Proteomické data

2D gelová elektroforéza

Analýza proteomu

- Od hmotnostních spekter – přeš komplexní struktury proteinových shluků - po analýzu funkce proteinů

  •Analýza **struktury**: Proteínová sekvenace – Edmanova degradace, hmotnostní spektrometrie

  •Analýza **exprese**: Hmotnostní spektrometrie, proteínové mikročipy, 2D gelová elektroforéza..

  •Analýza **funkce**: Modelovaní makromolekulárních systémů – odvozování vlastností z atomových interakcí

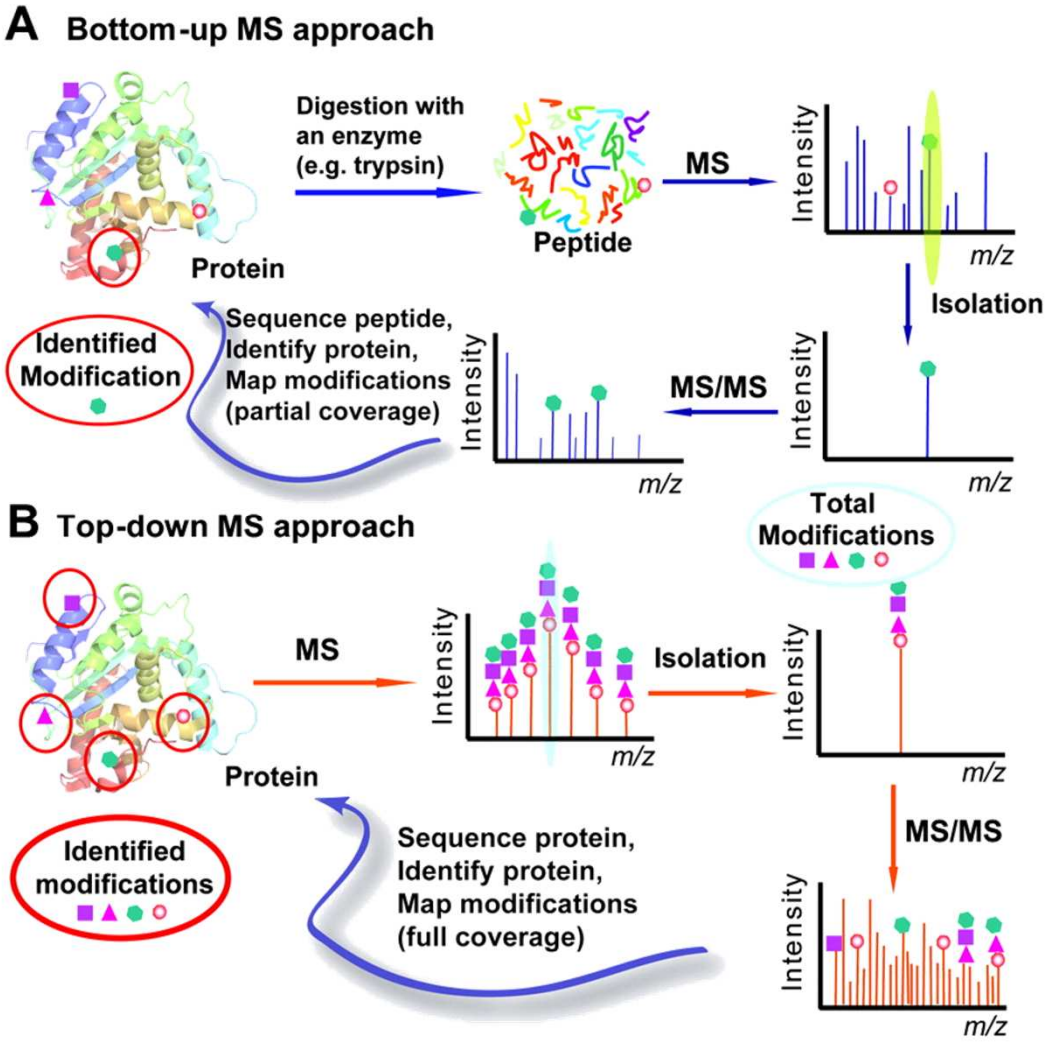
Dva hlavní přístupy

Bottom-up proteomika

- protein → peptidy
- analýza peptidů pomocí MS
- fragmentace peptidů v MS

Top-down proteomika

- analýza intaktních proteinů v MS
- fragmentace proteinu v MS

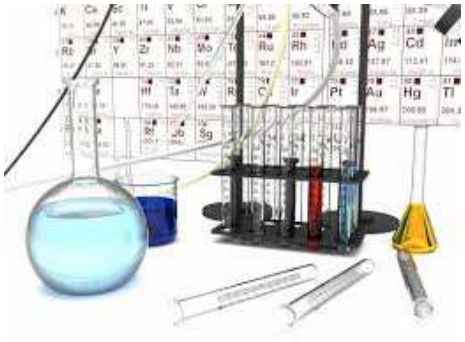


Zhang, H.; Ge, Y. Circ Cardiovasc Genet, 2011, 4, 711–711.

Experiment

Biologický materiál

Izolace, štěpení



LC MS/MS systém

Identifikace proteinů

Identification and relative comparison summary

	Prot Group	Accession	Description	Identification summary MW [kDa]	calc. pI	Sum(Coverage)	Identification threshold (pept 1) 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
685	3	Q15149-3	Isoform 3 of Plectin OS=Hom	518	5.7	58%	Y	Y	Y	Y
951	4	P09211;P0921	Glutathione S-transferase P (	23	5.6	63%	Y	Y	Y	Y
971	5	O75369-2	Isoform 2 of Filamin-B OS=Ho	276	5.8	47%	Y	Y	Y	Y
972	5	O75369	Filamin-B OS=Homo sapiens	278	5.7	46%	Y	Y	Y	Y
973	5	O75369-9	Isoform 9 of Filamin-B OS=Ho	277	5.7	46%	Y	Y	Y	Y

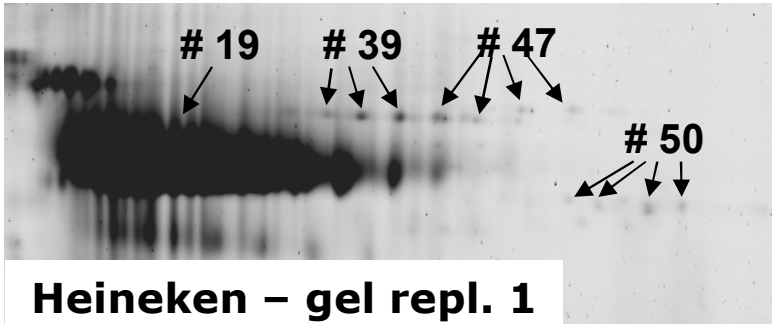
Sequence	# PSMs	# Proteins	# Protein Groups
AAGSGELGVTMK	8	6	1
AAGSGELGVTMKGPK	2	3	1
ADIEMPFDPSK	4	3	1
AEVSIQNNKDGTYAVTYVPLTA	2	3	1
AGGPGLER	2	9	2
AGLAPLEVR	2	3	1

Zpracování dat

Statistická a bioinformatická analýza

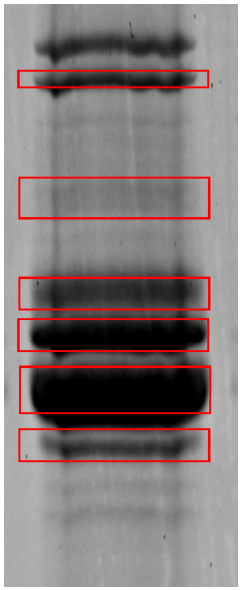
Typy experimentů I

- Podle komplexity vzorku (počtu očekávaných proteinů):
  - Identifikace proteinu(ů) z gelu 2D-SDS-PAGE (jednotky proteinů, desítky vzorků)



Typy experimentů I

- Podle komplexity vzorku (počtu očekávaných protinů):
  - Identifikace proteinu(ů) z gelu 2D-SDS-PAGE (jednotky proteinů, desítky vzorků)
  - Identifikace proteinu v proužkách z gelu 1D-SDS-PAGE (jednotky až stovky proteinů, jednotky vzorků)



Typy experimentů I

- Podle komplexity vzorku (počtu očekávaných protinů):
  - Identifikace proteinu(ů) z gelu 2D-SDS-PAGE (jednotky proteinů, desítky vzorků)
  - Identifikace proteinu v proužkách z gelu 1D-SDS-PAGE (jednotky až stovky proteinů, jednotky vzorků)
  - Relativní porovnání dvou bun. linií – wild-type vs. mutant (tisíce proteinů, jednotky vzorků)

Typy experimentů II

- Podle použitého značení:
  - *label-free* – bez značení
  - SILAC – stabilní izotopové značení aminokyselin v buněčných kulturách (2-3 vzorky)
  - iTRAQ – izobarické značky pro relativní a absolutní kvantifikaci (4 a nebo 8 vzorků)

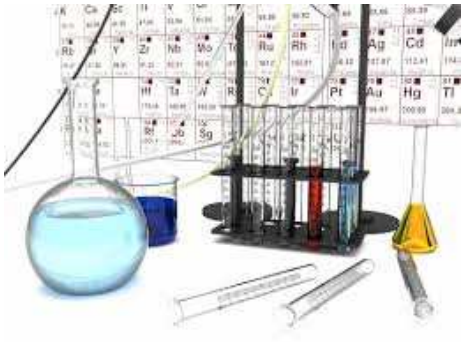
Získané informace

- Kvalitativní – primární cíl
  - Jaké proteiny se ve vzorku vyskytují?
- Kvantitativně
  - Hodnocení koncentrace proteinů (PSMs, AUC)
- Modifikace
  - Výskyt posttranslačních modifikací (fosforylace)

Experiment

Biologický materiál

Izolace, štěpení



LC MS/MS systém

Identifikace proteinů

	A	D	E	F	G	H	I	J	K	L	M	N
1					Identification and relative comparison summary							
2									1	Identification threshold (pept		
3	Prot Group	Accession	Description		Identification summary				Presence of the protein group in individ			
4					MW [kDa]	calc. pI	Sum(Coverage)		Line 1	Line 2	Line 3	Line 4
5												
6	1	Q09666	Neuroblast differentiation-ass		629	6.2	82%		Y	Y	Y	Y
7												
417	2	Q15149-4	Isoform 4 of Plectin OS=Hom		516	5.8	58%		Y	Y	Y	Y
418												
685	3	Q15149-3	Isoform 3 of Plectin OS=Hom		518	5.7	58%		Y	Y	Y	Y
686												
951	4	P09211;P0921	Glutathione S-transferase P (		23	5.6	63%		Y	Y	Y	Y
952												
971	5	O75369-2	Isoform 2 of Filamin-B OS=Ho		276	5.8	47%		Y	Y	Y	Y
972	5	O75369	Filamin-B OS=Homo sapiens		278	5.7	46%		Y	Y	Y	Y
973	5	O75369-9	Isoform 9 of Filamin-B OS=Ho		277	5.7	46%		Y	Y	Y	Y
974												
975			Sequence		# PSMs	# Proteins	# Protein Groups					
976			AAGSGELGVTMK		8	6	1					
977			AAGSGELGVTMKGPK		2	3	1					
978			ADIEMPFDPSK		4	3	1					
979			AEVSIQNNKDGTYAVTYVPLTA		2	3	1					
980			AGGPGLER		2	9	2					
981			AGLAPLEVR		2	3	1					

comments | report format | prot groups_orig_repl | line 1_combined | combined+peptides | line 2_combined | line 2_combined+pep

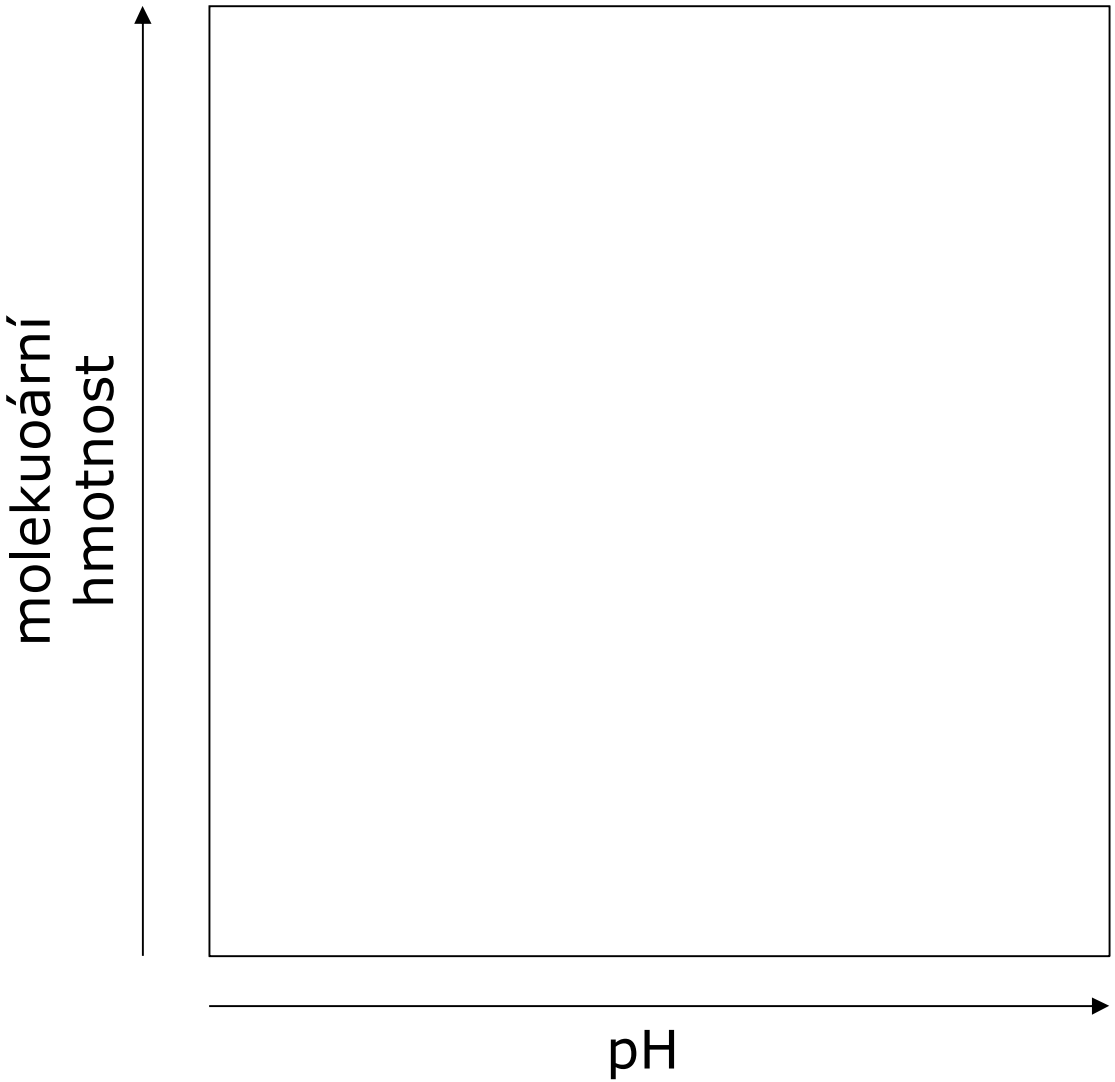
Zpracování dat

Statistická a bioinformatická analýza

2D gelová elektroforéza

Princip

- Proteiny jsou separované na gelu ve dvou dimenzích – na základě hmotnosti a na základě pH



Postup experimentu

1. **Proteiny** jsou **extrahované** ze vzorku

2. **Vzorky se umístí na gel** a proteiny migrují až dosáhnou izoelektronický bod (kdy je jejich náboj nula) - pH(I)
   - Důležitý je **výběr gelu**, který musí byt dostatečně pórovitý, aby umožnil proteinům pohyb (agaróza nebo polyakrylamidový gel)

3. Takto se **proteiny oddělí** vzhledem ke svému izoelektrickému bodu a k hmotnosti

4. Následně proteiny necháme se pohybovat na základě **hmotnosti** ve druhé dimenzi

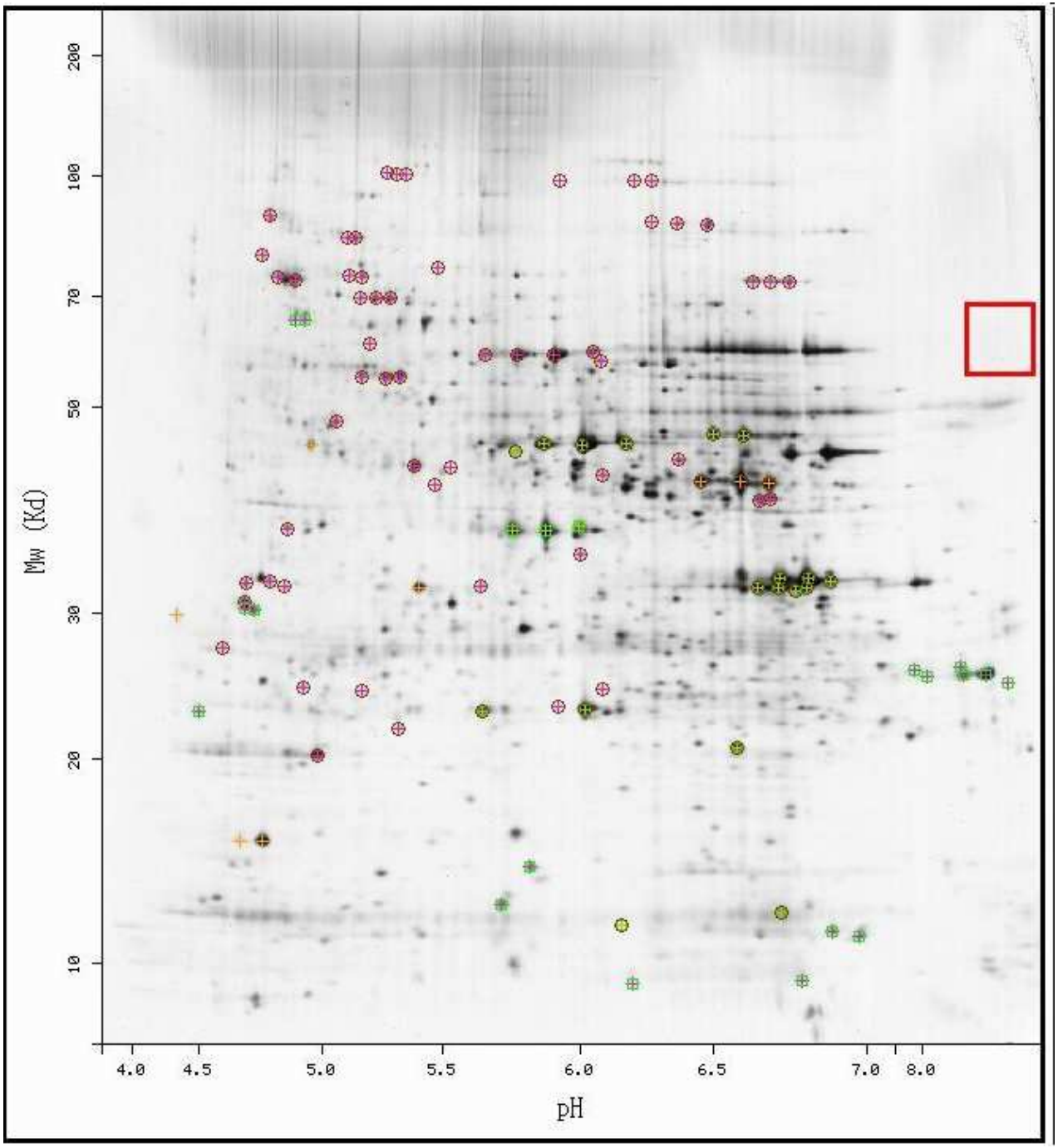
5. Nakonec je gel **zabarvený,** aby se detekovaly jednotlivé oblasti výskytu proteinů (spoty)

