

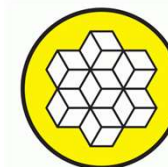
# Label Free kvantifikace, akvizční módy, databáze a FDR

Pavel Talacko

OMICS Servisní oddělení hmotnostní spektrometrie  
Přírodovědecká Fakulta UK, Biocev



PŘÍRODOVĚDECKÁ  
FAKULTA  
Univerzita Karlova



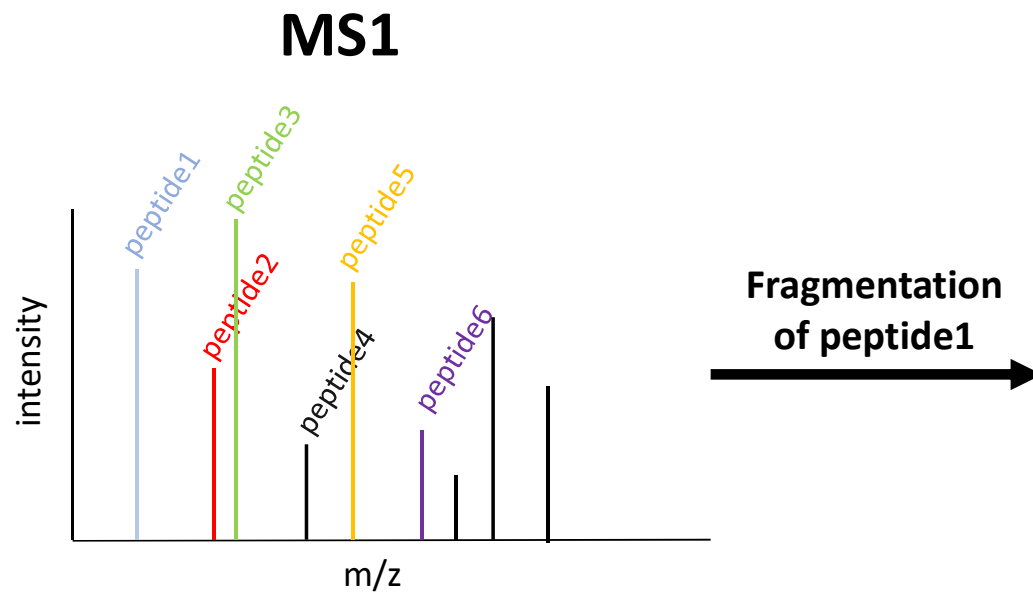
BIOCEV

# MS1 vs MS2 spectra

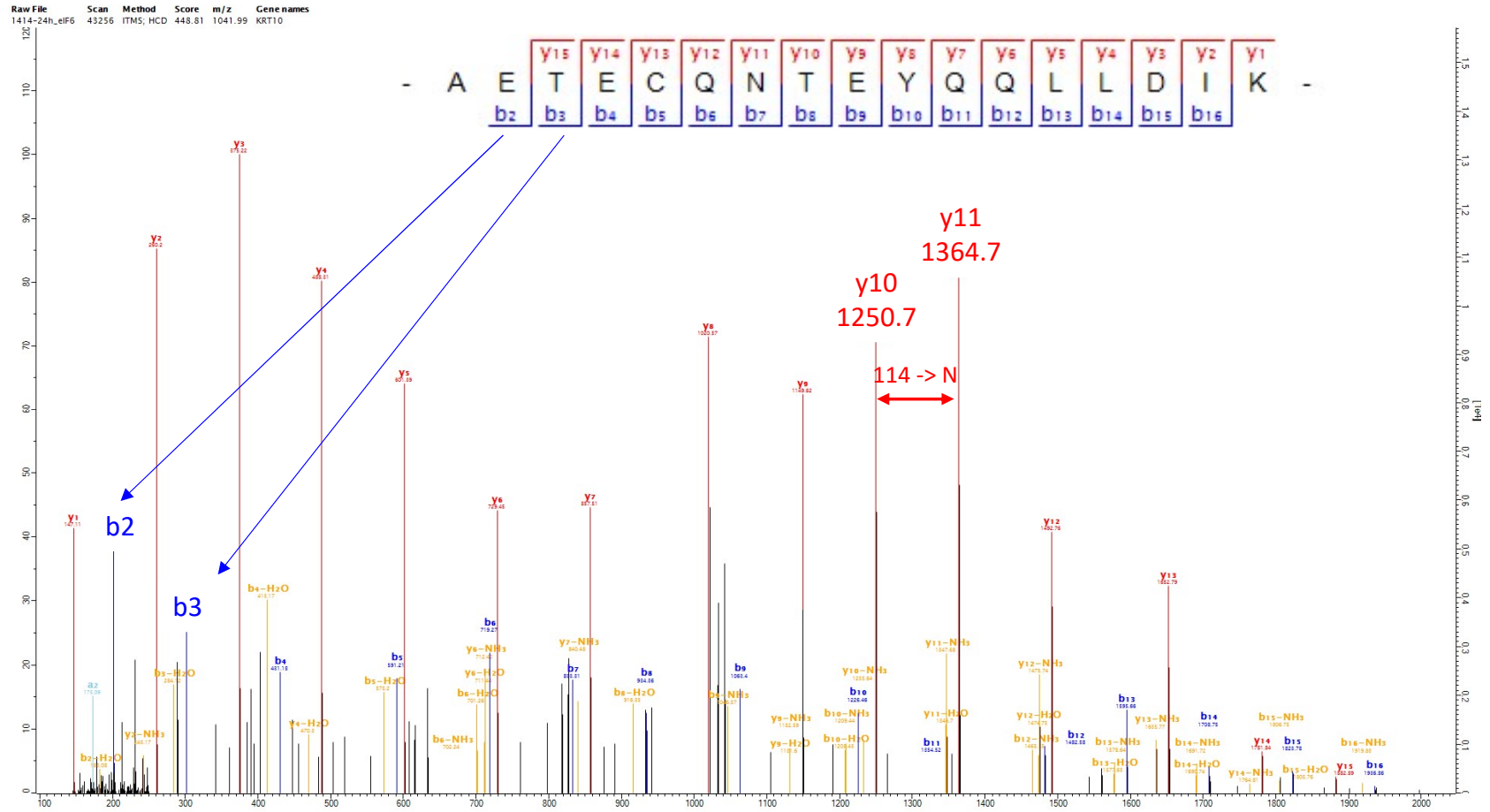
## MS1 spektrum (full scan):

Signál všech peptidů, které se v danou chvíli eluují z kolony

Peptid = prekurzor



# Fragmentační spektra peptidů

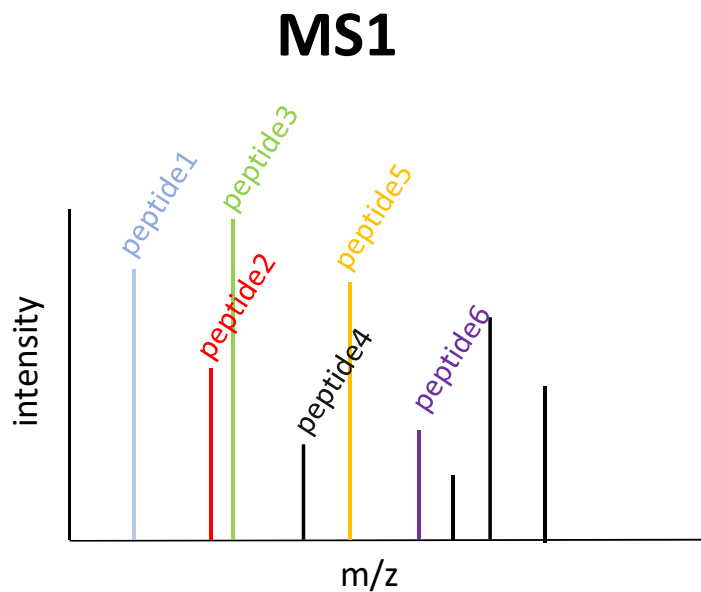


# MS1 vs MS2 spectra

## MS1 spektrum (full scan):

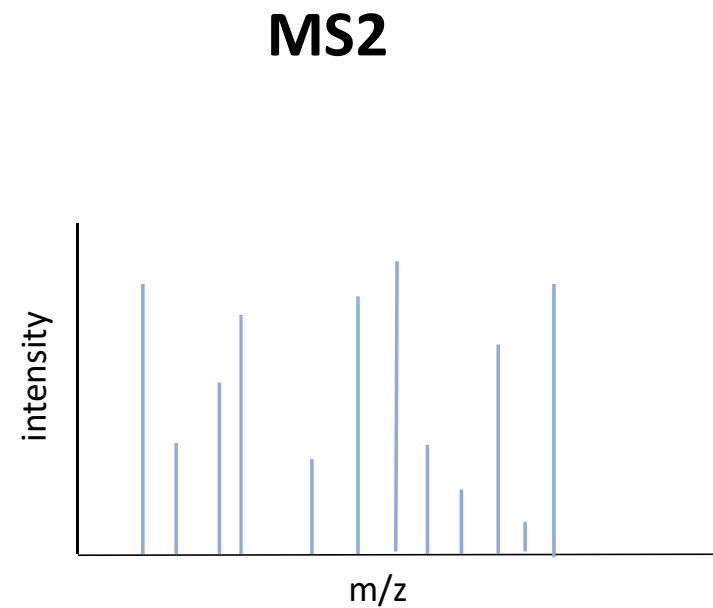
Signál všech peptidů, které se v danou chvíli eluují z kolony

Peptid = prekurzor



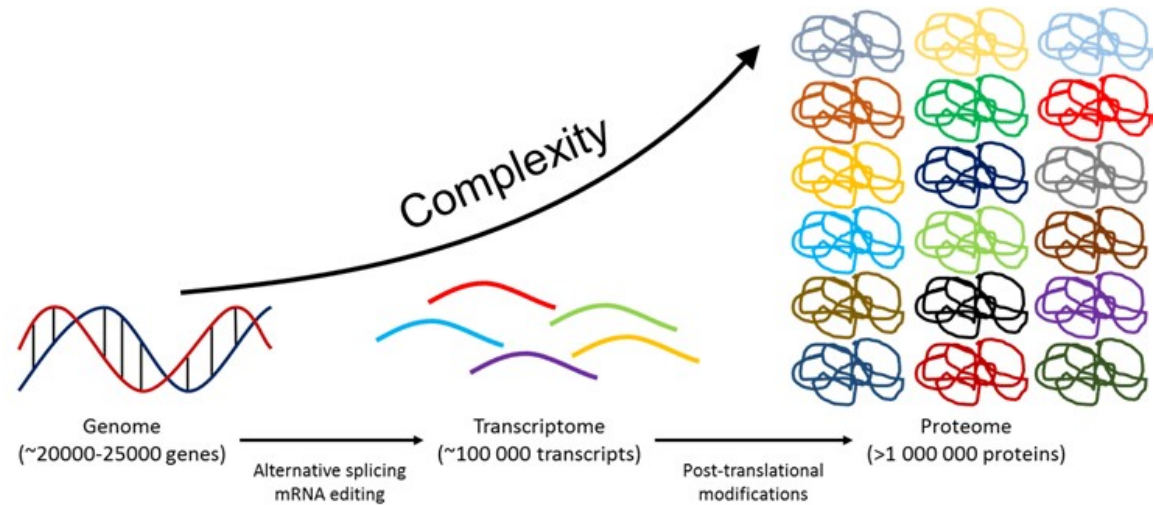
## MS2 spektrum (MS/MS, MS<sup>2</sup>)

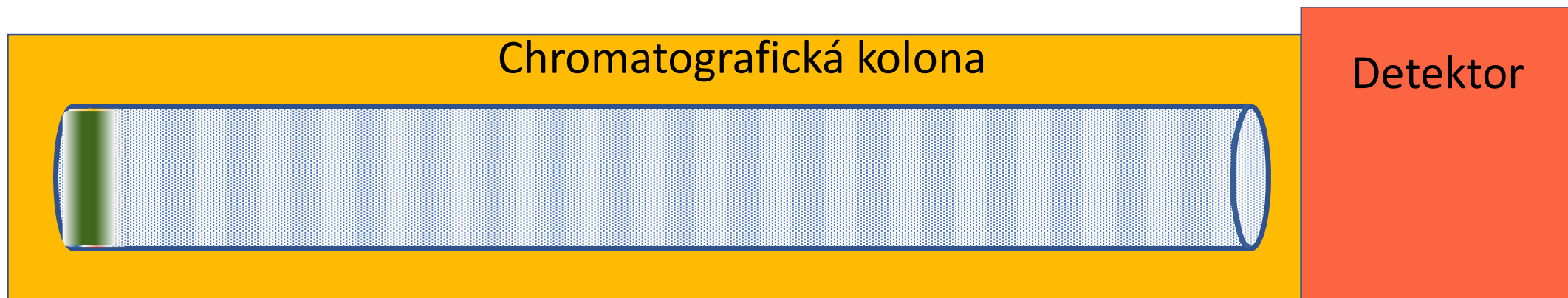
Signál fragmentů konkrétního peptidu (prekurzoru) = fragmentační spektrum



# Komplexita proteomu

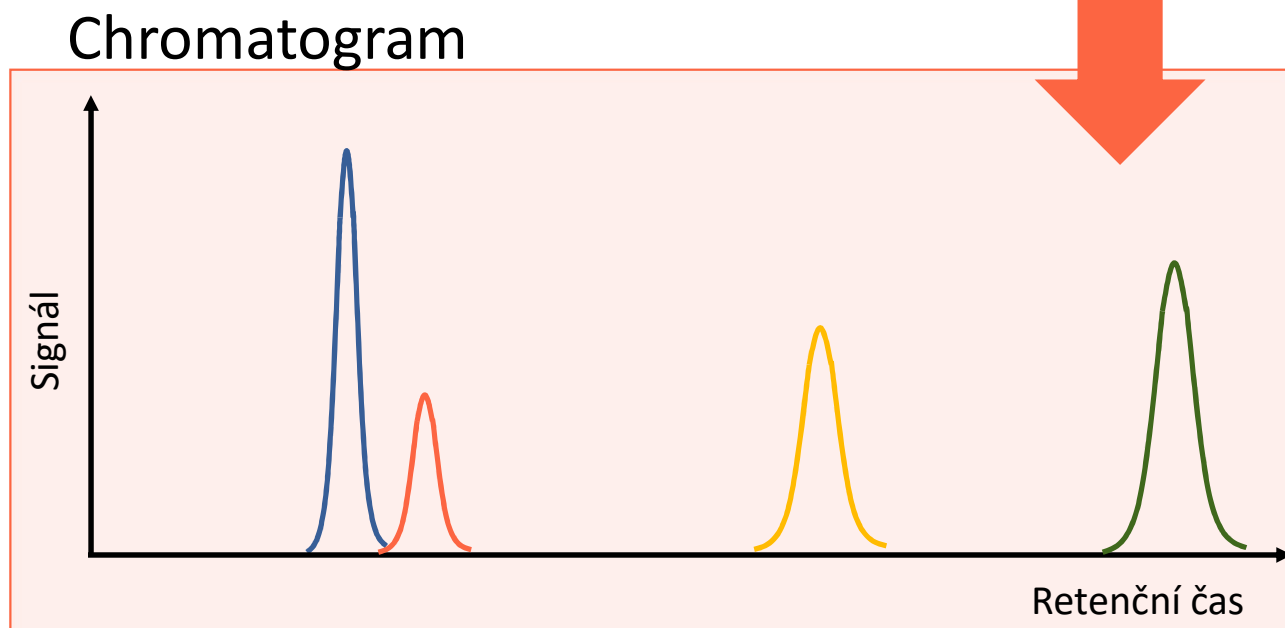
- Lidský genom obsahuje asi 20 000 kódujících sekvencí
- Průměrná lidská buňka exprimuje zhruba 12 000 genů
- Komplexita proteomu výrazně vyšší než komplexita genomu
- Navýšení komplexity v důsledku alternativního splicingu, alternativního startu translace, posttranslačních modifikací...
- Vzorek třeba nejprve separovat pro snížení komplexity

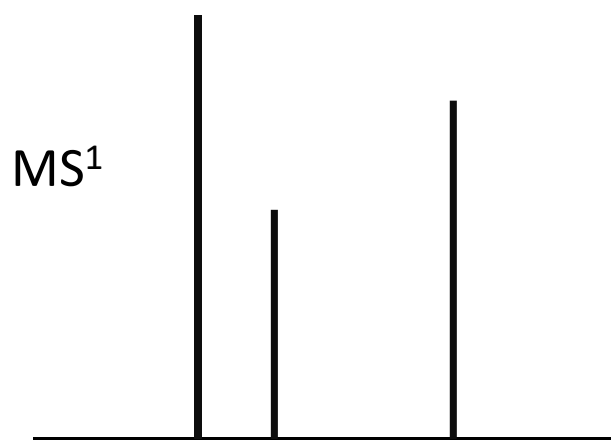
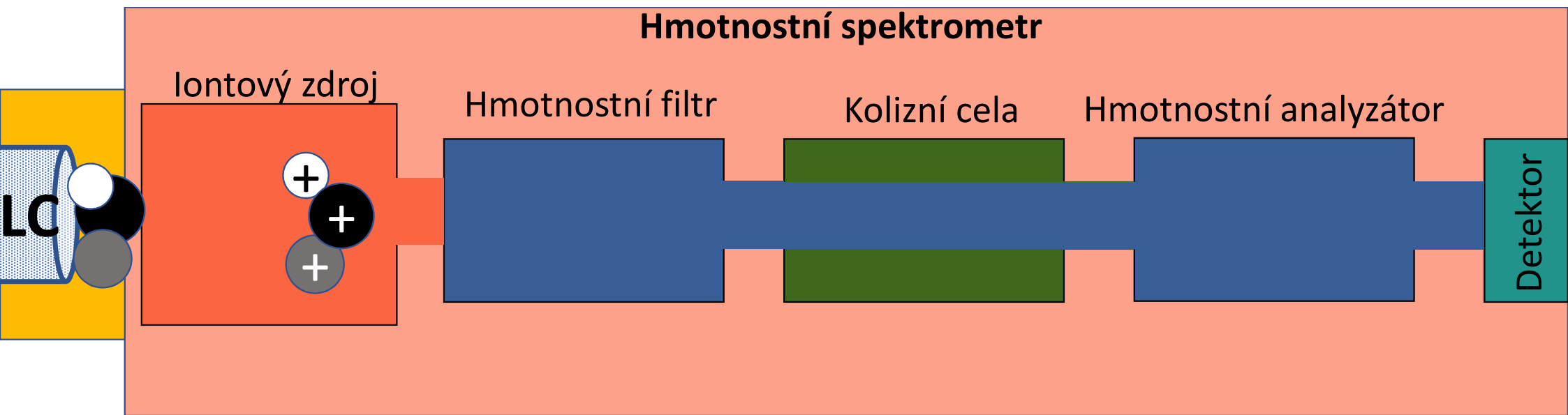


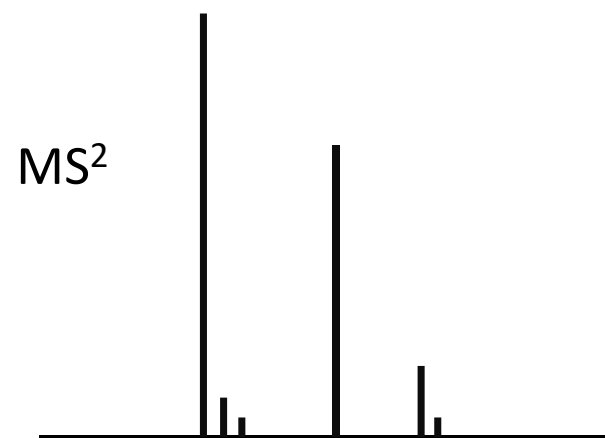
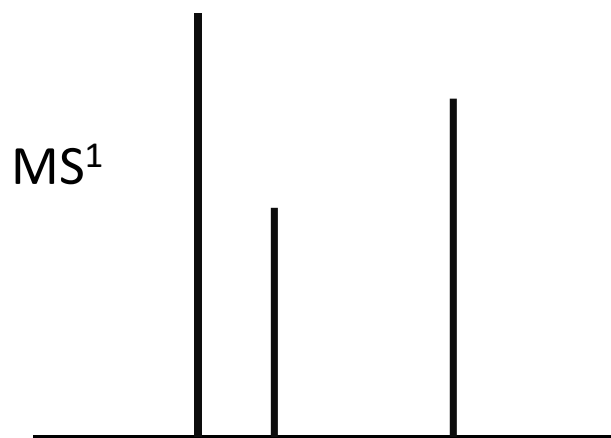
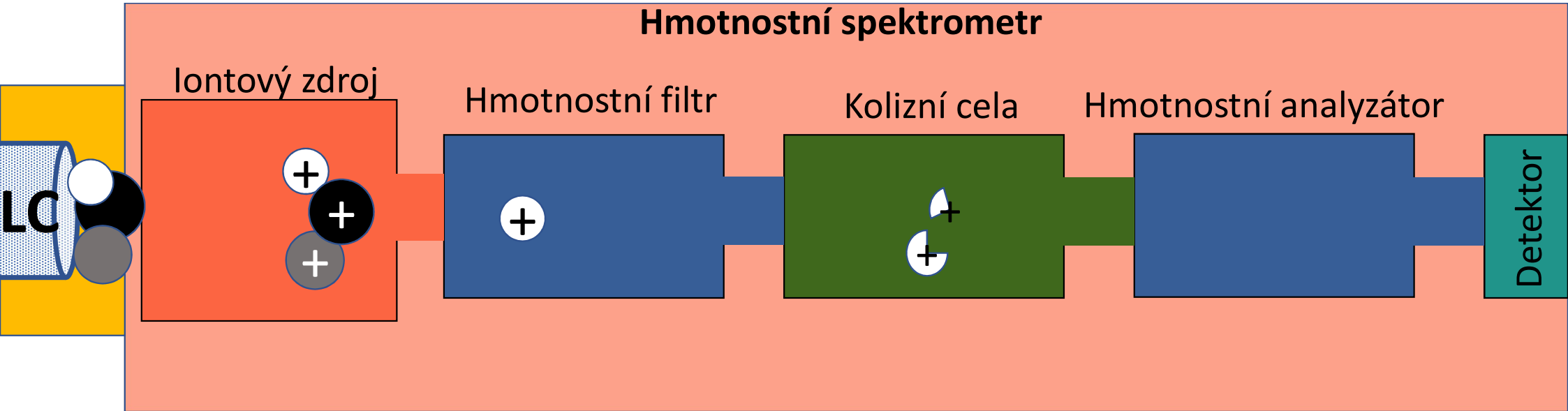


## Kapalinová chromatografie (LC)

- šířka chromatografického píku typicky cca 10 s
- chromatografický pík = signál konkrétního peptidu



