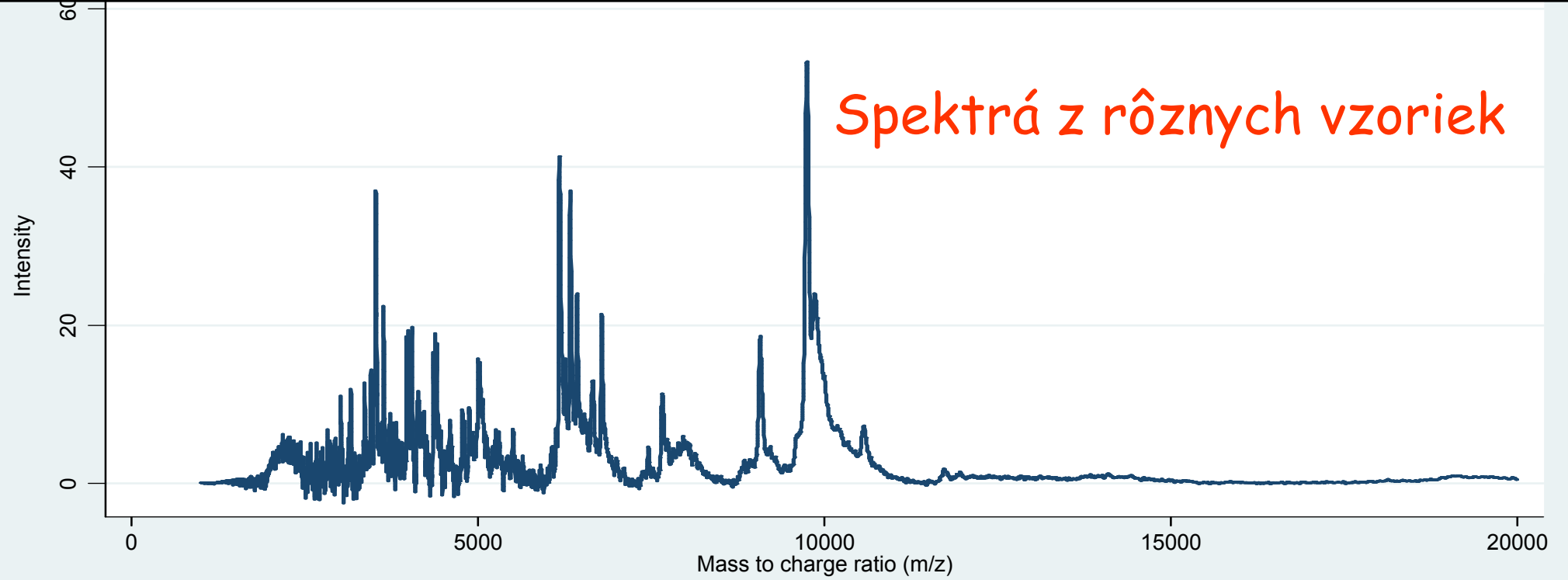
- **Najpoužívanéjšie:**
 - Matrix-Assisted Laser Desorption-Ionisation (MALDI)-TOF
 - Surface-Enhanced Laser Desorption-Ionisation (SELDI)-TOF
- **Způsob uchycení proteinů a ionizace**
 - Proteiny vzorku jsou před samotnou analýzou upevněny na podklad, který se v závislosti od typu hmotnostní spektrometrie liší.
 - Jeho úkolem je také absorbovat energii v ionizátoru a předat ji vzorku a tak usnadnit jeho ionizaci.
 - u MALDI se jedná o energii-absorbující matrici (matrix), co je nejčastěji organická kyselina s aromatickým jádrem
 - SELDI využívá proteinový čip (s několika - obvykle osmi - spoty), opatřen speciálním chromatografickým povrchem, takže se na povrch váží různé proteiny v závislosti na svých chemických vlastnostech a vlastnostech čipu. A až potom dojde k nanesení matrice, která se vzorkem vytvoří krystaly.

Ako vyzerá TOF MS profil



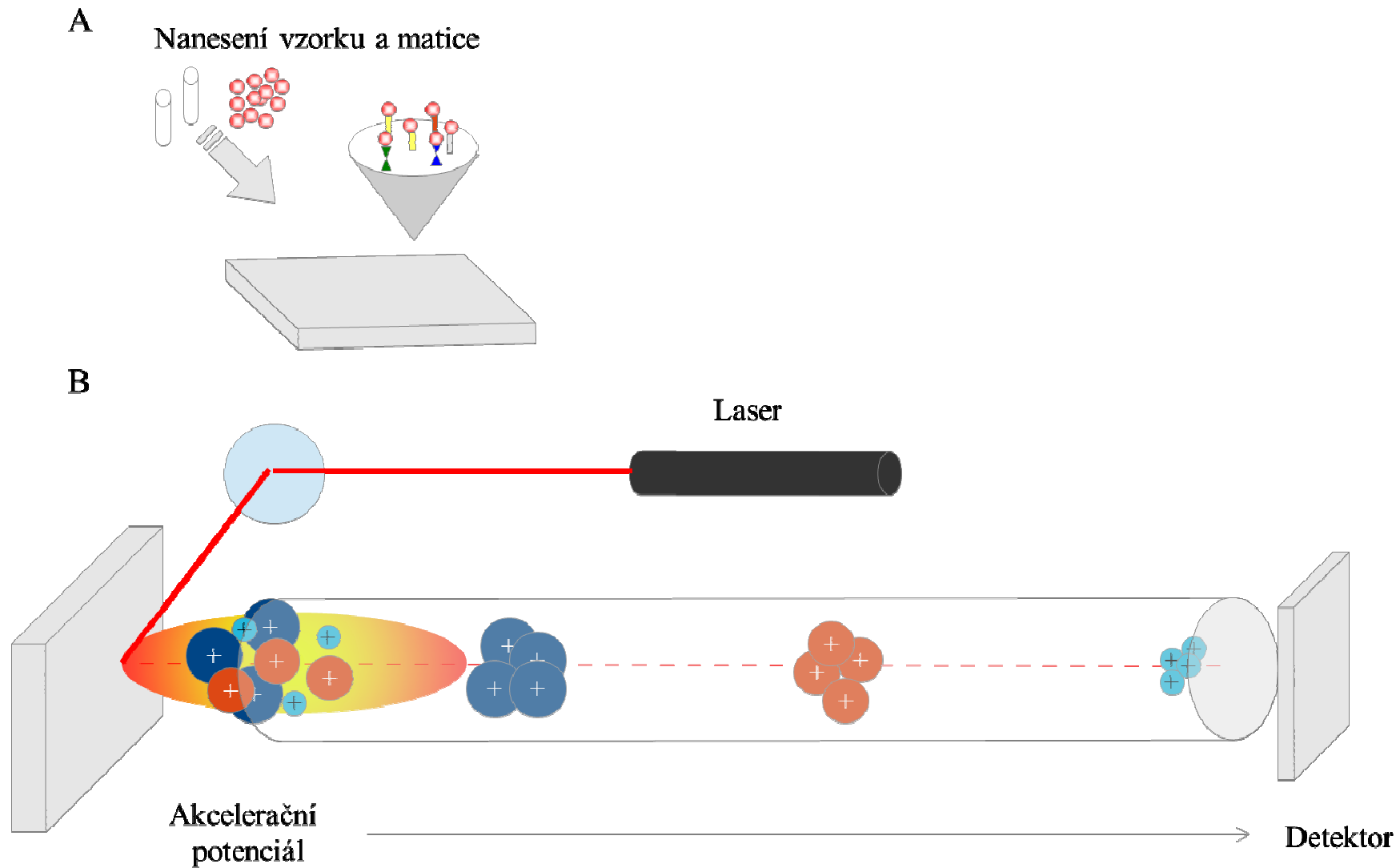


Spektrá z rôznych vzoriek



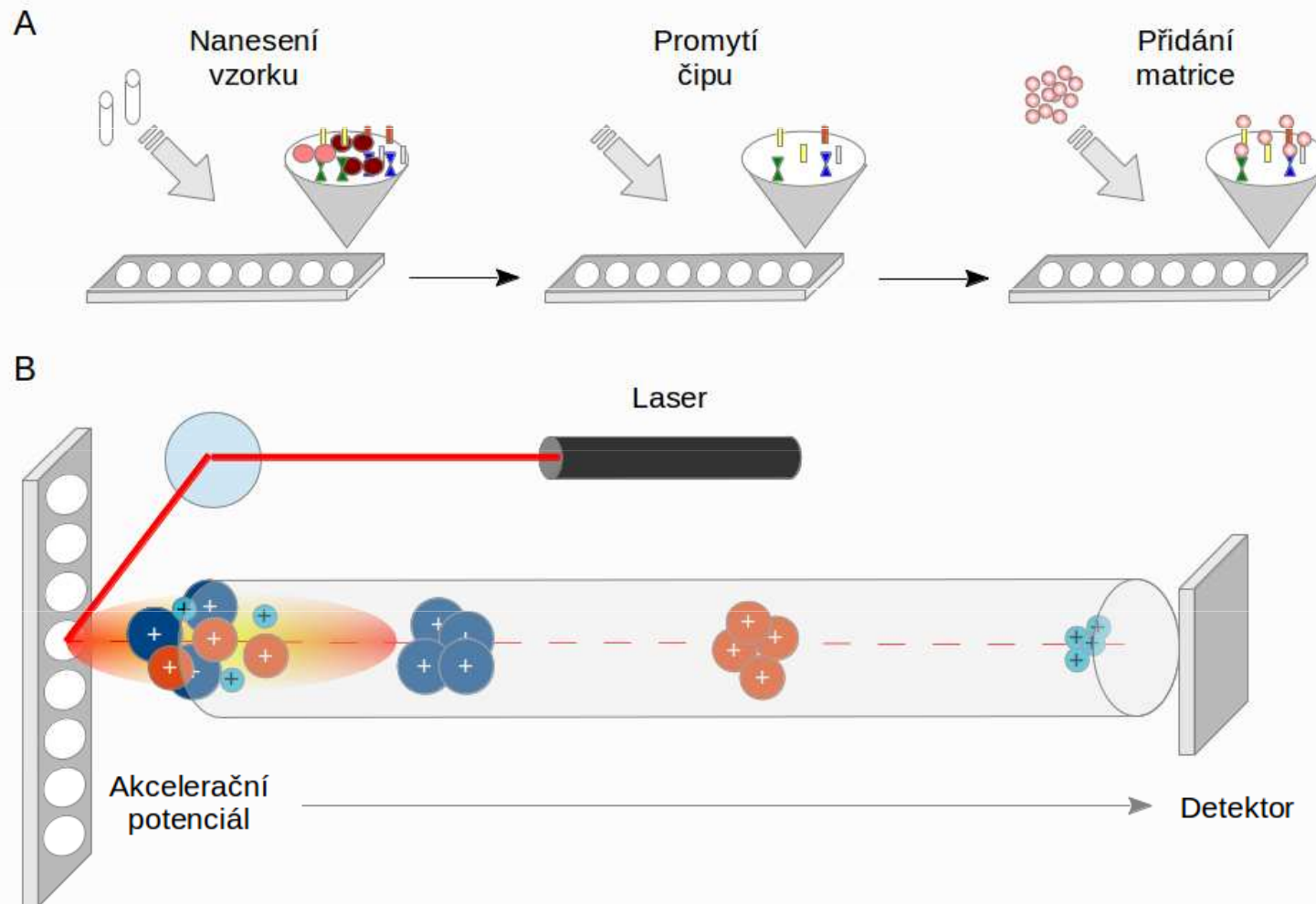
MALDI-TOF

- Matrix-Assisted Laser Desorption-Ionisation - TOF



SELDI-TOF

- **Surface-Enhanced Laser Desorption-Ionisation – TOF**
- Existuje několik druhů čipů (IMAC30, H50, NP20...), které se liší svým aktivním povrchem (anionický, kationický, kovový, normální fáze, hydrofobický, ...) a proto také přednostně vážou jiné molekuly.



Výhoda SELDI:

Možnost odmyt látky, které by jinak ovlivňovali spektrum vzorku (např. močovina používaná k přípravě vzorku, nebo Na^+ ionty přítomné fyziologicky ve vzorcích).