6. Zabarvený gel je pak digitalizován do obrazu (podobně jako mikročipy)

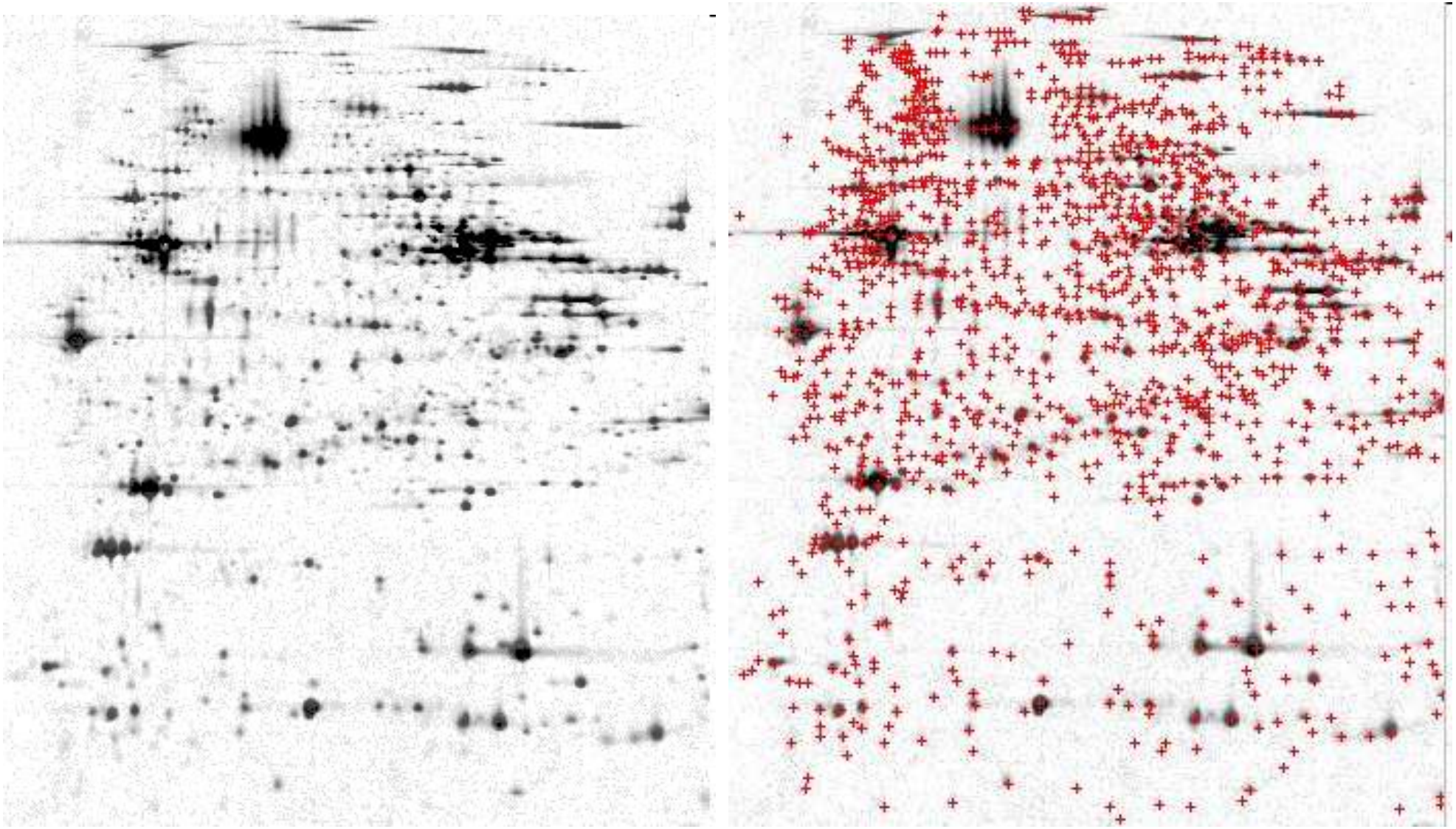
7. Intenzita pixelů koreluje **s množstvím proteinu**, používá se speciální SW pro analýzu spotů

IBA

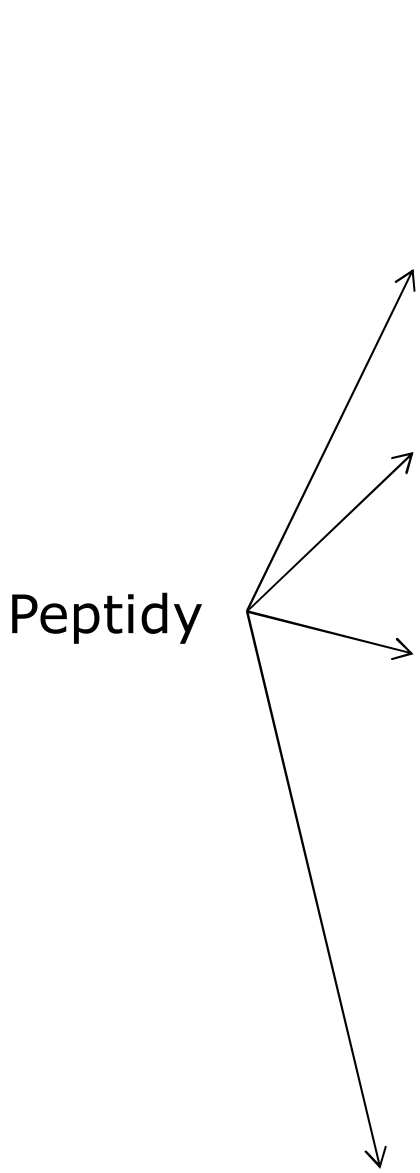
Jak vypadá obrázek



2-D gelová elektroforéza



Jak vypadají data

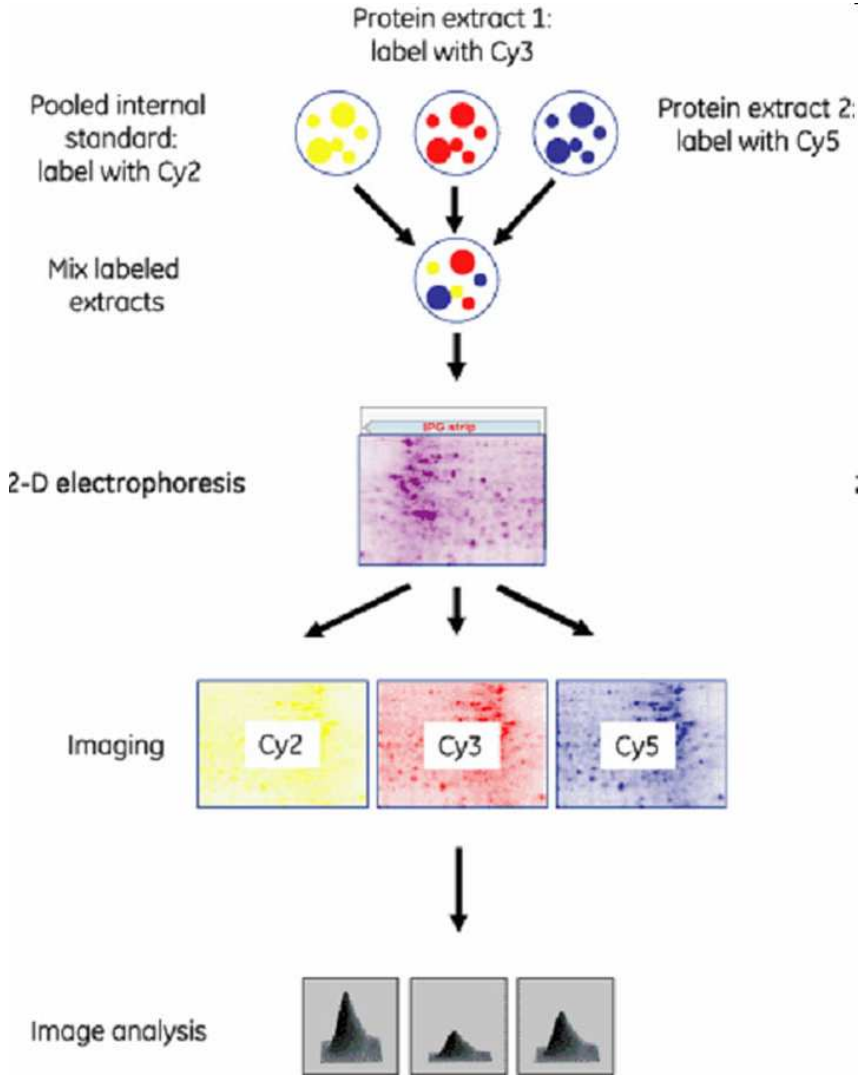


SSP	wt_A	wt_A	wt_A	wt_S
101	2338.84	2078.42	2625.1	2550.54
102	118.92	68.65	125.8	109.66
103	221.89	55.32	NA	NA
104	215.3	189.02	220.28	NA
105	106.56	NA	238.36	NA
202	328.32	226.46	522.52	1281.75
203	259.8	228.13	340.37	NA
205	1439.72	1213.28	1187.43	1353.14
206	1094.33	754.83	1291.89	1240.82
208	97.78	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

DIGE

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

- Proteiny se nejprve zabarví fluorescenčním barvivem
- Každé barvivo se skenuje pod jiným filtrem
- Takto se může porovnávat více vzorků

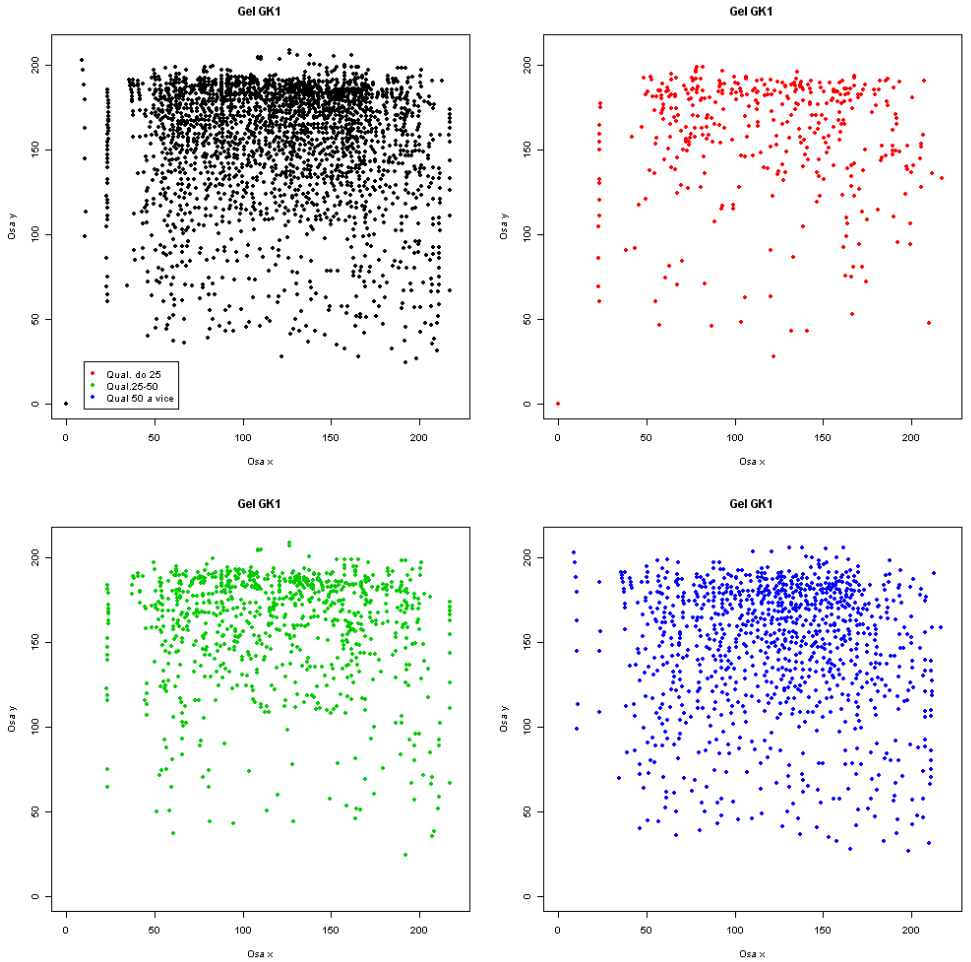


Nutnost úpravy dat

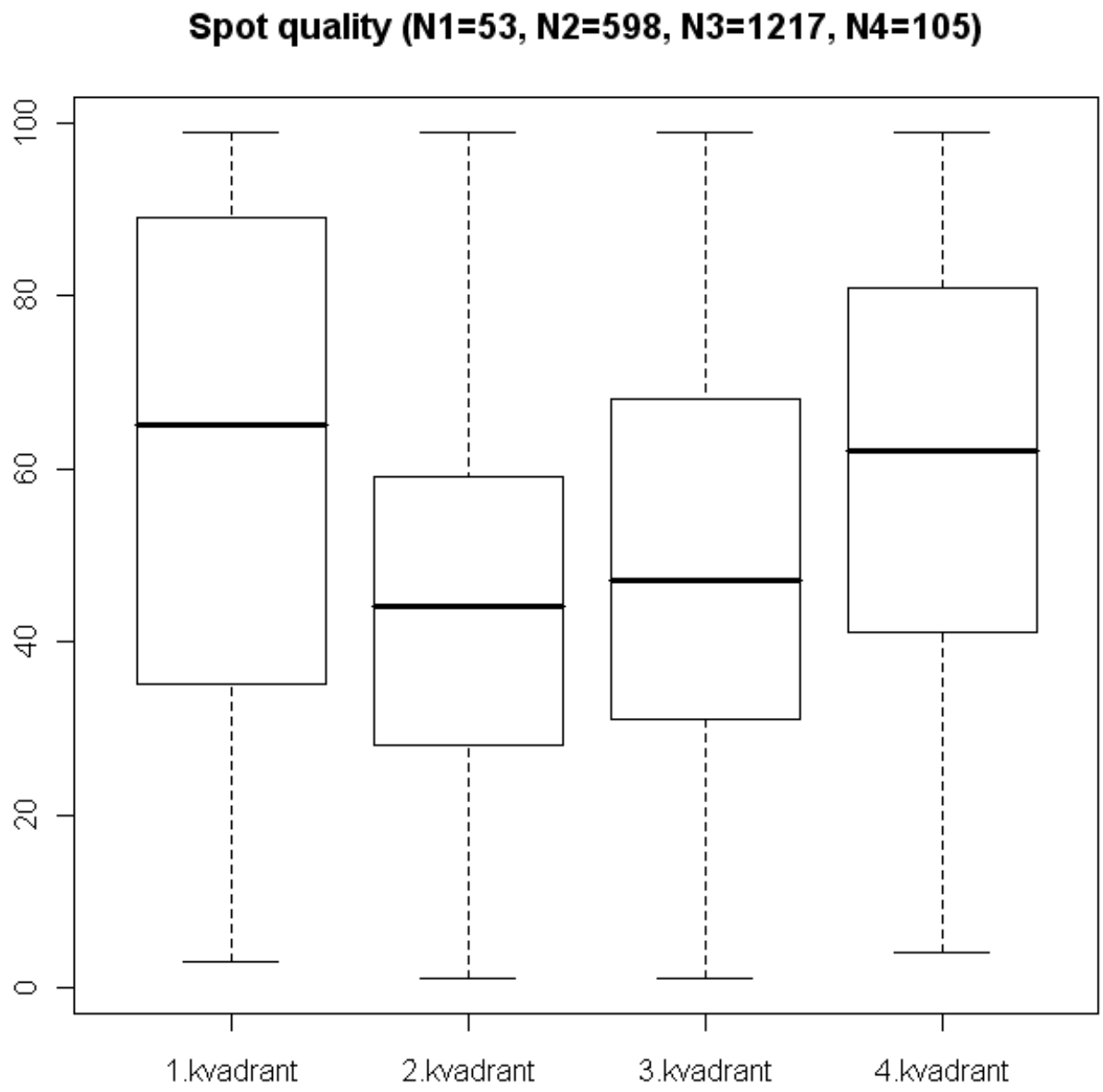
- Tak jako mikročipový experiment i 2-DE je vystavená experimentálním chybám, které jsou zdrojem šumu
- Je nutná úprava a normalizace dat
- Neexistuje tu ale taková automatická kvantifikace spotů tak jako u mikročipů, protože spoty nejsou fixně dané předem!
  - existující automatická kvantifikace vyžaduje manuální úpravu
  - proměnné kvality spotů
- Data z 2DE nejsou normálně rozložené – je nutná transformace (log)

Normalizace a úpravy dat

- Důležitým krokem v úpravě dat je ***kalibrace* všech expresních hodnot a gelů navzájem**
- V tomto procesu se odstraňuje prostorový efekt, i efekt barviva
- Na každém gelu jsou kontrolní proteiny, podle kterých se každý gel kalibruje (posouvá)