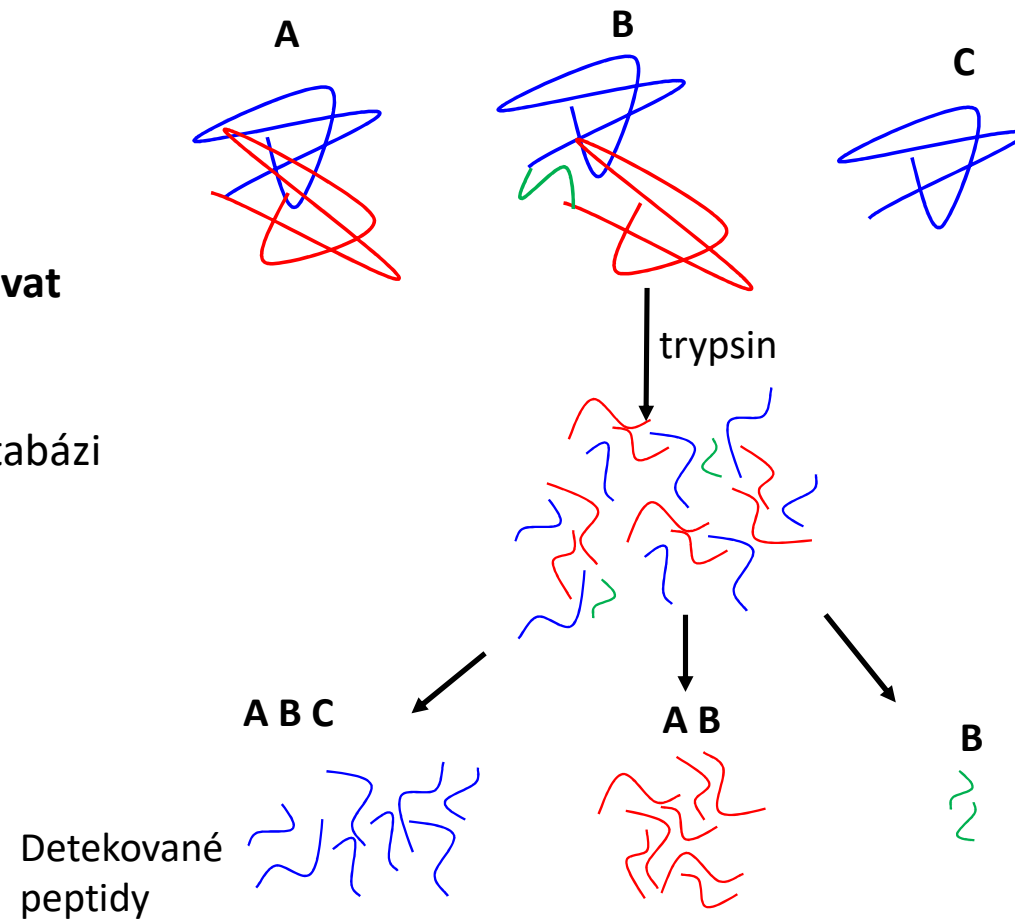






# Bottom up proteomika

- Detekujeme peptidy, ne proteiny
- Proteolytickým štěpením vzorků ztrácíme informaci o propojení peptid-protein
- **Seznam identifikovaných proteinů musíme rekonstruovat zpětně z identifikovaných peptidů = Protein inference**
- Vedle naměřených dat **potřebujeme** k vyhodnocení databázi proteinů – **neprobíhá de novo sekvenace**
- Problém v případě:
  - sekvenčních homologií
  - alternativního sestřihu mRNA
  - různých variant posttranslačního processingu
  - redundantní databáze



>sp|P60709|ACTB\_HUMAN Actin, cytoplasmic 1 OS=Homo sapiens OX=9606  
GN=ACTB PE=1 SV=1

MDDDIAALVVDNGSGMCKAGFAGDDAPRAVFPSIVGRPRHQGMVGMGQKDSYVGDE  
AQS**K**RGILTL**K**YPIEHGIVTNWDDME**K**IWHHTFYNEL**R**VAPEEHPVLLTEAPLN**P**KANREK  
MTQIMFETFNTPAMYVAIQAVLSLYASGR**T**TTGIVMDSGDGVTHTVPIYEGYALPHAIL**R**LDL  
AGRDLTDYLM**K**ILTERGYSFTTTAEREIV**R**DIKE**K**LCYVALDFEQEMATAASSSSLE**K**SY  
ELPDGQVITIGNER**F**RCPALFQPSFLGMESCGIHETTFNSIM**K**CDVDIR**K**DLYANTVLS  
GGTTMYPGIADRM**Q**KEITALAPSTM**K**IKI**I**APPER**K**YSVWIGGSILASLSTFQ**Q**MWISK**Q**  
EYDESGPSIVHR**K**CF

| Position of cleavage site | Name of cleaving enzyme(s) | Resulting peptide sequence (see <a href="#">explanations</a> ) | Peptide length [aa] | Peptide mass [Da] | <a href="#">Cleavage probability</a> |
|---------------------------|----------------------------|--|---------------------|-------------------|--------------------------------------|
| 18                        | Trypsin                    | MDDDIAALVVDNGSGMCK   | 18                  | 1854.095          | 100%                                 |
| 28                        | Trypsin                    | AGFAGDDAPR   | 10                  | 976.013           | 100%                                 |
| 37                        | Trypsin                    | AVFPSIVGR  | 9                   | 945.130           | 74.6%                                |
| 39                        | Trypsin                    | PR   | 2                   | 271.319           | 100%                                 |
| 50                        | Trypsin                    | HQGMVGMGQK   | 11                  | 1171.398          | 100%                                 |
| 61                        | Trypsin                    | DSYVGDEAQSK  | 11                  | 1198.208          | 65%                                  |
| 62                        | Trypsin                    | R  | 1                   | 174.203           | 95.3%                                |
| 68                        | Trypsin                    | GILTLK   | 6                   | 643.825           | 100%                                 |
| 84                        | Trypsin                    | YPIEHGIVTNWDDMEK   | 16                  | 1947.149          | 100%                                 |
| 95                        | Trypsin                    | IWHHTFYNELR  | 11                  | 1515.694          | 100%                                 |
| 113                       | Trypsin                    | VAPEEHPVLLTEAPLNPK   | 18                  | 1954.254          | 94.1%                                |
| 116                       | Trypsin                    | ANR  | 3                   | 359.385           | 100%                                 |
| 118                       | Trypsin                    | EK   | 2                   | 275.305           | 100%                                 |
| 147                       | Trypsin                    | MTQIMFETFNTPAMYVAIQAVLSLYASGR                                  | 29                  | 3254.825          | 100%                                 |
| 177                       | Trypsin                    | TTGIVMDSGDGVTHTVPIYEGYALPHAILR                                 | 30                  | 3184.614          | 100%                                 |
| 183                       | Trypsin                    | LDLAGR   | 6                   | 643.741           | 100%                                 |
| 191                       | Trypsin                    | DLTDYLMK   | 8                   | 998.159           | 100%                                 |
| 196                       | Trypsin                    | ILTER  | 5                   | 630.742           | 100%                                 |
| 206                       | Trypsin                    | GYSFTTTAER   | 10                  | 1132.195          | 78.4%                                |
| 210                       | Trypsin                    | EIVR   | 4                   | 515.610           | 100%                                 |
| 213                       | Trypsin                    | DIK  | 3                   | 374.437           | 93.1%                                |
| 215                       | Trypsin                    | EK   | 2                   | 275.305           | 94.7%                                |
| 238                       | Trypsin                    | LCYVALDFEQEMATAASSSSLEK  | 23                  | 2493.783          | 100%                                 |
| 254                       | Trypsin                    | SYELPDGQVITIGNER   | 16                  | 1790.947          | 100%                                 |
| 256                       | Trypsin                    | FR   | 2                   | 321.379           | 100%                                 |
| 284                       | Trypsin                    | CPEALFQPSFLGMESCGIHETTFNSIMK                                   | 28                  | 3118.602          | 100%                                 |
| 290                       | Trypsin                    | CDVDIR   | 6                   | 719.811           | 84.6%                                |
| 291                       | Trypsin                    | K  | 1                   | 146.189           | 100%                                 |
| 312                       | Trypsin                    | DLYANTVLSGGTTMYPGIADR  | 21                  | 2215.463          | 100%                                 |
| 315                       | Trypsin                    | MQK  | 3                   | 405.513           | 84.3%                                |
| 326                       | Trypsin                    | EITALAPSTMK  | 11                  | 1161.379          | 100%                                 |
| 328                       | Trypsin                    | IK   | 2                   | 259.349           | 100%                                 |
| 335                       | Trypsin                    | IAPPER   | 7                   | 794.949           | 88.4%                                |
| 336                       | Trypsin                    | K  | 1                   | 146.189           | 84.3%                                |
| 359                       | Trypsin                    | YSVWIGGSILASLSTFQ <b>Q</b> MWISK                               | 23                  | 2603.031          | 100%                                 |
| 372                       | Trypsin                    | QEYDESGPSIVHR  | 13                  | 1516.587          | 100%                                 |
| 373                       | Trypsin                    | K  | 1                   | 146.189           | 100%                                 |
| 375                       | end of sequence            | CF   | 2                   | 268.331           | -                                    |

# Actin B vs Actin C

Sequence identity: 93.9 %

```
sp|P60709|ACTB_HUMAN      --MDDDDIAALVVDNGSGMCKKAGFAGDDAPRAVFPSIVGRPRHQGVMVGMGQKDSYVGDEA
sp|P68032|ACTC_HUMAN      MCDDEETTALVCDNGSGLVKAGFAGDDAPRAVFPSIVGRPRHQGVMVGMGQKDSYVGDEA
                              *:: :*** *****: *****:*****:*****:*****:*****:*****
```

```
sp|P60709|ACTB_HUMAN      QSKRGILTLKYPIEHGIVTNWDDMEKIWHHTFYNELRVAPEEHPVLLTEAPLNPKANREK
sp|P68032|ACTC_HUMAN      QSKRGILTLKYPIEHGIITNWDDMEKIWHHTFYNELRVAPEEHPTLLTEAPLNPKANREK
                              *****:*****:*****:*****:*****:*****:*****:*****
```

```
sp|P60709|ACTB_HUMAN      MTQIMFETFNTPAMYVAIQAVLSLYASGRTTGIVMDSGDGVTHTVPIYEGYALPHAILRL
sp|P68032|ACTC_HUMAN      MTQIMFETFNVPAMYVAIQAVLSLYASGRTTGIVLDSGDGVNVPIYEGYALPHAIMRL
                              *****:*****:*****:*****:*****:*****:*****:*****
```

```
sp|P60709|ACTB_HUMAN      DLAGRDLTDYLMKILTERGYSFTTAEREIVRDIKEKLCYVALDFEQEMATAASSSSLEK
sp|P68032|ACTC_HUMAN      DLAGRDLTDYLMKILTERGYSFVTTAEREIVRDIKEKLCYVALDFENEMATAASSSSLEK
                              *****:*****:*****:*****:*****:*****:*****:*****
```

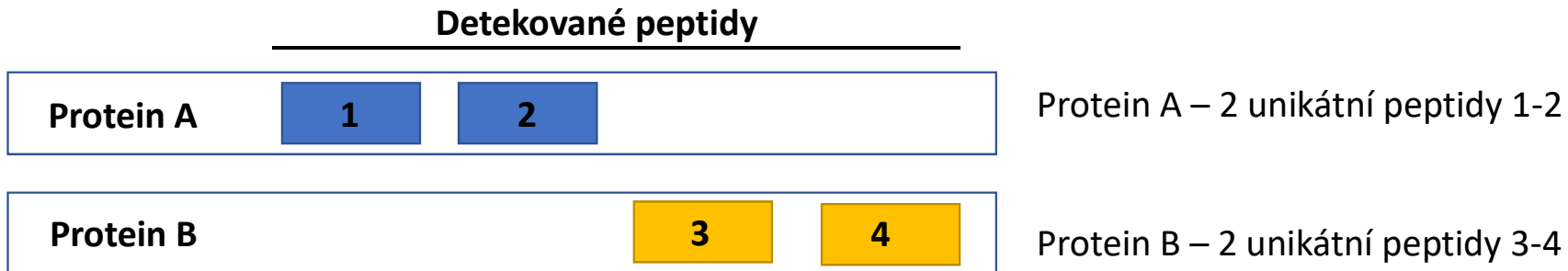
```
sp|P60709|ACTB_HUMAN      SYELPDGQVITIGNERFRCPEALFQPSFLGMESCGIHETTFNSIMKCDVDIRKDLYANTV
sp|P68032|ACTC_HUMAN      SYELPDGQVITIGNERFRCPETLFQPSFIGMESAGIHETTYNSIMKCDIDIRKDLYANNV
                              *****:*****:*****:*****:*****:*****:*****:*****
```

```
sp|P60709|ACTB_HUMAN      LSGGTTMYPGIADRMQKEITALAPSTMKIKIIAPPERKYSVWIGGSILASLSTFQQMWIS
sp|P68032|ACTC_HUMAN      LSGGTTMYPGIADRMQKEITALAPSTMKIKIIAPPERKYSVWIGGSILASLSTFQQMWIS
                              *****:*****:*****:*****:*****:*****:*****:*****
```

```
sp|P60709|ACTB_HUMAN      KQEYDESGPSIVHRKCF
sp|P68032|ACTC_HUMAN      KQEYDEAGPSIVHRKCF
                              *****:*****
```



# Protein inference – od peptidů zpět k proteinům



- Dva jednoznačně identifikovatelné proteiny
- **Unikátní (proteotypický) peptid** – lze jednoznačně přiřadit k jednomu konkrétnímu proteinu

Interpretation of Shotgun Proteomic Data - THE PROTEIN INFERENCE PROBLEM, Alexey I. Nesvizhskii and Ruedi Aebersold, Molecular & Cellular Proteomics 2005

# Protein inference – od peptidů zpět k proteinům

## Detekované peptidy

|           | 1 | 2 | 3 | 4 |
|-----------|---|---|---|---|
| Protein A | 1 | 2 | 3 | 4 |
| Protein B |   |   | 3 | 4 |

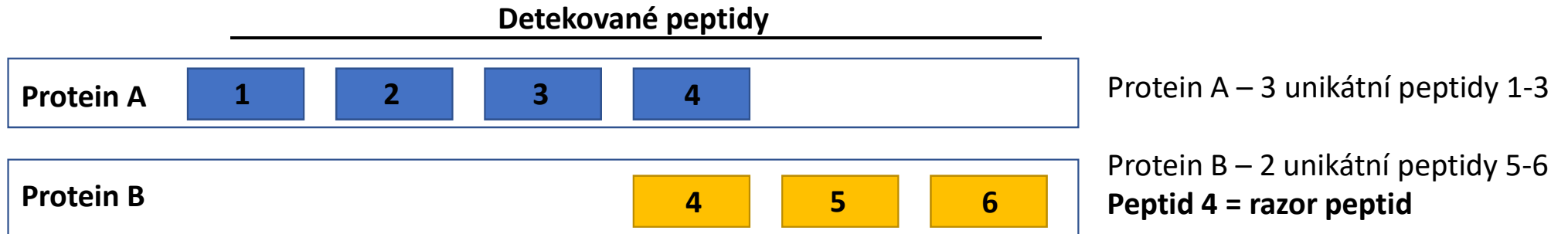
Protein A – 2 unikátní peptidy 1-2

Protein B – žádné unikátní peptidy  
Peptidy 3 – 4 sdílené proteiny A a B

- Protein A určitě přítomen, přítomnost proteinu B nelze vyloučit
- Ve výsledku proteiny A a B spojeny do jedné **protein group**

Interpretation of Shotgun Proteomic Data - THE PROTEIN INFERENCE PROBLEM, Alexey I. Nesvizhskii and Ruedi Aebersold, Molecular & Cellular Proteomics 2005

# Protein inference – od peptidů zpět k proteinům

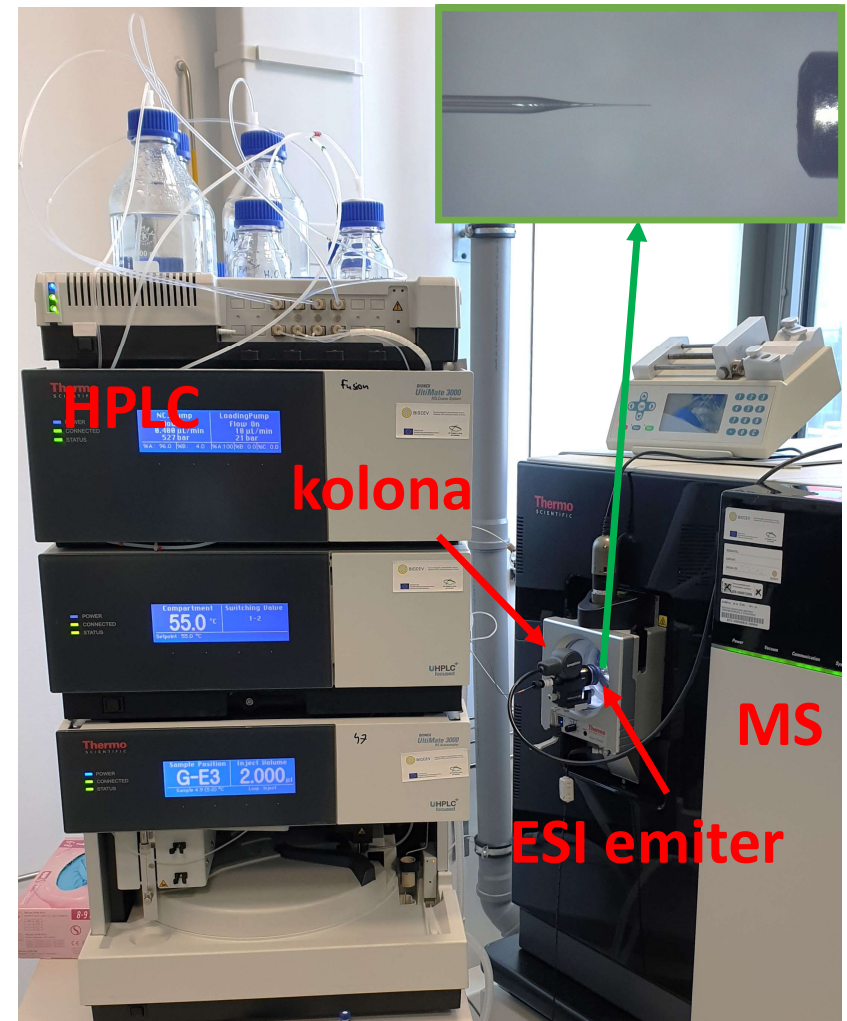


- Proteiny A a B jsou identifikovány na základě unikátních peptidů a nemůžeme je tak zařadit do jedné protein group – jsou uvedeny jako jednotlivé identifikace
- **Kvantitativní hodnota razor peptidu bude přičtena proteinu o vyšším počtu unikátních peptidů**
- Razor peptid nemá vliv z hlediska identifikace, slouží **pouze pro kvantifikaci**

Interpretation of Shotgun Proteomic Data - THE PROTEIN INFERENCE PROBLEM, Alexey I. Nesvizhskii and Ruedi Aebersold, Molecular & Cellular Proteomics 2005

# Analýza komplexních vzorků v praxi

- Současná technika umožňuje detekci a kvantifikaci asi 10 000 proteinů z jednoho nástřiku lyzátu savčích buněk – zhruba 80 % – 90 % celkového proteomu
- HPLC separace online propojená s MS = LC-MS (LC-MS/MS)
- **Nutná MS detekce v časové škále odpovídající LC separaci** – šířka chromatografického píku obvykle kolem 10 s (v závislosti na koloně, gradientu atd.)

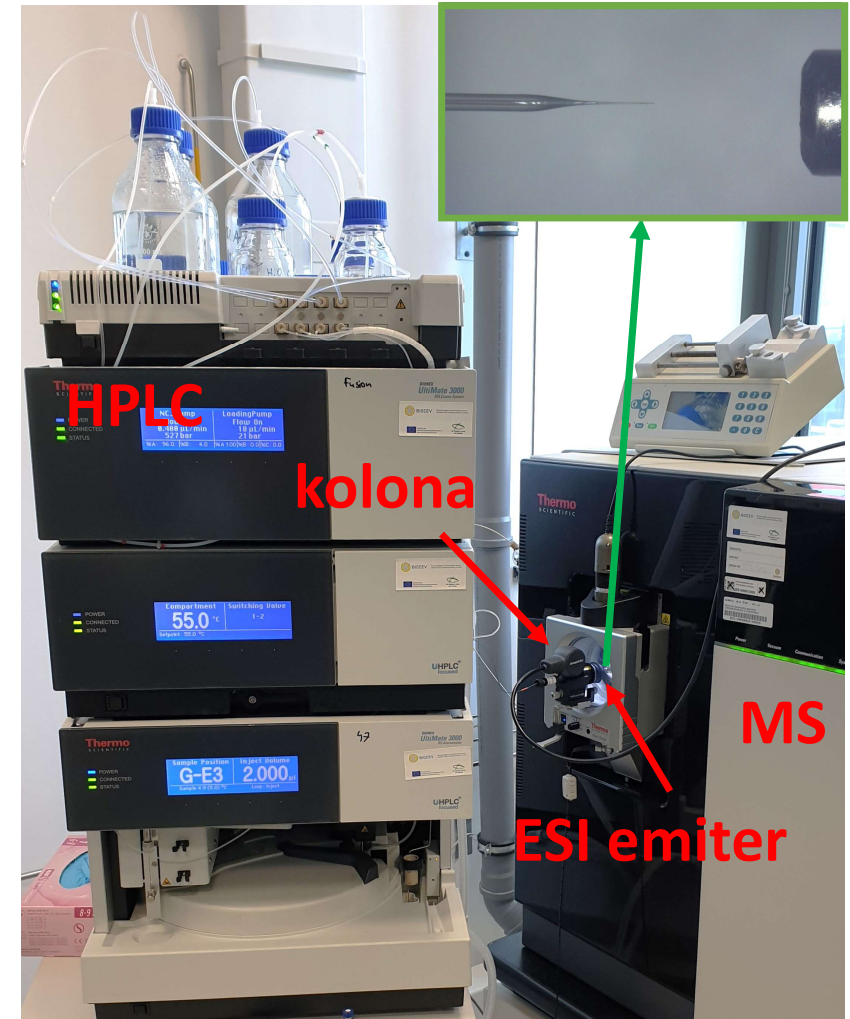
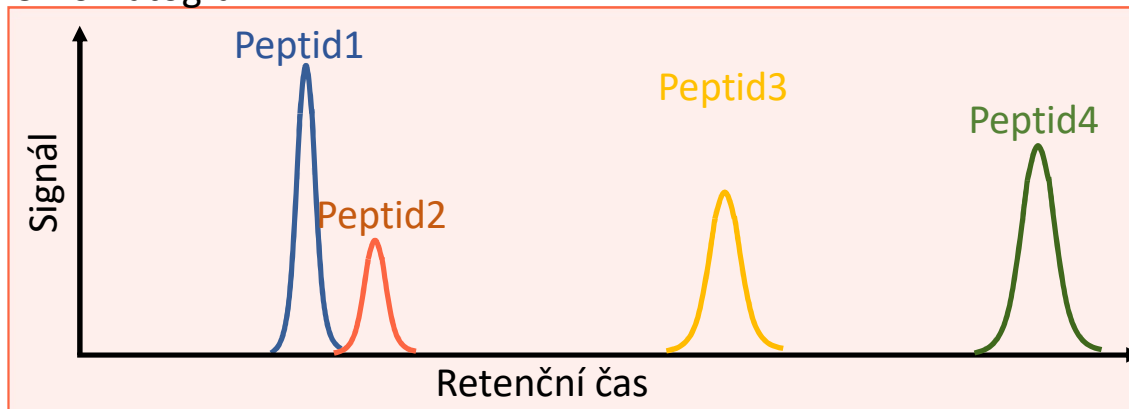




# Analýza komplexních vzorků v praxi

- Současná technika umožňuje detekci a kvantifikaci asi 10 000 proteinů z jednoho nástřiku lyzátu savčích buněk – zhruba 80 % – 90 % celkového proteomu
- HPLC separace online propojená s MS = LC-MS (LC-MS/MS)
- **Nutná MS detekce v časové škále odpovídající LC separaci** – šířka chromatografického píku obvykle kolem 10 s (v závislosti na koloně, gradientu atd.)