Čipy pre SELDI hmotnostnú spektrometriu

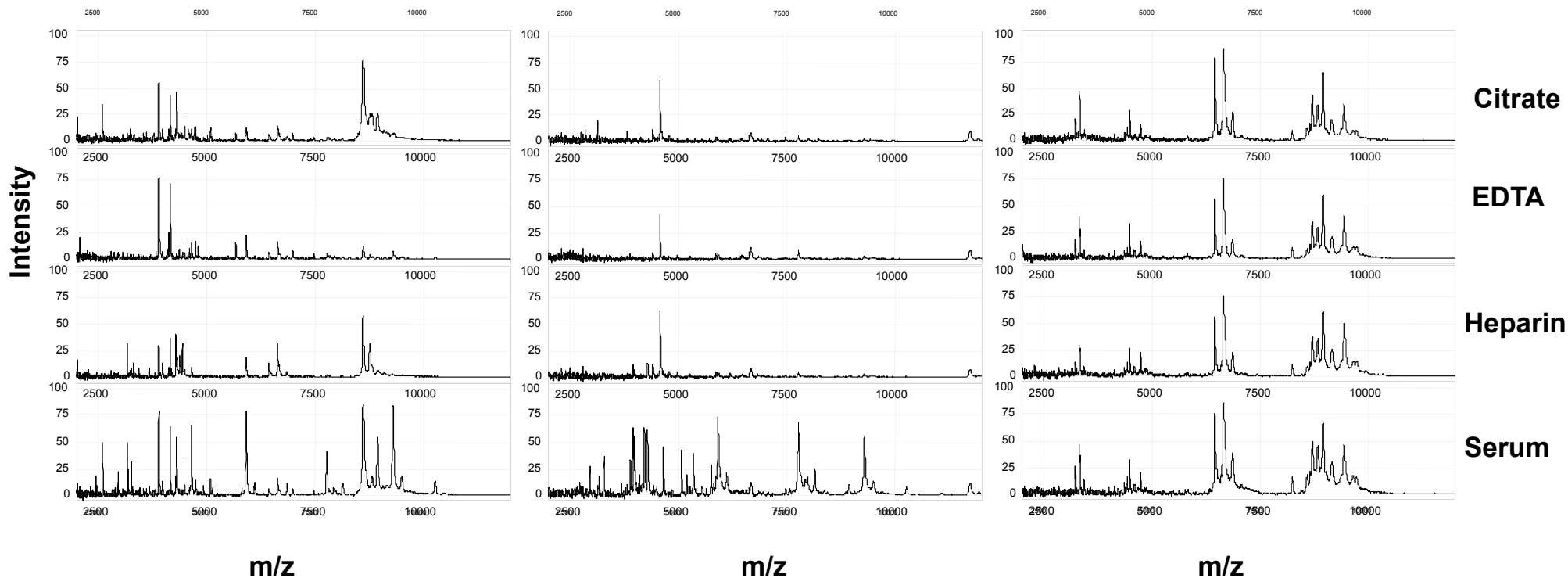
- Kvantitatívne hodnoty proteómu sú takisto ovplyvnené rôznymi zdrojmi variability (experimentálne aj biologické)
- Veľmi veľké rozdiely medzi typmi použitého čipu!

	<u>H50</u>		<u>IMAC30</u>		<u>NP20acid</u>		<u>NP20alkaline</u>	
	<u>N</u>	<u>%</u>	<u>N</u>	<u>%</u>	<u>N</u>	<u>%</u>	<u>N</u>	<u>%</u>
<u>H50</u>	75	100.0	19	47.5	24	52.2	56	59.6
<u>IMAC30</u>	19	25.3	40	100.0	19	41.3	21	22.3
<u>NP20acid</u>	24	32.0	19	47.5	46	100.0	30	31.9
<u>NP20alkaline</u>	56	74.7	21	52.5	30	65.2	94	100.0
<u>separate M/Z</u>	15	20.0%	15	37.5%	12	30.0%	28	29.8%

CM10

IMAC-Cu

H50

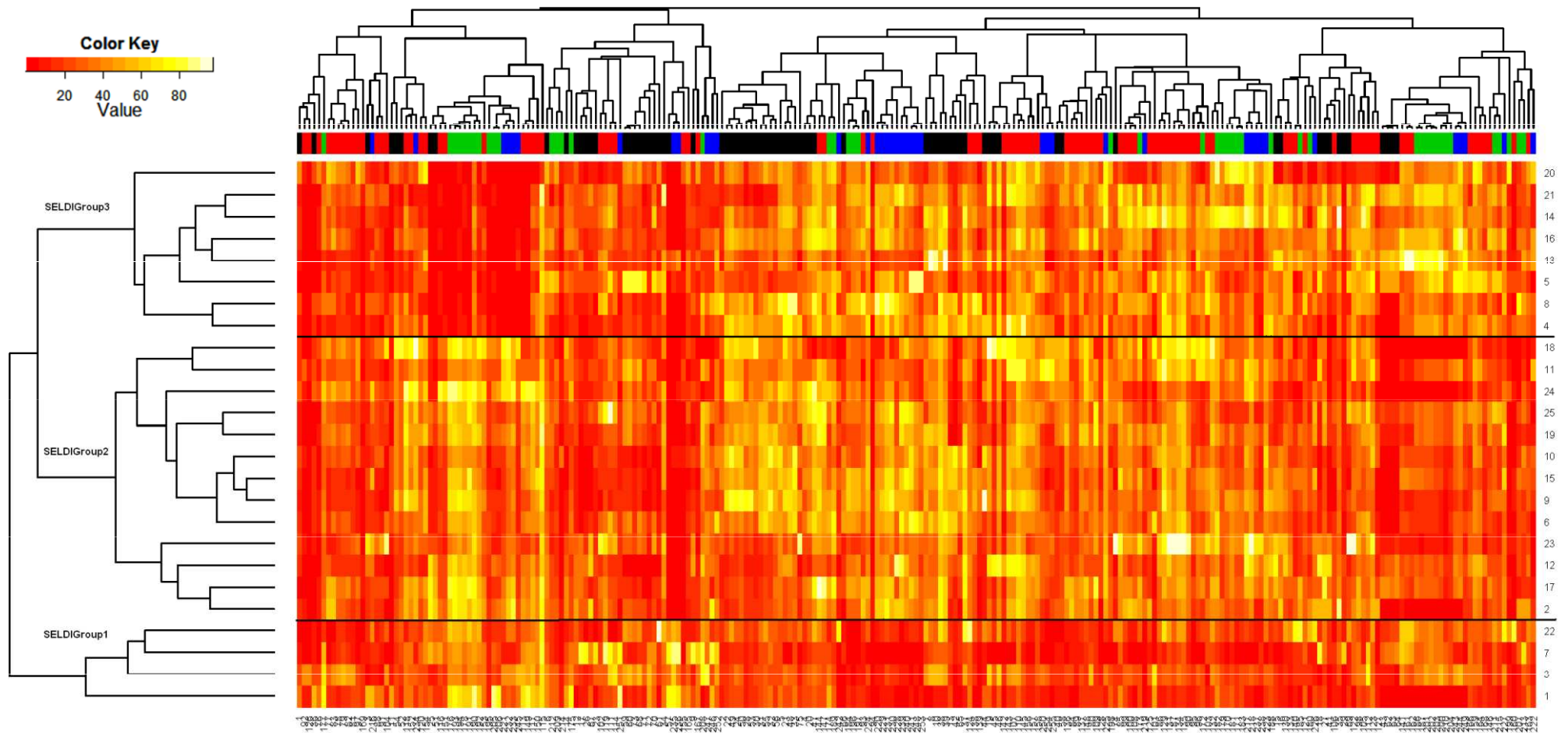


Peaky profilov 3 odlišných SELDI čipov
Vzorky spracované 4 rôznymi spôsobmi
Banks et al, *Clinical Chemistry* 2005

Príklad

- Zhlukovanie profilov tých istých vzoriek zo 4 typov SELDI sklíčok:

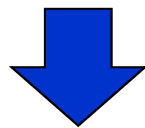
■ IMAC30, ■ H50, ■ NP20zas, ■ NP20kys



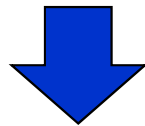
Úpravy dát

- **Kalibrácia**

- *TOF* premenený na škálu m/z pomocou množstva kalibračných proteínov so známou m/z hodnotou. Toto sa deje ešte v laboratóriu