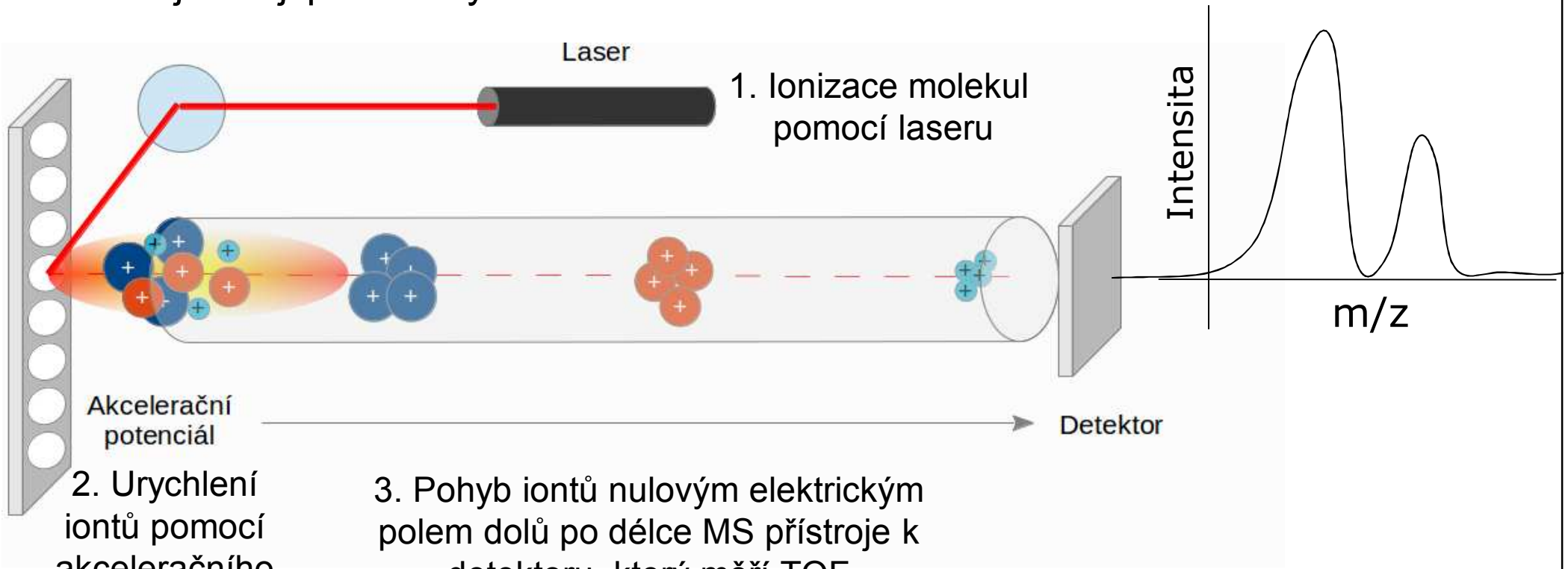
Kontrola kvality spotů



Hmotnostní spektrometrie

Hmotnostní spektrometrie

- Technika používaná pro charakterizaci proteomu v biologickém vzorku (plasma, sérum, . . .)
- Založená 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.
- Najčastěji používaný - TOF



Hmotnostní spektrometr TOF - princip

- 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:

$$m / z = B (t_D - A)\text{^}2$$

- A a B jsou stanoveny pomocí kalibrace

Hmotnostní spektrometr TOF - druhy

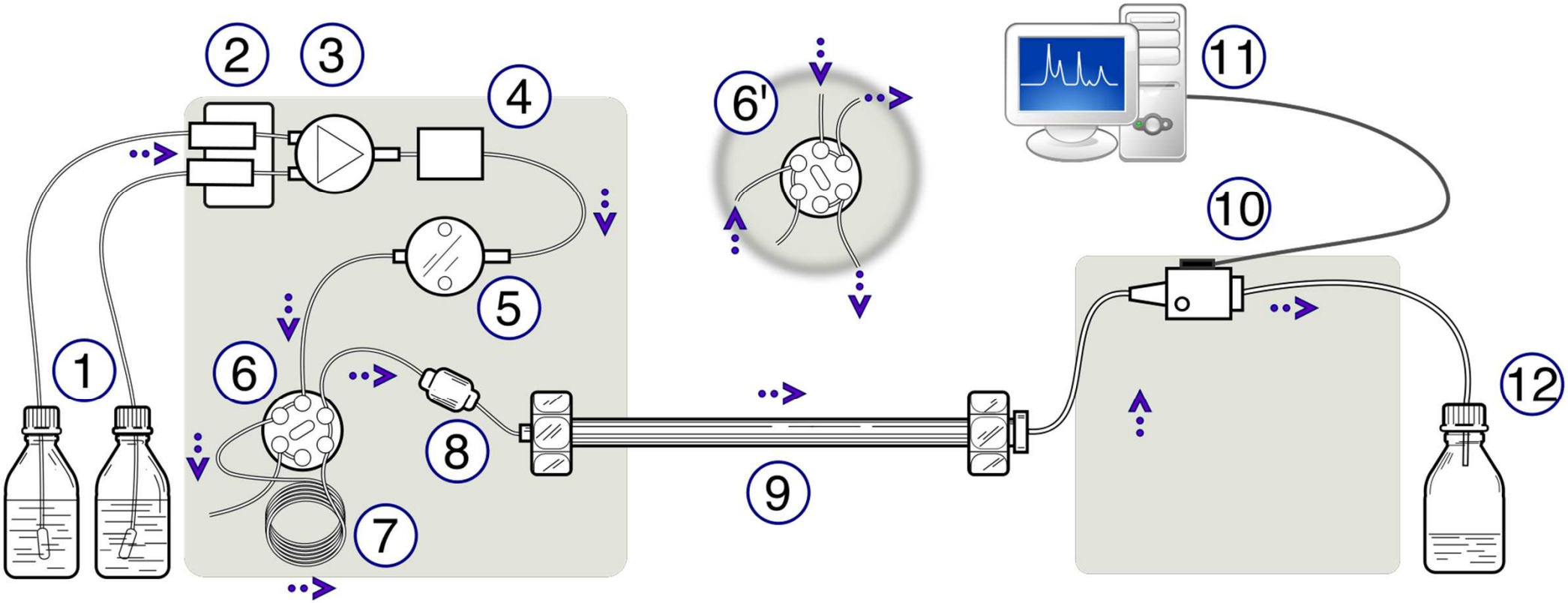
- **Nejpoužívanější:**
  - Matrix-Assisted Laser Desorption-Ionisation (MALDI)-TOF
  - Surface-Enhanced Laser Desorption-Ionisation (SELDI)-TOF

- **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.

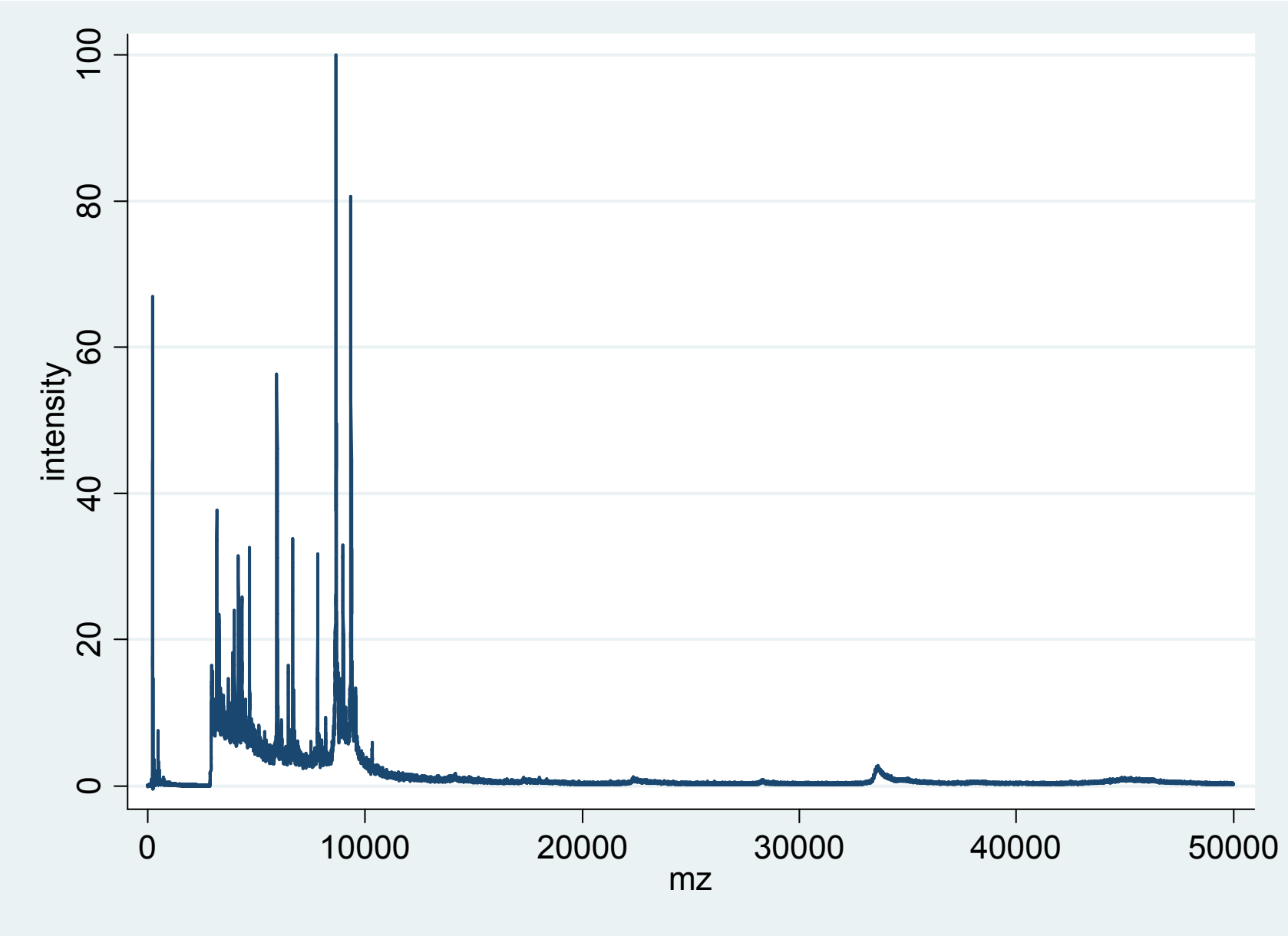
IBA

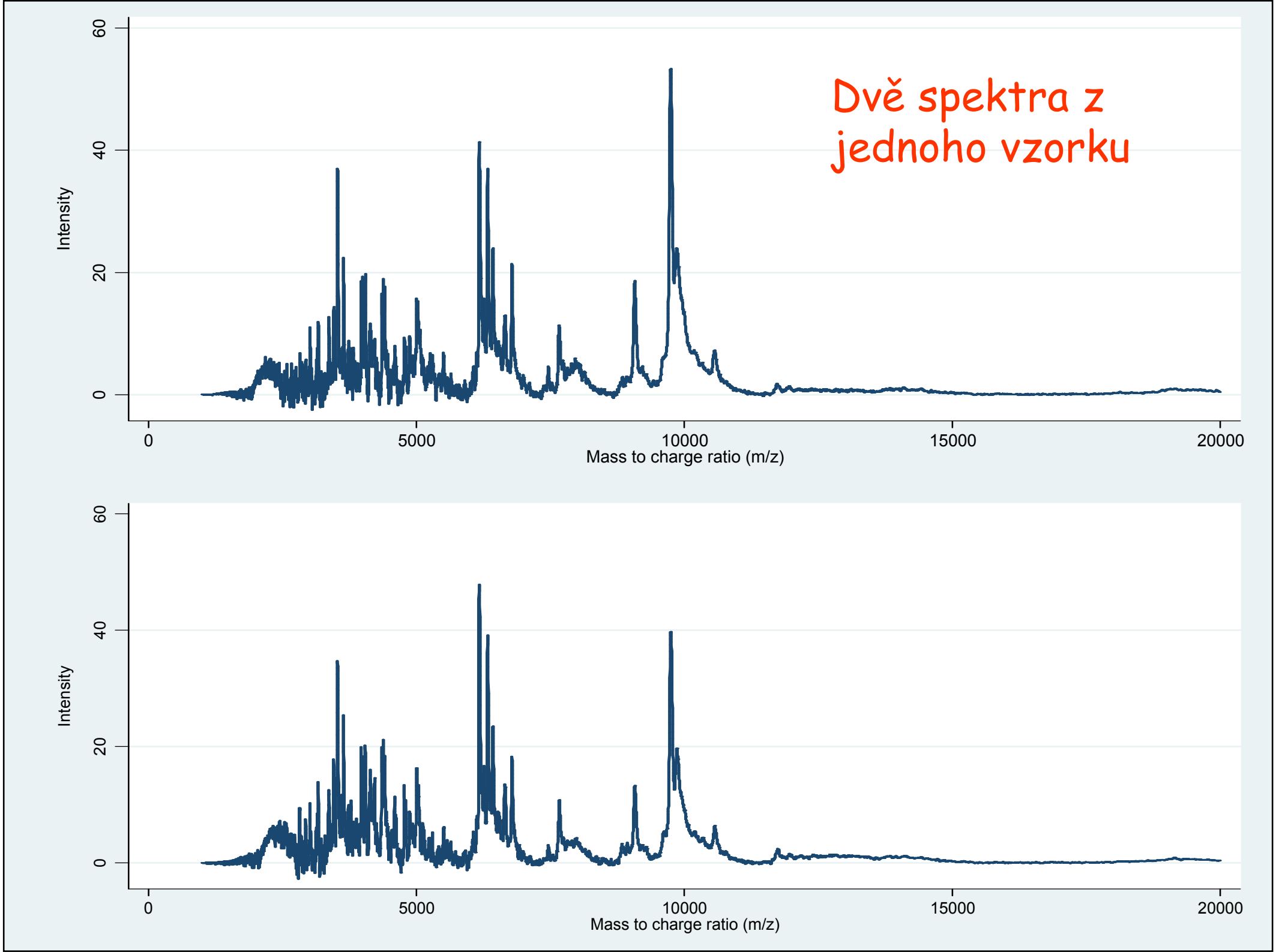
Kapalinová chromatografie – MS/MS

- Další druh hmotnostní spektrometrie
- Vzorky nejsou na matrici jako u MALDI nebo SELDI, ale v kapalině

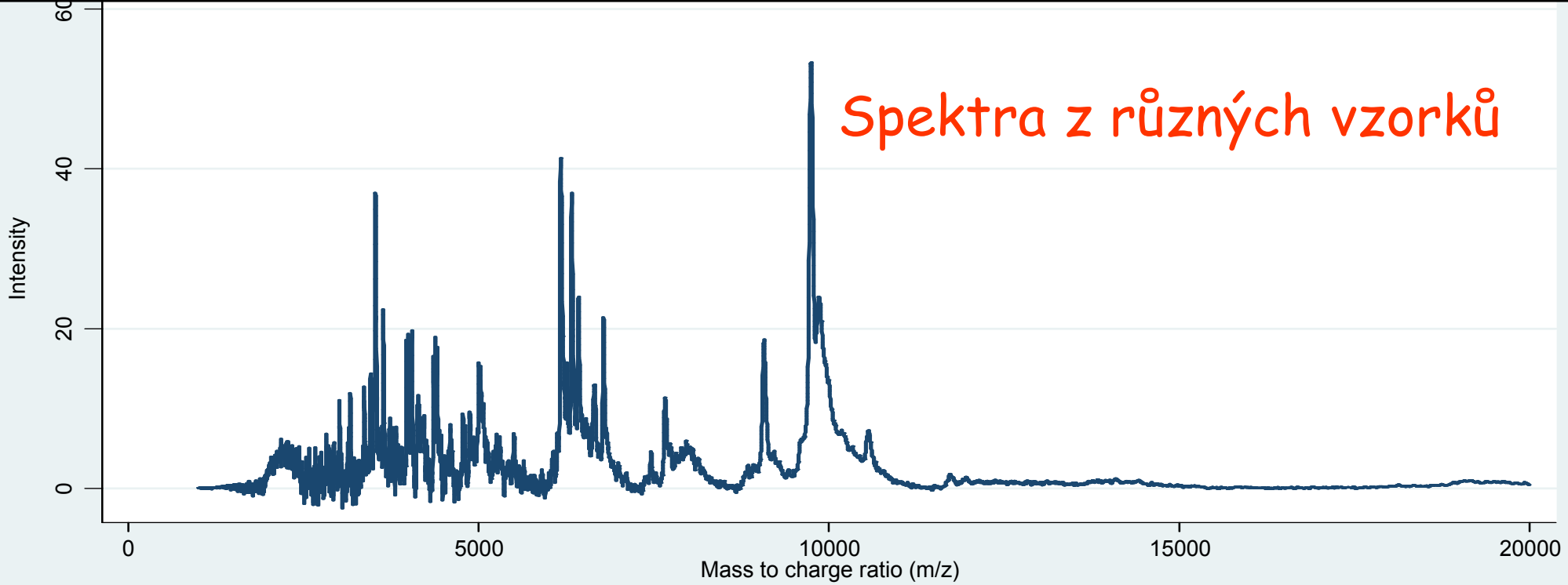


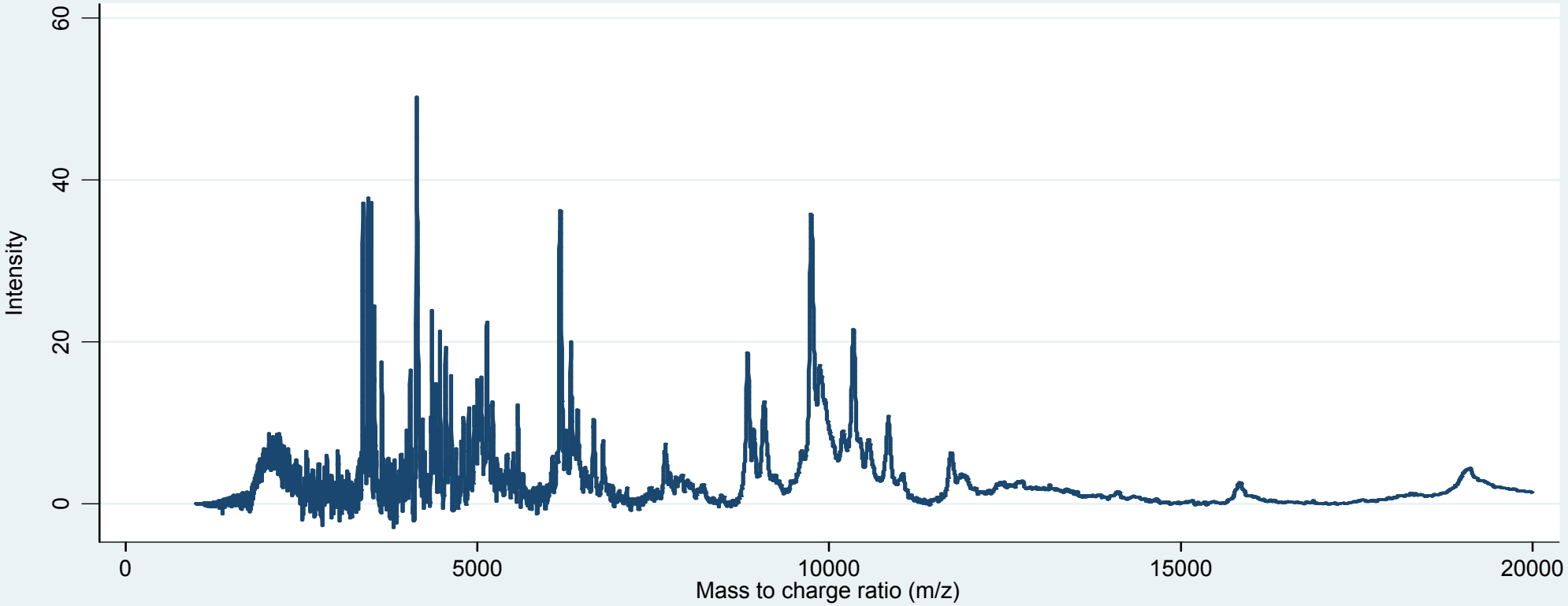
Jak vypadá TOF MS *profil*




Dvě spektra z jednoho vzorku
Intensity
0
20
40
60
0
5000
10000
15000
20000
Mass to charge ratio (m/z)
Intensity
0
20
40
60
0
5000
10000
15000
20000
Mass to charge ratio (m/z)