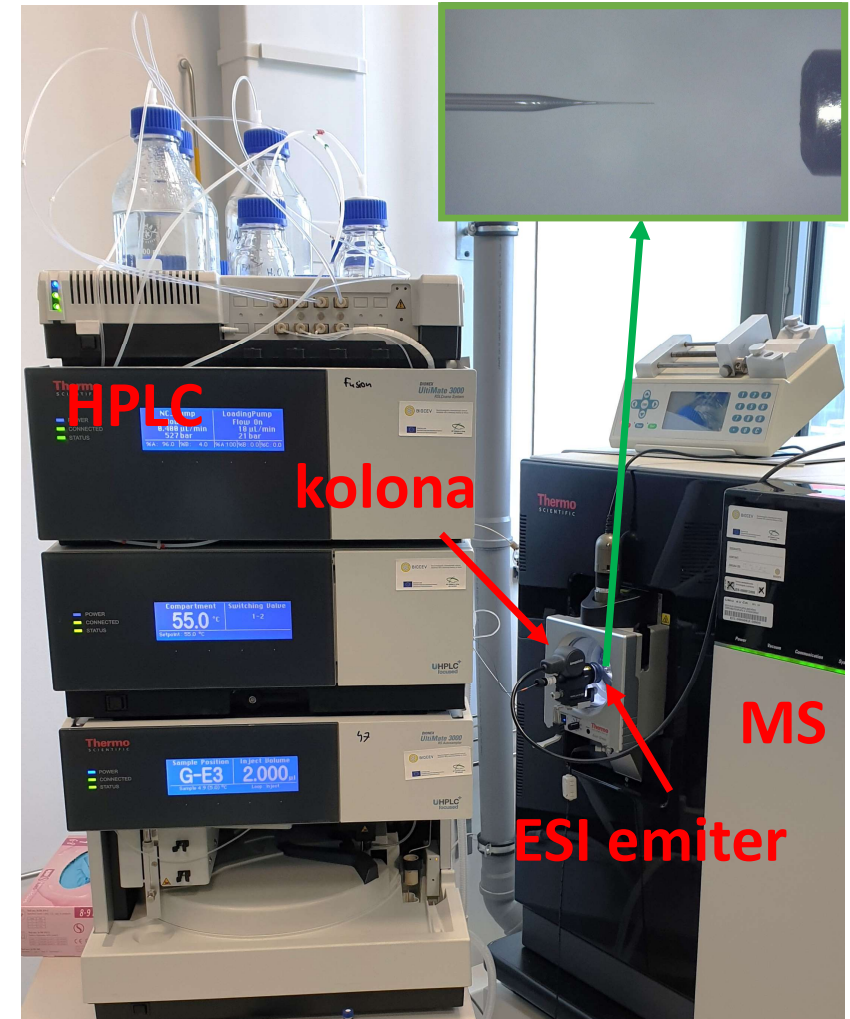
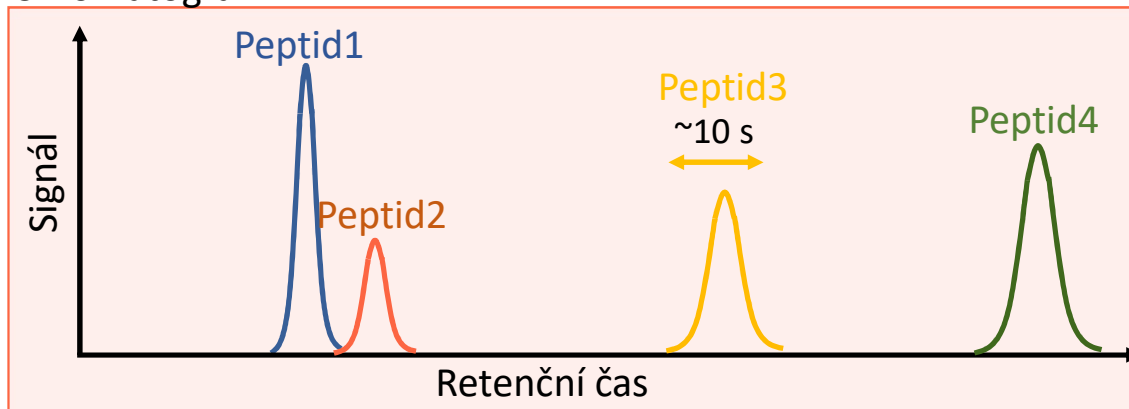
Chromatogram



# Analýza komplexních vzorků v praxi

- Současná technika umožňuje detekci a kvantifikaci asi 10 000 proteinů z jednoho nástřiku lyzátu savčích buněk – zhruba 80 % – 90 % celkového proteomu
- HPLC separace online propojená s MS = LC-MS (LC-MS/MS)
- **Nutná MS detekce v časové škále odpovídající LC separaci** – šířka chromatografického píku obvykle kolem 10 s (v závislosti na koloně, gradientu atd.)

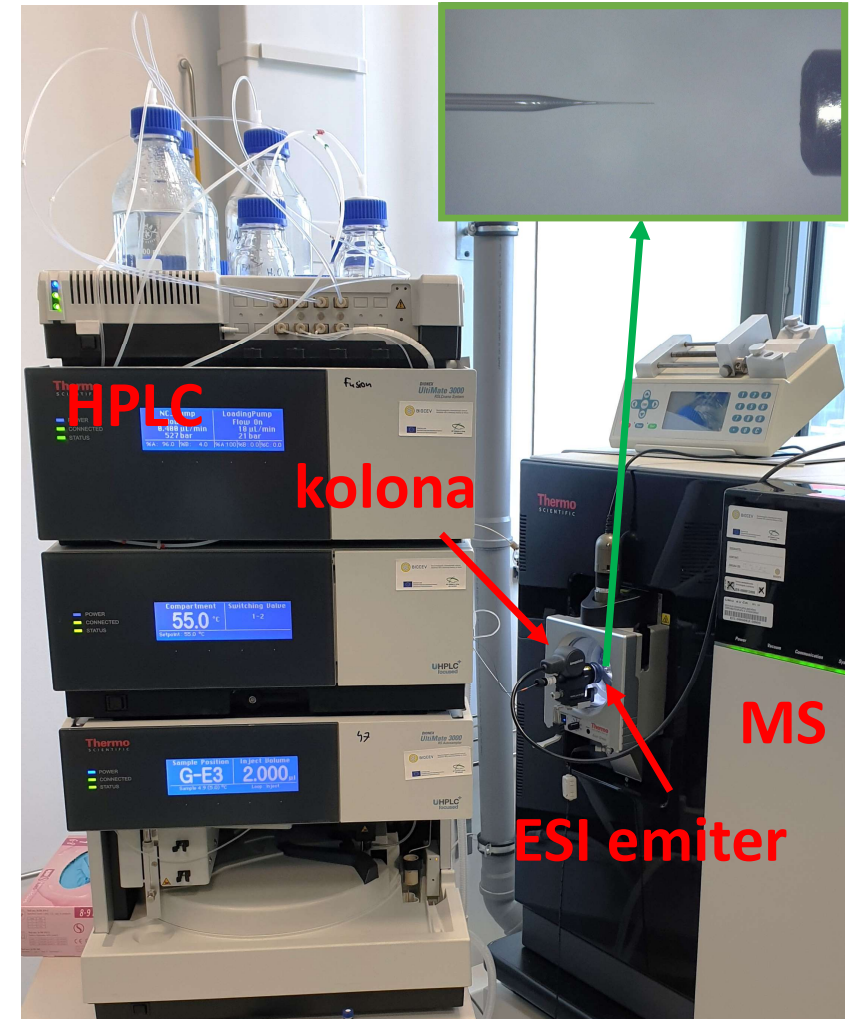
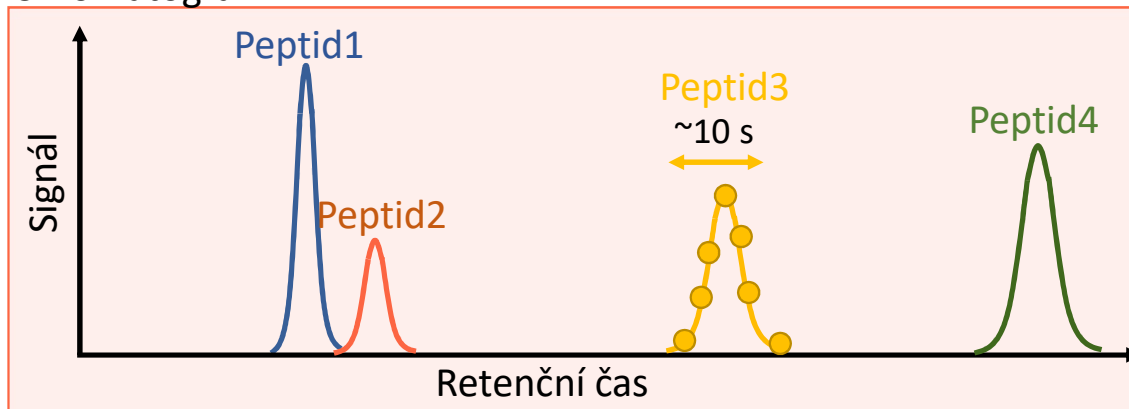
Chromatogram



# Analýza komplexních vzorků v praxi

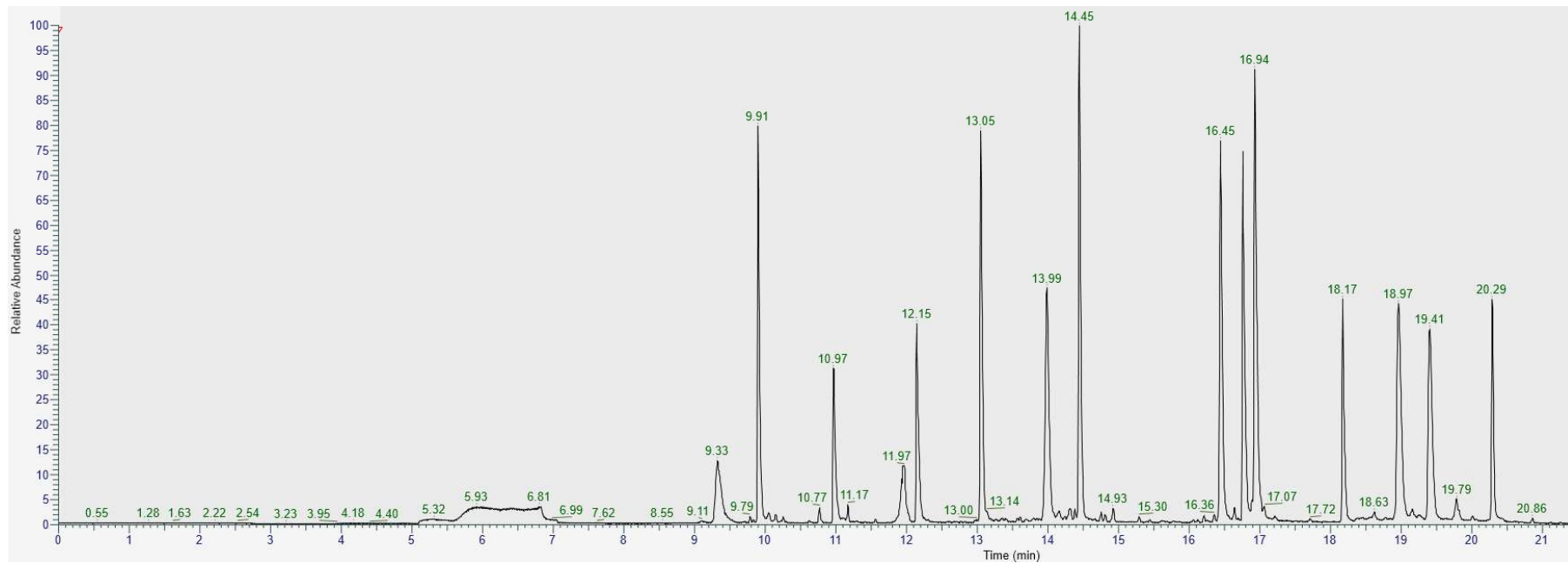
- Současná technika umožňuje detekci a kvantifikaci asi 10 000 proteinů z jednoho nástřiku lyzátu savčích buněk – zhruba 80 % – 90 % celkového proteomu
- HPLC separace online propojená s MS = LC-MS (LC-MS/MS)
- **Nutná MS detekce v časové škále odpovídající LC separaci** – šířka chromatografického píku obvykle kolem 10 s (v závislosti na koloně, gradientu atd.)

Chromatogram



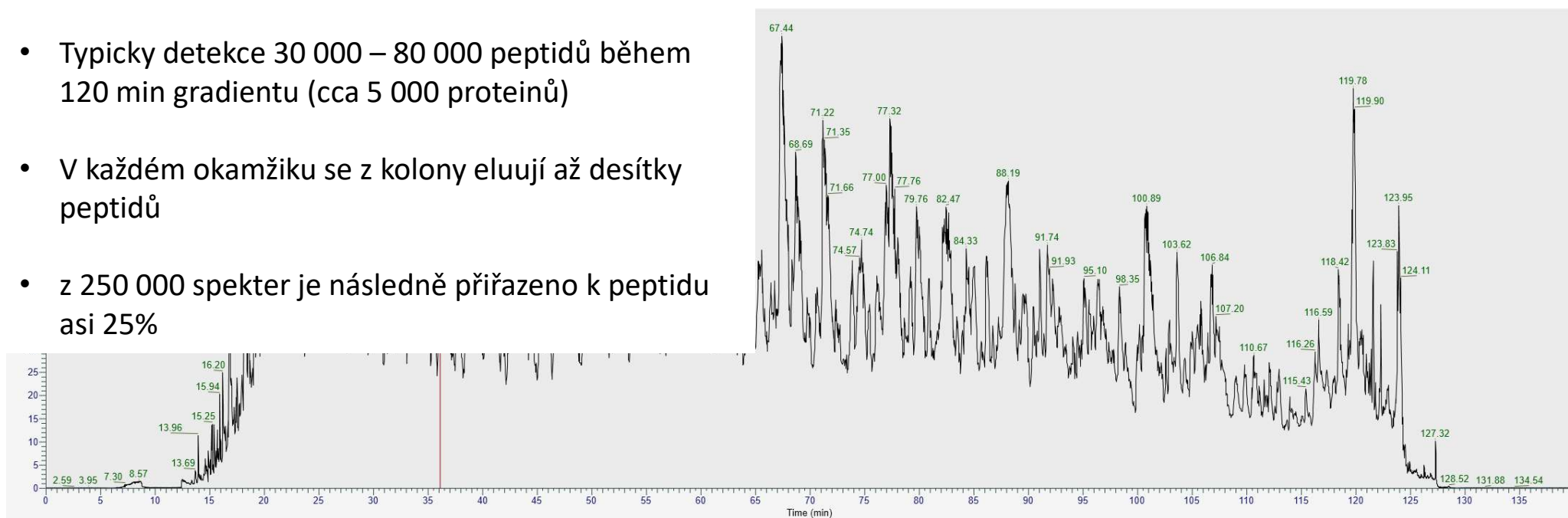
# Analýza komplexních vzorků

Směs 14 peptidů – málo komplexní vzorek



# Analýza komplexních vzorků

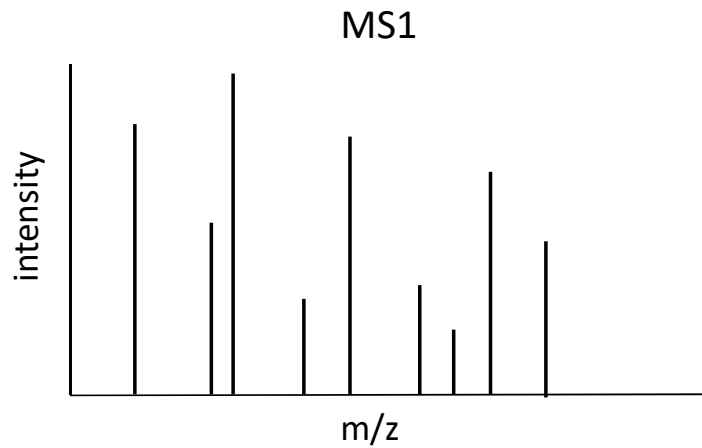
- Analýza 1 µg buněk ovariálního karcinomu (Hela)
- Typicky detekce 30 000 – 80 000 peptidů během 120 min gradientu (cca 5 000 proteinů)
- V každém okamžiku se z kolony eluují až desítky peptidů
- z 250 000 spekter je následně přiřazeno k peptidu asi 25%



# Akviziční módy

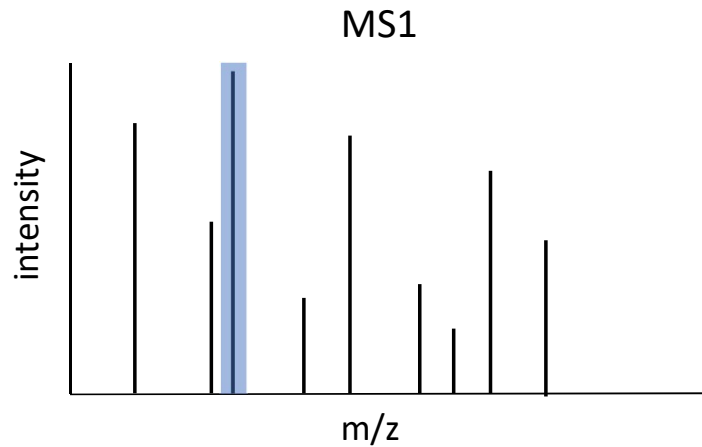
- **MS1 a MS2 (MS/MS, MS<sup>2</sup>)** scany poskytují různé druhy informací
- Cyklické střídání MS1 a MS2 scanů podle předem daných pravidel
- Cílem měření je:
  - Identifikovat co nejvíce peptidů (naměřit co nejvíce spekter)
  - Identifikované peptidy co nejlépe kvantifikovat
- **Necílené akviziční módy:**
  - **DDA – data dependent acquisition**
  - **DIA – data independent acquisition**
- **Cílené metody:**
  - SRM – selected reaction monitoring
  - PRM – parallel reaction monitoring

# Data Dependent Acquisition



1. MS1 scan (full scan) – detekce prekurzorů s vysokou přesností a rozlišením (300-1600 m/z)

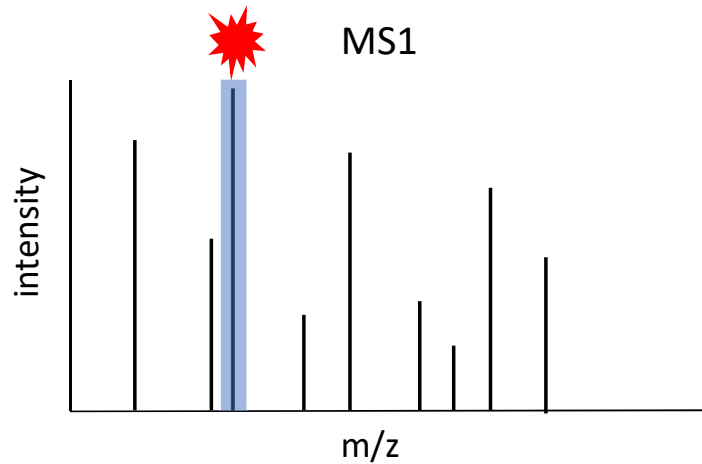
# Data Dependent Acquisition



1. MS1 scan (full scan) – detekce prekurzorů s vysokou přesností a rozlišením (300-1600 m/z)
2. Izolace vybraného prekurzoru – kvadrupól, iontová past

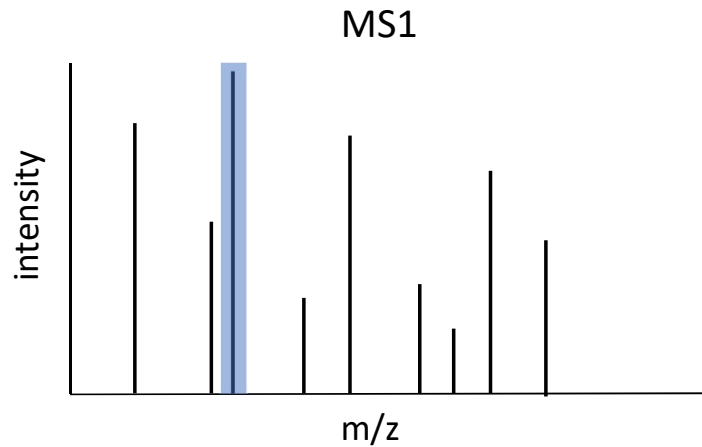


# Data Dependent Acquisition

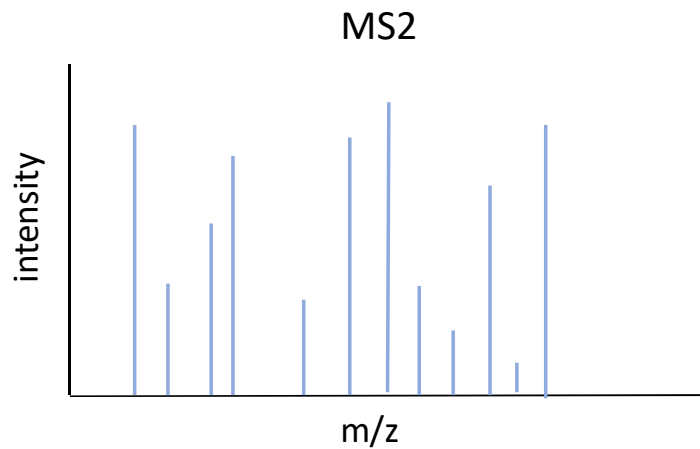


1. MS1 scan (full scan) – detekce prekurzorů s vysokou přesností a rozlišením (300-1600 m/z)
2. Izolace vybraného prekurzoru – kvadrupól, iontová past
3. Fragmentace