Základné dáta

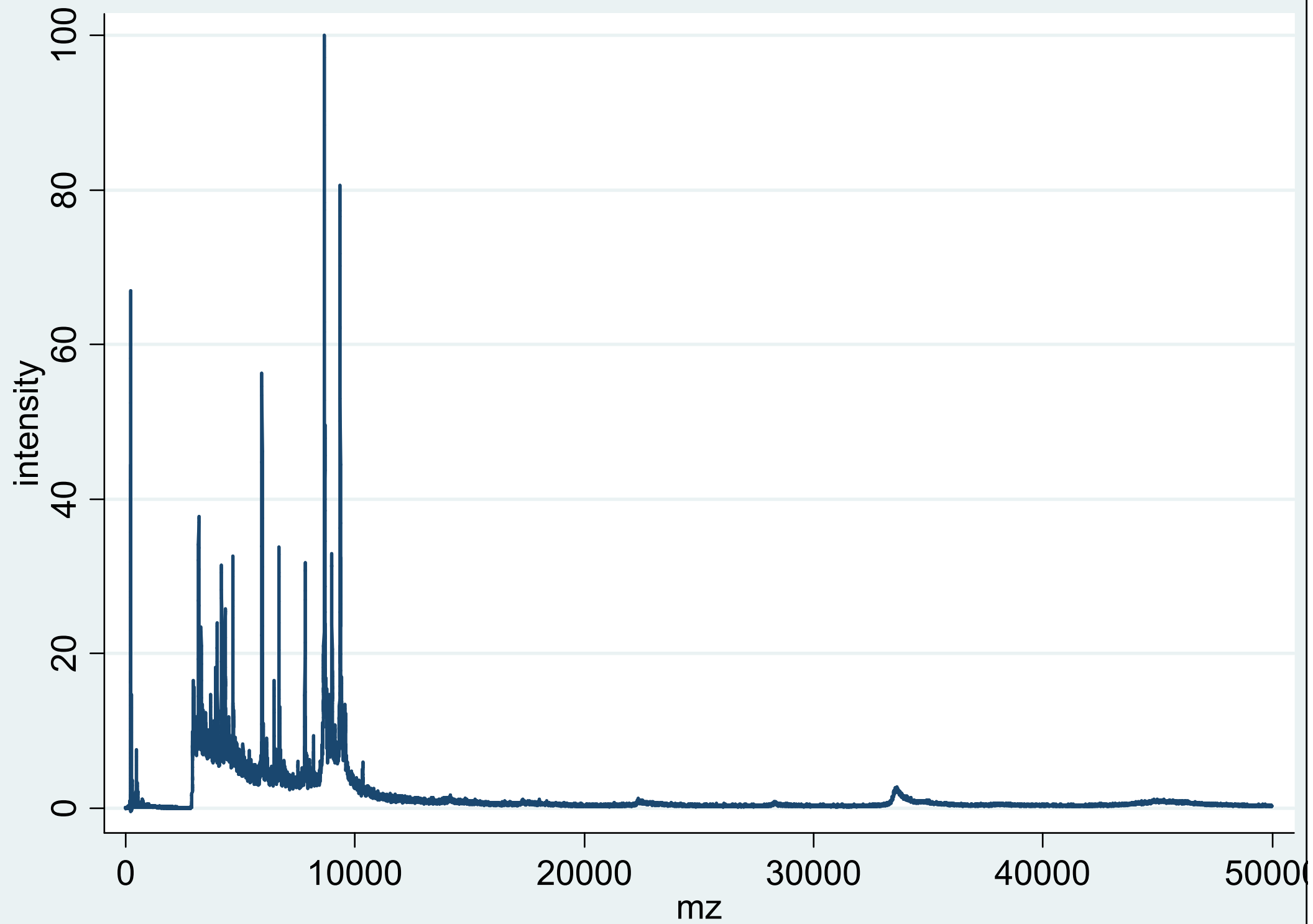


- **Baseline subtraction**

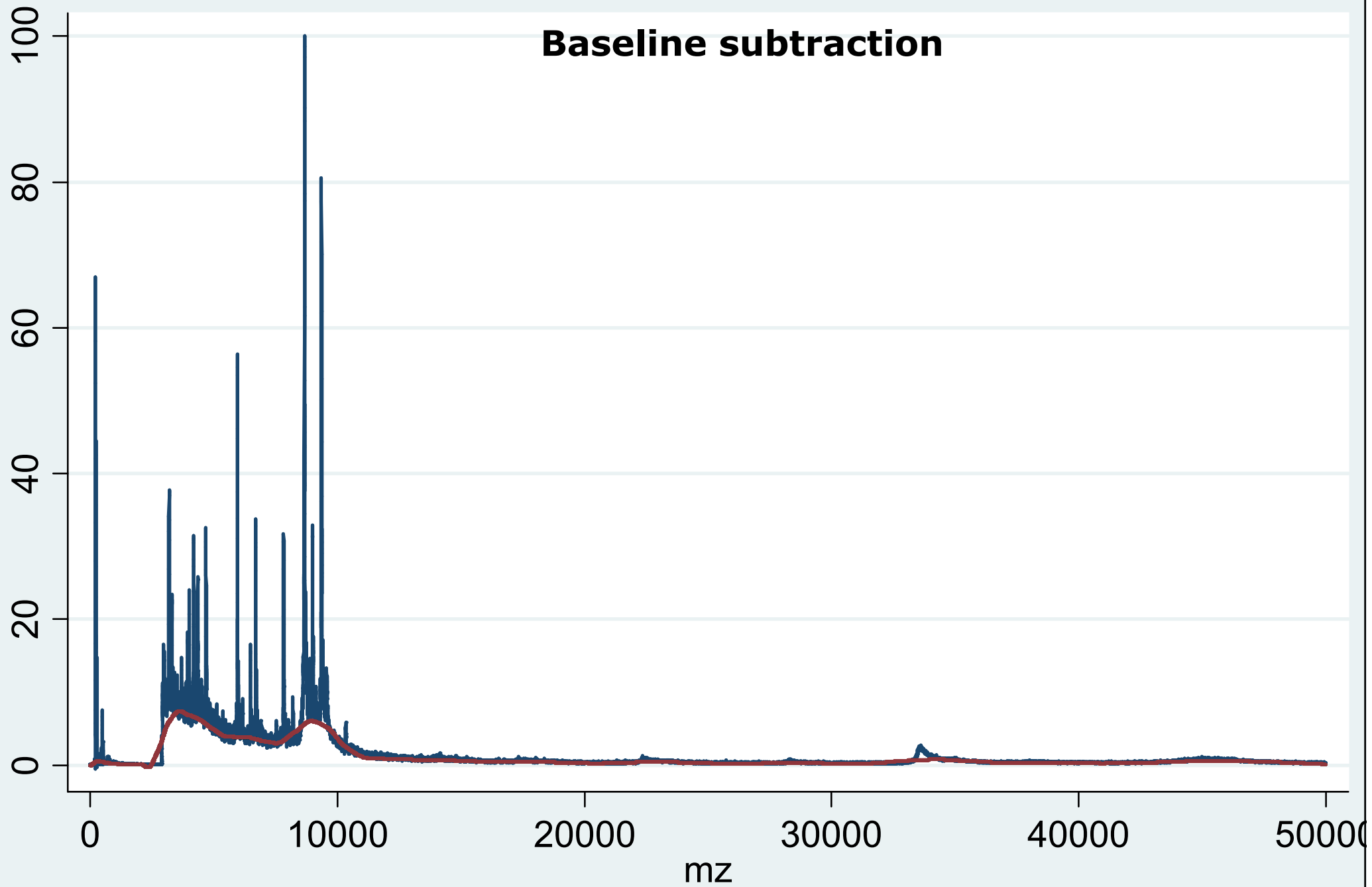
- Odstránenie baseline šumu z profilu, napríklad pomocou **loess**

- **Normalizácia**

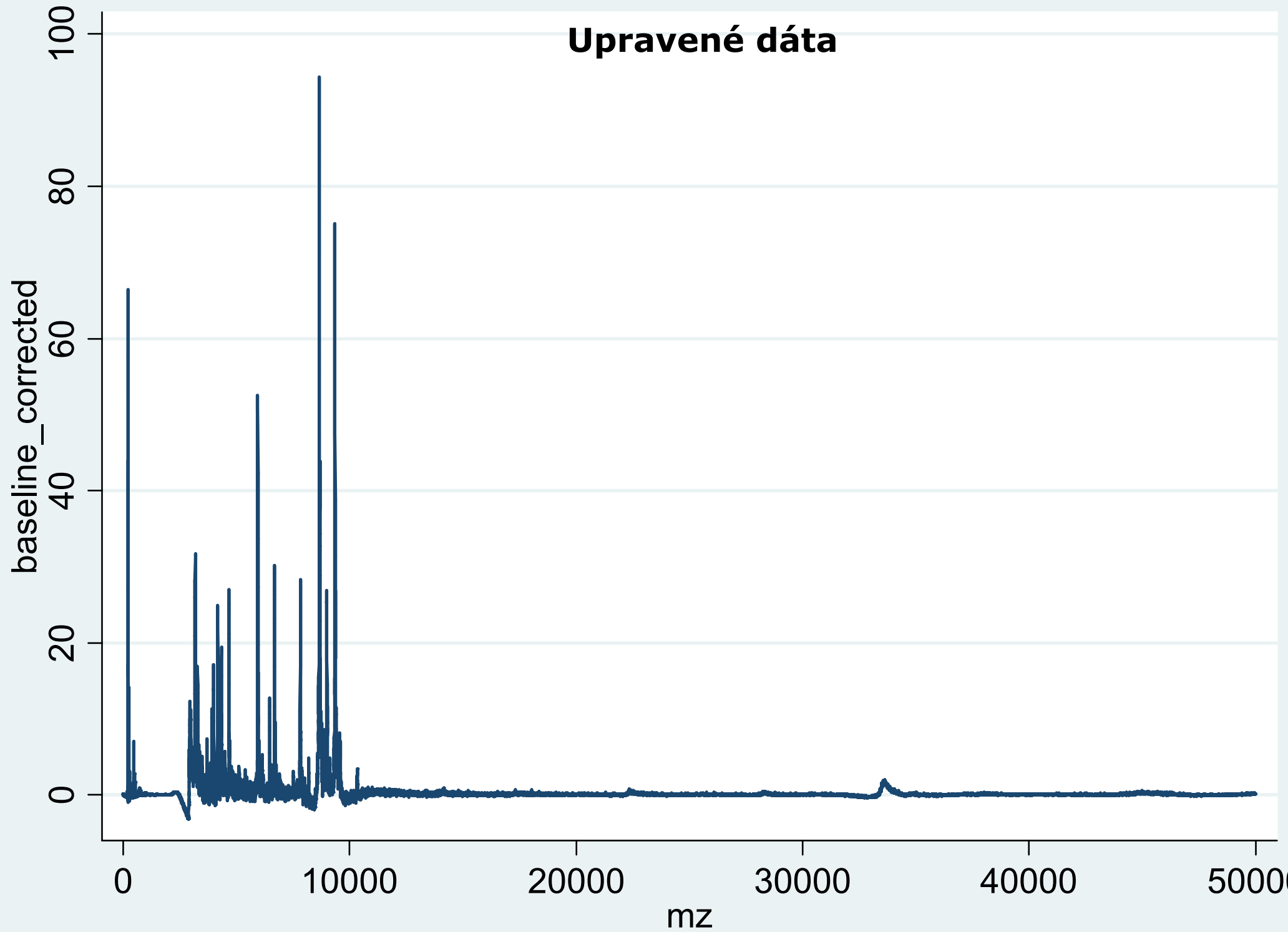
- Aby sme mohli porovnať spektrá medzi vzorkami



Baseline subtraction

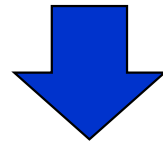


Upravené dáta



Normalizácia

- Odstraňujeme technickú variabilitu (prístrojové chyby, odlišné množstvo vzorky)
- Koncentrácia proteínu sa odhaduje ako plocha pod píkrom (Area Under Curve – AUC)

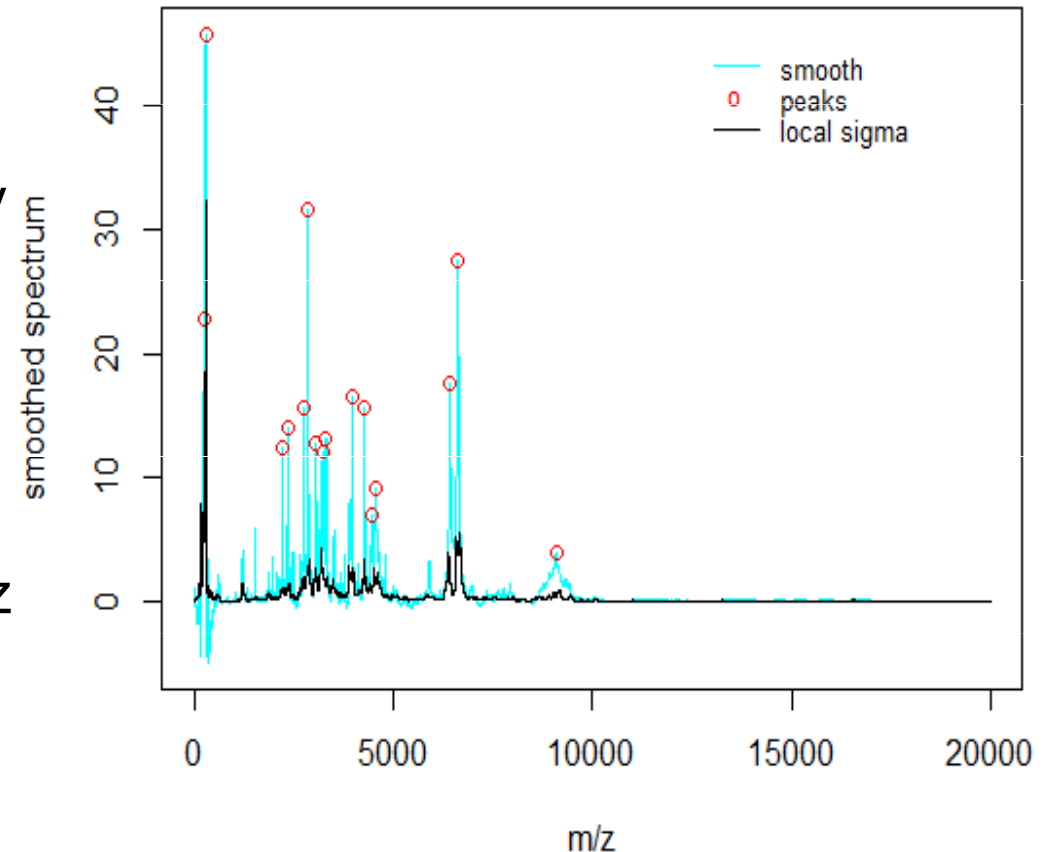


- Normalizácia pomocou *priemernej AUC (TIC – total ion current)*

AUC celého spektra / priemerná AUC všetkých spektier

Detekcia píkov a ich zarovnanie

- Pík ~ peptid/proteín, definuje sa ako lokálne maximum na základe porovnania variability v okolí
- Existujú nepresnosti na x (m/z) a y (množstvo) osiach
- Píky ktoréhokoľvek spektra môžu byť definované ako body ktoré sú maximálne +/- N bodov v okolí m/z
 - first, second, estimated..
- Dôležité je brať do úvahy *signal-to-noise* ratio - peaky musia prekročiť nejakú bežnú hranicu šumu, aby boli označené za píky