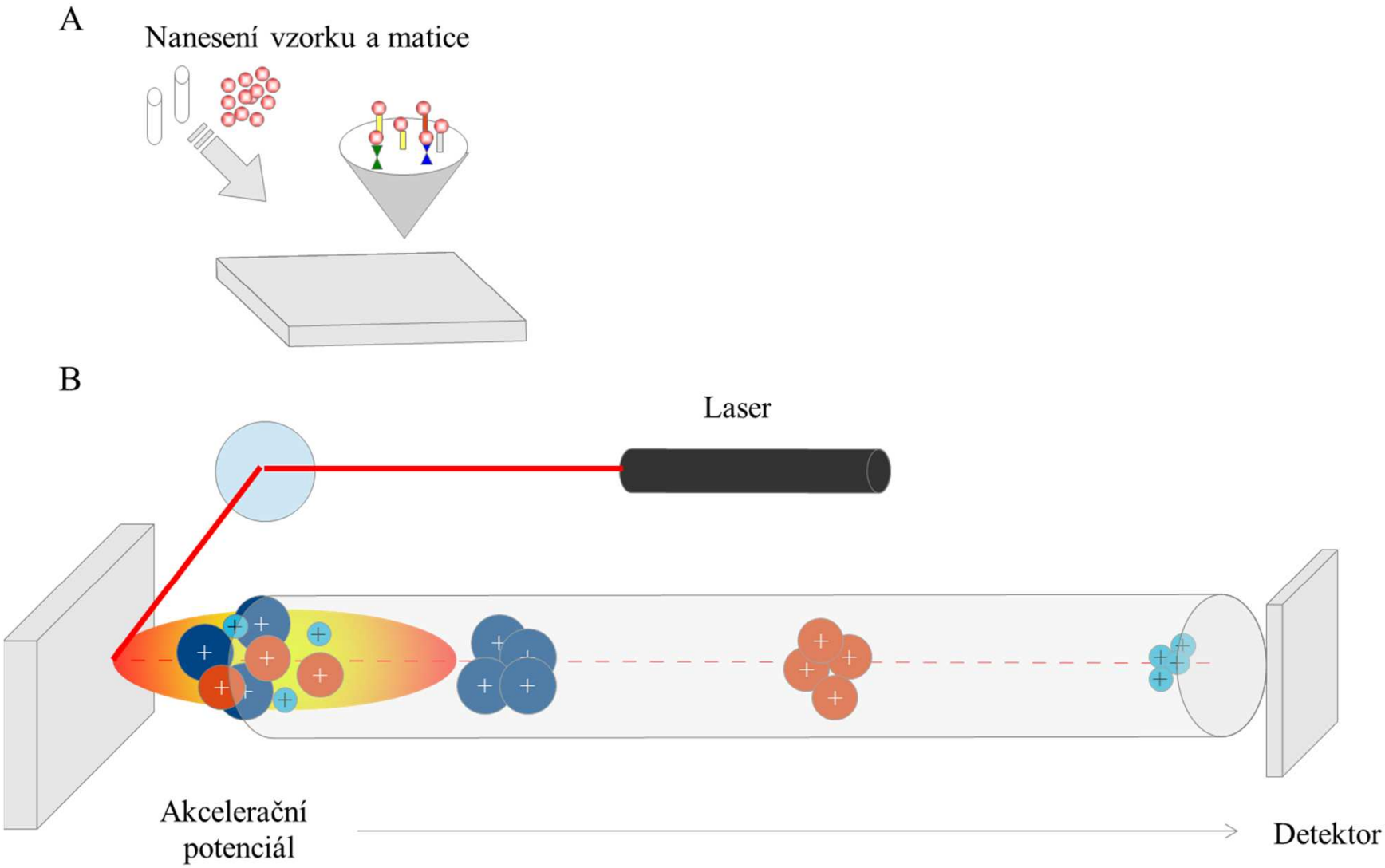
Spektra z různých vzorků





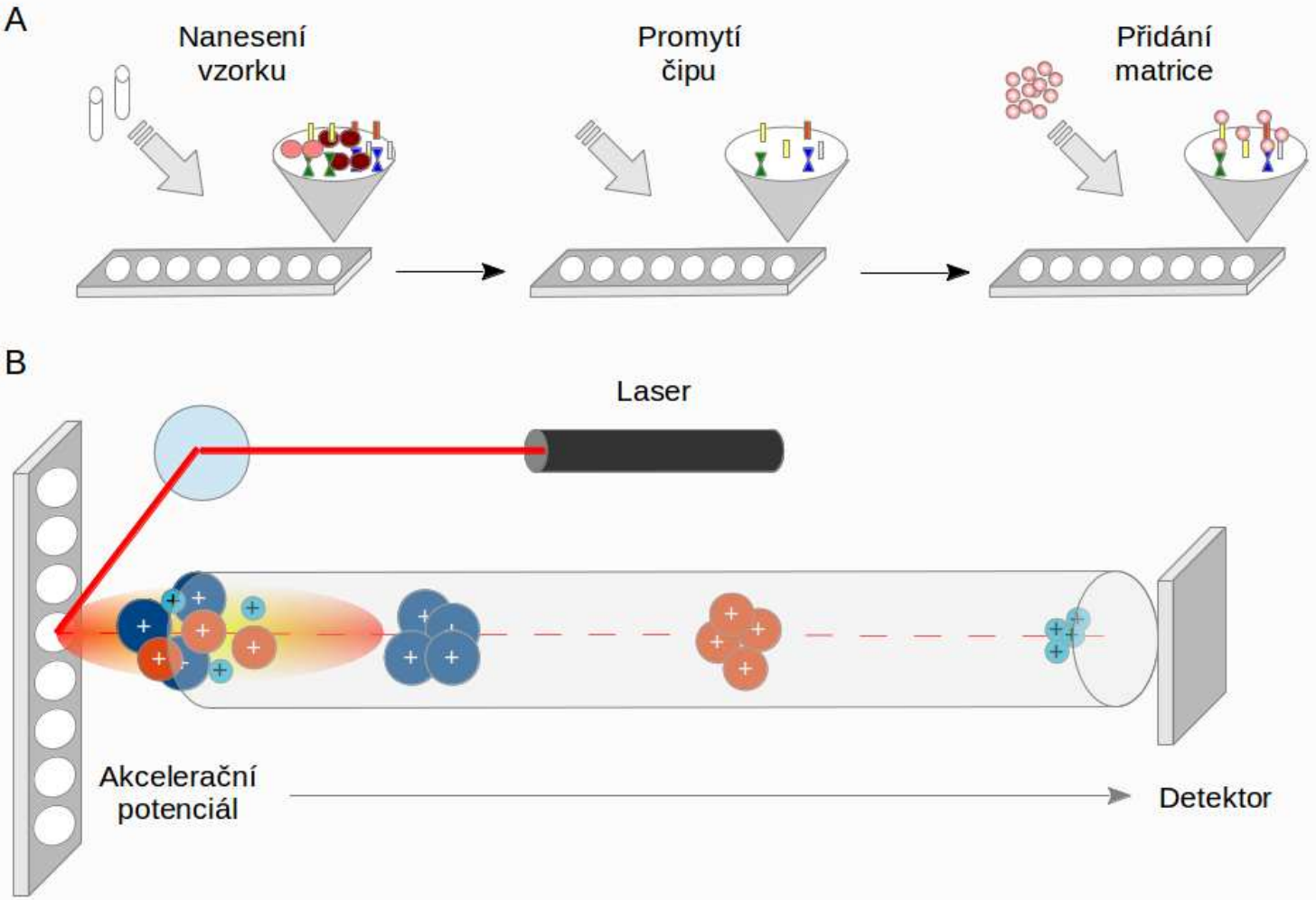
MALDI-TOF

- **M**atrix-**A**ssisted **L**aser **D**esorption-**I**onisation - TOF



SELDI-TOF

- **S**urface-**E**nhanced **L**aser **D**esorption-**I**onisation – 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.



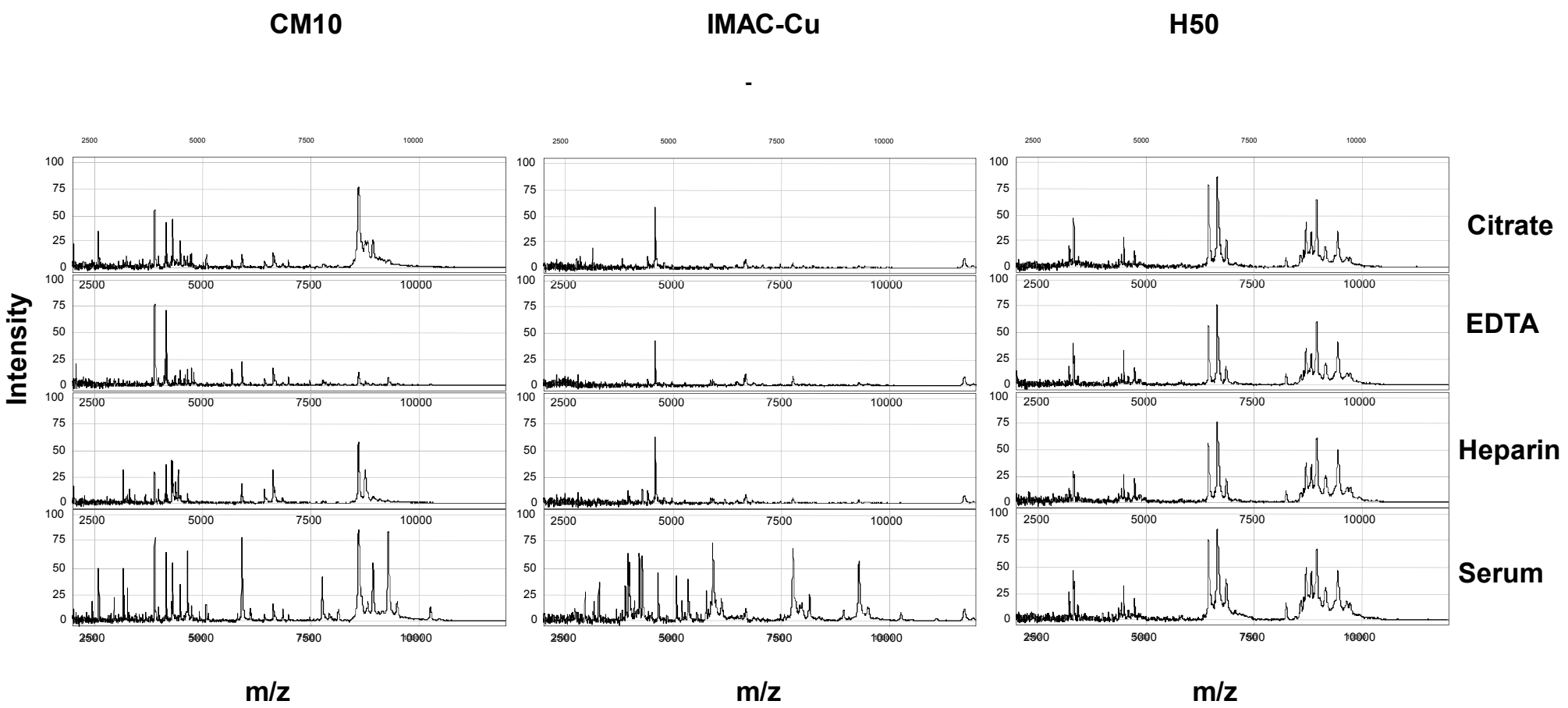
Výhoda SELDI:

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

Čipy pro SELDI hmotnostní spektrometrii

- Kvantitativní hodnoty proteomu jsou také ovlivněné různými zdroji variability (experimentální i biologické)
- Velmi velké rozdíly mezi typy použitého čipu!

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

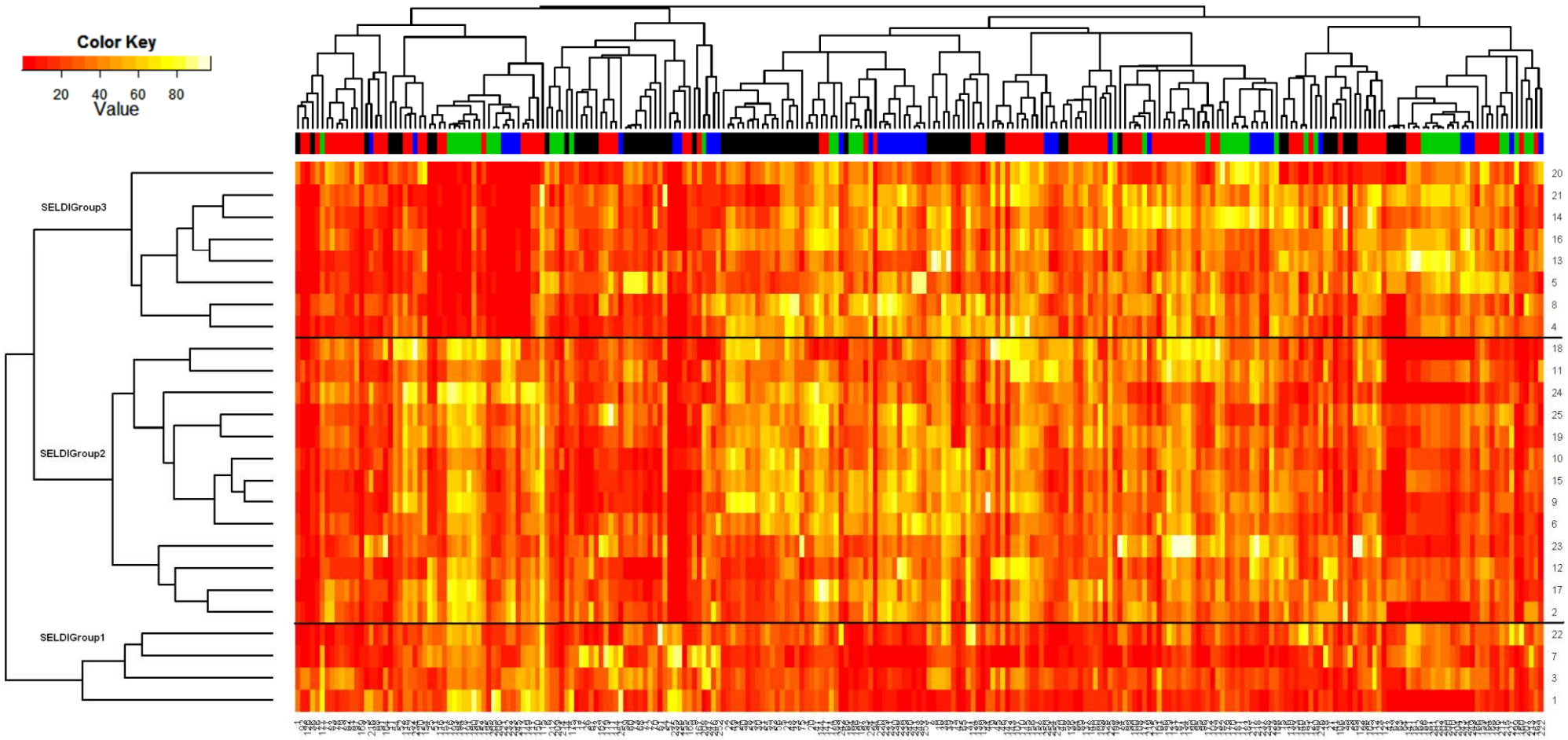


Peaky profilů 3 odlišných SELDI čipů
Vzorky zpracované 4 různými způsoby
Banks et al, Clinical Chemistry 2005

Příklad

- Shlukování profilů stejných vzorků ze 4 typů SELDI sklíček:

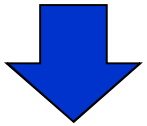
IMAC30, H50, NP20zas, NP20kys



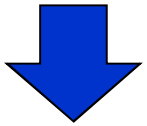
IBA

Úpravy dat

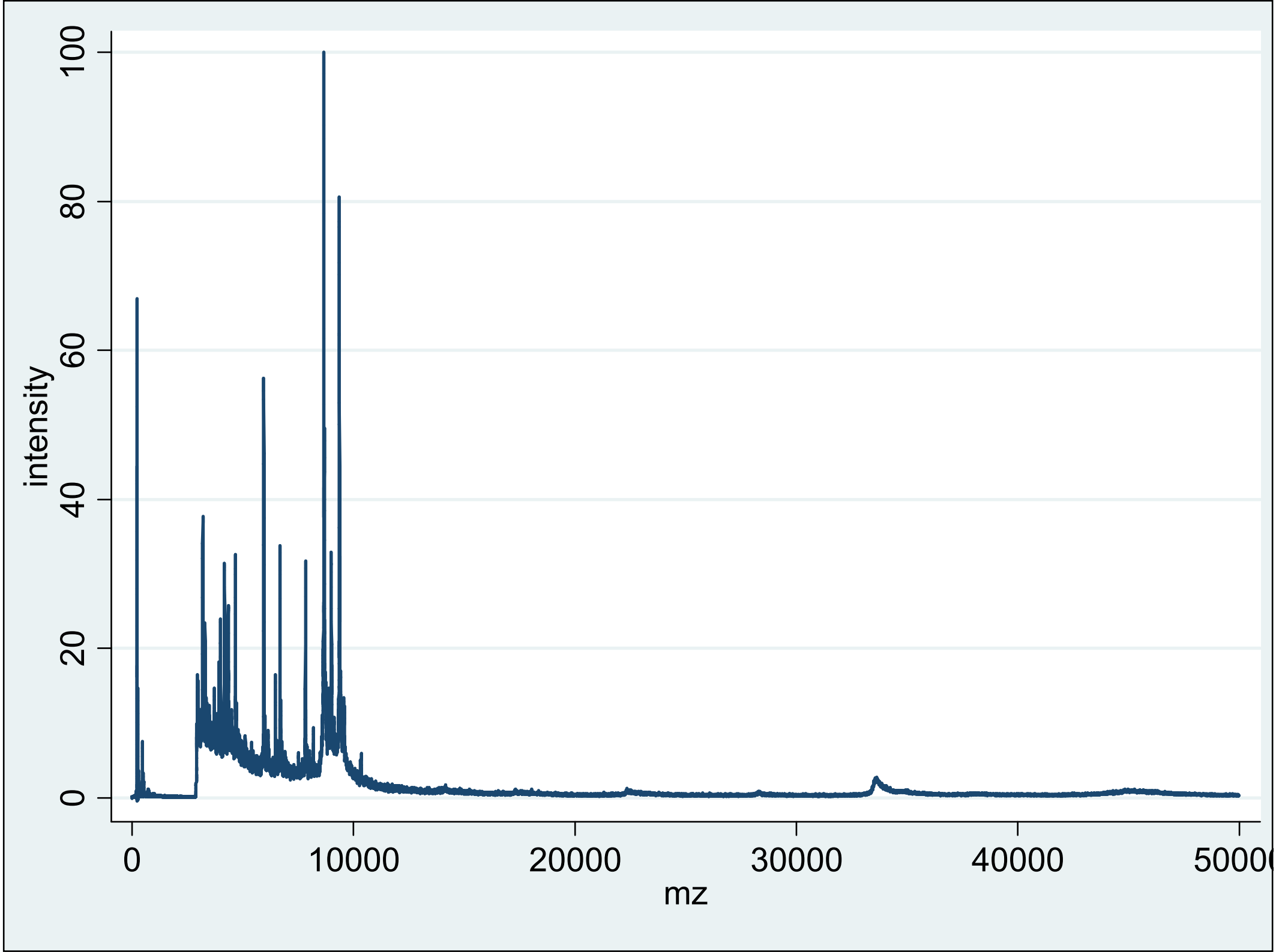
- **Kalibrace**
  - *TOF* přeměněný na škálu m/z pomocí množství kalibračních proteinů ze známou m/z hodnotou. Toto se děje ještě v laboratoři

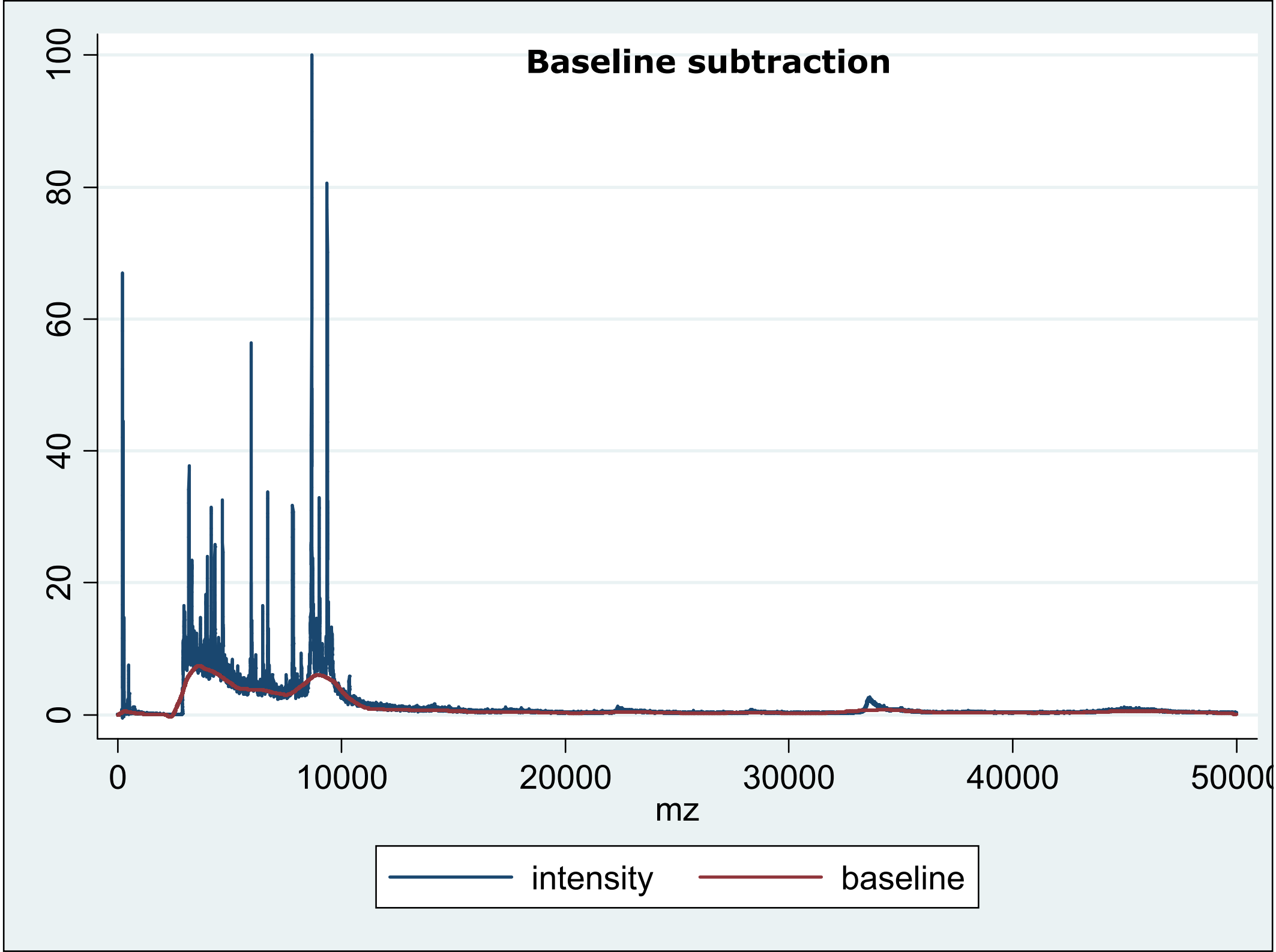


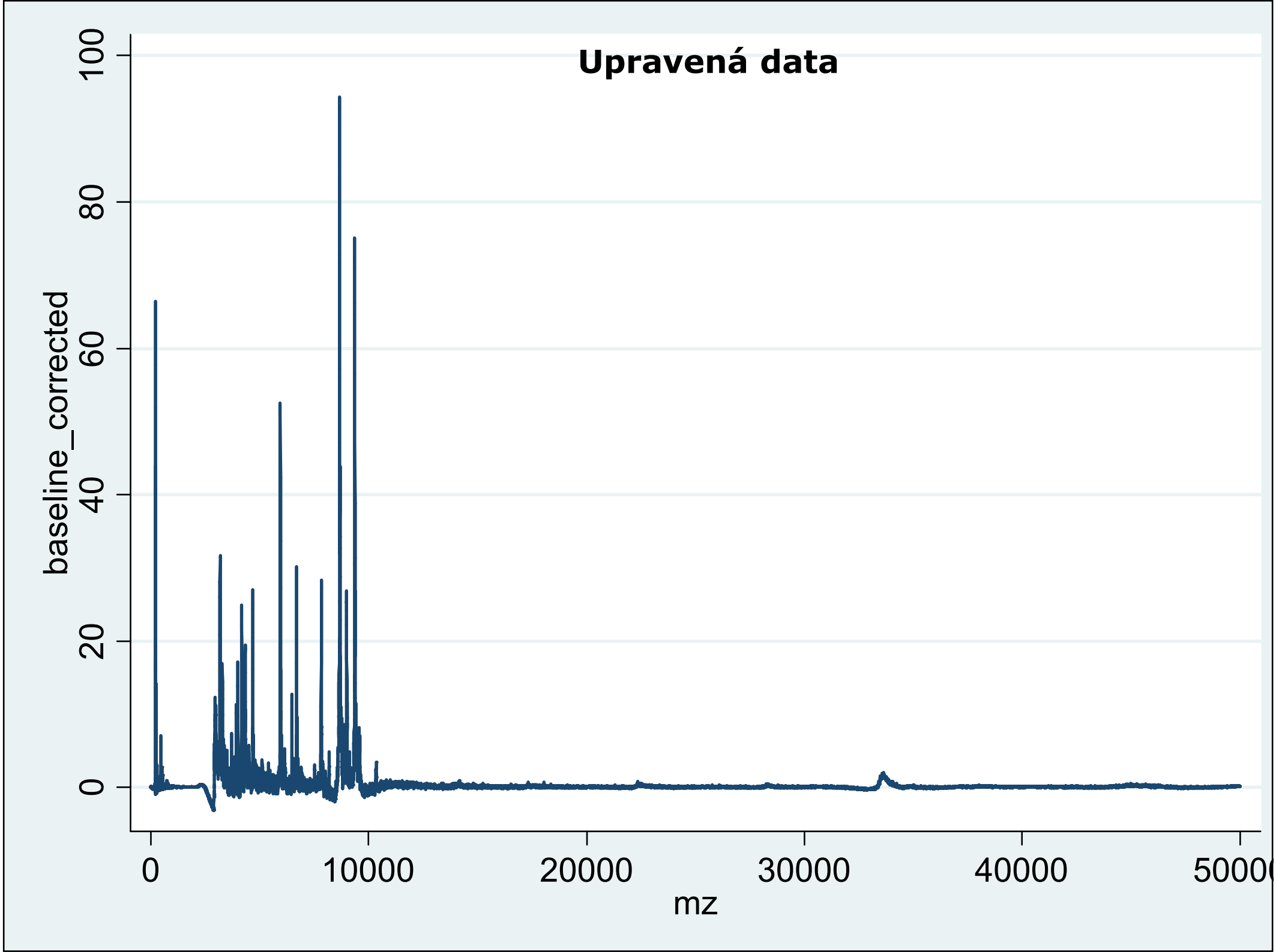
Základní data



- **Odstranění baseline**
  - Odstranění baseline šumu z profilu, například pomocí **loess**
- **Normalizace**
  - Abychom mohli porovnat spektra mezi vzorky


intensity
0
20
40
60
80
100
0
10000
20000
30000
40000
50000
mz

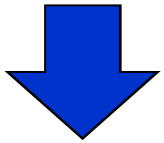

Baseline subtraction
100
80
60
40
20
0
0
10000
20000
30000
40000
50000
mz
intensity
baseline


Upravená data
baseline_corrected
0
20
40
60
80
100
0
10000
20000
30000
40000
mz

Normalizace

- Odstraňujeme technickou variabilitu (přístrojové chyby, odlišné množství vzorku)

- Koncentrace proteinu se odhaduje jako plocha pod píkem (Area Under Curve – AUC)

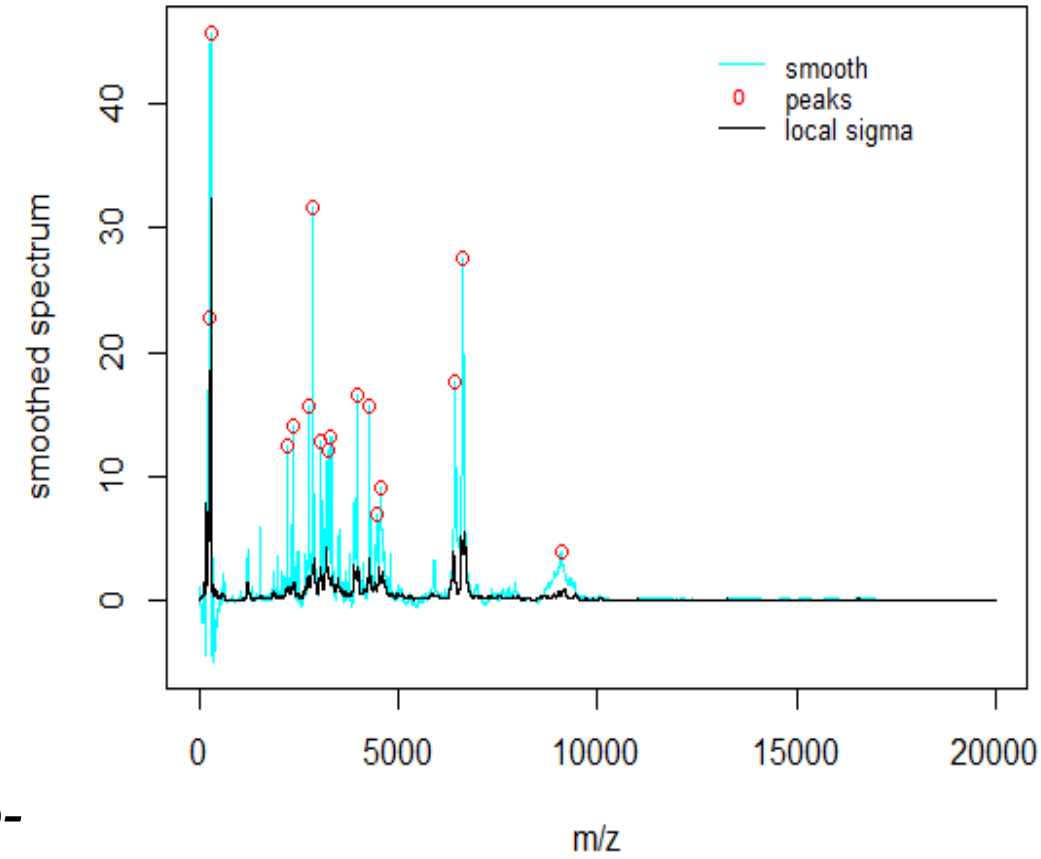


- Normalizace pomocí *průměrné AUC* (*TIC – total ion current)*

  AUC celého spektra / průměrná AUC všech spekter

Detekce píků a jejich zarovnání

- Pík ~ peptid/proteín, definuje se jako lokální maximum na základě porovnání variability v okolí
- Existují nepřesnosti na x (m/z) a y (signál) osách
- Píky každého spektra můžou být definované jako body které jsou maximálně +/- N bodů v okolí m/z
  - first, second, estimated..
- Důležité je brát do úvahy *signal-to-noise* ratio – píky musí překročit nějakou běžnou hranici šumu