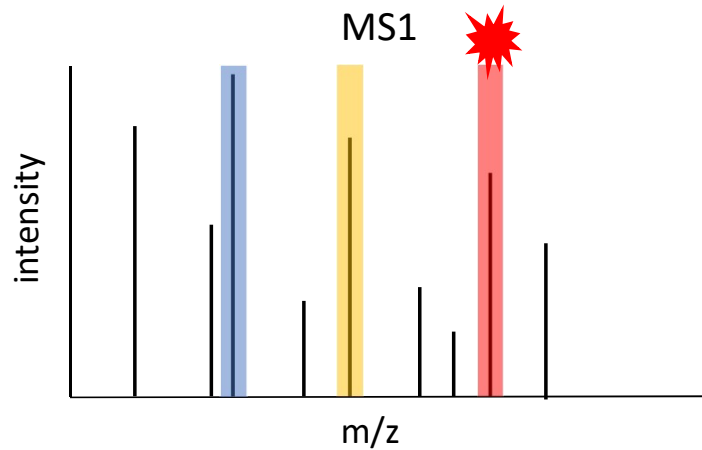
# Data Dependent Acquisition



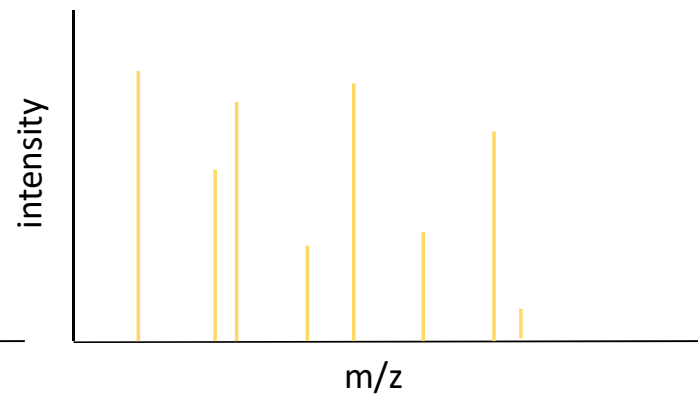
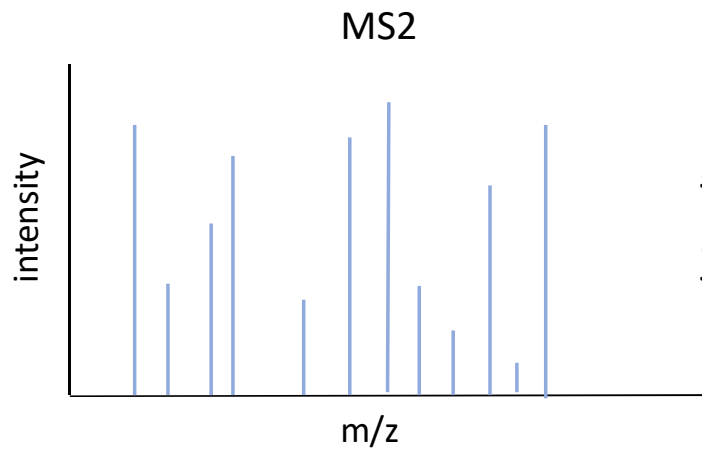
1. MS1 scan (full scan) – detekce prekurzorů s vysokou přesností a rozlišením (300-1600 m/z)
2. Izolace vybraného prekurzoru – kvadrupól, iontová past
3. Fragmentace
4. Detekce vzniklých fragmentů



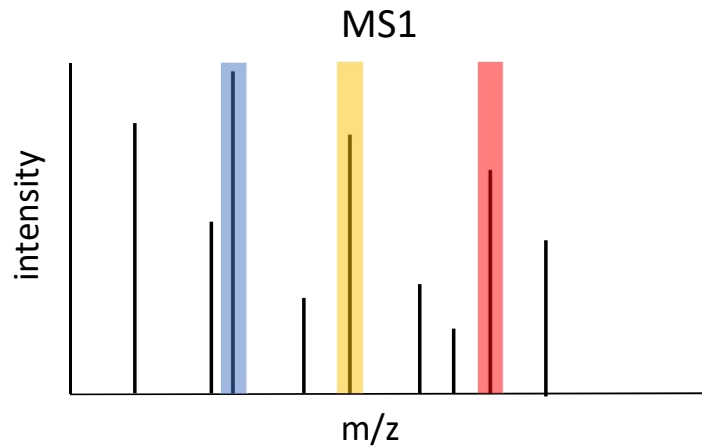
# Data Dependent Acquisition



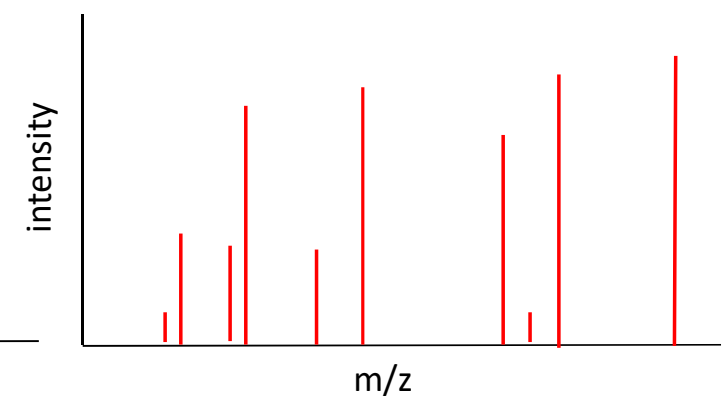
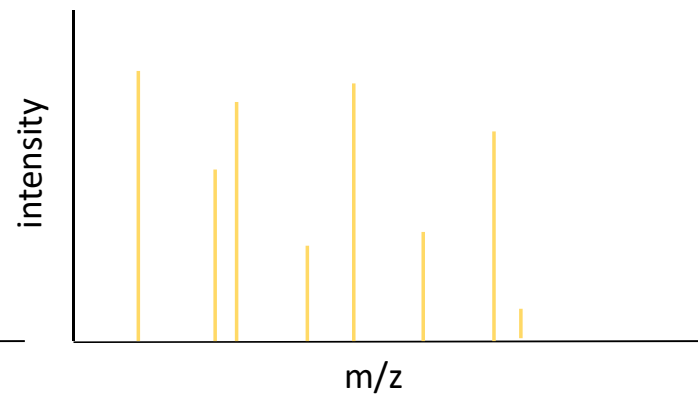
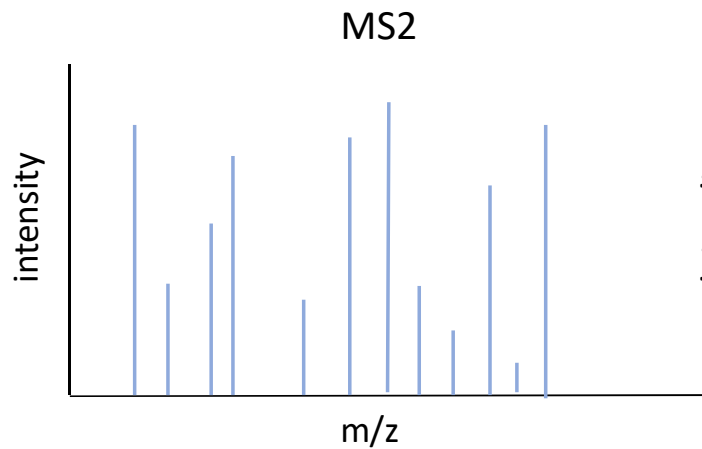
1. MS1 scan (full scan) – detekce prekurzorů s vysokou přesností a rozlišením (300-1600 m/z)
2. Izolace vybraného prekurzoru – kvadrupól, iontová past
3. Fragmentace
4. Detekce vzniklých fragmentů



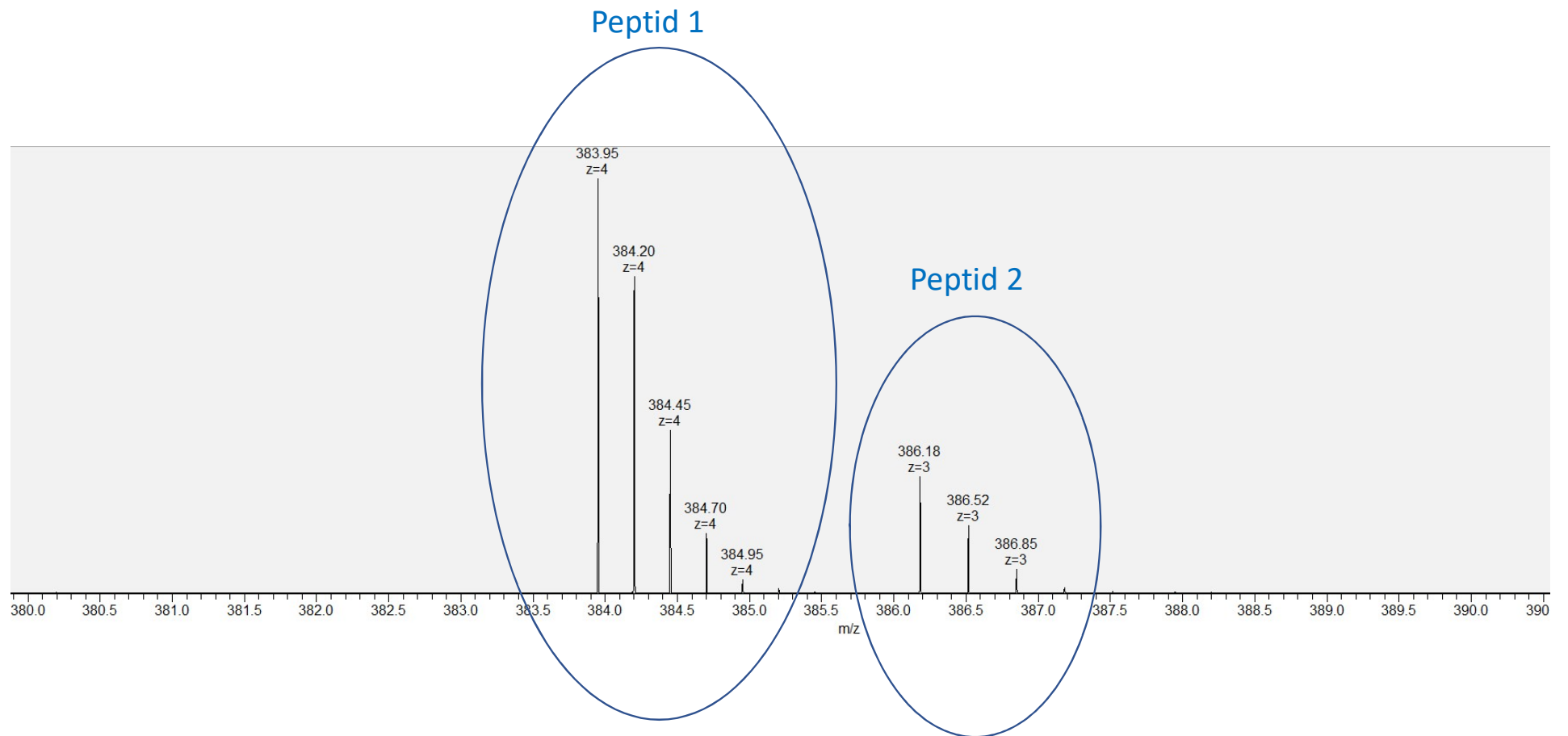
# Data Dependent Acquisition



1. MS1 scan (full scan) – detekce prekurzorů s vysokou přesností a rozlišením (300-1600 m/z)
2. Izolace vybraného prekurzoru – kvadrupól, iontová past
3. Fragmentace
4. Detekce vzniklých fragmentů



# Izolace peptidu



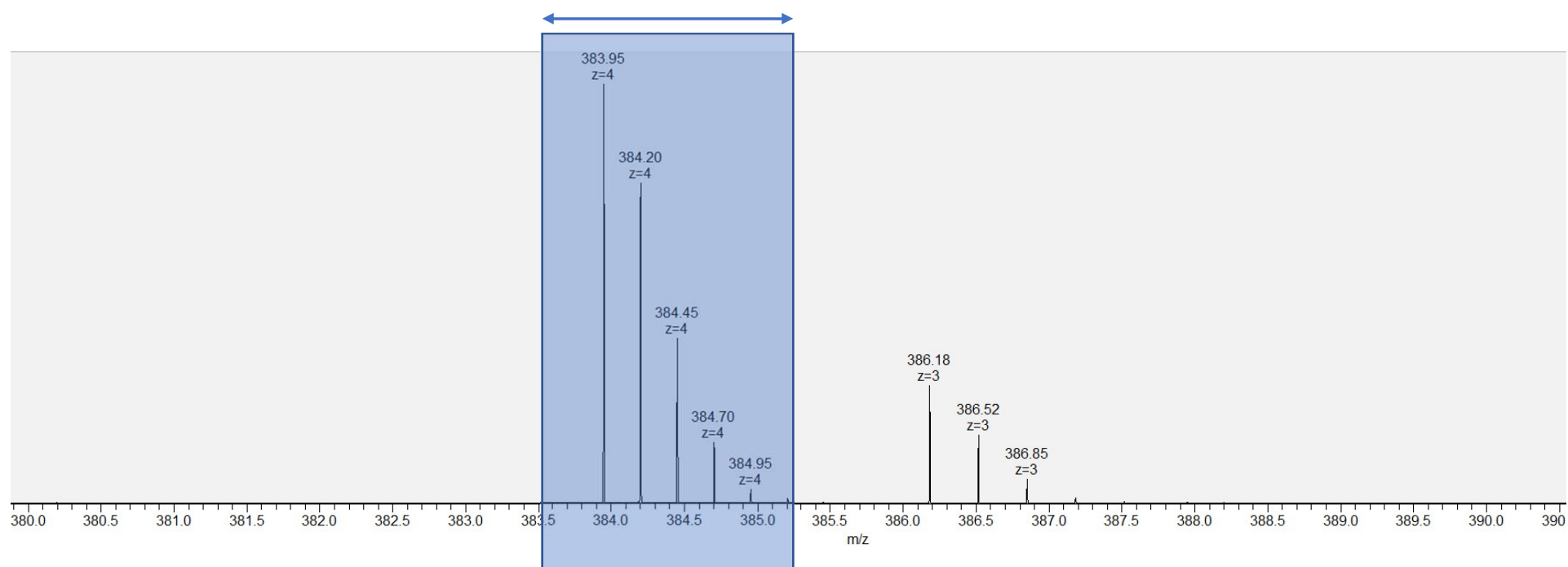
# Izolace peptidu

## Hmotnostní filtr:

- Kvadrupól
- Iontová past
  
- Elektronika nastaví takové parametry (DC napětí, frekvence a amplituda RF napětí atd.), aby prošel jenom peptid o požadovaném  $m/z$
- Ostatní ionty jsou odfiltrovány pryč a dále nejsou součástí analýzy

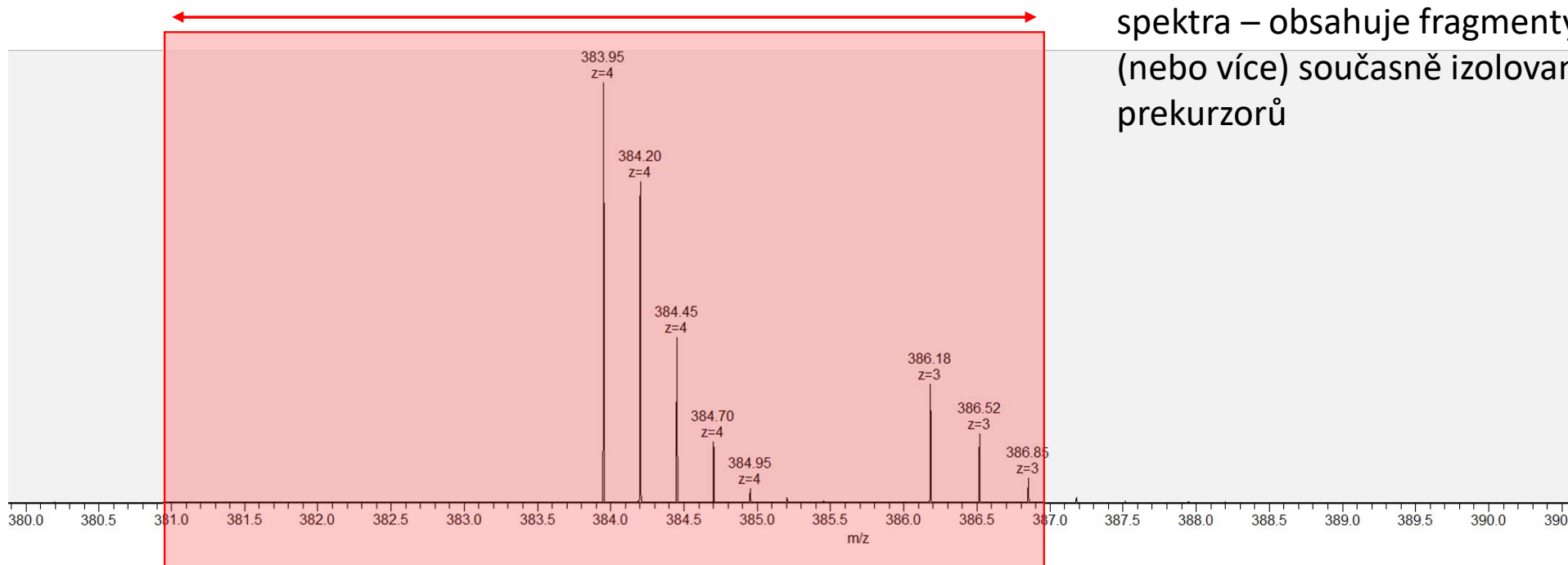
# Izolace peptidu

Šířka izolačního okna



# Izolace peptidu

Šířka izolačního okna



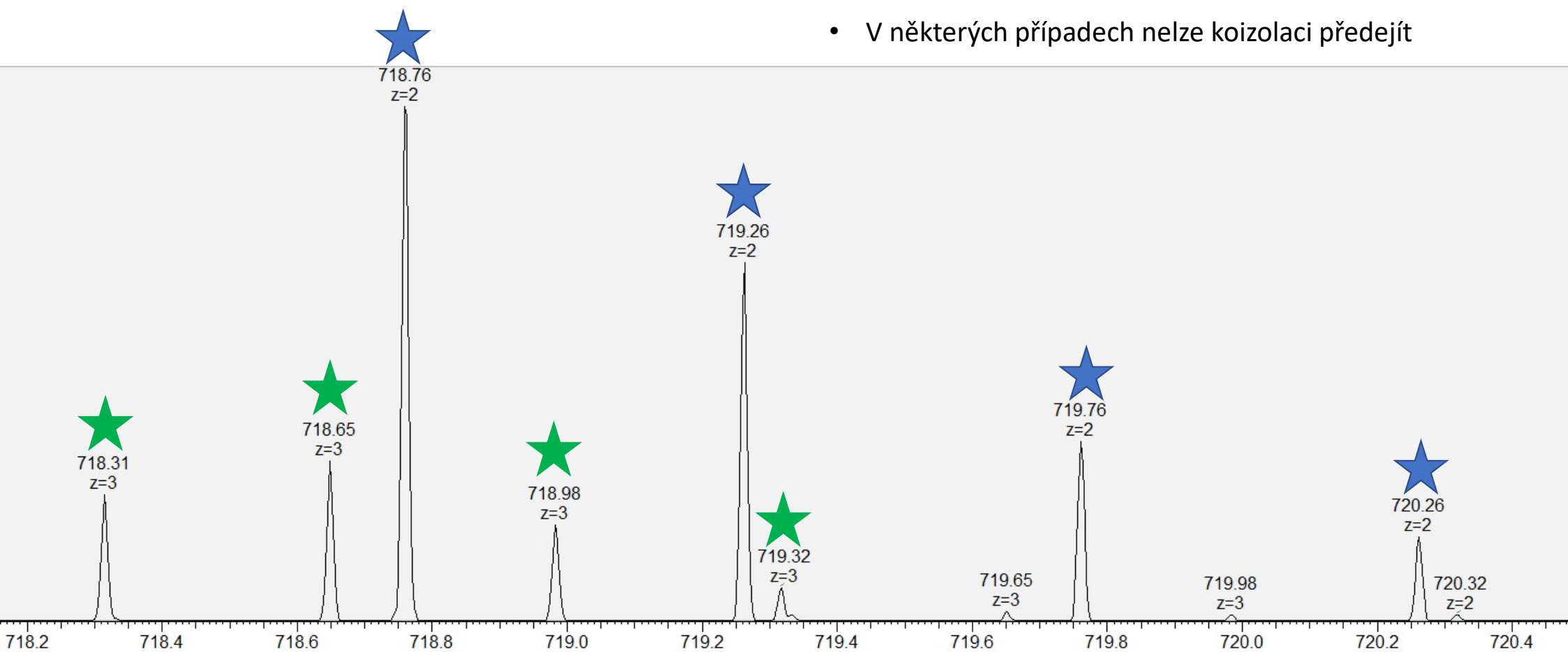
## Koizolace

- Vede ke vzniku směsného MSMS spektra – obsahuje fragmenty dvou (nebo více) současně izolovaných prekurzorů



# Izolace peptidu

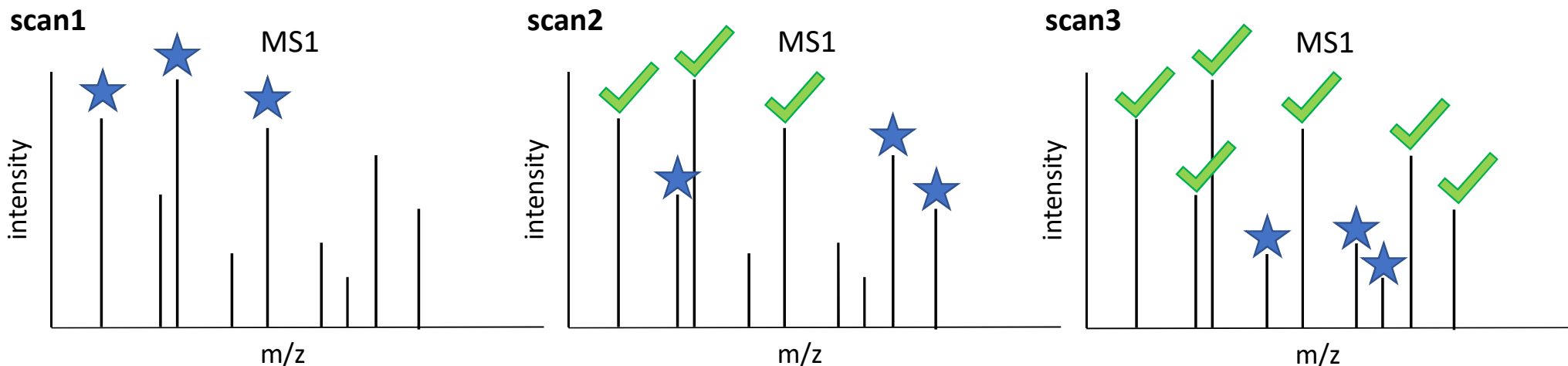
- V některých případech nelze koizolaci předejít



# Dynamická exkluze

- Obvykle vybíráme pro fragmentaci několik nejintenzivnějších peptidů ve spektru
- Dynamická exkluze předchází opakované fragmentaci stejných peptidů
- Fragmentovaný peptid je po daný čas vyloučen z výběru pro fragmentaci – nedochází k opakované fragmentaci v rámci jednoho chromatografického píku