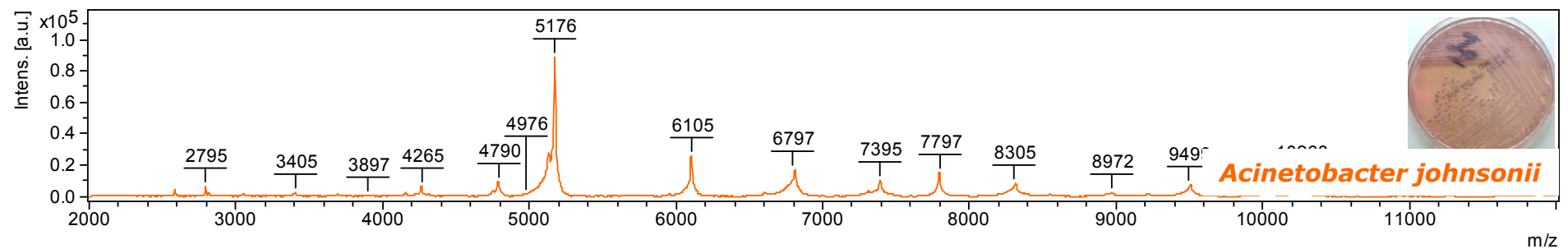
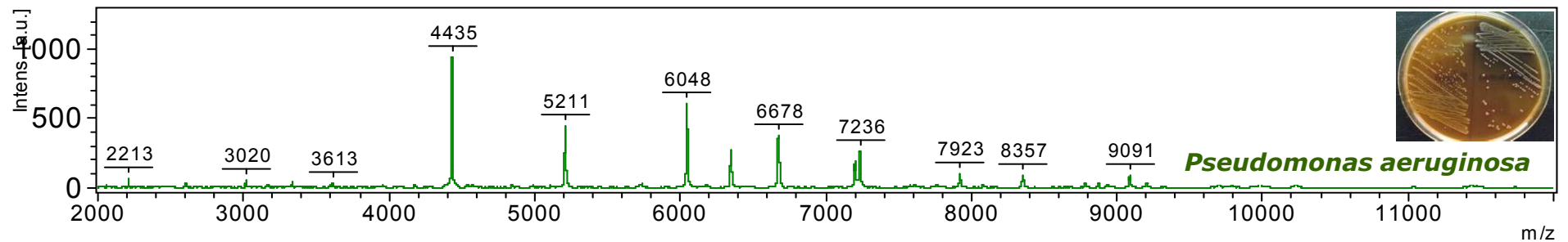
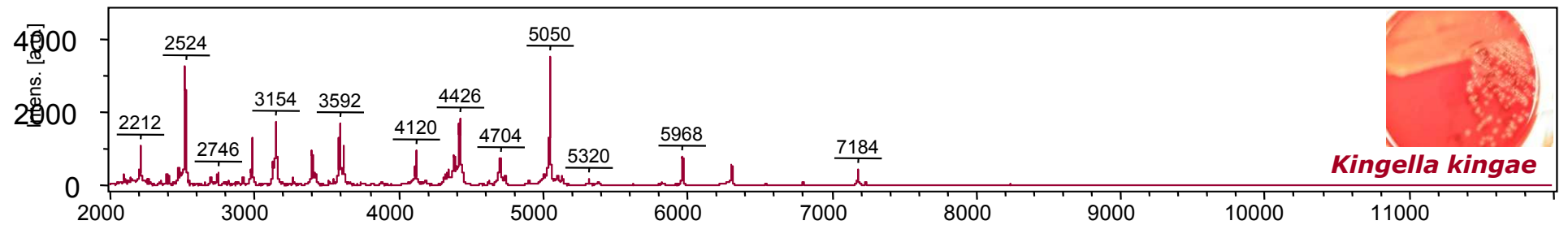
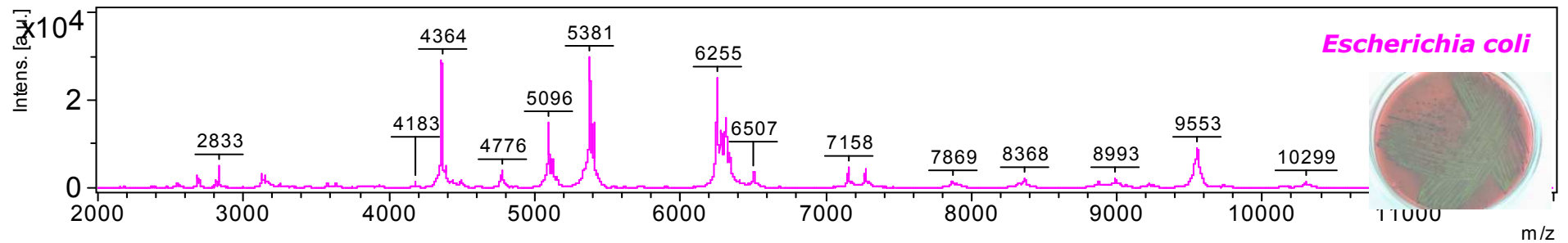


Ako vyzerajú dáta po zarovnaní a detekcii pík

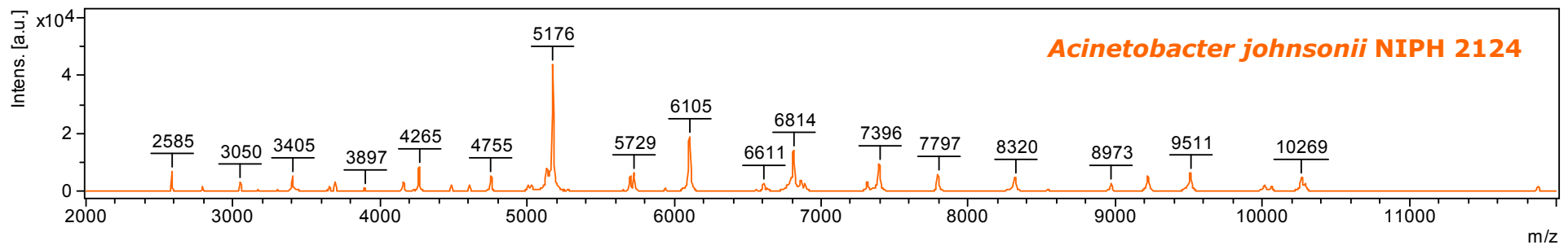
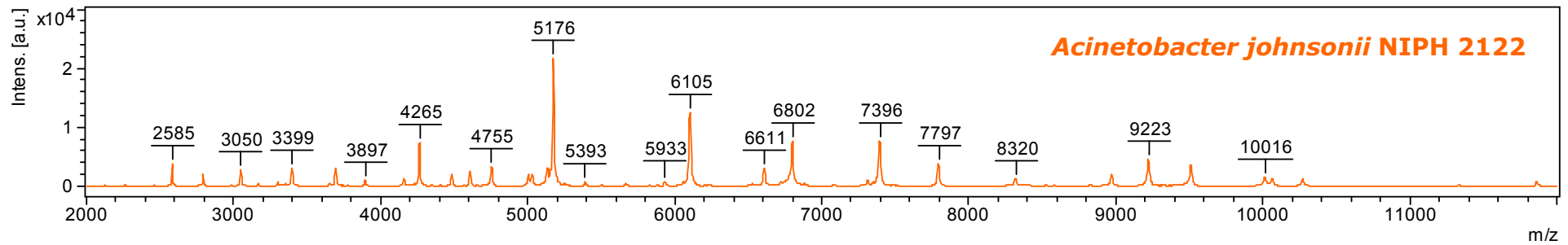
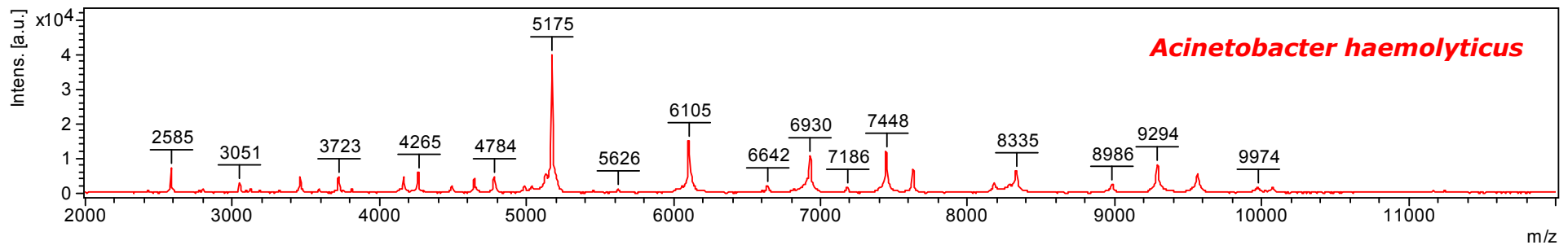
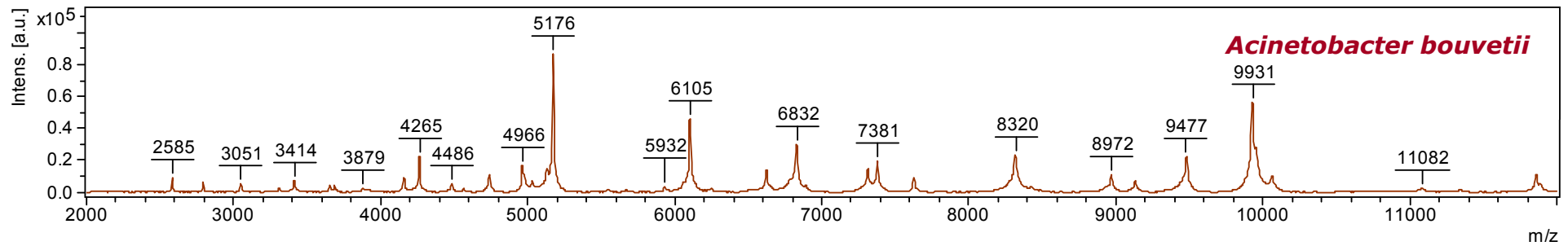
■ SELDI-TOF

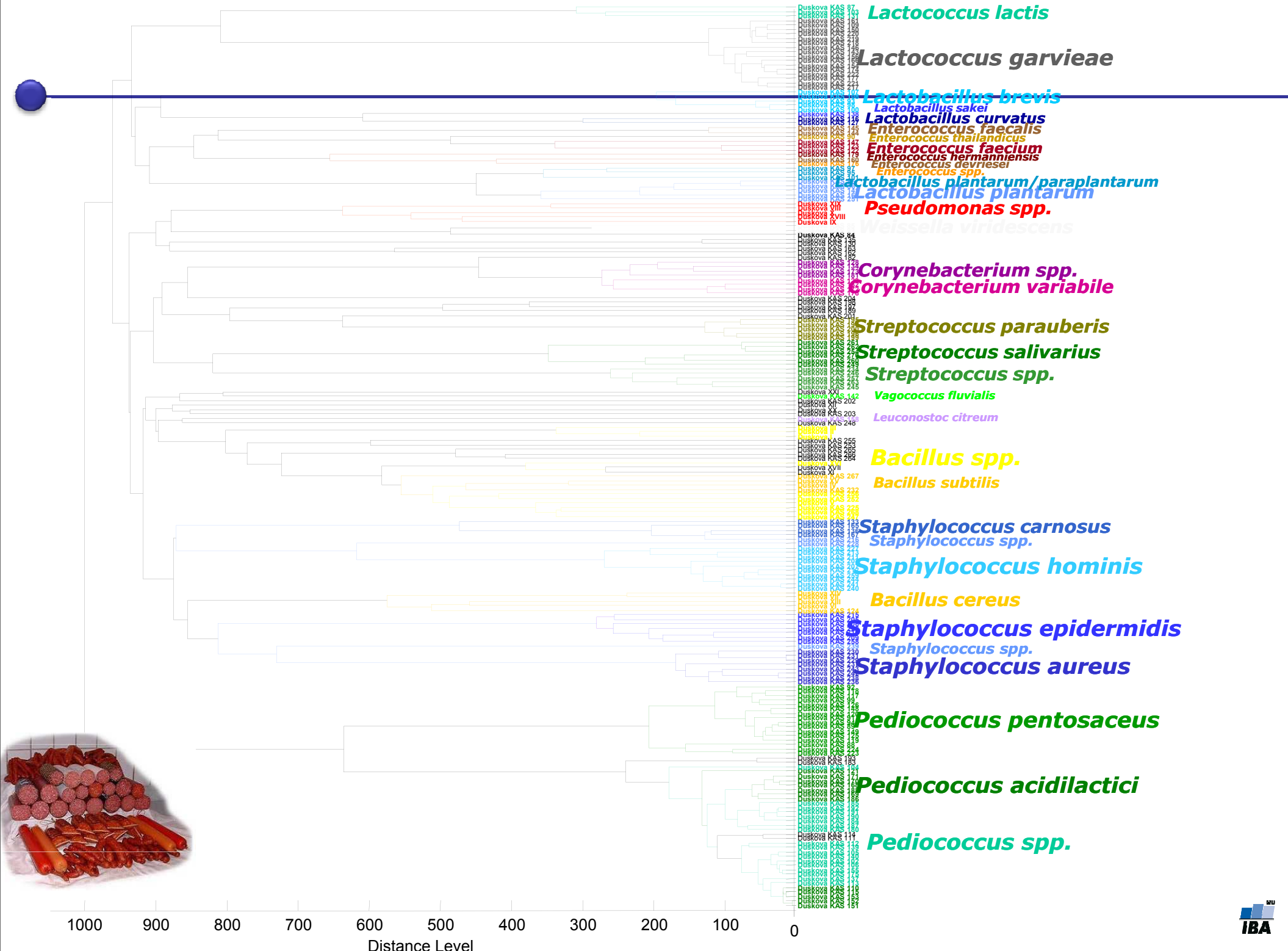
Cluster	Group	Norm. Log Intensity	M/Z	Intensity	Norm. Linear Intensity	Type	Mass Dev.
1	chemoresistentni	0.581550	2392.84	3.058176	30.578211	estimated	0.000007
1	chemoresistentni	-0.072123	2392.84	1.943959	12.984676	estimated	0.000007
1	chemoresistentni	0.023116	2392.84	2.076621	15.079403	estimated	0.000007
1	chemoresistentni	0.160910	2392.84	2.284742	18.365652	estimated	0.000007
1	chemoresistentni	0.199591	2392.84	2.346828	19.345988	estimated	0.000007
1	chemoresistentni	0.161331	2392.82	2.285410	18.376190	first	-0.000004

Aplikácie I – identifikácia baktérii



Aplikácie I – identifikácia baktérii





Lactococcus lactis

Lactococcus garvieae

Lactobacillus brevis

Lactobacillus sakei

Lactobacillus curvatus

Enterococcus faecalis

Enterococcus thailandicus

Enterococcus faecium

Enterococcus hermamiensis

Enterococcus devriesei

Enterococcus spp.