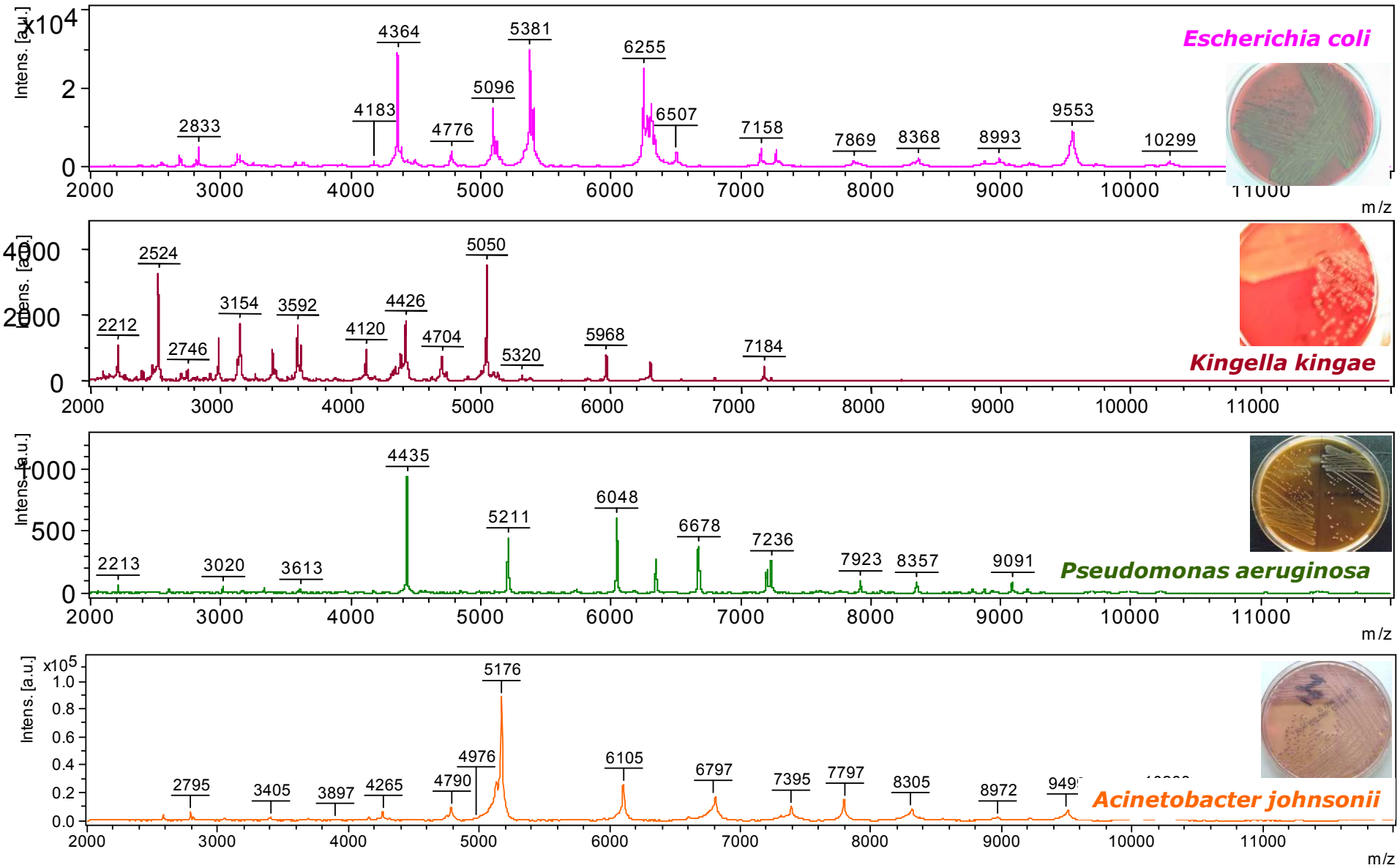


Jak vypadají data po zarovnání a detekci píků

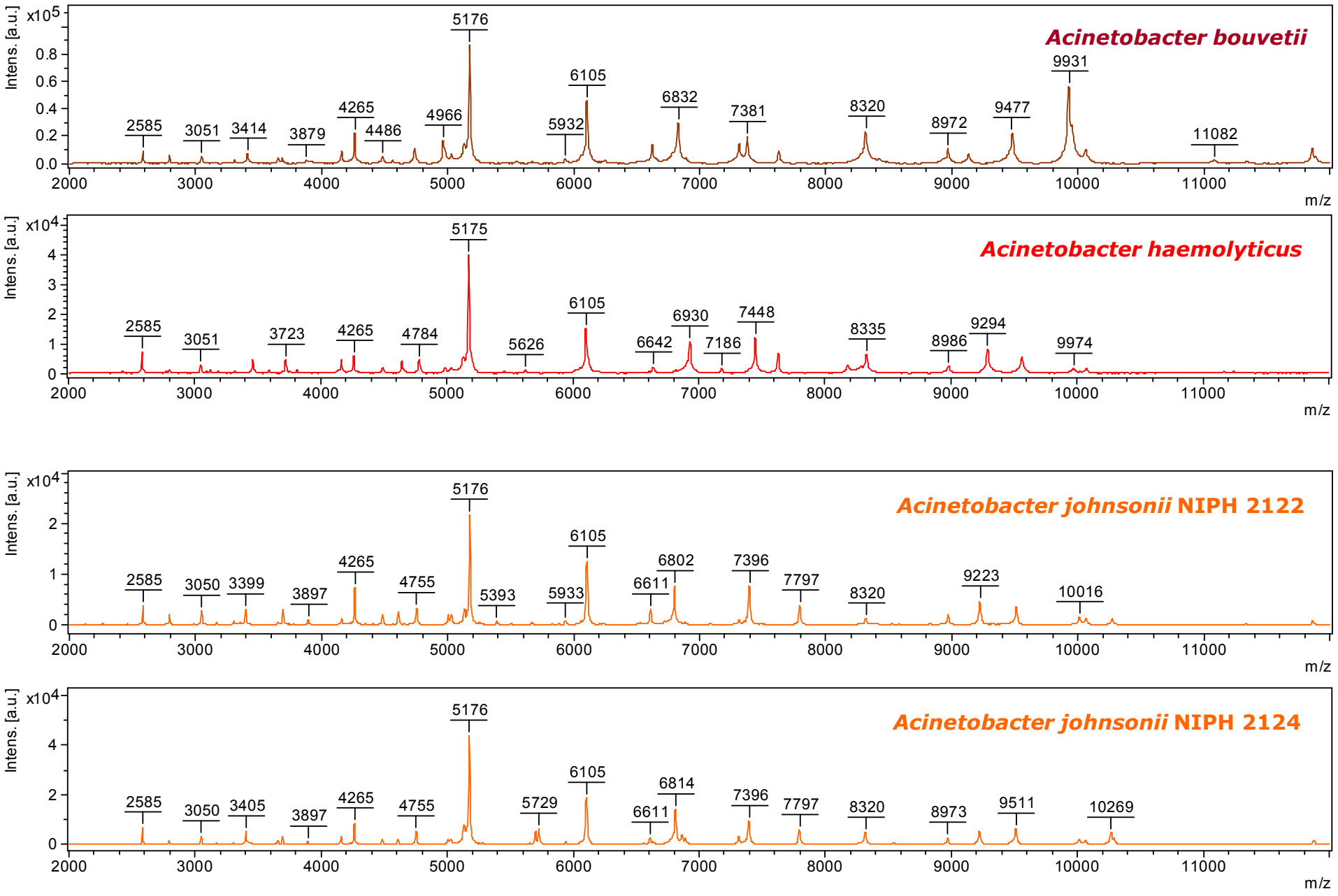
- SELDI-TOF

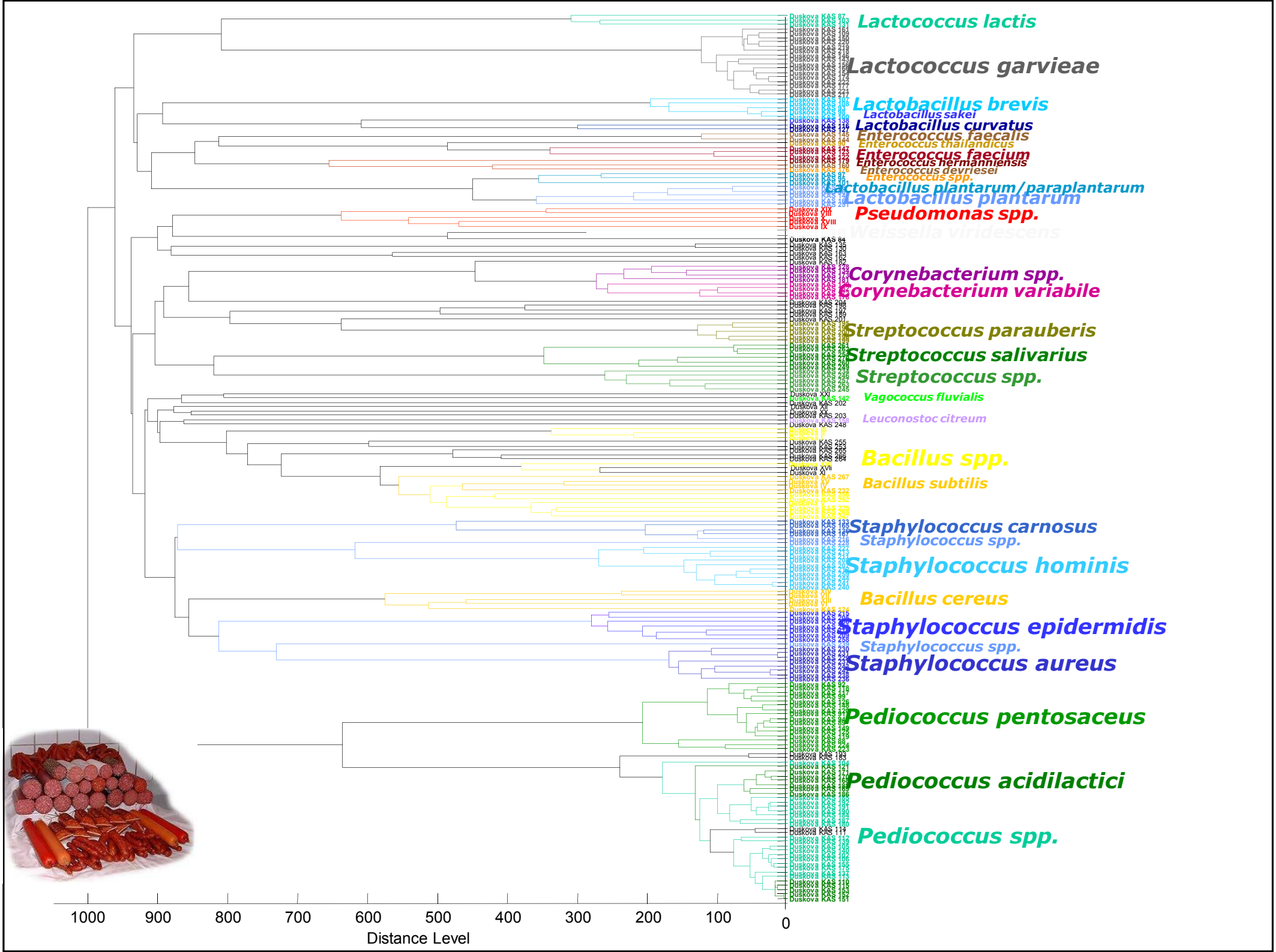
Cluster	Group	Norm. Log Intensity	M/Z	Intensity	Norm. Linear Intensity	Type	Mass Dev.
1	chemoresistentni	0.581550	2392.84	3.058176	30.578211	estimated	0.000007
1	chemoresistentni	-0.072123	2392.84	1.943959	12.984676	estimated	0.000007
1	chemoresistentni	0.023116	2392.84	2.076621	15.079403	estimated	0.000007
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

Aplikace I – identifikace bakterií

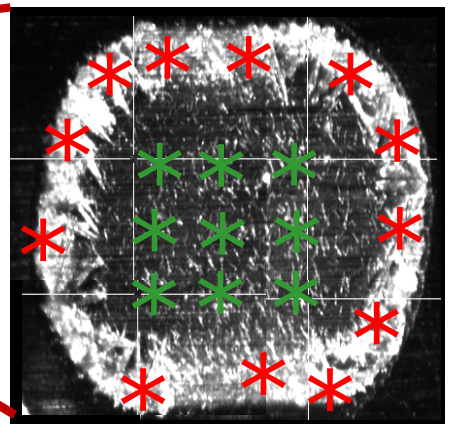


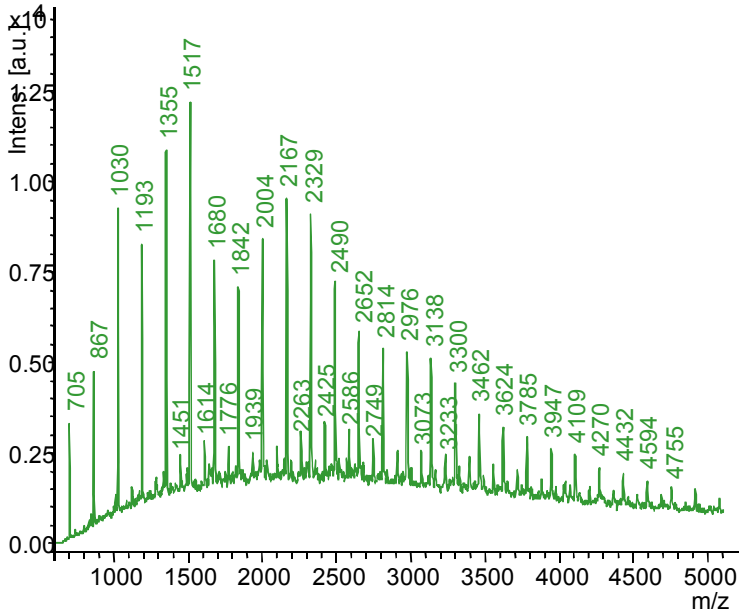
Aplikace I – identifikace bakterií



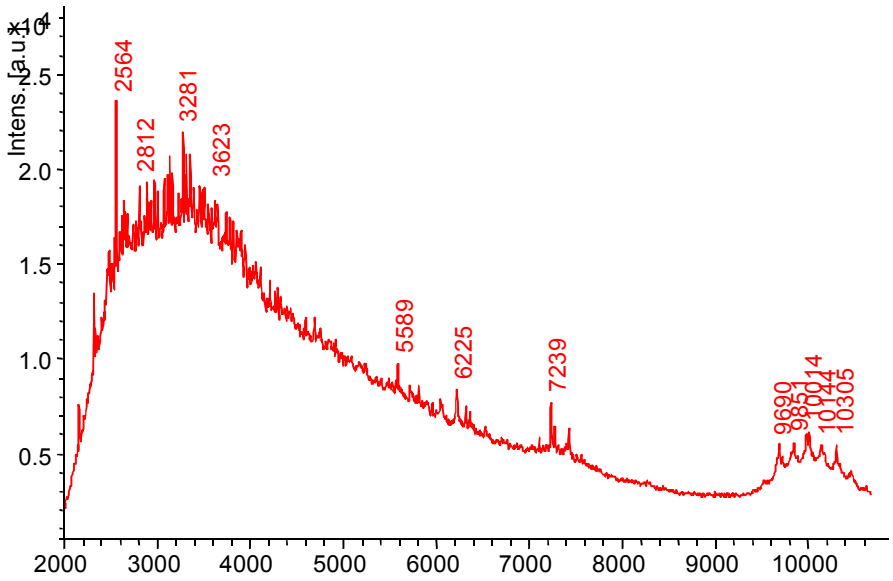

Lactococcus lactis
Lactococcus garvieae
Lactobacillus brevis
Lactobacillus sakei
Lactobacillus curvatus
Enterococcus faecalis
Enterococcus thailandicus
Enterococcus faecium
Enterococcus hermanniensis
Enterococcus devriesei
Enterococcus spp.
Lactobacillus plantarum/paraplantarum
Lactobacillus plantarum
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

Aplikace II





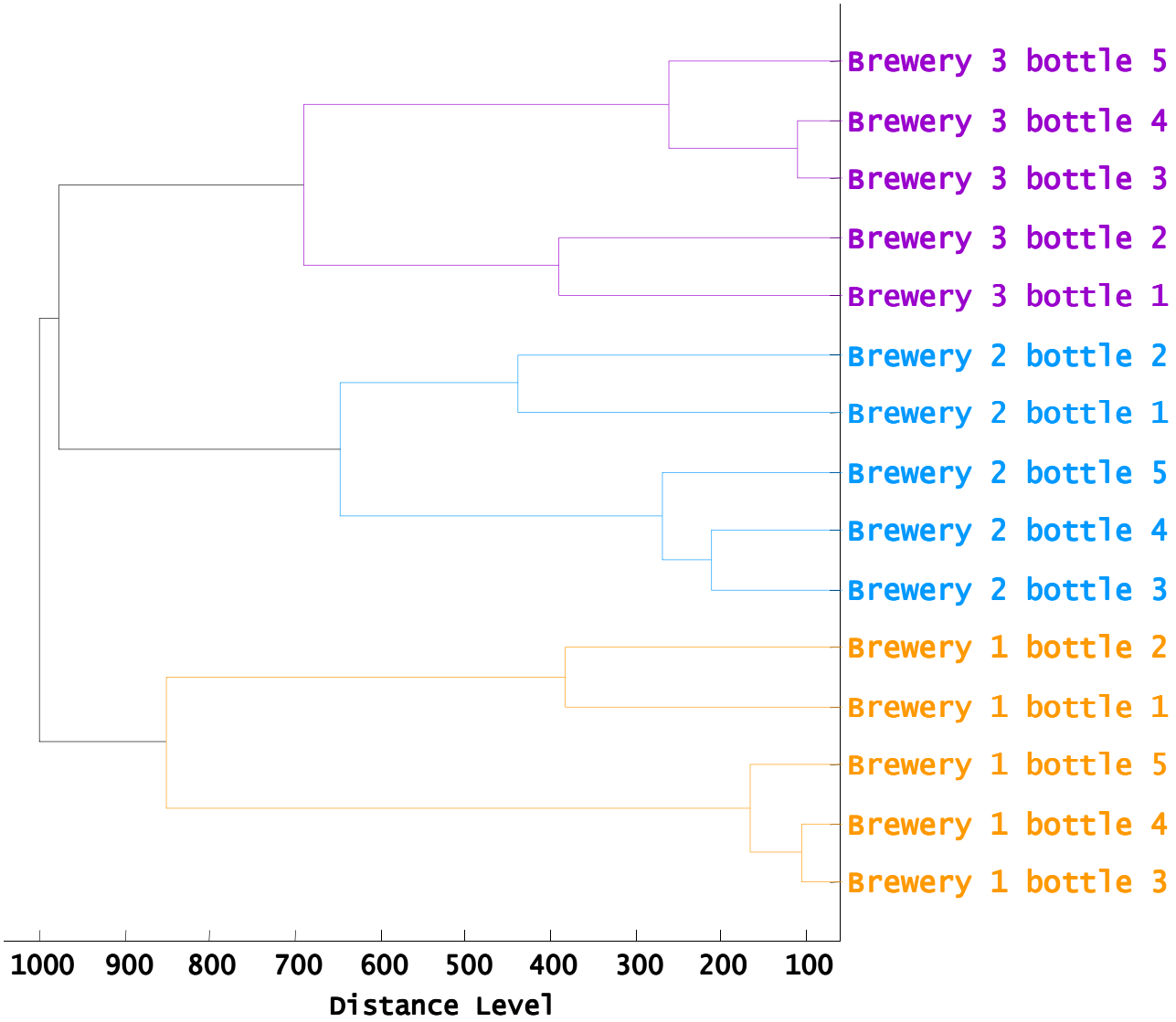
MALDI-TOF MS fingerprint containing maltooligosaccharides