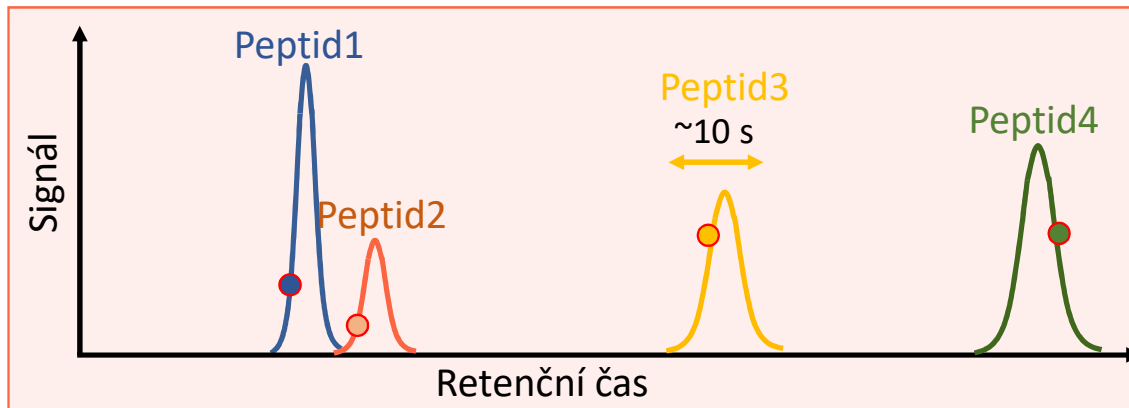
| Dynamic Exclusion Properties                                     |                                     |
|--|-------------------------------------|
| Exclude after n times  | 1                                   |
| Exclusion duration (s)   | 60                                  |
| Mass Tolerance   | ppm                                 |
| Low  | 10                                  |
| High   | 10                                  |
| Exclude Isotopes   | <input checked="" type="checkbox"/> |
| Perform dependent scan on single charge state per precursor only | <input type="checkbox"/>            |
| Exclude Within Cycle   | <input checked="" type="checkbox"/> |



# Dynamická exkluze

- Obvykle vybíráme pro fragmentaci několik nejintenzivnějších peptidů ve spektru
- Dynamická exkluze předchází opakované fragmentaci stejných peptidů
- Fragmentovaný peptid je po daný čas vyloučen z výběru pro fragmentaci – nedochází k opakované fragmentaci v rámci jednoho chromatografického píku

Chromatogram

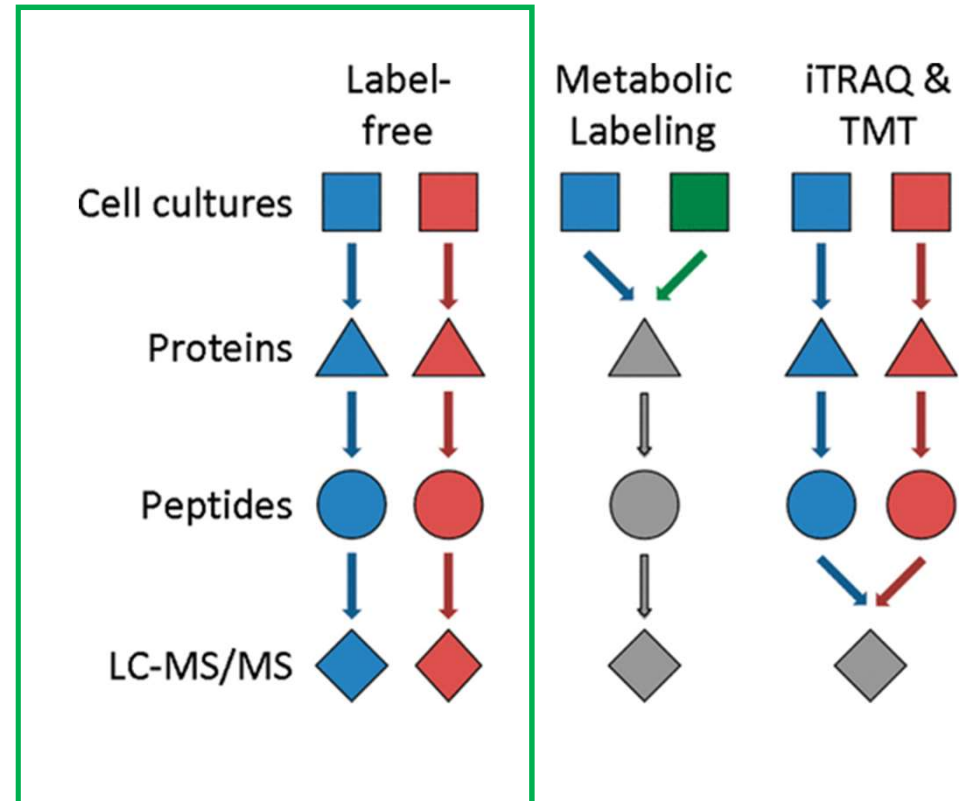


| Dynamic Exclusion Properties                                     |                                     |
|--|-------------------------------------|
| Exclude after n times  | 1                                   |
| Exclusion duration (s)   | 60                                  |
| Mass Tolerance   | ppm                                 |
| Low  | 10                                  |
| High   | 10                                  |
| Exclude isotopes   | <input checked="" type="checkbox"/> |
| Perform dependent scan on single charge state per precursor only | <input type="checkbox"/>            |
| Exclude Within Cycle   | <input checked="" type="checkbox"/> |

Kvantifikace

# Label Free Quantification

- Pro kvantifikaci se využívá výhradně signál samotného peptidu (výška píku/plocha pod píkem)
- Není třeba žádného dodatečného značení na úrovni proteinu/peptidu (SILAC, isobarické značení – viz. následující přednáška)
- Není třeba investovat do metabolických nebo isobarických značek
- Přesnost kvantifikace nižší než u značených přístupů
- Nelze brát v potaz menší kvantitativní rozdíly (10% není z pohledu label free přístupu žádný rozdíl)



Li et al, Systematic Comparison of Label-Free, Metabolic Labeling, and Isobaric Chemical Labeling for Quantitative Proteomics on LTQ Orbitrap Velos, J Prot Res, 2012

# Label Free Quantification

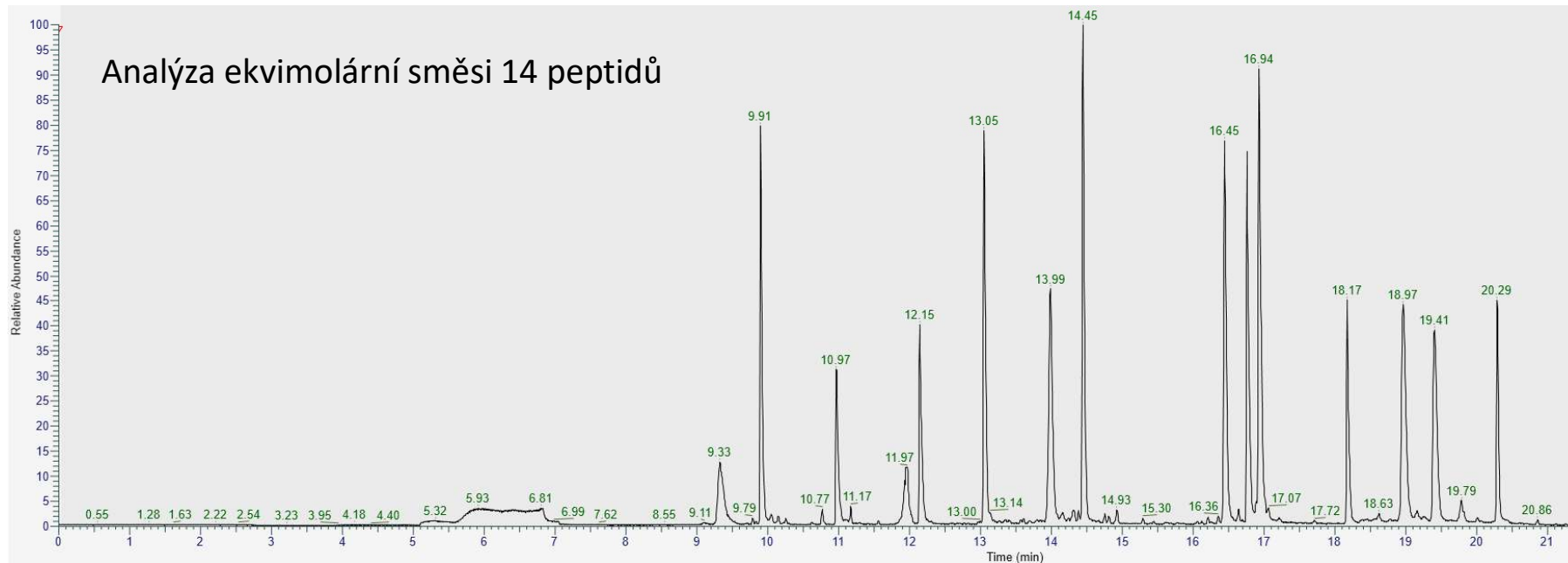
- **MS/MS count (spectral count)**
  - Počet fragmentačních spekter úměrný množství peptidu
  - Používáno v minulosti
- **Intenzita**
  - **Výška chromatografického píku nebo plocha pod píkem.** Množství proteinu odpovídá intenzitám signálu jednotlivých peptidů v MS1.

# Label Free Quantification

- **MS/MS count (spectral count)**
  - Počet fragmentačních spekter úměrný množství peptidu.
  - Používáno v minulosti
- **Intenzita**
  - **Výška chromatografického píku nebo plocha pod píkem.** Množství proteinu odpovídá intenzitám signálu jednotlivých peptidů v MS1.

# Kvantifikace pomocí MS

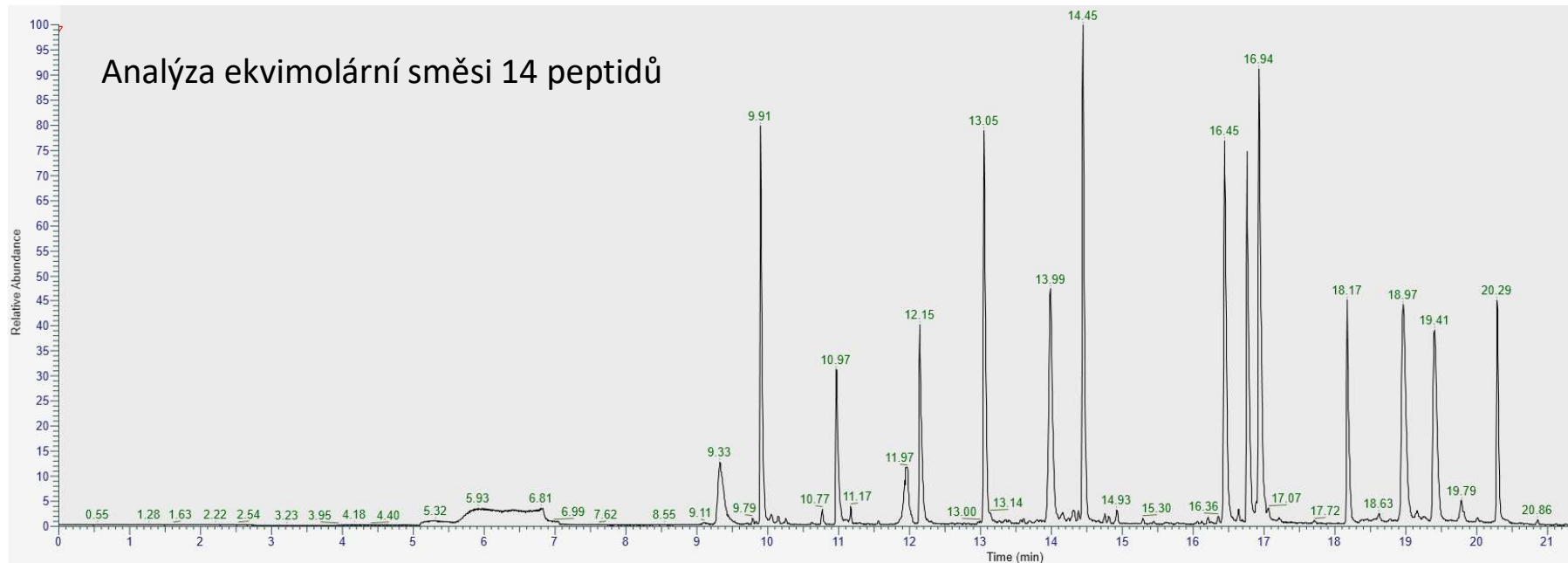
- Samotná MS analýza nemusí být kvantitativní
- Nejintenzivnější signál ve spektru nemusí vždy patřit nejvíce zastoupené látce
- Ovlivněno: přípravou vzorku, schopností ionizace, iontovou supresí, stabilitou iontu ve vakuu, matričními efekty



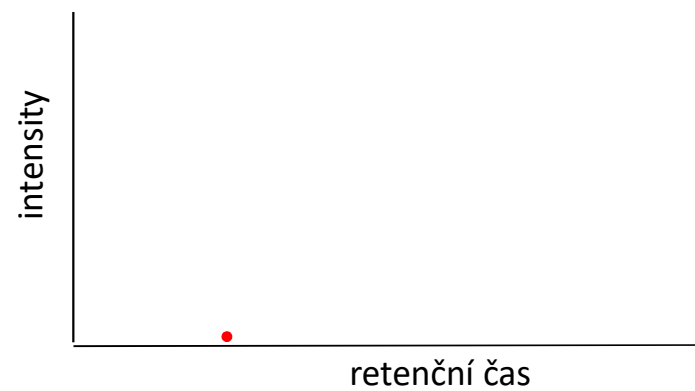
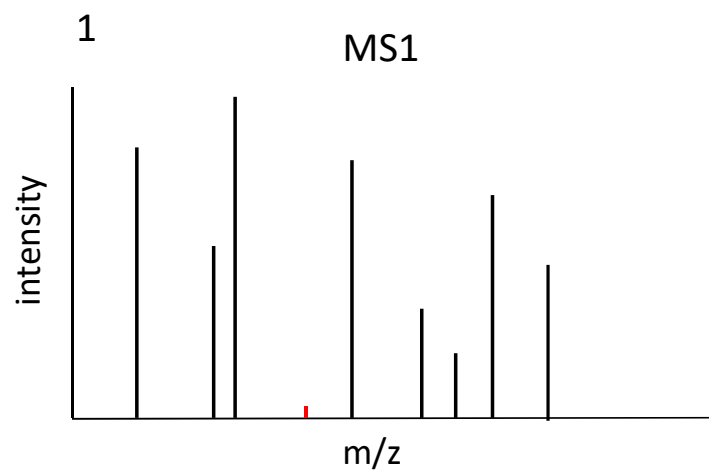


# Kvantifikace pomocí MS

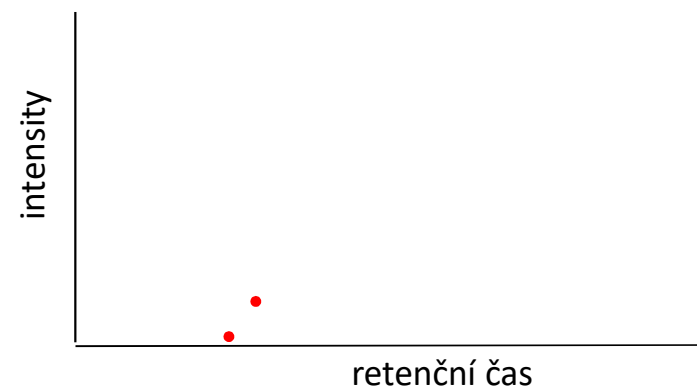
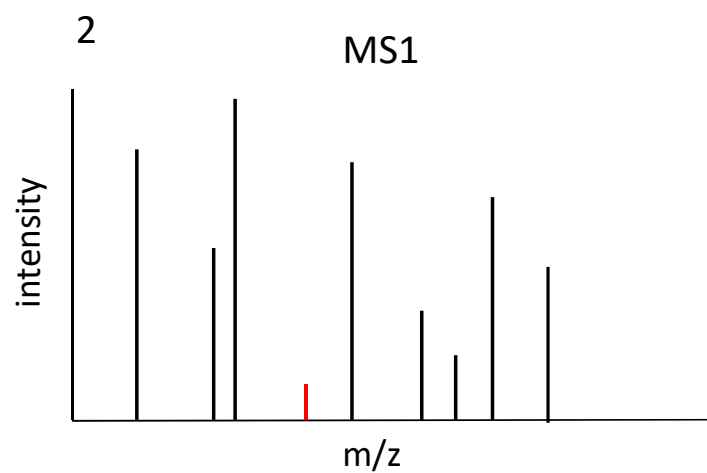
- **Pro kvantitativní porovnání musí být vzorky:**
  - Porovnatelné pouze vzorky téměř identického složení
  - Připravované identicky (ideálně jedním člověkem)
  - Měřeny ve stejný čas, stejnou metodou
  - Reproducibilní chromatografie (stabilní gradient, dostatečná ekvilibrace)



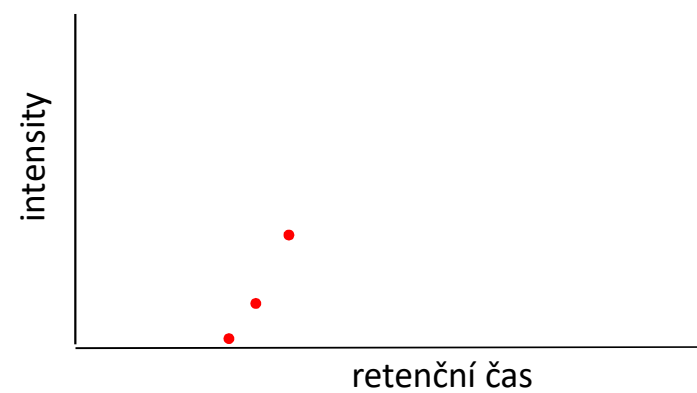
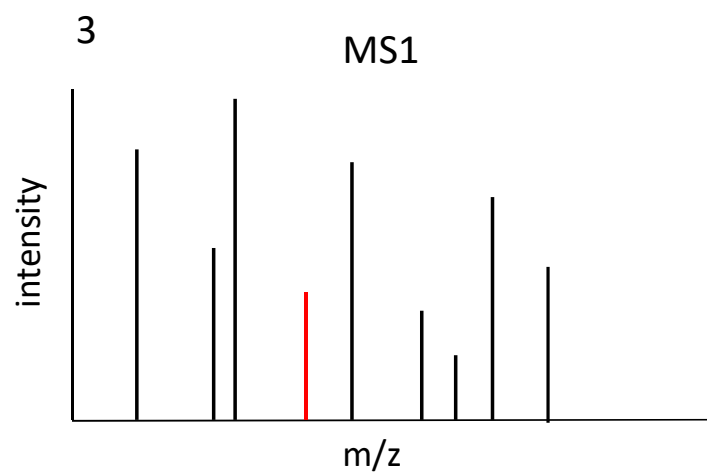
# Kde se u DDA bere kvantitativní informace?



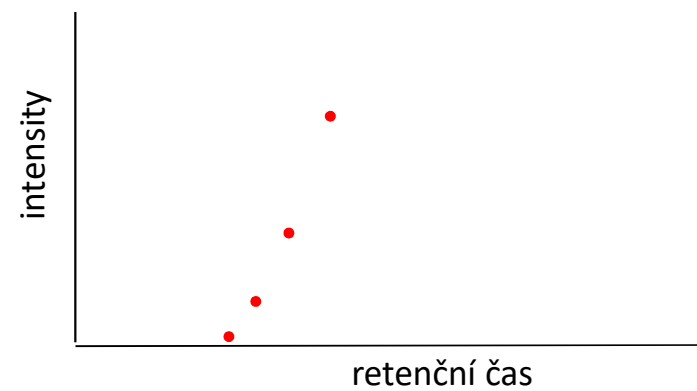
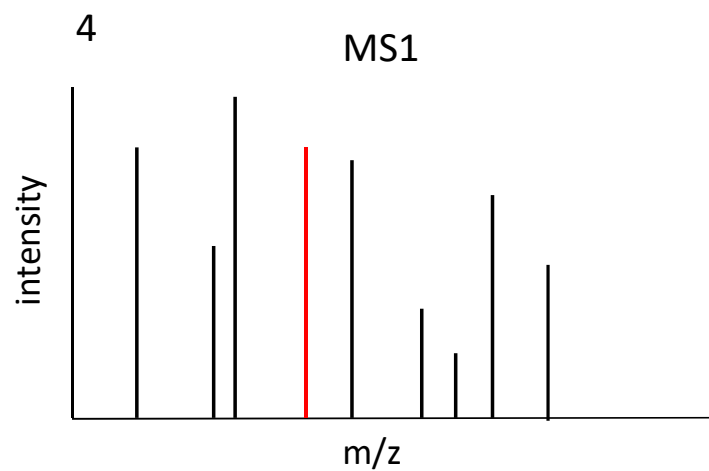
# Kde se u DDA bere kvantitativní informace?



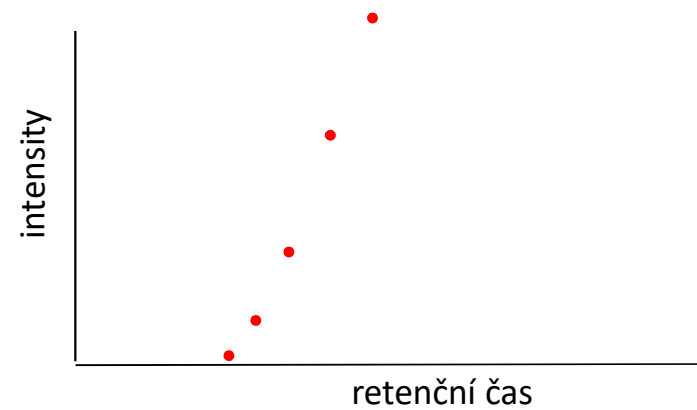
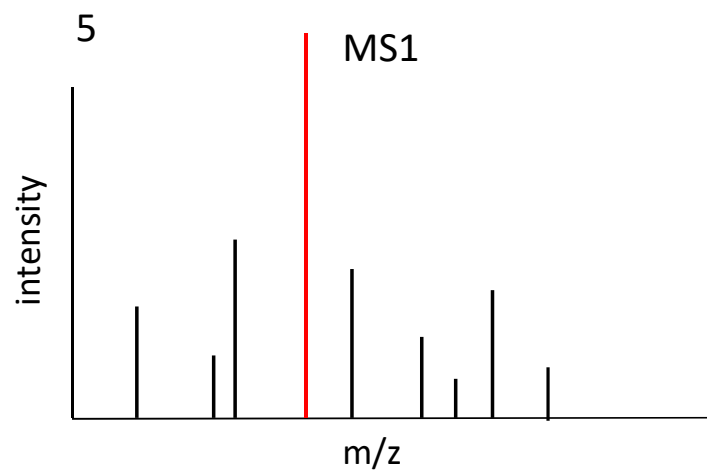
# Kde se u DDA bere kvantitativní informace?