Lactobacillus plantarum/paraplantarum

Lactobacillus plantarum

Pseudomonas spp.

Weissella viridescens

Corynebacterium spp.

Corynebacterium variabile

Streptococcus parauberis

Streptococcus salivarius

Streptococcus spp.

Vagococcus fluvialis

Leuconostoc citreum

Bacillus spp.

Bacillus subtilis

Staphylococcus carnosus

Staphylococcus spp.

Staphylococcus hominis

Bacillus cereus

Staphylococcus epidermidis

Staphylococcus spp.

Staphylococcus aureus

Pediococcus pentosaceus

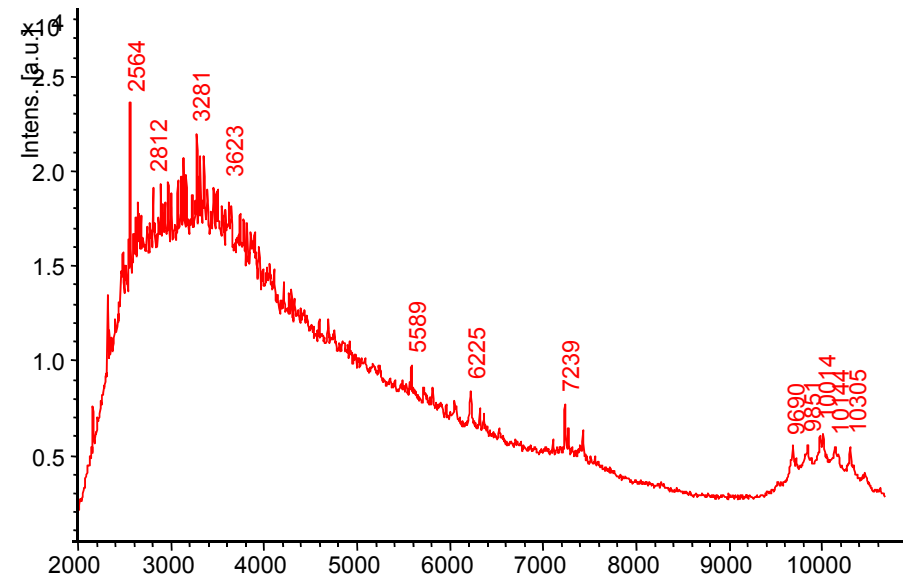
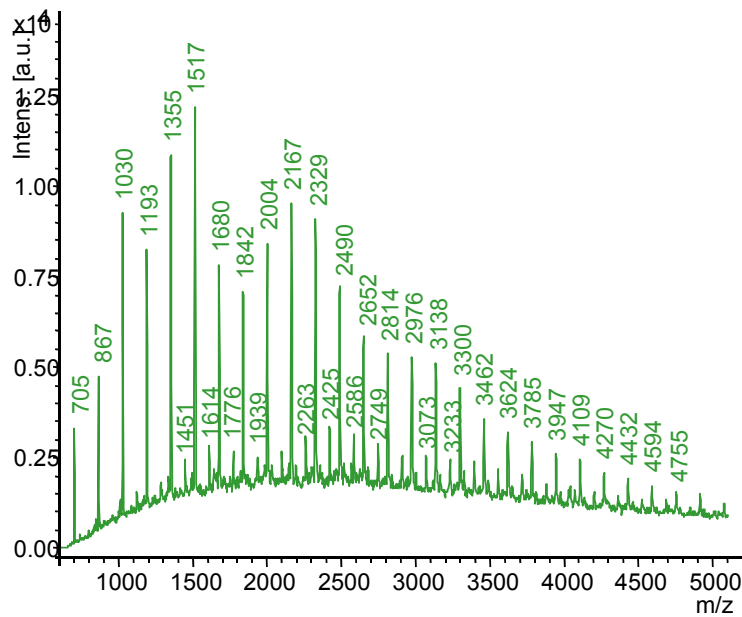
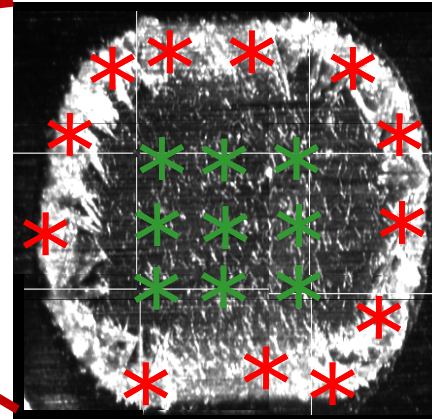
Pediococcus acidilactici

Pediococcus spp.

1000 900 800 700 600 500 400 300 200 100 0

Distance Level

Aplikácia II

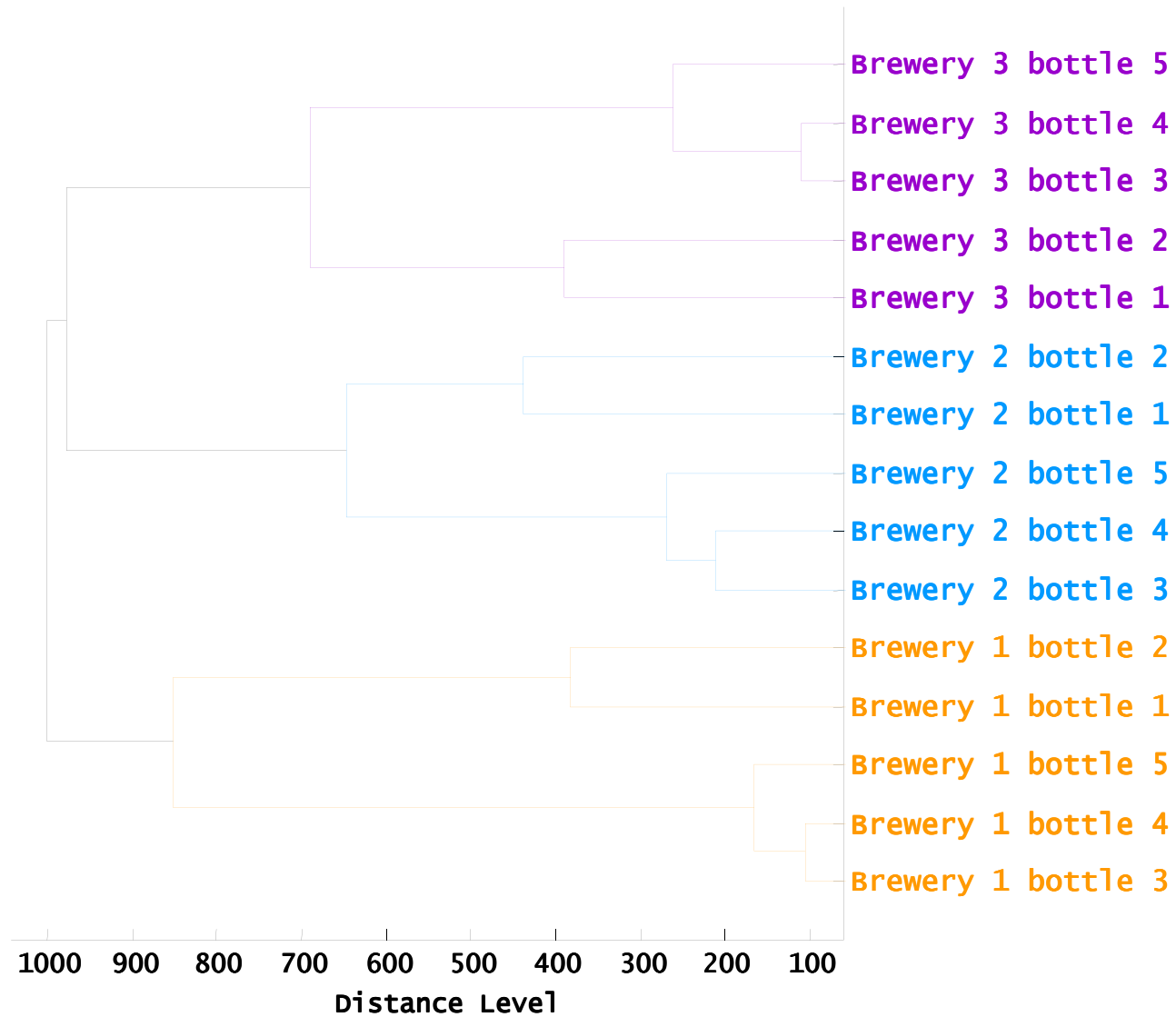


MALDI-TOF MS fingerprint containing maltoligosaccharides

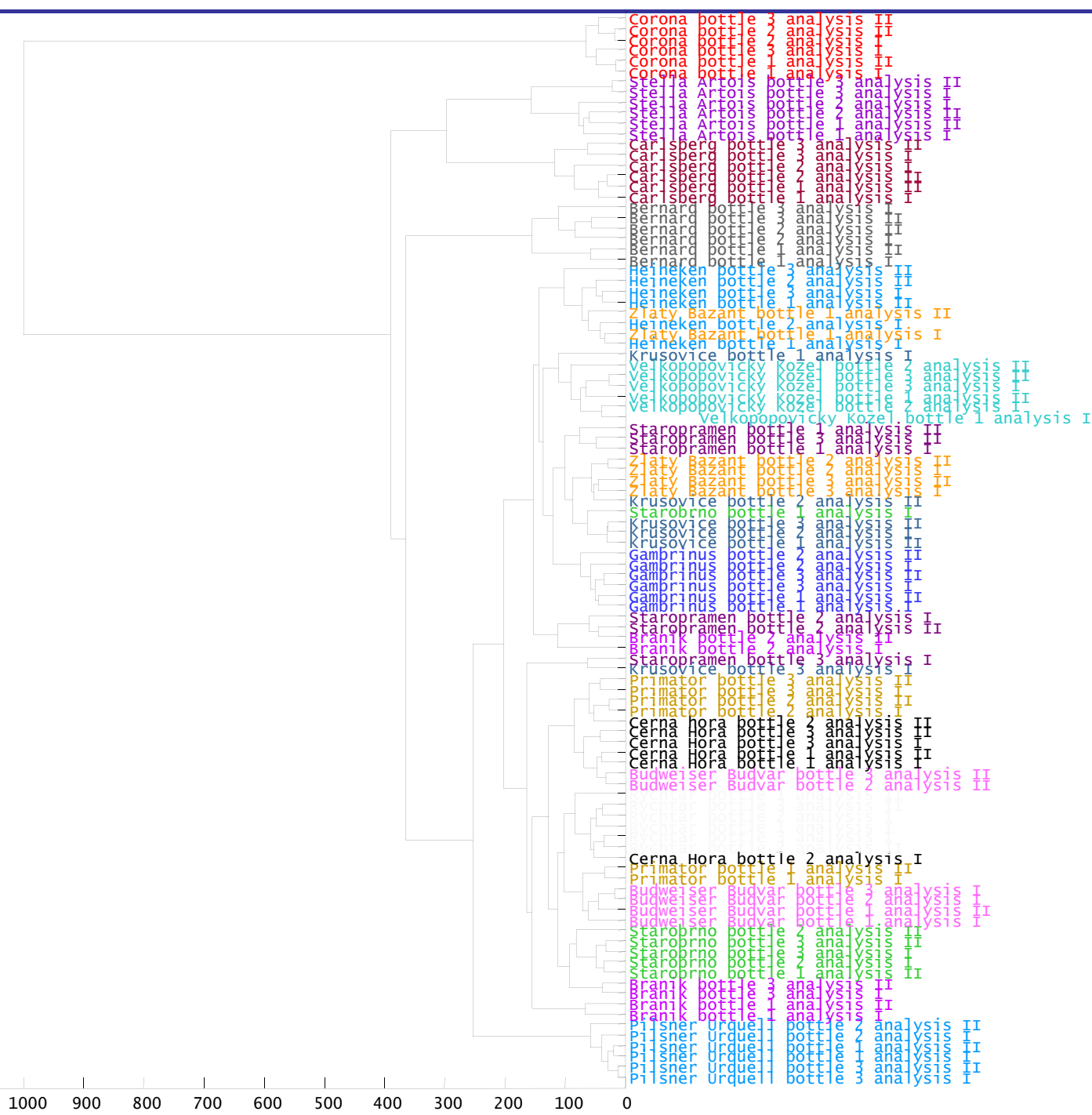
MALDI-TOF MS fingerprint containing proteins