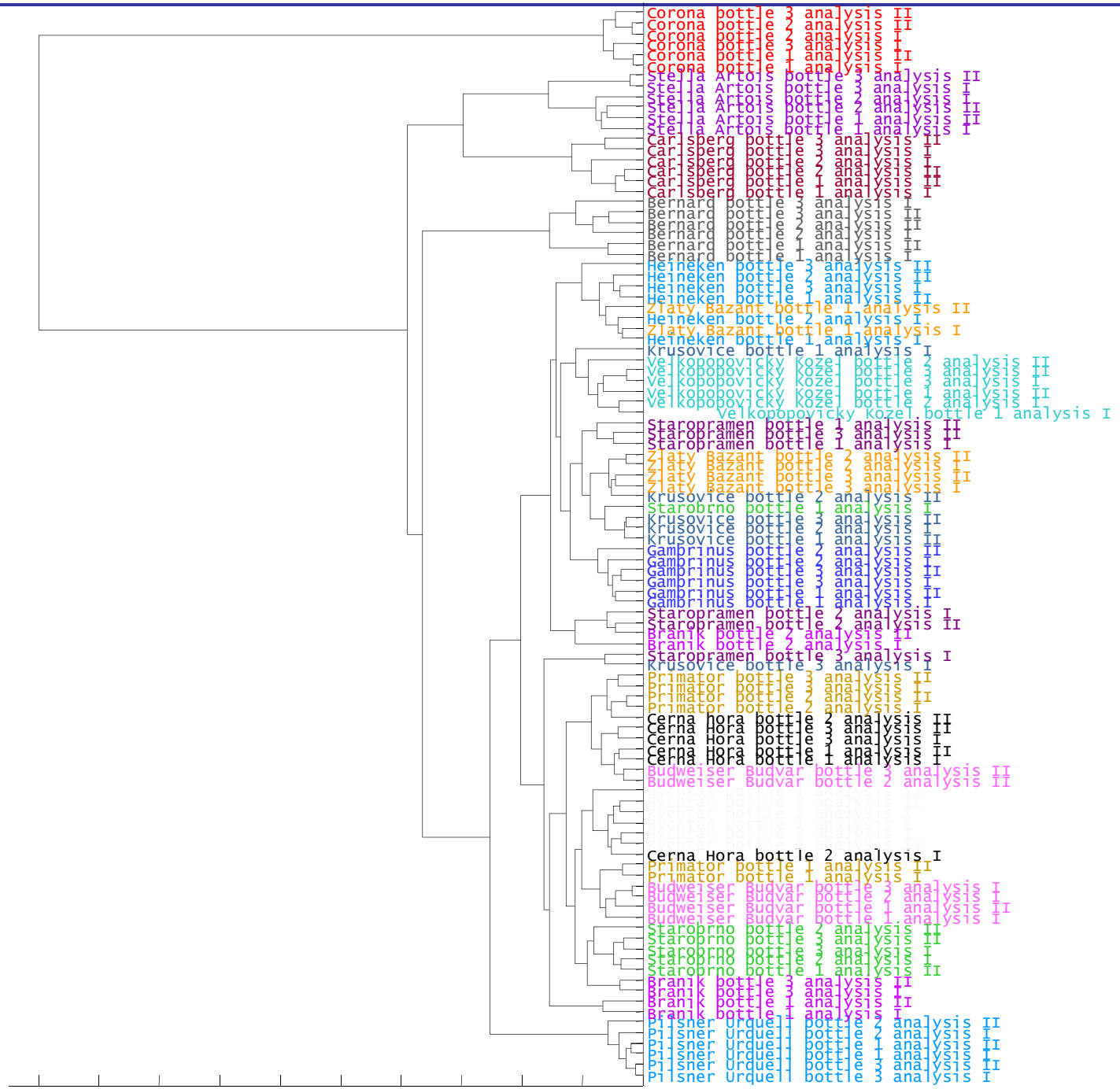
MALDI-TOF MS fingerprint containing proteins

Šedo et al., 2012

Aplikace II


Brewery 3 bottle 5
Brewery 3 bottle 4
Brewery 3 bottle 3
Brewery 3 bottle 2
Brewery 3 bottle 1
Brewery 2 bottle 2
Brewery 2 bottle 1
Brewery 2 bottle 5
Brewery 2 bottle 4
Brewery 2 bottle 3
Brewery 1 bottle 2
Brewery 1 bottle 1
Brewery 1 bottle 5
Brewery 1 bottle 4
Brewery 1 bottle 3
1000 900 800 700 600 500 400 300 200 100
Distance Level

Aplikace II


Corona bottle 3 analysis II
Corona bottle 2 analysis II
Corona bottle 2 analysis I
Corona bottle 3 analysis I
Corona bottle 1 analysis II
Corona bottle 1 analysis I
Stella Artois bottle 3 analysis II
Stella Artois bottle 3 analysis I
Stella Artois bottle 2 analysis I
Stella Artois bottle 2 analysis II
Stella Artois bottle 1 analysis II
Stella Artois bottle 1 analysis I
Carlsberg bottle 3 analysis II
Carlsberg bottle 3 analysis I
Carlsberg bottle 2 analysis I
Carlsberg bottle 2 analysis II
Carlsberg bottle 1 analysis II
Carlsberg bottle 1 analysis I
Bernard bottle 3 analysis I
Bernard bottle 3 analysis II
Bernard bottle 2 analysis II
Bernard bottle 2 analysis I
Bernard bottle 1 analysis II
Bernard bottle 1 analysis I
Heineken bottle 3 analysis II
Heineken bottle 2 analysis II
Heineken bottle 3 analysis I
Heineken bottle 1 analysis II
Zlaty Bazant bottle 1 analysis II
Heineken bottle 2 analysis I
Zlaty Bazant bottle 1 analysis I
Heineken bottle 1 analysis I
Krusovice bottle 1 analysis I
Velkopopovicky Kozel bottle 2 analysis II
Velkopopovicky Kozel bottle 3 analysis II
Velkopopovicky Kozel bottle 3 analysis I
Velkopopovicky Kozel bottle 1 analysis II
Velkopopovicky Kozel bottle 2 analysis I
Velkopopovicky Kozel bottle 1 analysis I
Staropramen bottle 1 analysis II
Staropramen bottle 3 analysis II
Staropramen bottle 1 analysis I
Zlaty Bazant bottle 2 analysis II
Zlaty Bazant bottle 2 analysis I
Zlaty Bazant bottle 3 analysis II
Zlaty Bazant bottle 3 analysis I
Krusovice bottle 2 analysis II
Starobrno bottle 1 analysis I
Krusovice bottle 3 analysis II
Krusovice bottle 2 analysis I
Krusovice bottle 1 analysis II
Gambrinus bottle 2 analysis II
Gambrinus bottle 2 analysis I
Gambrinus bottle 3 analysis II
Gambrinus bottle 3 analysis I
Gambrinus bottle 1 analysis II
Gambrinus bottle 1 analysis I
Staropramen bottle 2 analysis I
Staropramen bottle 2 analysis II
Branik bottle 2 analysis II
Branik bottle 2 analysis I
Staropramen bottle 3 analysis I
Krusovice bottle 3 analysis I
Primator bottle 3 analysis II
Primator bottle 3 analysis I
Primator bottle 2 analysis II
Primator bottle 2 analysis I
Cerna Hora bottle 2 analysis II
Cerna Hora bottle 3 analysis II
Cerna Hora bottle 3 analysis I
Cerna Hora bottle 1 analysis II
Cerna Hora bottle 1 analysis I
Budweiser Budvar bottle 3 analysis II
Budweiser Budvar bottle 2 analysis II
Cerna Hora bottle 2 analysis I
Primator bottle 1 analysis II
Primator bottle 1 analysis I
Budweiser Budvar bottle 3 analysis I
Budweiser Budvar bottle 2 analysis I
Budweiser Budvar bottle 1 analysis II
Budweiser Budvar bottle 1 analysis I
Starobrno bottle 2 analysis II
Starobrno bottle 3 analysis II
Starobrno bottle 3 analysis I
Starobrno bottle 2 analysis I
Starobrno bottle 1 analysis II
Branik bottle 3 analysis II
Branik bottle 3 analysis I
Branik bottle 1 analysis II
Branik bottle 1 analysis I
Pilsner Urquell bottle 2 analysis II
Pilsner Urquell bottle 2 analysis I
Pilsner Urquell bottle 1 analysis II
Pilsner Urquell bottle 1 analysis I
Pilsner Urquell bottle 3 analysis II
Pilsner Urquell bottle 3 analysis I
1000
900
800
700
600
500
400
300
200
100
0

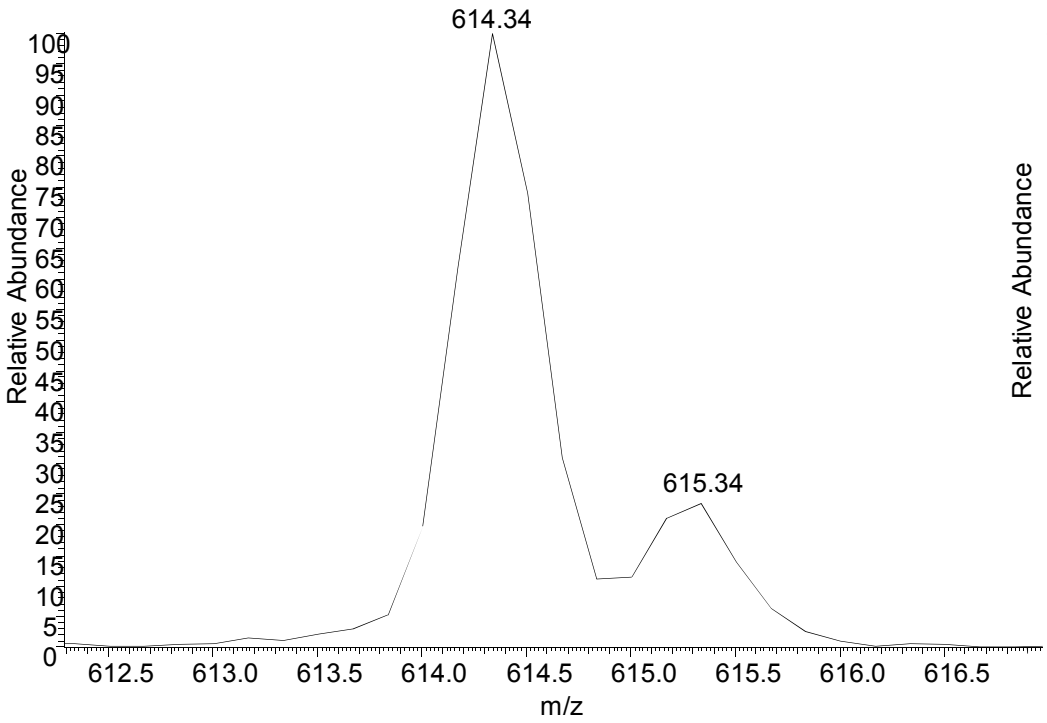
Zpracování dat

1. úprava hrubých dat (MS/MS i MS)
2. příprava dat pro databázové vyhledávání
3. příprava proteinové databáze, databázové vyhledávání
4. zpracování výsledků z pohledu FDR
5. výběr peptidových identifikací (PSMs)
6. rekonstrukce seznamu „identifikovaných“ proteinů
7. interpretace proteinového seznamu(ů)

Úprava hrubých dat

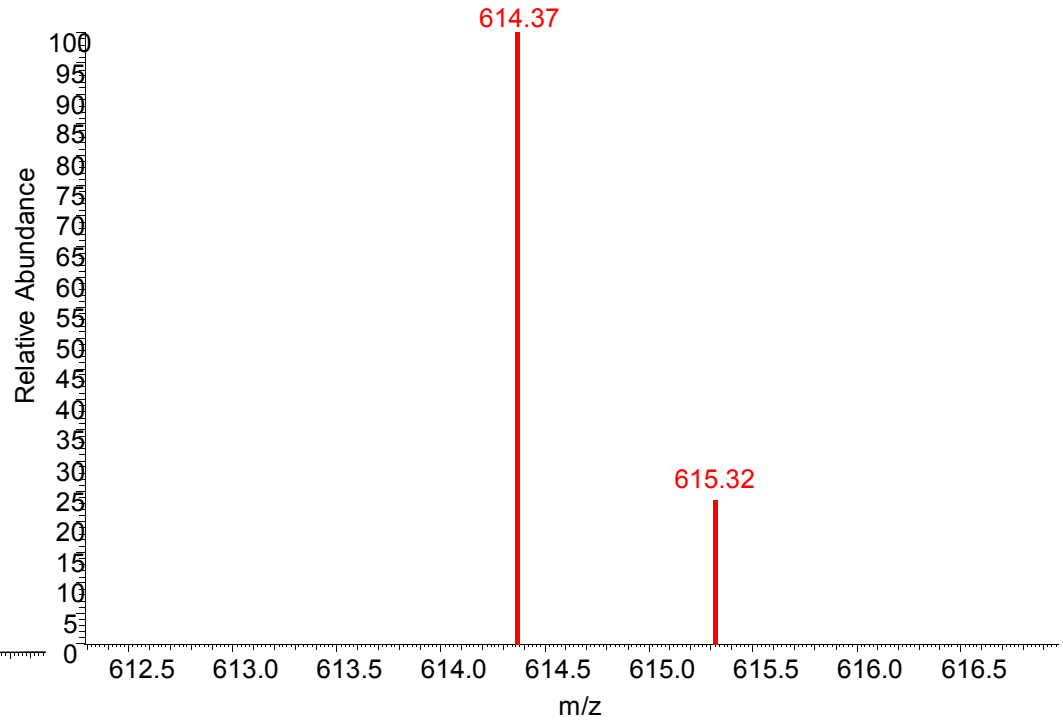
Profilové spektrum

- Získané z experimentu

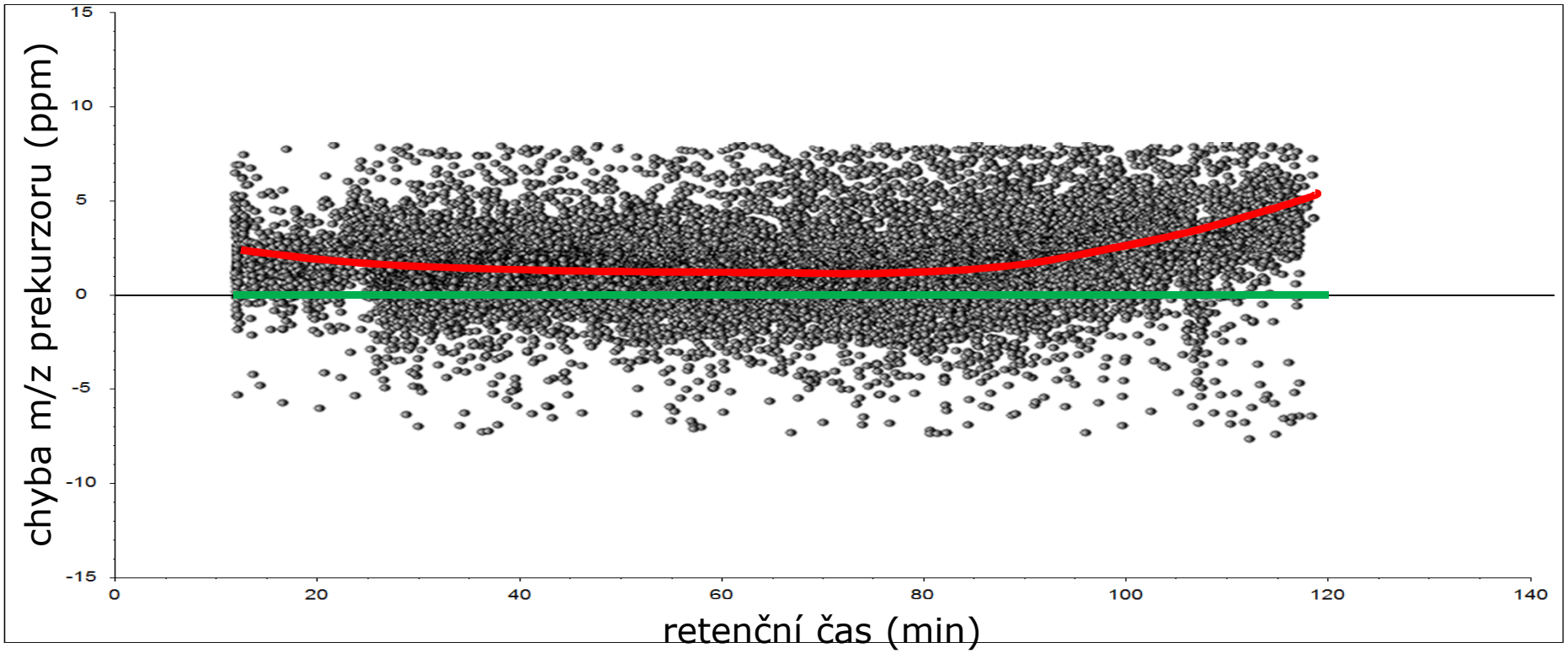


Čárové spektrum

- Vypočítané z profilového



Rekalibrace



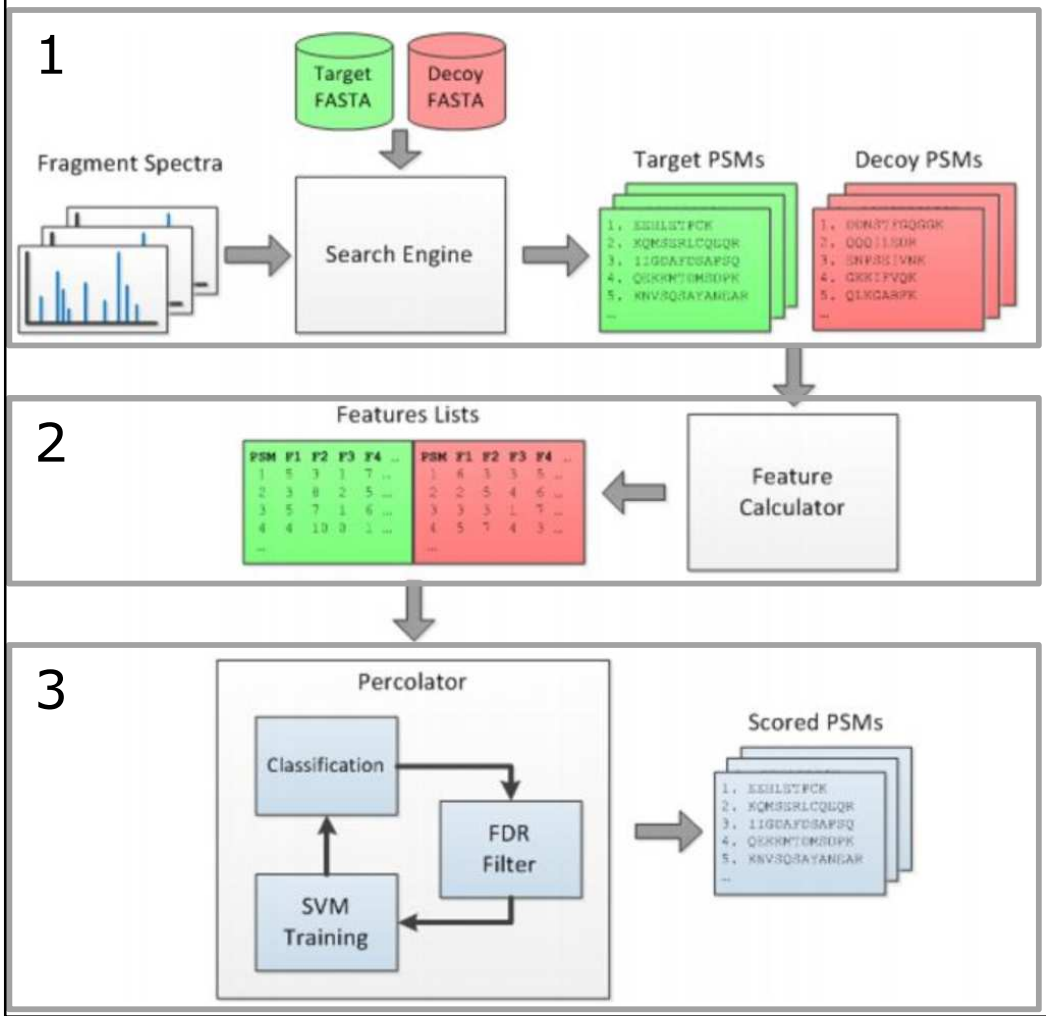
- Interní rekalibrace – stejná na celý záznam
- lockmass – každé spektrum podle opakující se kontaminanty
- Podle identifikovaných peptidů – odhad závislosti z identifikovaných peptidů (polynom)