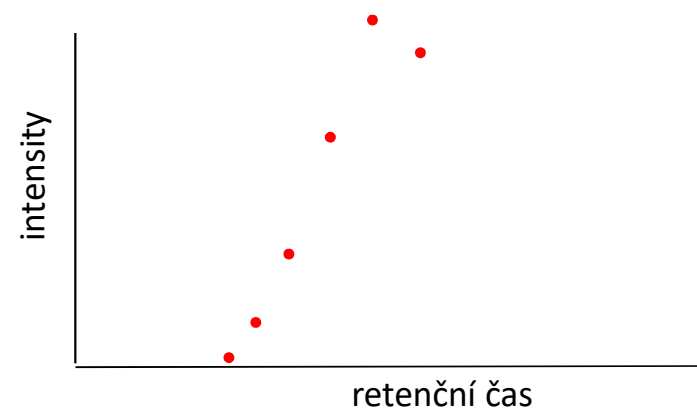
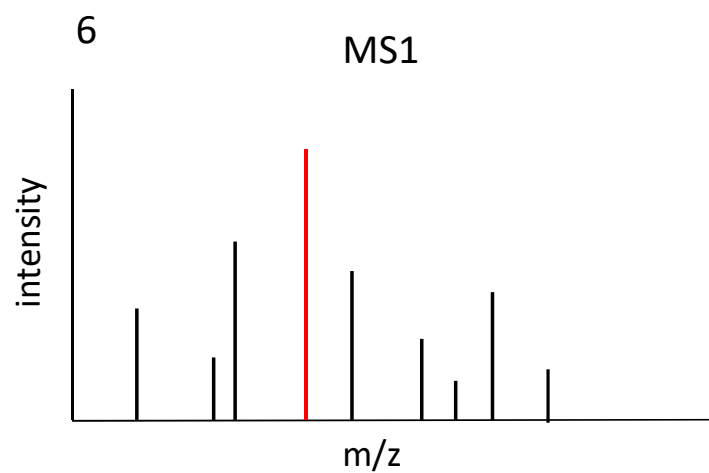
# Kde se u DDA bere kvantitativní informace?



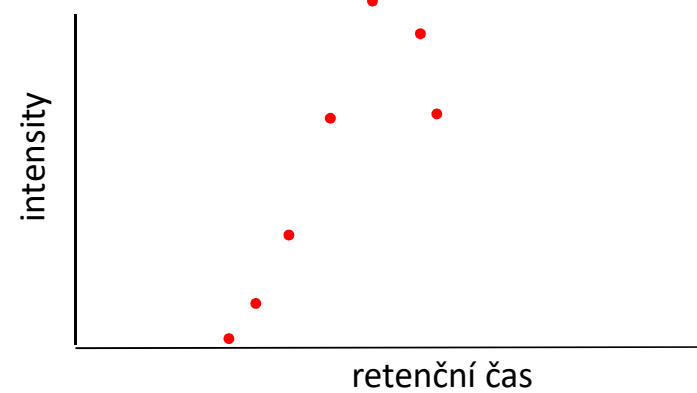
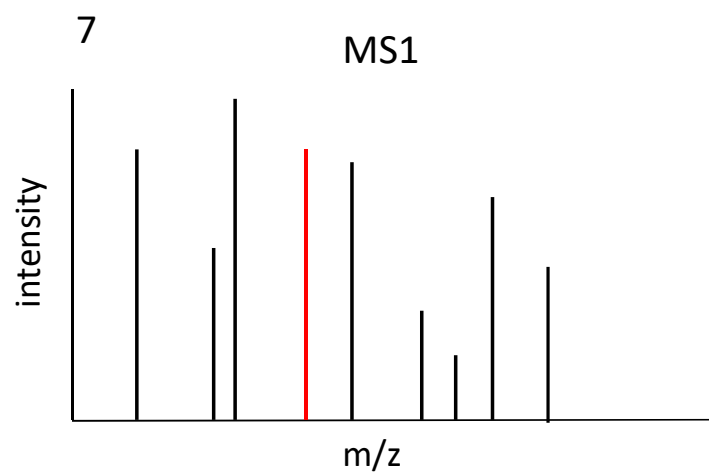
# Kde se u DDA bere kvantitativní informace?



# Kde se u DDA bere kvantitativní informace?

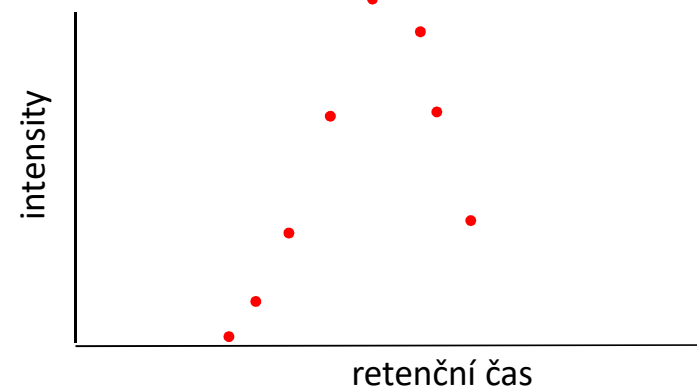
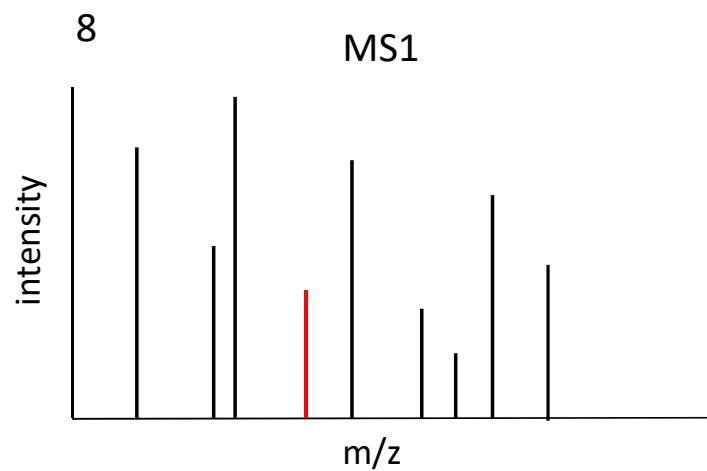


# Kde se u DDA bere kvantitativní informace?

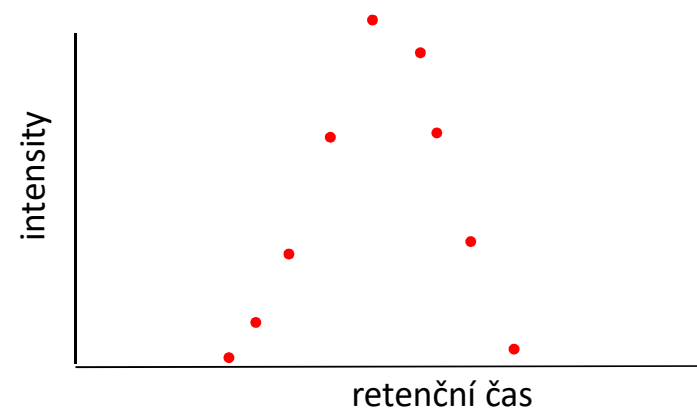
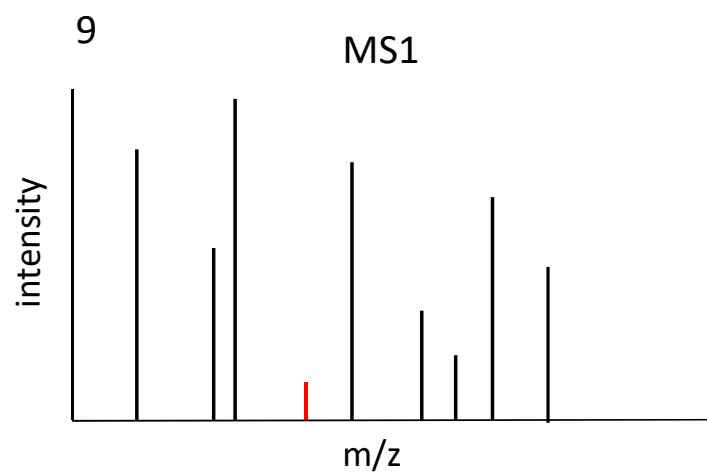




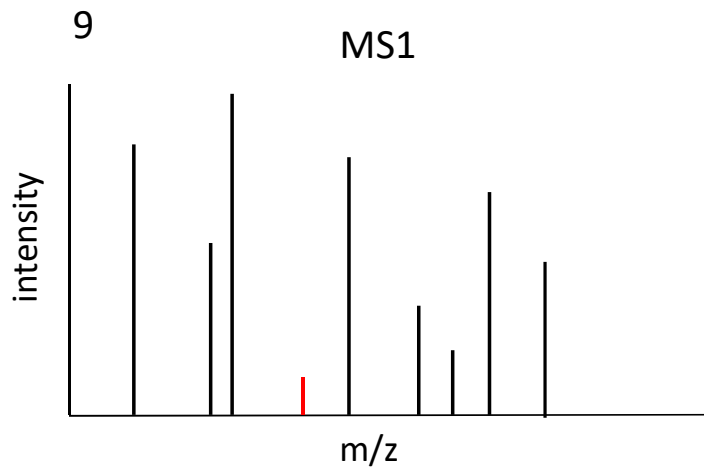
# Kde se u DDA bere kvantitativní informace?



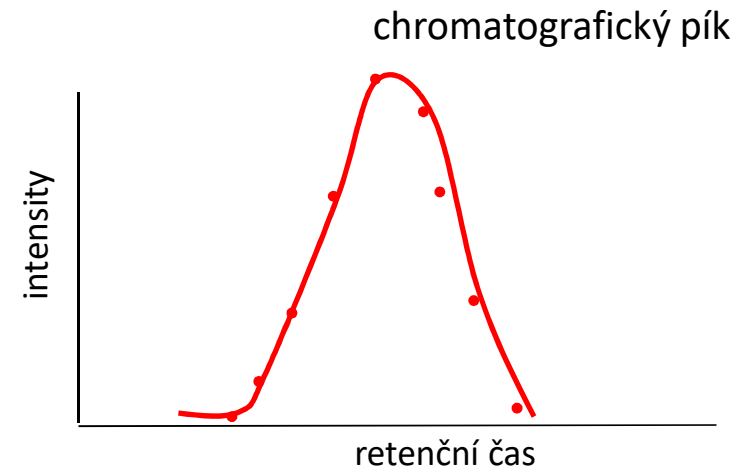
# Kde se u DDA bere kvantitativní informace?



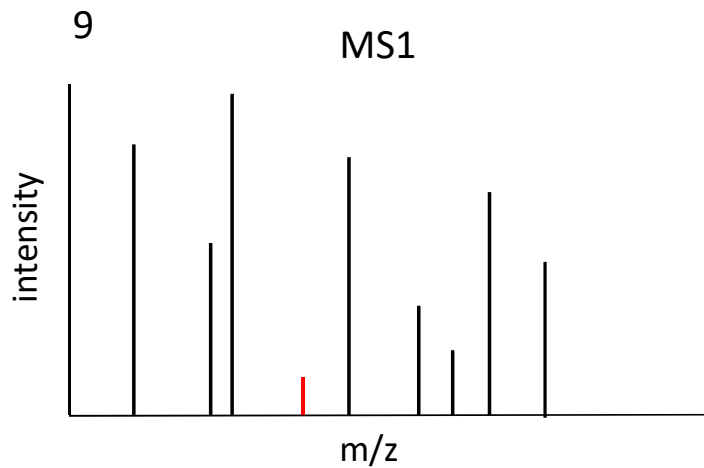
# Kde se u DDA bere kvantitativní informace?



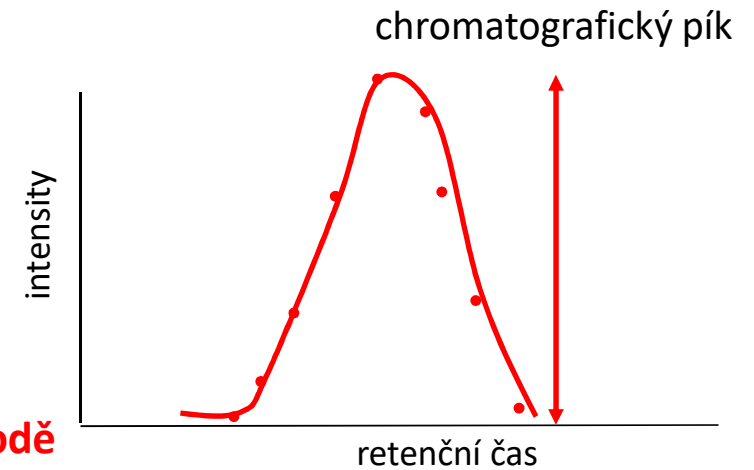
- **Kvantitativní informace obsažena v MS1 signálu**
- V případě DDA analýzy **MS2 spektra neobsahují** kvantitativní informaci
- Základ pro tzv. **Label Free Quantification (LFQ)**



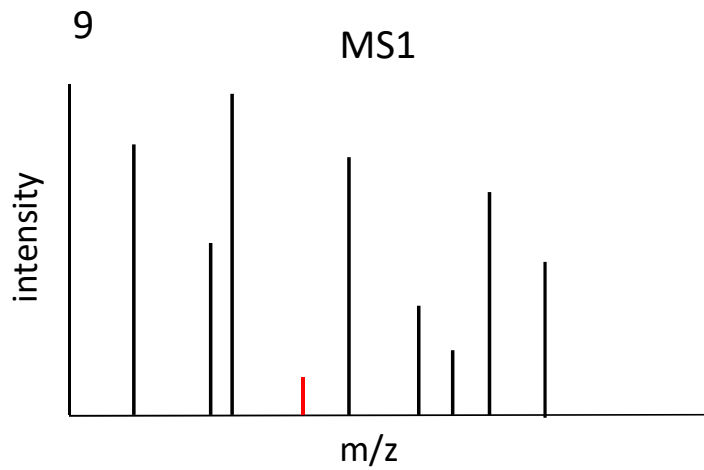
# Kde se u DDA bere kvantitativní informace?



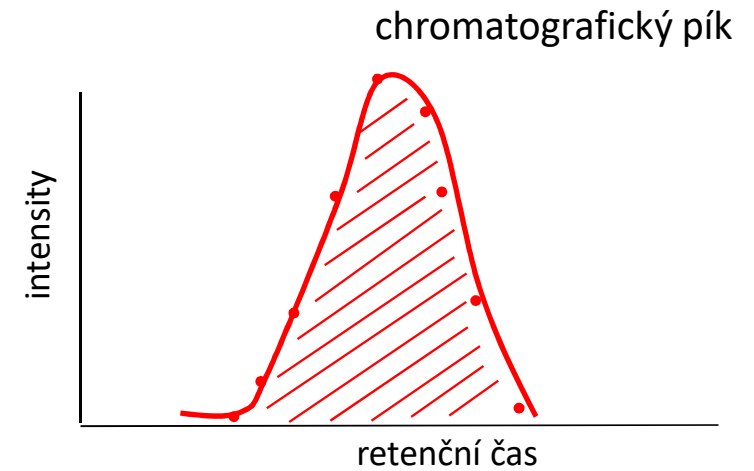
- **Výška píku v nejvyšším bodě**
  - **necílená proteomika**
  - **vyšší reprodukovatelnost**



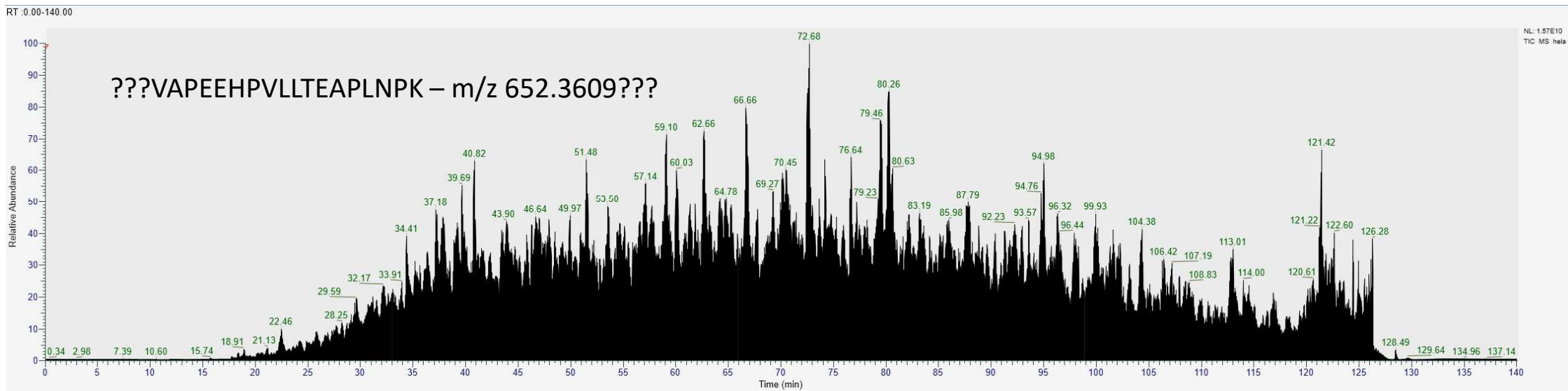
# Kde se u DDA bere kvantitativní informace?



- **Plocha pod píkem**
  - použití u cílených metod (SRM, PRM)

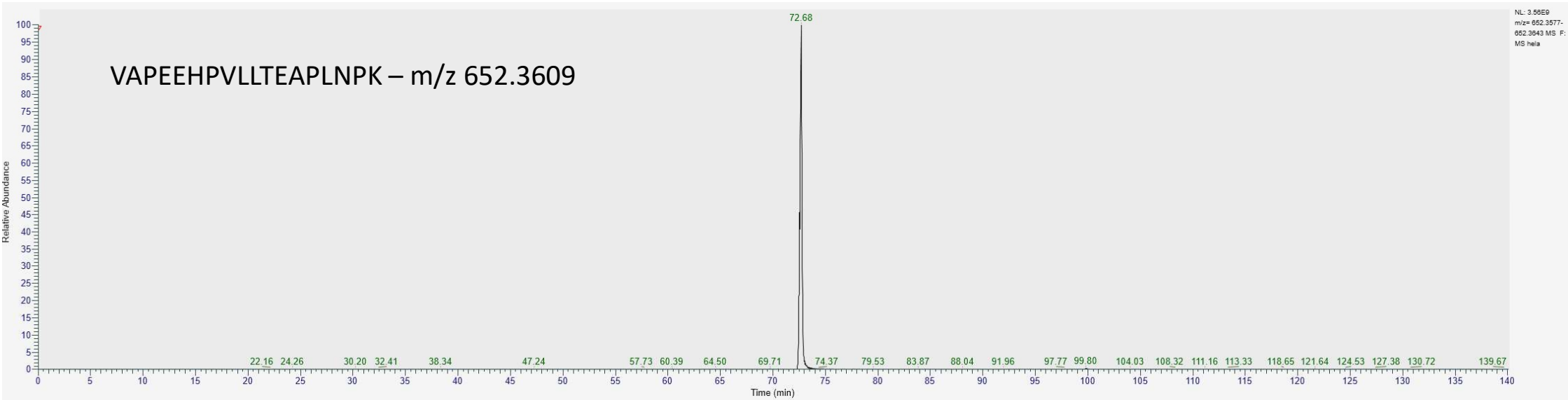


# Kde se u DDA bere kvantitativní informace?



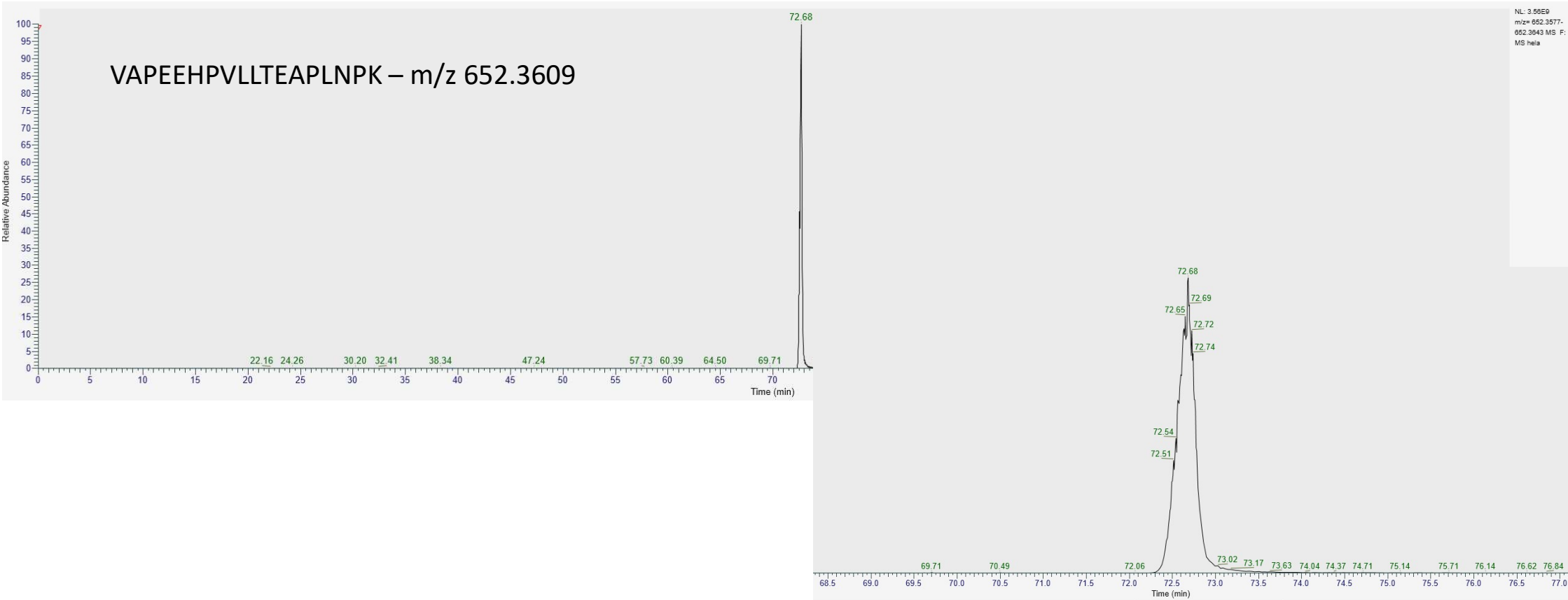
# Kde se u DDA bere kvantitativní informace?