Šedo et al., 2012

Aplikácia II



Aplikácia II



Liquid Chromatography MS/MS

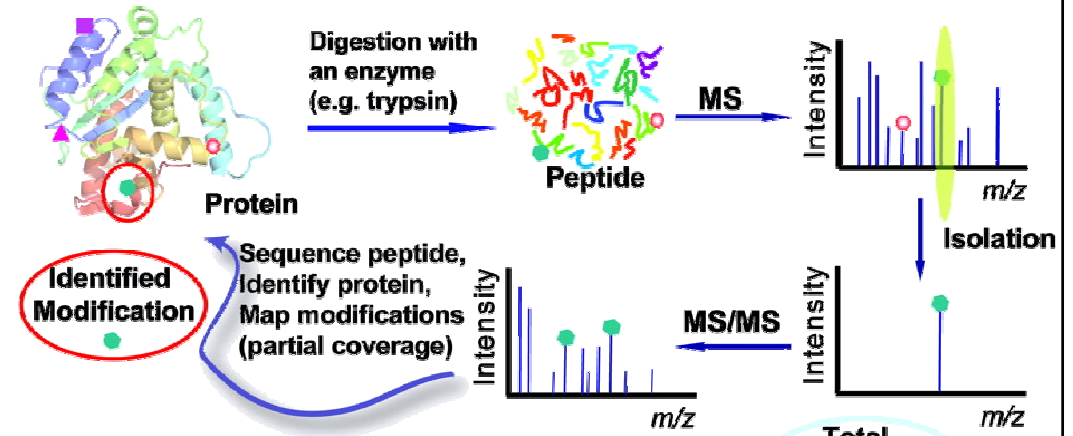
Z pohľadu centrálného laboratória

Dva hlavné prístupy

Bottom-up proteomika

- protein → peptidy
- analýza peptidov pomocou MS
- fragmentácia peptidov v MS

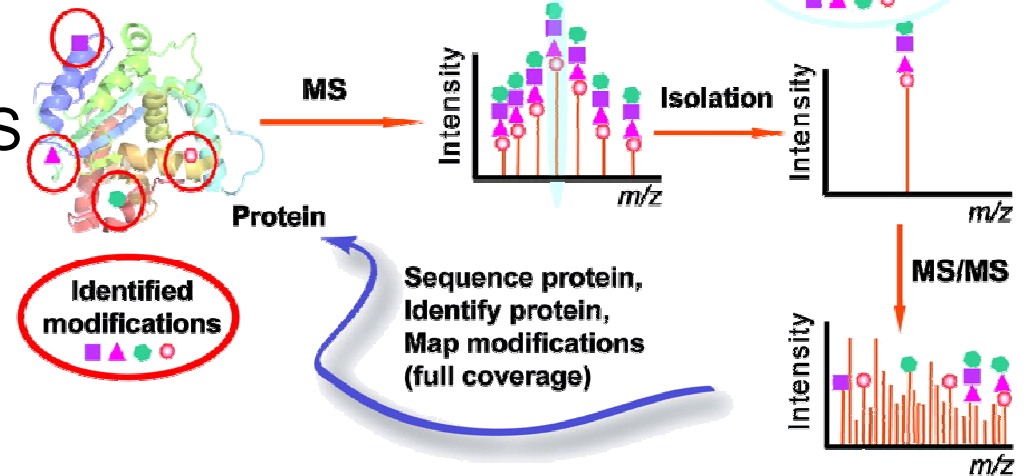
A Bottom-up MS approach



Top-down proteomika

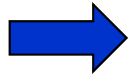
- analýza intaktných proteínov v MS
- fragmentácia proteínu v MS

B Top-down MS approach

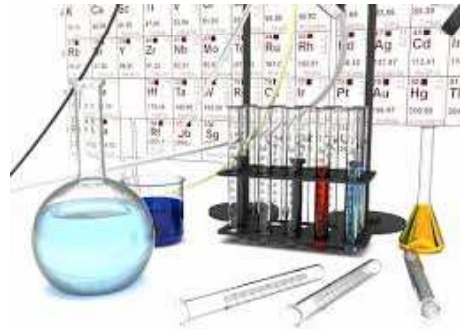


Experiment

Biologický materiál



Izolácia, štiepenie

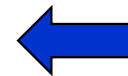


LC MS/MS systém



Identifikácie proteínov

Identification and relative comparison summary										
Identification summary			Identification summary				1 Identification threshold (pept)			
Prot Group	Accession	Description	MW [kDa]	calc. pI	Sum(Coverage)	Presence of the protein group in individ				
						Line 1	Line 2	Line 3	Line 4	
6	1	Q09666	Neuroblast differentiation-ass	629	6.2	82%	Y	Y	Y	Y
417	2	Q15149-4	Isoform 4 of Plectin OS=Hom	516	5.8	58%	Y	Y	Y	Y
418	3	Q15149-3	Isoform 3 of Plectin OS=Hom	518	5.7	58%	Y	Y	Y	Y
885	4	P09211;P0921	Glutathione S-transferase P	23	5.6	63%	Y	Y	Y	Y
886	5	O75369-2	Isoform 2 of Filamin-B OS=Hc	276	5.8	47%	Y	Y	Y	Y
951	5	O75369	Filamin-B OS=Homo sapiens	278	5.7	46%	Y	Y	Y	Y
952	5	O75369-9	Isoform 9 of Filamin-B OS=Hc	277	5.7	46%	Y	Y	Y	Y
		Sequence	# PSMs	# Proteins	# Protein Groups					
974		AAGSGELGVTMK	8	6	1					
975		AAGSGELGVTMKGPK	2	3	1					
976		ADIEMPFDPSK	4	3	1					
977		AEVSIQNKDGYAVTYVPLTA	2	3	1					
978		AGGPGLER	2	9	2					
979		AGLAPLEVR	2	3	1					
980		AGDPTSYTECDK	6	3	4					
981										

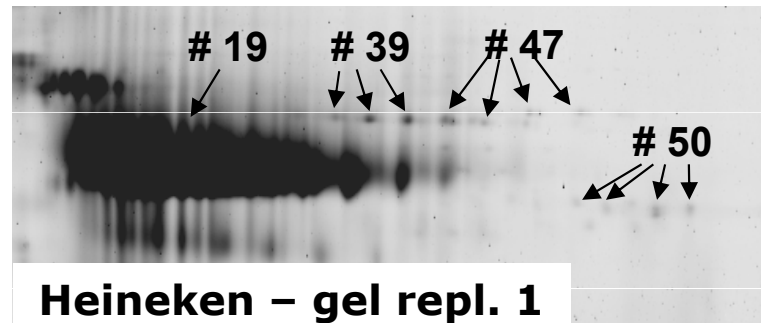


Spracovanie dát

Štatistická a bioinformatická analýza

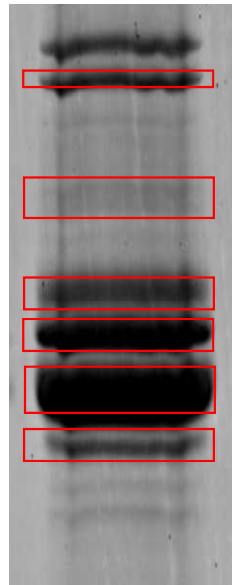
Typy experimentov I

- Podľa komplexity vzorky (počtu očakávaných proteínov):
 - Identifikácia proteínu(ov) z gelu 2D-SDS-PAGE (jednotky proteínov, desiatky vzoriek)



Typy experimentov I

- Podľa komplexity vzorky (počtu očakávaných proteínov):
 - Identifikácia proteínu(ov) z gelu 2D-SDS-PAGE (jednotky proteínov, desiatky vzoriek)
 - Identifikácia proteínov v prúžkoch z gelu 1D-SDS-PAGE (jednotky až stovky proteínov, jednotky vzoriek)



Typy experimentov I

- Podľa komplexity vzorky (počtu očakávaných proteínov):
 - Identifikácia proteínu(ov) na gele 2D-SDS-PAGE (jednotky proteínov, desiatky vzoriek)
 - Identifikácia proteínov v prúžkoch z gelu 1D-SDS-PAGE (jednotky až stovky proteínov, jednotky vzoriek)
 - Relatívne porovnanie dvou bun. linií – wild-type vs. mutant (tisíce proteínov, jednotky vzoriek)

Typy experimentov II

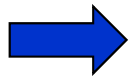
- Podľa použitého značenia:
 - *label-free* – bez značenia
 - SILAC – stabilné izotopové značenie aminokyselín v bunkových kultúrach (2-3 vzorky)
 - iTRAQ – izobarické značky pre relatívnu a absolútnu kvantifikáciu (4 alebo 8 vzoriek)

Získané informácie

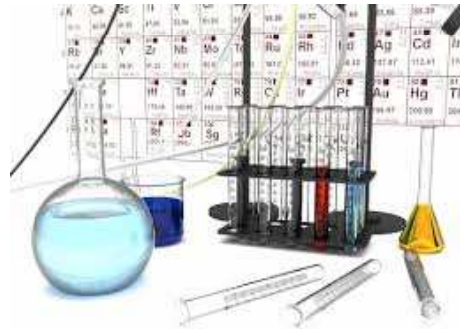
- Kvalitatívne – primárny cieľ
 - Aké proteíny sa vo vzorke vyskytujú?
- Kvantitatívne
 - Hodnotenie koncentrácie proteínov (PSMs, AUC)
- Modifikácie
 - Výskyt posttranslačných modifikácií (fosforylácie)

Experiment

Biologický materiál



Izolácia, štiepenie

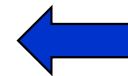


LC MS/MS systém



Identifikácie proteínov

Identification and relative comparison summary										
Identification summary			Identification threshold (pept)							
Prot Group	Accession	Description	MW [kDa]	calc. pI	Sum(Coverage)	Line 1	Line 2	Line 3	Line 4	
1	Q09666	Neuroblast differentiation-ass	629	6.2	82%	Y	Y	Y	Y	
2	Q15149-4	Isoform 4 of Plectin OS=Hom	516	5.8	58%	Y	Y	Y	Y	
3	Q15149-3	Isoform 3 of Plectin OS=Hom	518	5.7	58%	Y	Y	Y	Y	
4	P09211;P0921	Glutathione S-transferase P	23	5.6	63%	Y	Y	Y	Y	
5	O75369-2	Isoform 2 of Filamin-B OS=Hc	276	5.8	47%	Y	Y	Y	Y	
5	O75369	Filamin-B OS=Homo sapiens	278	5.7	46%	Y	Y	Y	Y	
5	O75369-9	Isoform 9 of Filamin-B OS=Hc	277	5.7	46%	Y	Y	Y	Y	
Sequence	# PSMs	# Proteins	# Protein Groups							
AAGSGELGVTMK	8	6	1							
AAGSGELGVTMKGPK	2	3	1							
ADIEMPFDPSK	4	3	1							
AEVSIQNKDGTAVTYVPLTA	2	3	1							
AGGPGLER	2	9	2							
AGLAPLEVR	2	3	1							
AGDPTSVTECDPK	6	3	1							