Databázové vyhledávání

1. Příprava dat
   - Výběr „reprezentativních“ signálů MS/MS
   - Odstranění „méně kvalitních“ spekter MS/MS
     - Top N (z okna), dekonvoluce signálu a šumu
   - Získáme tabulku m/z hodnot a intenzit
2. Příprava databáze
   - *in silico* štěpení sekvencí z databáze
   - Přiřadění jednoho a nebo více peptidů k jednomu spektru (*decoy* databáze, FDR, Percolator)
3. Výběr peptidových identifikací

Percolator

1. prohledání dat MS/MS
2. výpočet „vlastností" peptidů
3. propočítání skóre peptidů



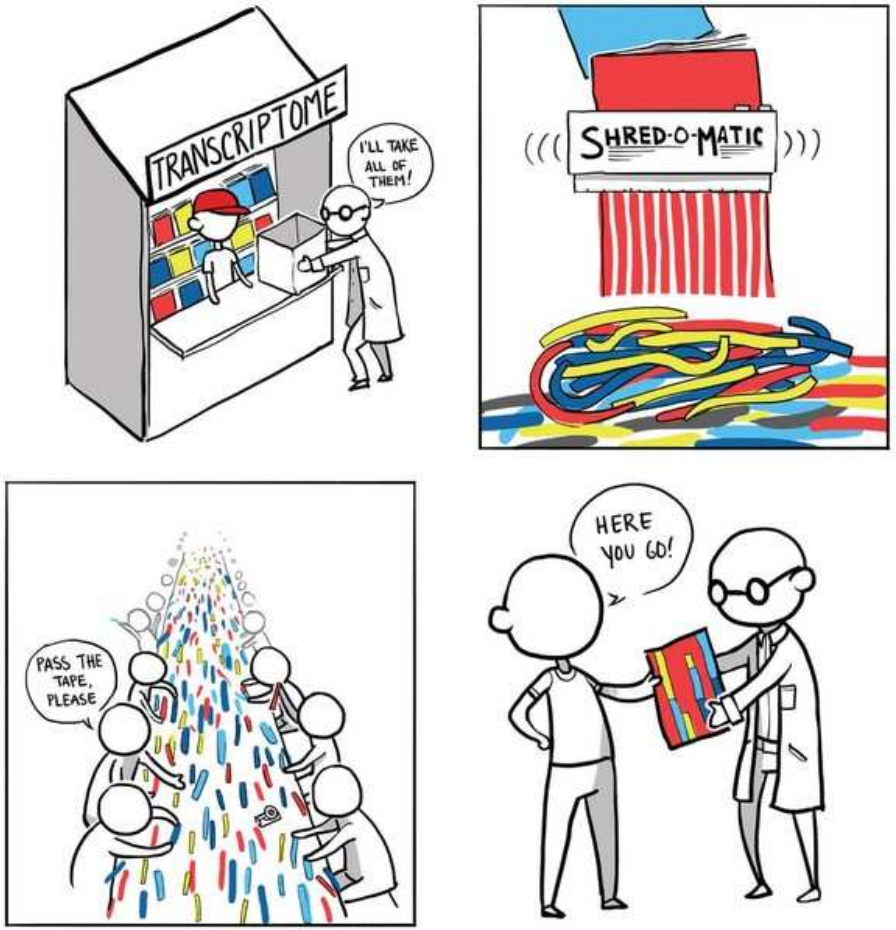
Propočítání skóre peptidů

- Použití *support vector machines* (SVM)
- sady identifikací
  - falešně pozitivní – *decoy* databáze
  - pozitivní – původní databáze (↑skóre)
- přiřazení vah vlastnostem v SVM
  - např. skóre; chyba hmotnosti intenzita, modifikace, ...

⇒ víc identifikovaných peptidů

IBA

Rekonstrukce sady proteinů



Analogie puzzle, ALE:

- Tisíce kousků:
  - Stejné
  - Poškozené
  - Chýbějící
  - Z jiných skládaček
- Pasují na stejná místa

Korf, Nat Methods, 2013

Vybrané přístupy

- N – peptidové pravidlo
  - Proteiny, u kterých pozorujeme alespoň N peptidů
  - Vysoká falešná pozitivita
  - Používané na sekvenční homologické proteiny

- Pravděpodobnostní přístupy
  - *ProteinProphet, Nested mixtures, Fido*

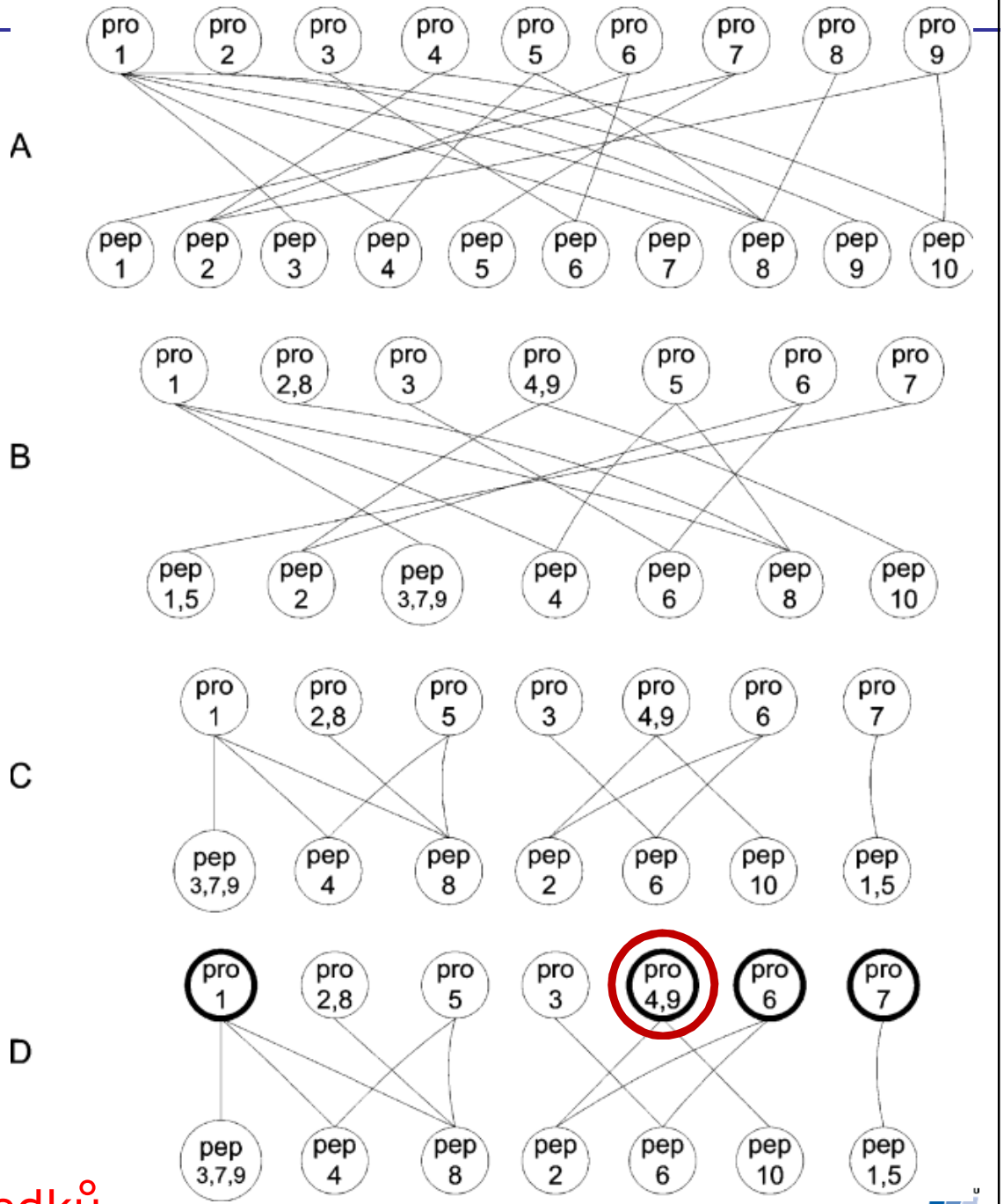
Princip parsimonie a Occamové břitvy

A. Vytvoření biparitního grafu: peptidy - možné proteiny

B. Sloučení proteinů a peptidů do skupin (např. pep 3,7,9; pro 4,9)

C. Rozdělení skupin

D. Výběr minimální sady proteinů



Důsledek: falešná negativita výsledků

Zhang, B. et al. *J. Proteome Res.*, 2007, *6*, 3549–3557.

Co s identifikovanými proteiny?

- Závisí od původního experimentu

- Typicky doplnění anotace proteinů z databáze (GO, KEGG, TAIR) a použití metod analýzy genových sad

Identifikace proteinů

- NCBI Protein -http://www.ncbi.nlm.nih.gov/protein
  - jen pro proteinové sekvence odvozeneé ch
- RefSeq - http://www.ncbi.nlm.nih.gov/RefSeq/
- UniProt– ze; kompozit SwissProt,TrEMBL a PIR-PSD–http://www.uniprot.org

Databáze dat

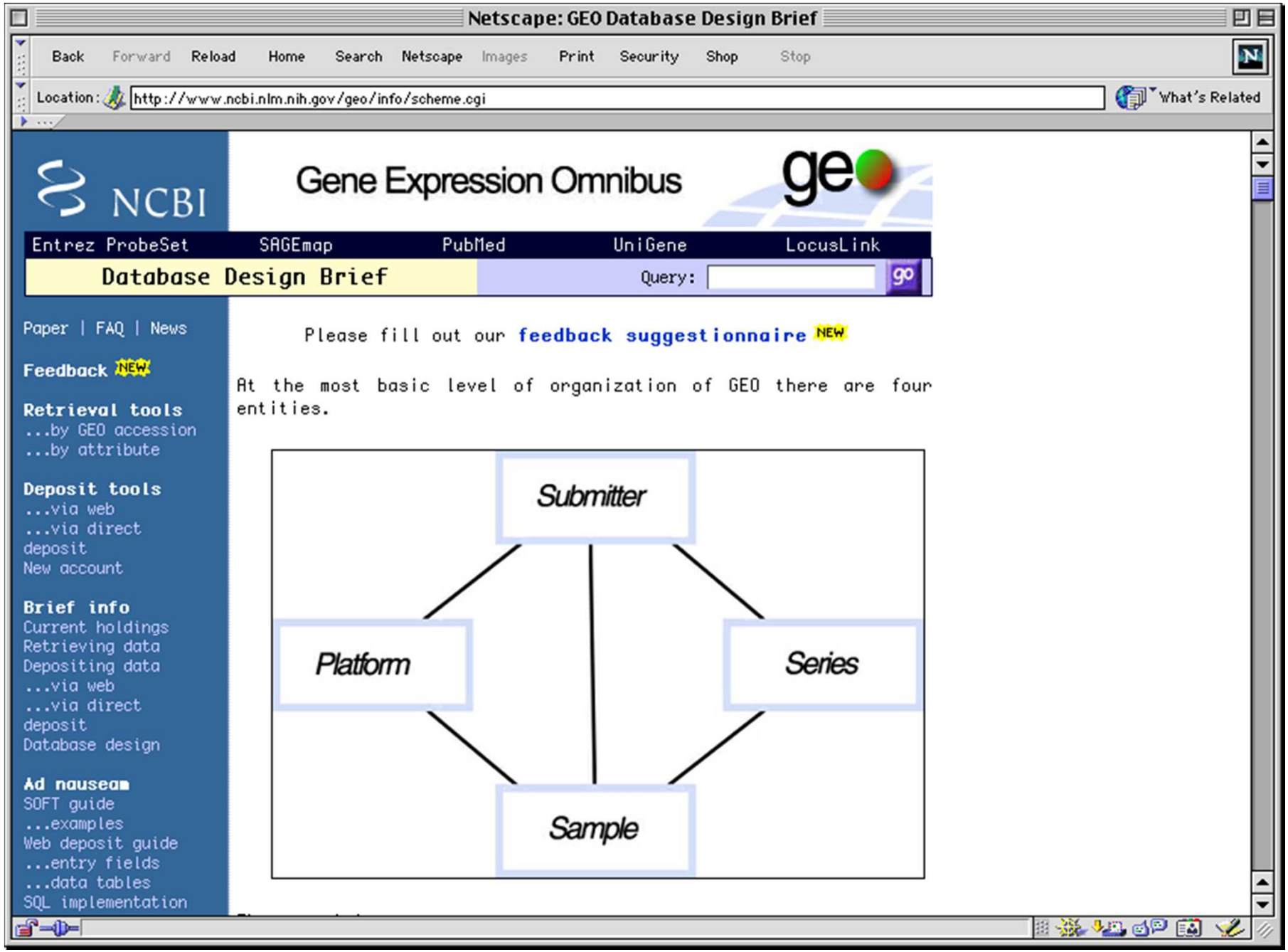
Veřejně přístupné databáze

- Velké experimenty mají až stovky, a nebo tisíce vzorků, v každé se studují desetitisíce až stovky genů

- Pro publikaci výsledků je vyžadované vložit data ve standardizovaném formátu (MIAME – Minimal Information About a Microarray Experiment) do jedné z veřejně přístupných databází tak, aby kdokoliv byl schopný výsledky zreprodukovat

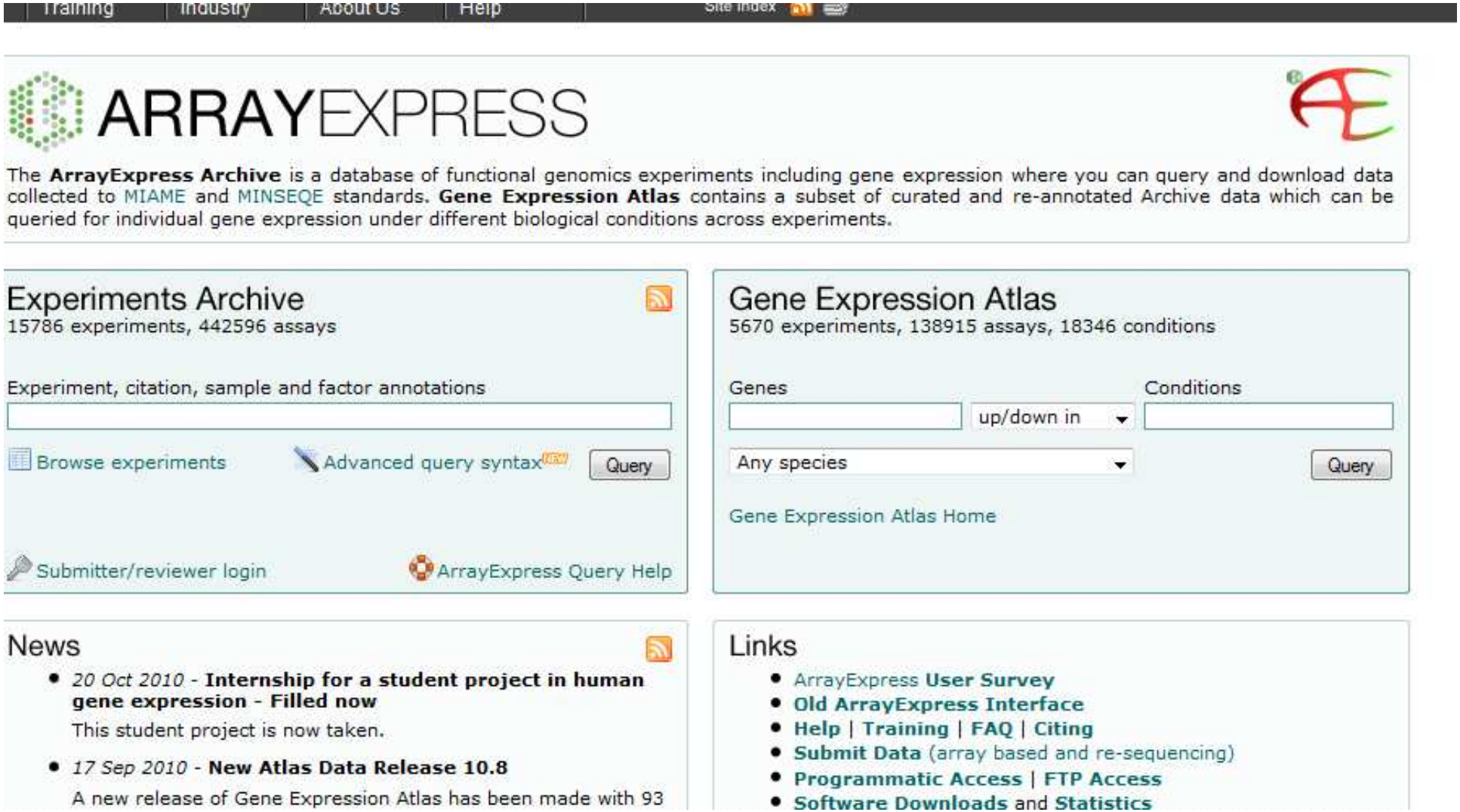
- Toto přináší velkou výhodu:
  - Můžeme data podrobit meta-analýze (simultánně porovnat data z různých experimentů)
  - Díky standardnímu formátu můžeme vyhledávat soubory s parametry, které potřebujeme
  - Data můžeme automaticky stahovat

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.
Submitter
Platform
Series
Sample

Array Express na EBI

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



- E-learningová skripta analýzy dat IBA
- http://portal.matematickabiologie.cz/index.php?pg=analyza-genomickych-a-proteomickych-dat--analyza-genomickych-a-proteomickych-dat