## Extracted Ion Chromatogram – XIC



# Kde se u DDA bere kvantitativní informace?

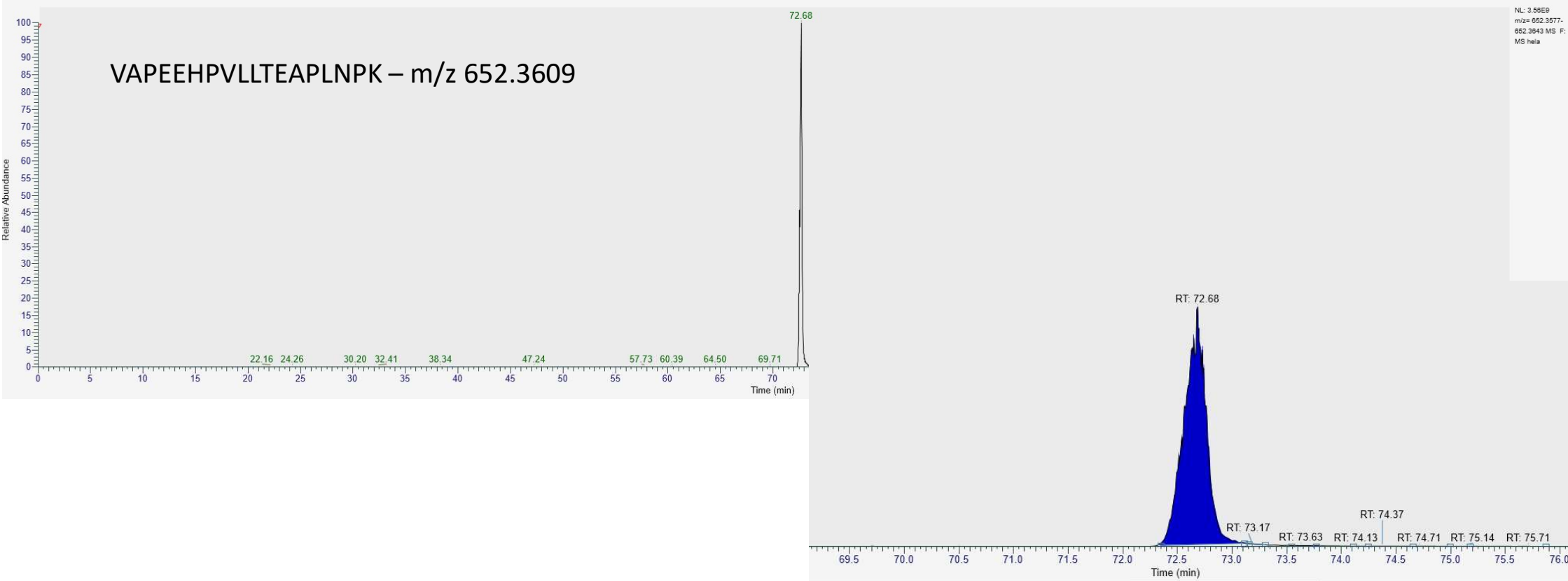
## Extracted Ion Chromatogram – XIC





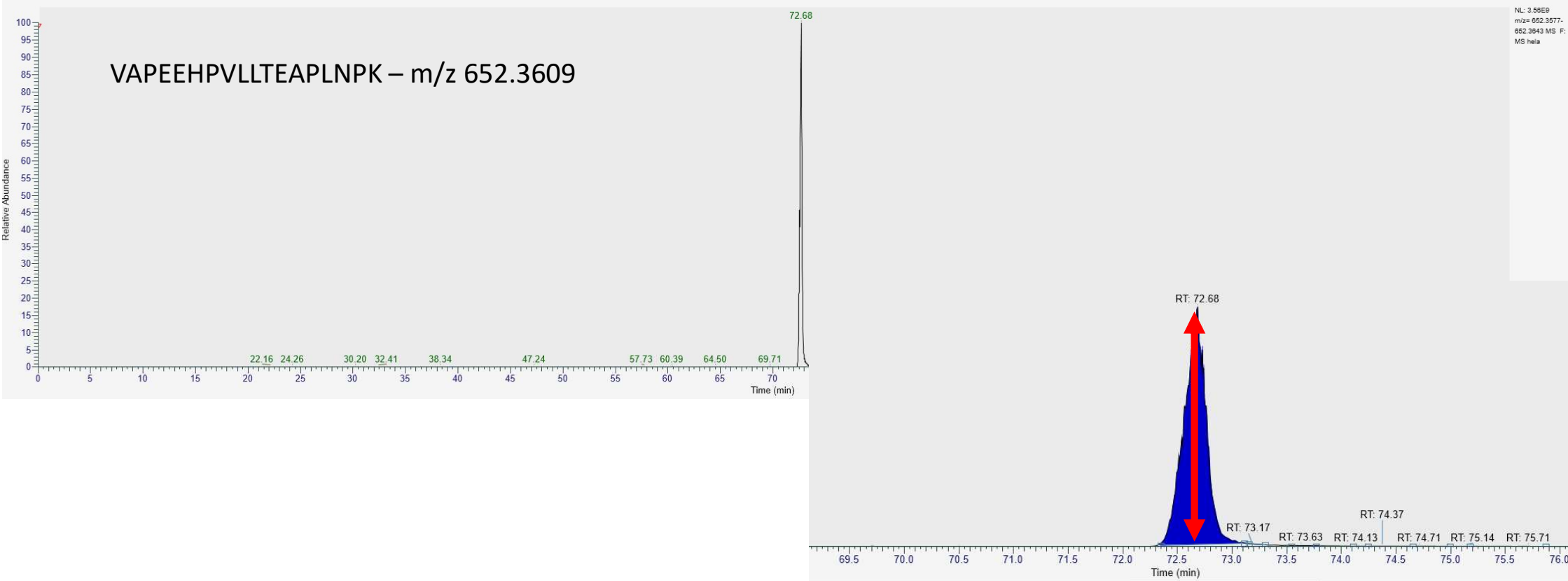
# Kde se u DDA bere kvantitativní informace?

## Extracted Ion Chromatogram – XIC



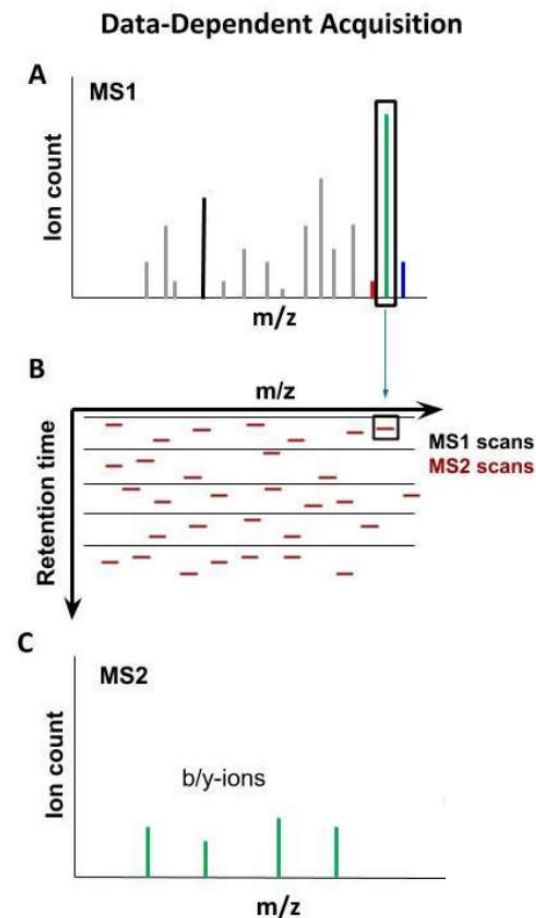
# Kde se u DDA bere kvantitativní informace?

## Extracted Ion Chromatogram – XIC



# Data Dependent Acquisition

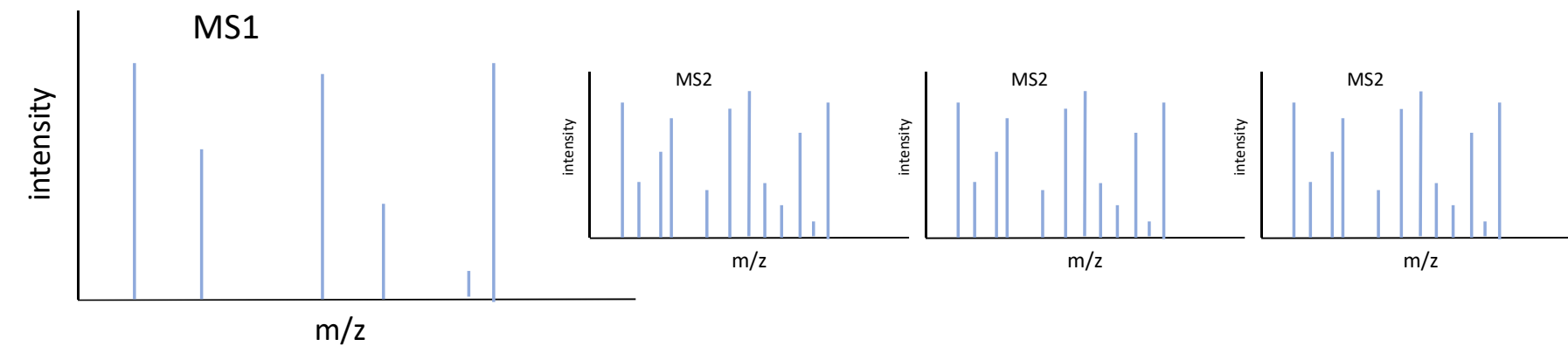
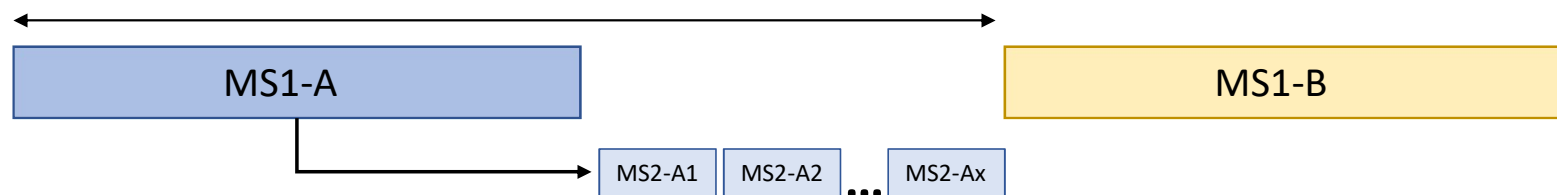
- Stochastický/semistochastický princip – výběr prekurzorů je v podstatě náhodný
- Detekce zhruba 5 000 proteinů ve vzorku savčích buněk při 120 min gradientu
- **Máme jednoznačné propojení prekurzor – fragmentační spektrum**
- **Kvantitativní informace uložena pouze v MS1**
- **Informace o peptidech, které nebyly vybrány pro fragmentaci je nevratně ztracena**
- CV asi 25 %



# Délka cyklu vs. šířka píku

- **Skenovací cyklus**
- Celkový čas věnovaný jednomu cyklu - od počátku jednoho MS1 scanu do začátku následujícího MS1 scanu = cycle time

# Délka cyklu vs. šířka píku



# Délka cyklu vs. šířka píku

- Délka cyklu v závislosti na šířce chromatografického píku definuje počet bodů přes pík
- Přímo ovlivňuje přesnost kvantifikace – čím lepší pokrytí píku, tím přesnější kvantifikace

- **Delší cyklus**
  - nasbíráme více MS/MS spekter
  - máme horší pokrytí píku
- **Kratší cyklus**
  - lepší pokrytí píku
  - méně MS/MS spekter

Máme 10 s chromatografický pík

Při délce cyklu 2,5 s



4 body přes pík

Při délce cyklu 1,25 s

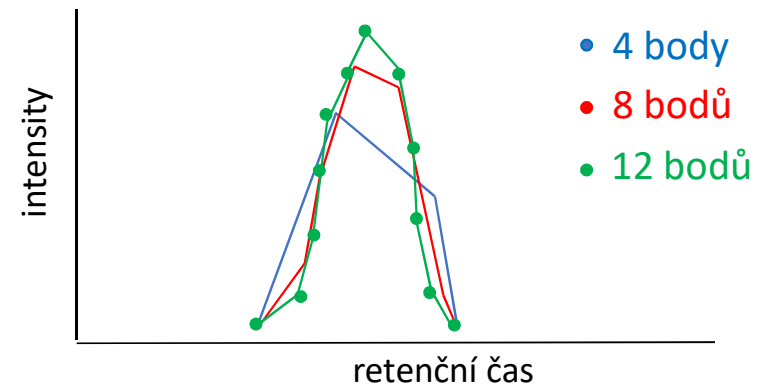


8 bodů přes pík

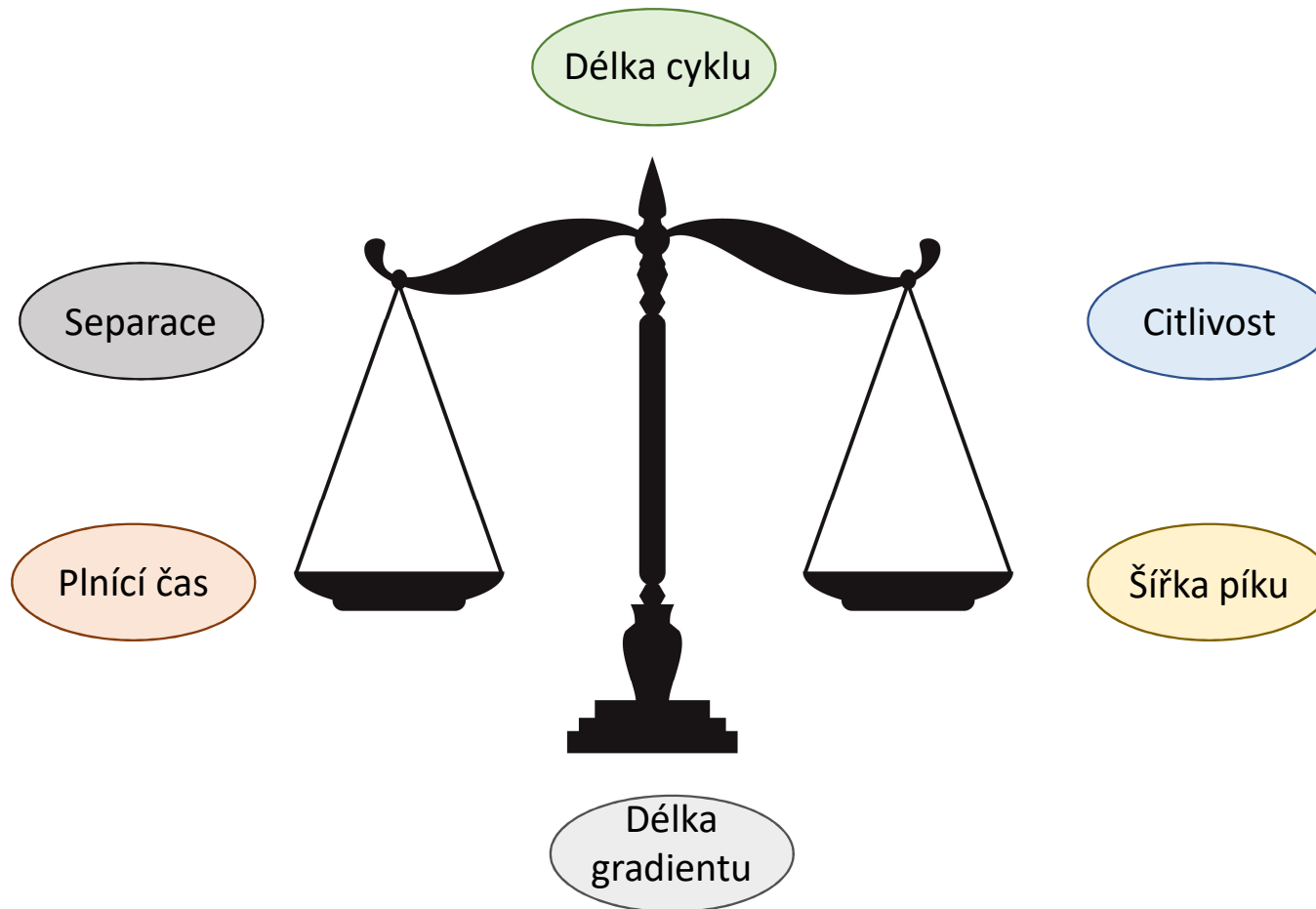
Při délce cyklu 0,83 s



12 bodů přes pík



# Délka cyklu vs. šířka píku

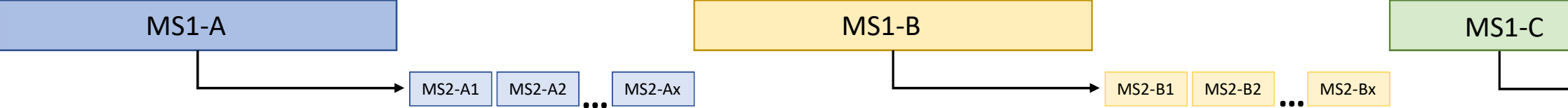


# Paralelizace - tribridy

MS1 sken v orbitrapu asi 250 ms (rozlišení 120 K)

MS2 sken v iontové pasti asi 25 ms

## Bez paralelizace



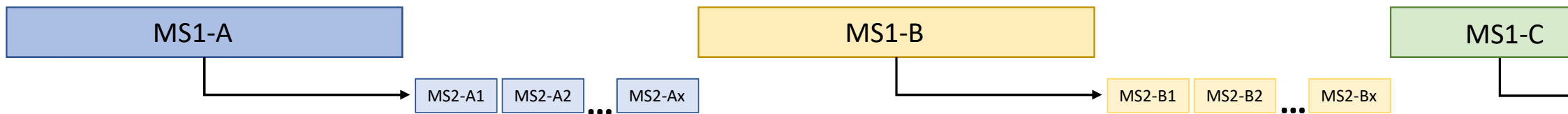


# Paralelizace – oddělení MS1 a MS2 v prostoru

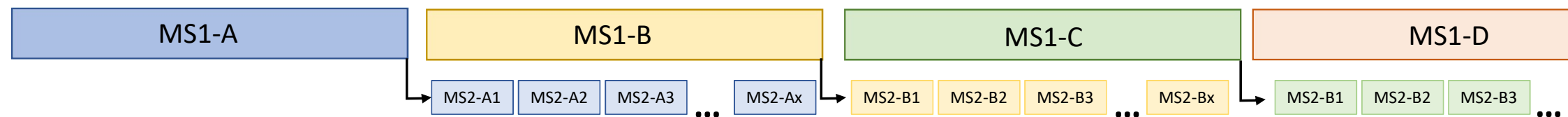
MS1 sken v orbitrapu asi 250 ms (rozlišení 120 K)

MS2 sken v iontové pasti asi 25 ms

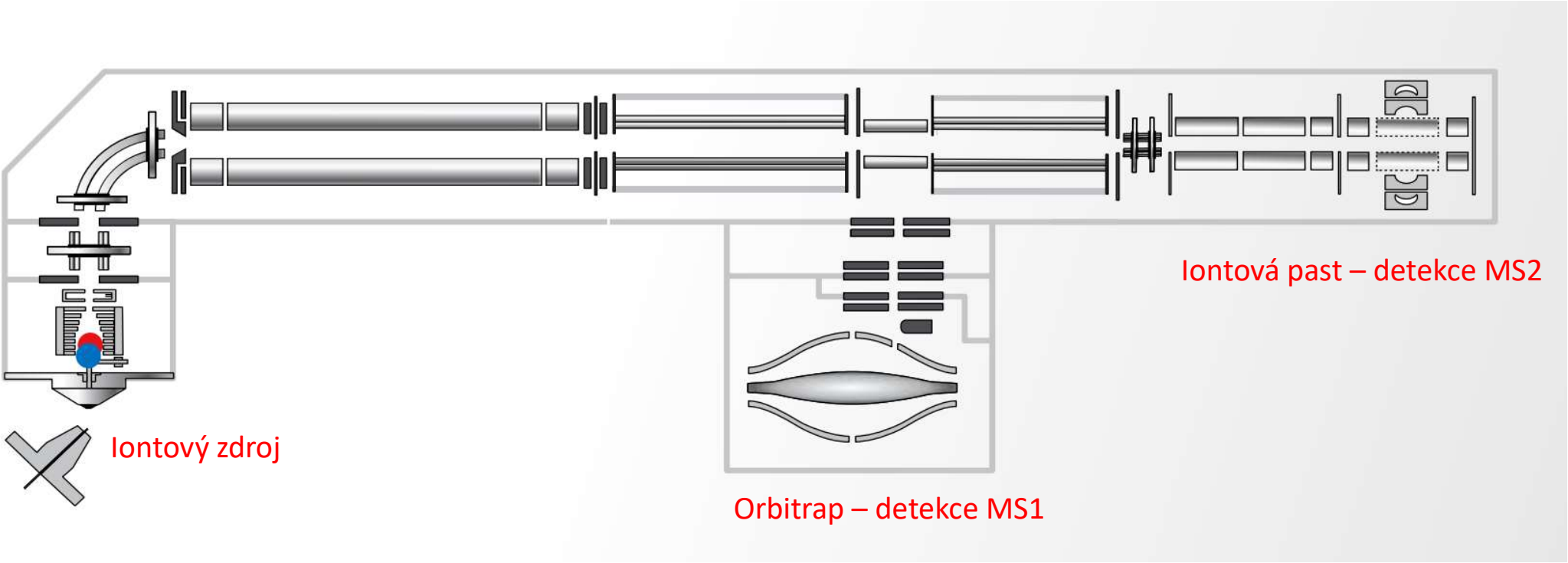
## Bez paralelizace



## S paralelizací

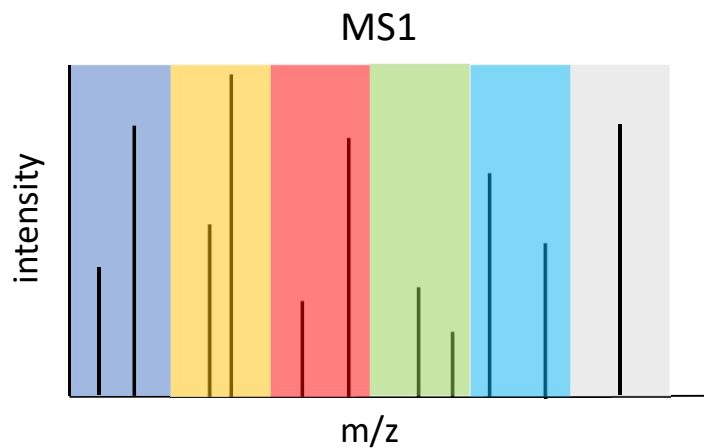


# Paralelizace - tribridy



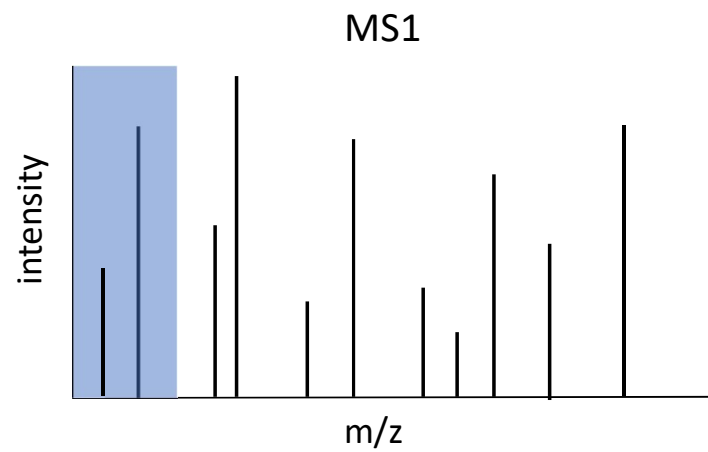
# **Data Independent Acquisition - DIA**