Spracovanie dát

Štatistická a bioinformatická analýza

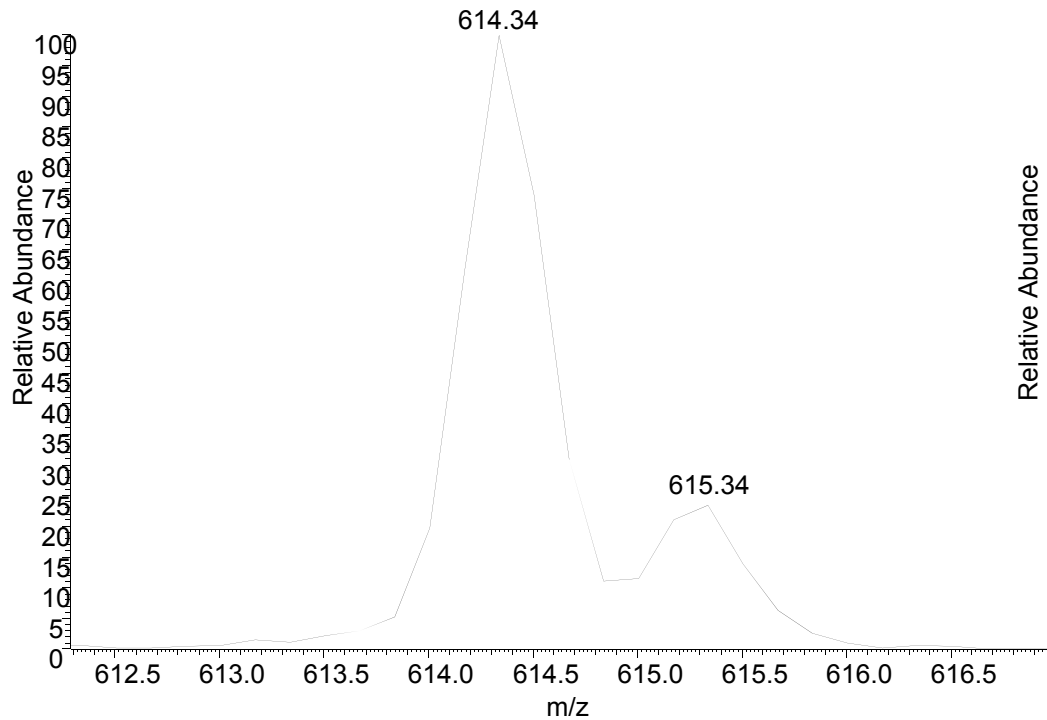
Spracovanie dát

1. úprava hrubých dát (MS/MS i MS)
2. príprava dát pre databázové vyhľadávanie
3. príprava proteínovej databázy, databázové vyhľadávanie
4. spracovanie výsledkov z pohľadu FDR
5. výber peptidových identifikácií (PSMs)
6. rekonštrukcia zoznamu „identifikovaných“ proteínov
7. interpretácia proteínového zoznamu(ov)

Úprava hrubých dat

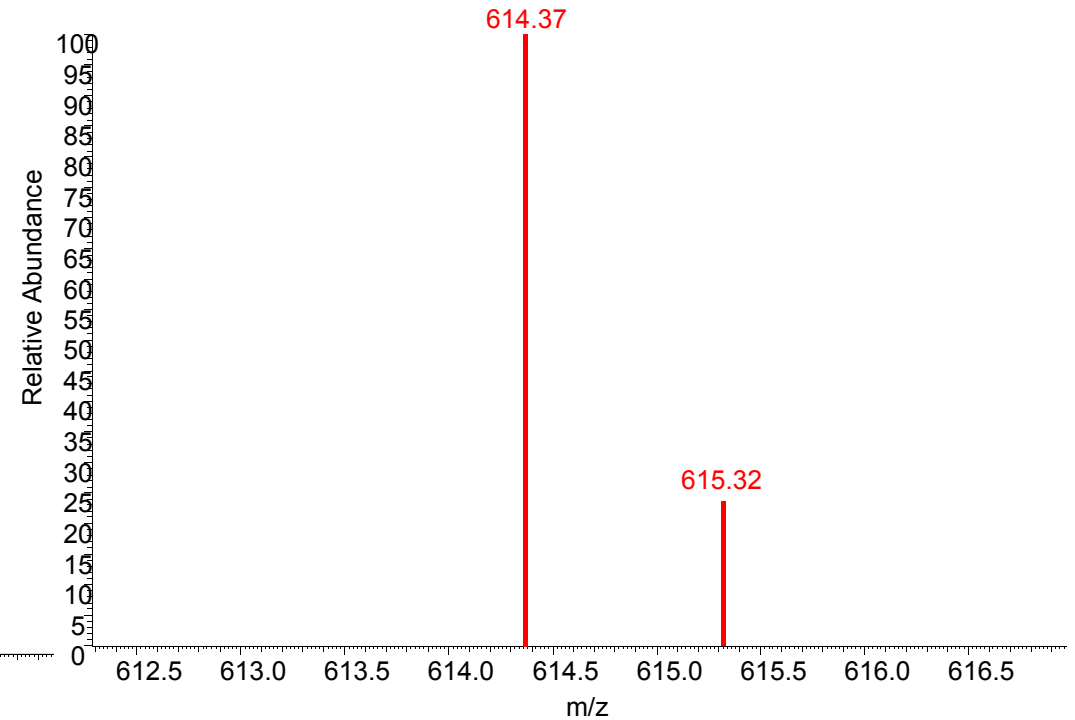
Profilové spektrum

- Získané z experimentu

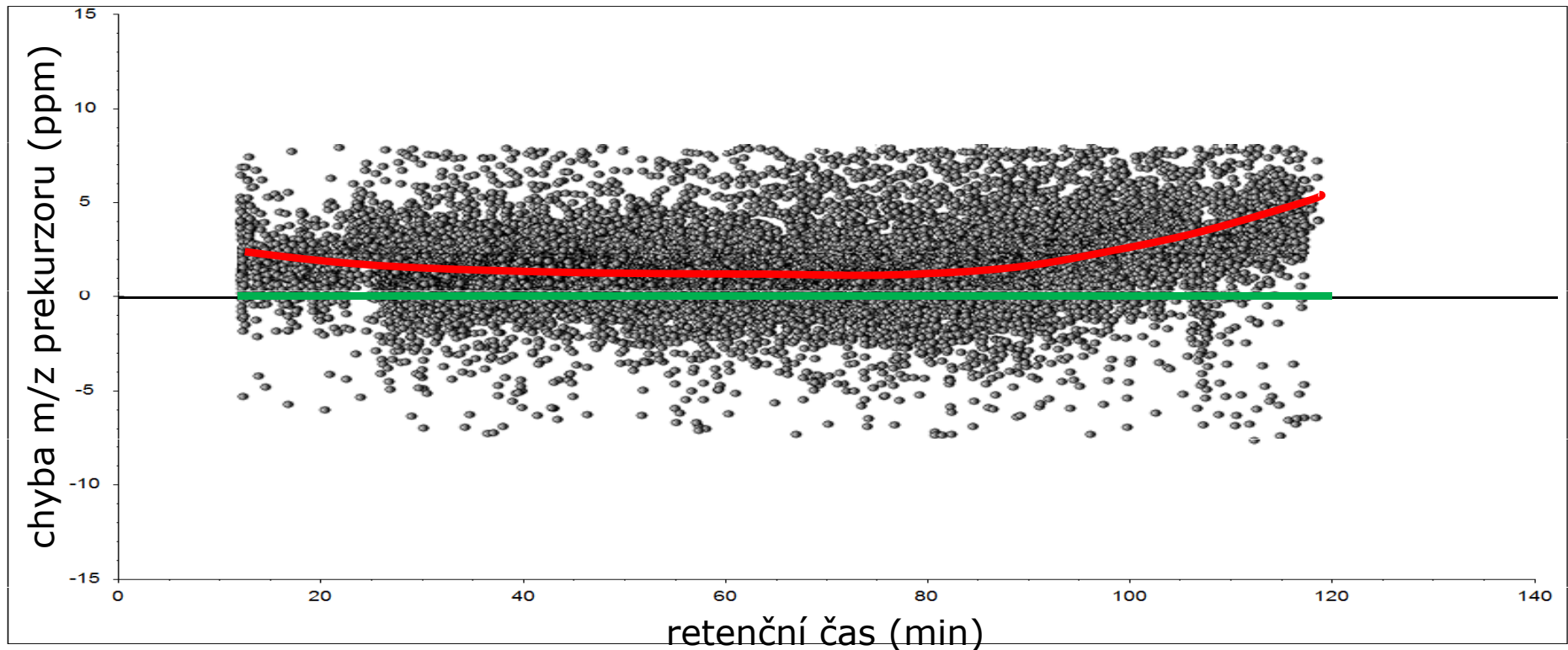


Čiarové spektrum

- Vypočítané z profilového



Rekalibrácia



- Interná rekalibrácia – rovnaká na celý záznam
- lockmass – každé spektrum podľa opakujúcej sa kontaminanty
- Podľa identifikovaných peptidov – odhad závislosti z identifikovaných peptidov (polynóm)

Databázove vyhľadávanie

1. Príprava dat

- Výber „reprezentatívnych“ signálov MS/MS
- Odstránenie „menej kvalitných“ spektier MS/MS
 - Top N (z okna), dekonvolúcia signálu a šumu
- Získame tabuľku m/z hodnôt a intenzít

2. Príprava databázy

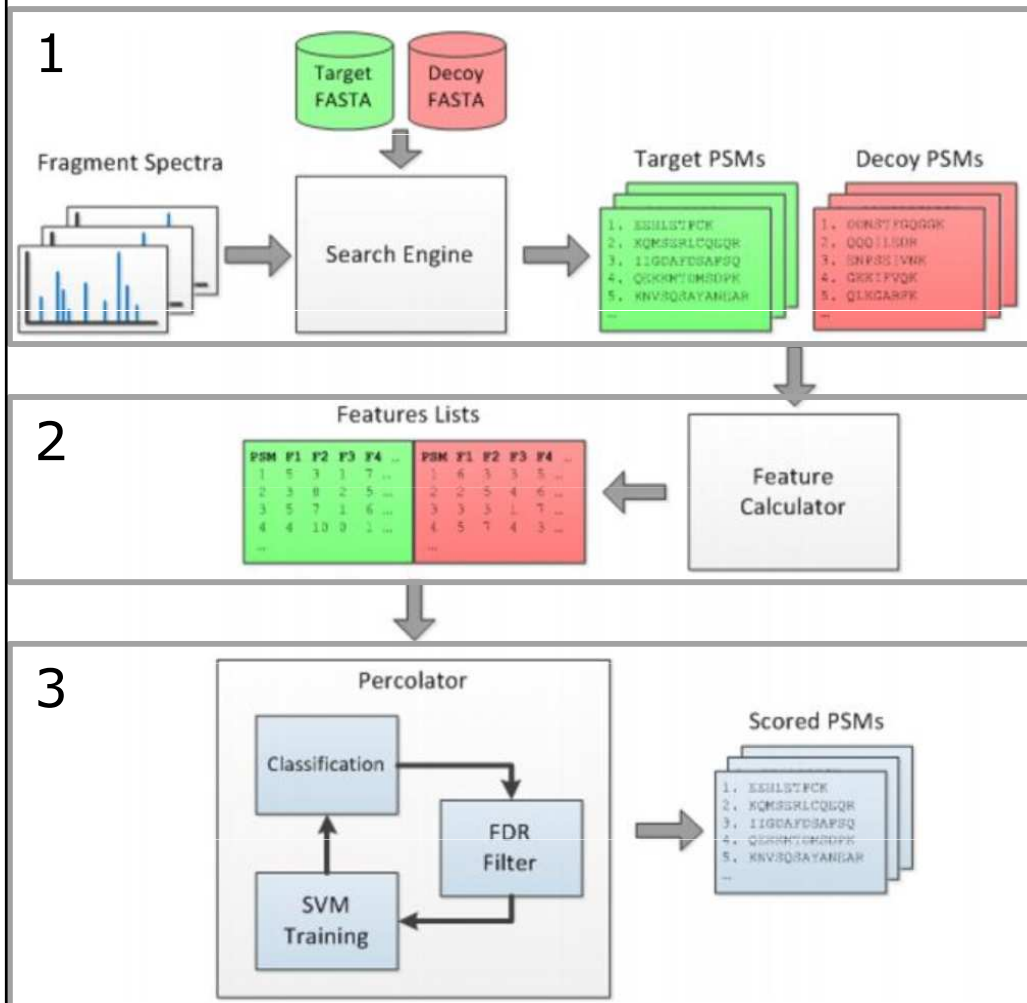
- *in silico* štiepenie sekvencií z databázy
- Priradenie jedného alebo viacerých peptidov k jednému spektru (*decoy* databáza, FDR, Percolator)

3. Výber peptidových identifikácií

Percolator



1. prehľadanie dát MS/MS
2. výpočet „vlastností“ peptidov
3. prepočítanie skór peptidov

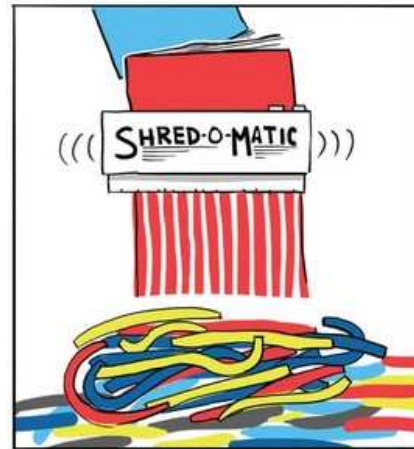


Prepočítanie skór peptidov

- Použitie *support vector machines* (SVM)
- sady identifikací
 - falošne pozitívne – *decoy* databáza
 - pozitívne – pôvodná databáza (↑skóre)
- priradenie váh vlastnostiam v SVM
 - např. skóre; chyba hmotnosti intenzita, modifikácie, ...

⇒ viac identifikovaných peptidov

Rekonštrukcia sady proteínov



Analógia puzzle, ALE:

- Tisíce kúskov:
 - Rovnaké
 - Poškodené
 - Chýbajúce
 - Z iných skladačiek
- Pasujú na rovnaké miesta

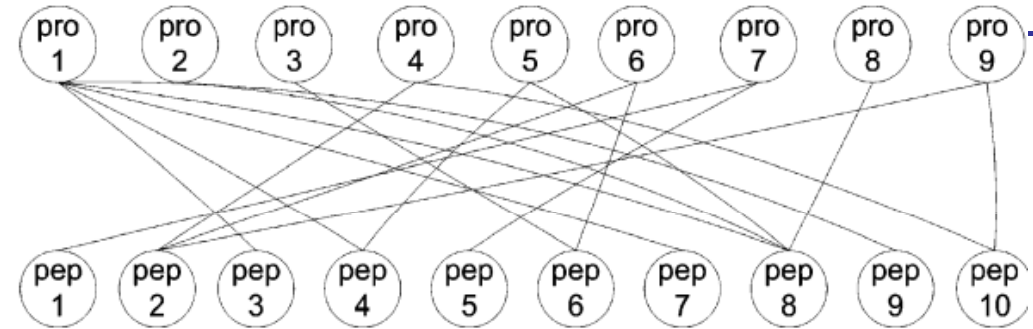
Vybrané přístupy

- N – peptidové pravidlo
 - Proteíny, u kterých pozorujeme aspoň N peptidov
 - Vysoká falošná pozitivita
 - Používané na sekvenčně homologické proteiny
- Pravdepodobnostné přístupy
 - *ProteinProphet, Nested mixtures, Fido*

Princíp parsimónie a Occamovej britvy

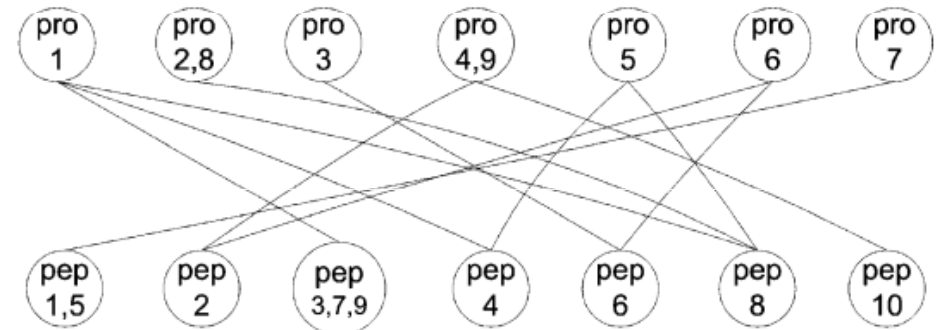
A. Vytvorenie biparitného grafu:
peptidy - možné proteíny

A



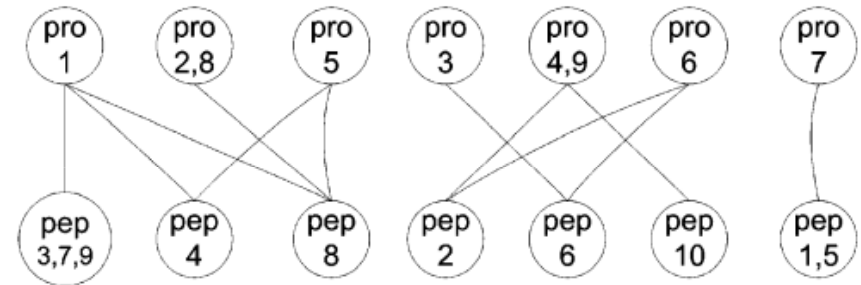
B. Zlúčenie proteínov a peptidov
do skupín (napr. pep 3,7,9;
pro 4,9)

B



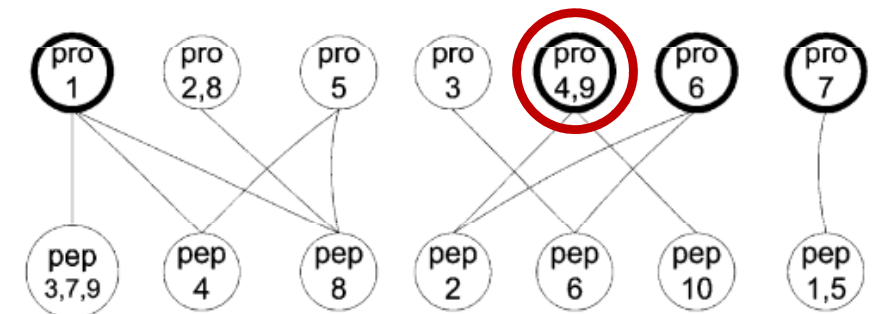
C. Rozdelenie skupín

C



D. Výber minimálnej sady
proteínov

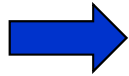
D



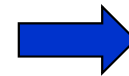
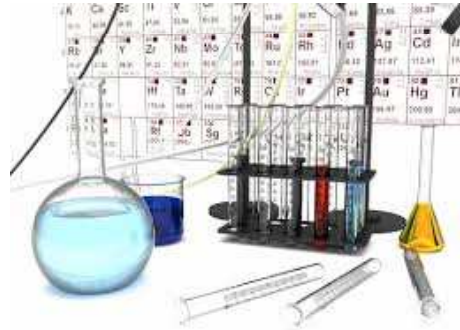
Dôsledok: falošná negativita výsledkov

Experiment

Biologický materiál



Izolácia, štiepenie

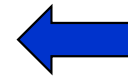


LC MS/MS systém



Identifikácie proteínov

Identification and relative comparison summary											
Prot Group	Accession	Description	Identification summary	1				Identification threshold (pept			
			MW [kDa]	calc. pI	Sum(Coverage)	Presence of the protein group in individ	Line 1	Line 2	Line 3	Line 4	
6	1	Q09666	Neuroblast differentiation-ass	629	6.2	82%	Y	Y	Y	Y	
417	2	Q15149-4	Isoform 4 of Plectin OS=Hom	516	5.8	58%	Y	Y	Y	Y	
418	3	Q15149-3	Isoform 3 of Plectin OS=Hom	518	5.7	58%	Y	Y	Y	Y	
885	4	P09211;P0921	Glutathione S-transferase P	23	5.6	63%	Y	Y	Y	Y	
886	5	O75369-2	Isoform 2 of Filamin-B OS=Hc	276	5.8	47%	Y	Y	Y	Y	
951	5	O75369	Filamin-B OS=Homo sapiens	278	5.7	46%	Y	Y	Y	Y	
952	5	O75369-9	Isoform 9 of Filamin-B OS=Hc	277	5.7	46%	Y	Y	Y	Y	
Sequence	# PSMs	# Proteins	# Protein Groups								
AAGSGELGVTMK	8	6	1								
AAGSGELGVTMKGPK	2	3	1								
ADIEMPFDPSK	4	3	1								
AEVSIQNKDGYAVTYVPLTA	2	3	1								
AGGPGLER	2	9	2								
AGLAPLEVR	2	3	1								
AGDPTSYTICDEK	6	3	1								



Spracovanie dát

Štatistická a bioinformatická analýza

Čo s identifikovanými proteínmi?

- Závisí od pôvodného experimentu
- Typicky, doplnenie anotácie proteínov z databázy (GO, KEGG, TAIR) a použitie metód analýzy génových sad

Kapitola X

Databázy dát

Verejne prístupné databázy

- Veľké experimenty majú až stovky, alebo tisíce vzoriek, v každej sa študujú desaťtisíce, až stotisíce génov
- Pre publikáciu výsledkov je vyžadované vložiť dáta v štandardizovanom formáte (MIAME - Minimal Information About a Microarray Experiment) do jednej z verejne prístupných databáz, tak aby ktokoľvek bol schopný výsledky zreprodukovať
- Toto prináša veľkú výhodu:
 - môžeme dáta podrobovať meta-analýze (simultánne porovnávať dáta z rôznych experimentov)
 - vďaka štandardnému formátu môžeme vyhľadávať súbory s parametrami, ktoré potrebujeme
 - dáta môžeme automaticky sťahovať

GEO na NCBI

Netscape: GEO Database Design Brief

Back Forward Reload Home Search Netscape Images Print Security Shop Stop

Location: <http://www.ncbi.nlm.nih.gov/geo/info/scheme.cgi> What's Related

NCBI Gene Expression Omnibus **geo**

Entrez ProbeSet SAGEmap Pubmed UniGene LocusLink

Database Design Brief Query: go

Paper | FAQ | News

Feedback NEW

Retrieval tools
...by GEO accession
...by attribute

Deposit tools
...via web
...via direct deposit
New account

Brief info
Current holdings
Retrieving data
Depositing data
...via web
...via direct deposit
Database design

Ad nauseam
SOFT guide
...examples
Web deposit guide
...entry fields
...data tables
SQL implementation

Please fill out our [feedback suggestionnaire](#) NEW

At the most basic level of organization of GEO there are four entities.

```
graph TD; Submitter[Submitter] --- Platform[Platform]; Submitter --- Series[Series]; Platform --- Sample[Sample]; Series --- Sample;
```

IBA

Array Express na EBI

<http://www.ebi.ac.uk/arrayexpress/>

[Training](#)

[Industry](#)

[About Us](#)

[Help](#)

[Site Index](#)



The **ArrayExpress Archive** is a database of functional genomics experiments including gene expression where you can query and download data collected to **MIAME** and **MINSEQE** standards. **Gene Expression Atlas** contains a subset of curated and re-annotated Archive data which can be queried for individual gene expression under different biological conditions across experiments.

Experiments Archive

15786 experiments, 442596 assays



Experiment, citation, sample and factor annotations

[Browse experiments](#)

[Advanced query syntax](#) ^{NEW}

[Submitter/reviewer login](#)

[ArrayExpress Query Help](#)

Gene Expression Atlas

5670 experiments, 138915 assays, 18346 conditions

Genes

up/down in

Conditions

Any species

▼

[Gene Expression Atlas Home](#)

News



- **20 Oct 2010 - Internship for a student project in human gene expression - Filled now**
This student project is now taken.
- **17 Sep 2010 - New Atlas Data Release 10.8**
A new release of Gene Expression Atlas has been made with 93

Links

- [ArrayExpress User Survey](#)
- [Old ArrayExpress Interface](#)
- [Help](#) | [Training](#) | [FAQ](#) | [Citing](#)
- [Submit Data](#) (array based and re-sequencing)
- [Programmatic Access](#) | [FTP Access](#)
- [Software Downloads](#) and [Statistics](#)

- E-learningové skriptá analýzy dát IBA
- <http://portal.matematickabiologie.cz/index.php?pg=analiza-genomickych-a-proteomickych-dat--analiza-genomickych-a-proteomickych-dat>