# Data Independent Acquisition - DIA

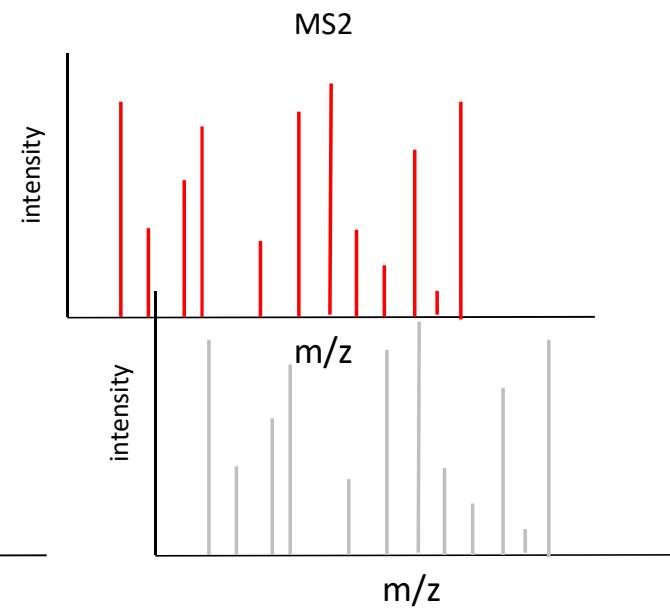
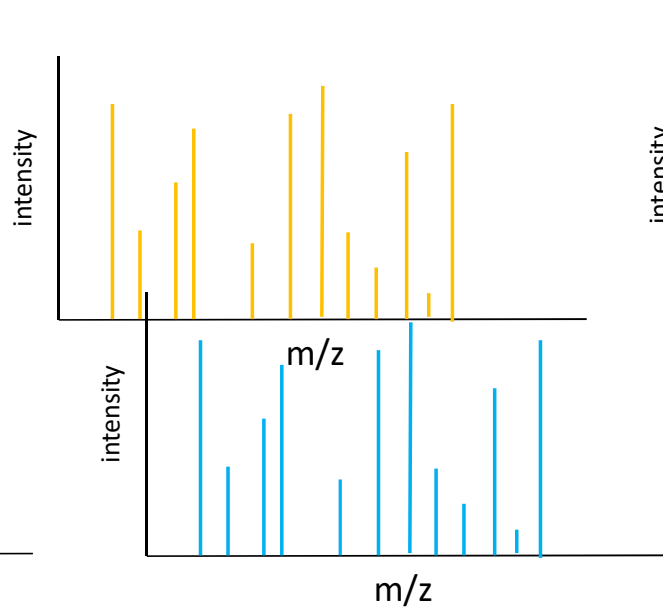
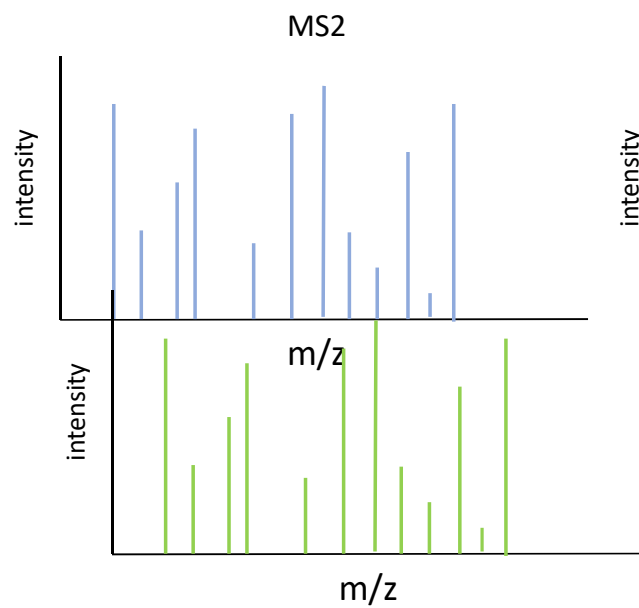
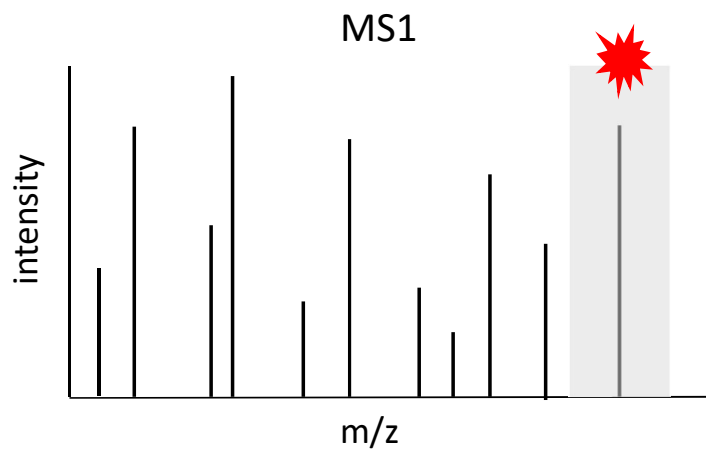


- Určitý počet izolačních oken pokrývajících celý hmotnostní rozsah
- Šířka okna obvykle 10 – 20 Da – v závislosti na konkrétním typu MS instrumentu
- **Izolace a fragmentace všech prekurzorů v daném hmotnostním okně**

# Data Independent Acquisition - DIA

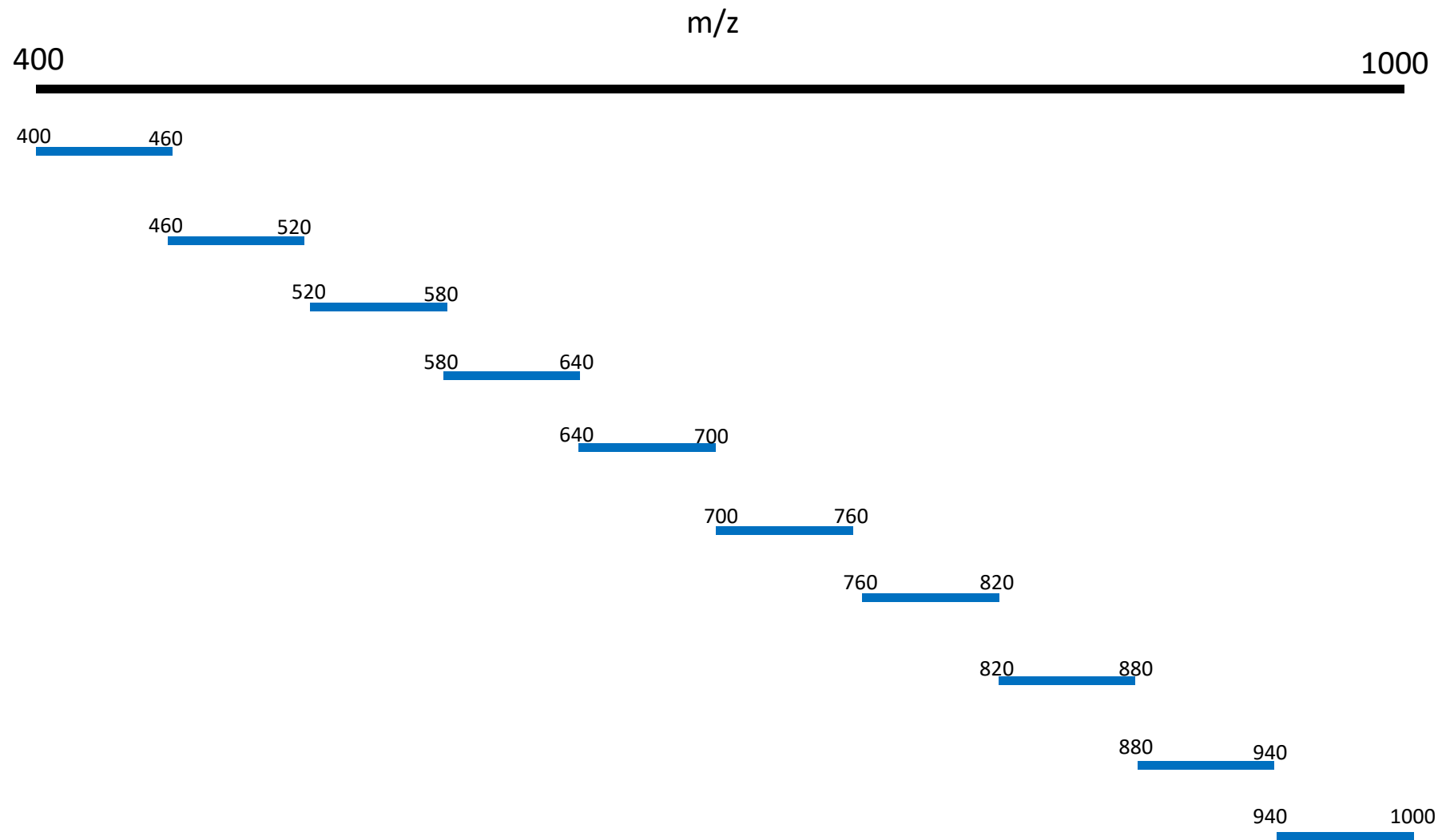


# Data Independent Acquisition - DIA



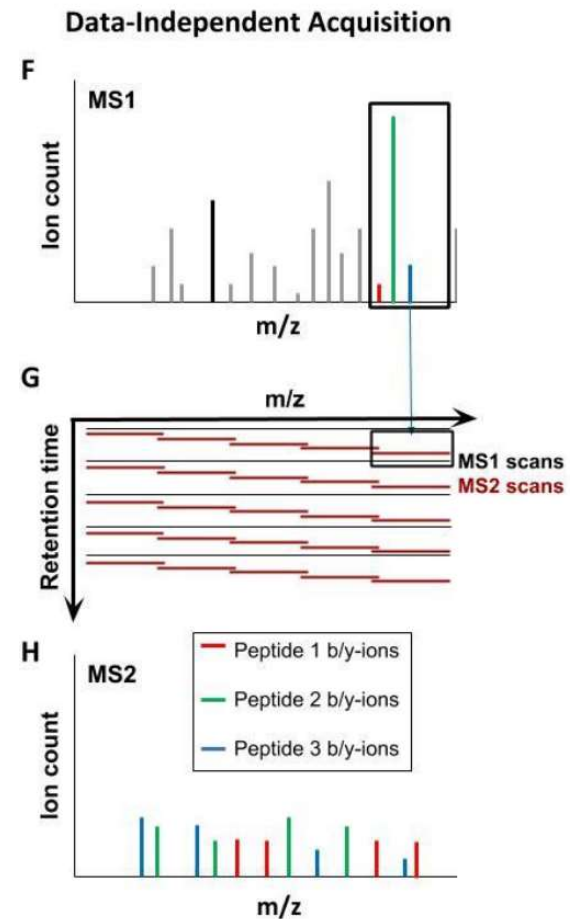
Hmotnostní rozsah – 400 – 1000 m/z

Šířka okna 60 Da = 10 oken



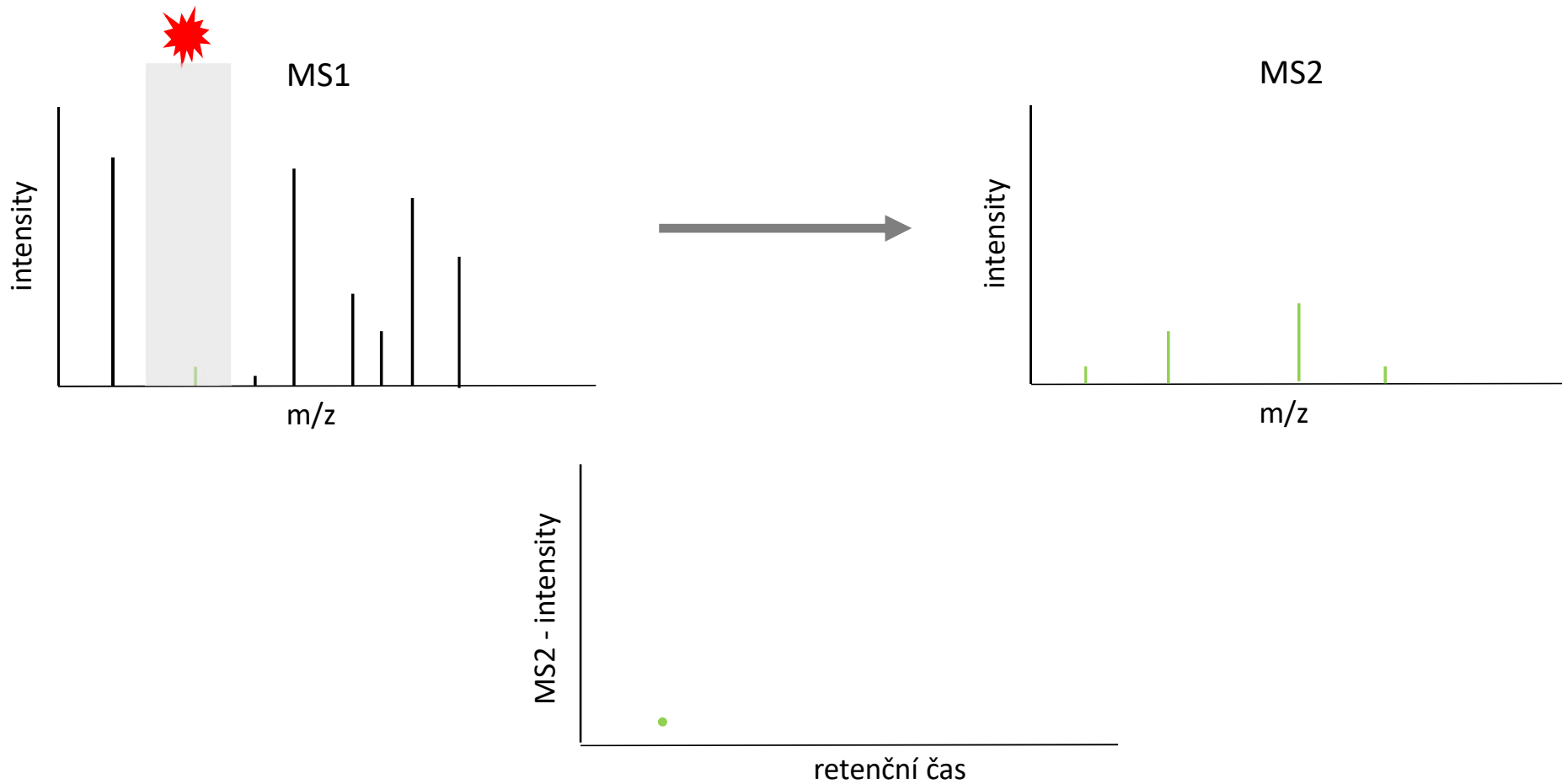
# Data Independent Acquisition - DIA

- Fragmentační spektrum obsahuje fragmenty všech peptidů izolovaných v daném hmotnostním okně
- zachovaná informace o všech peptidech
- kvantitativní informace uložena v MS2
- detekce zhruba 8 000 proteinů z lyzátu savčích buněk při 60 min gradientu
- přesnější než DDA – CV kolem 5 – 10 %
- menší počet missing values
  - DDA – 30 – 40 %
  - **DIA – 15 %**
- MS2 spektra výrazně složitější než u DDA

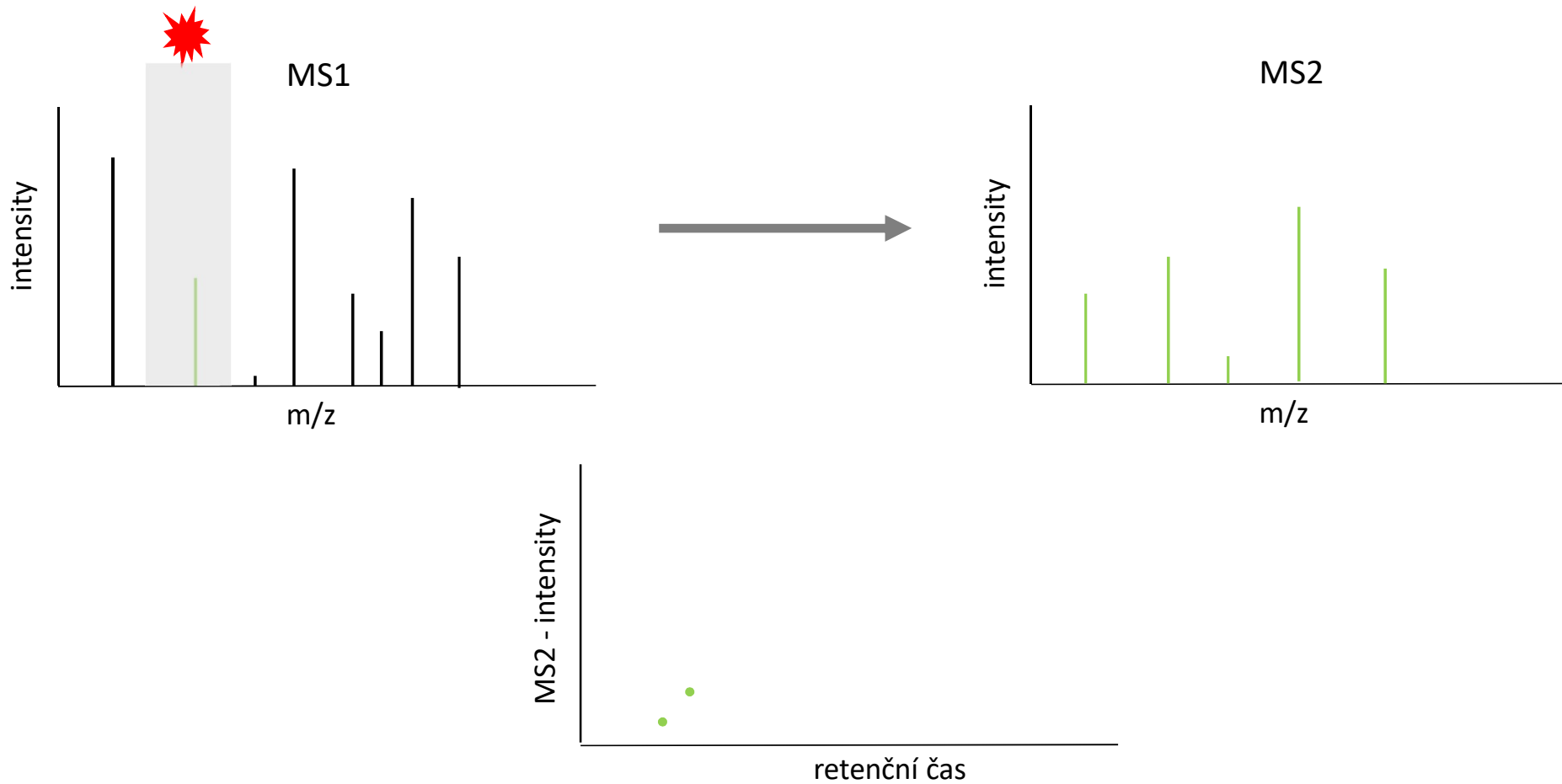




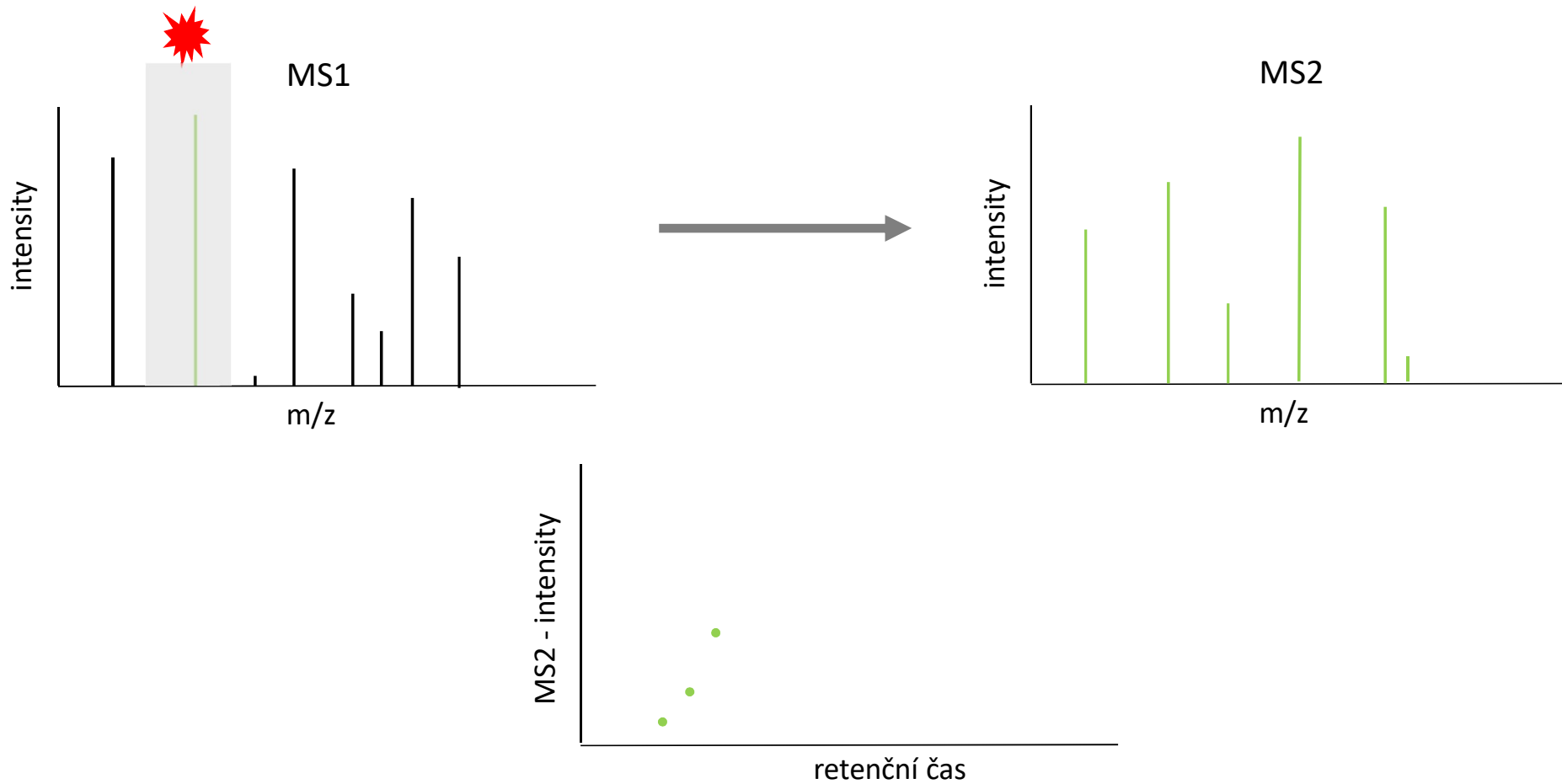
# Kde se u DIA bere kvantitativní informace?



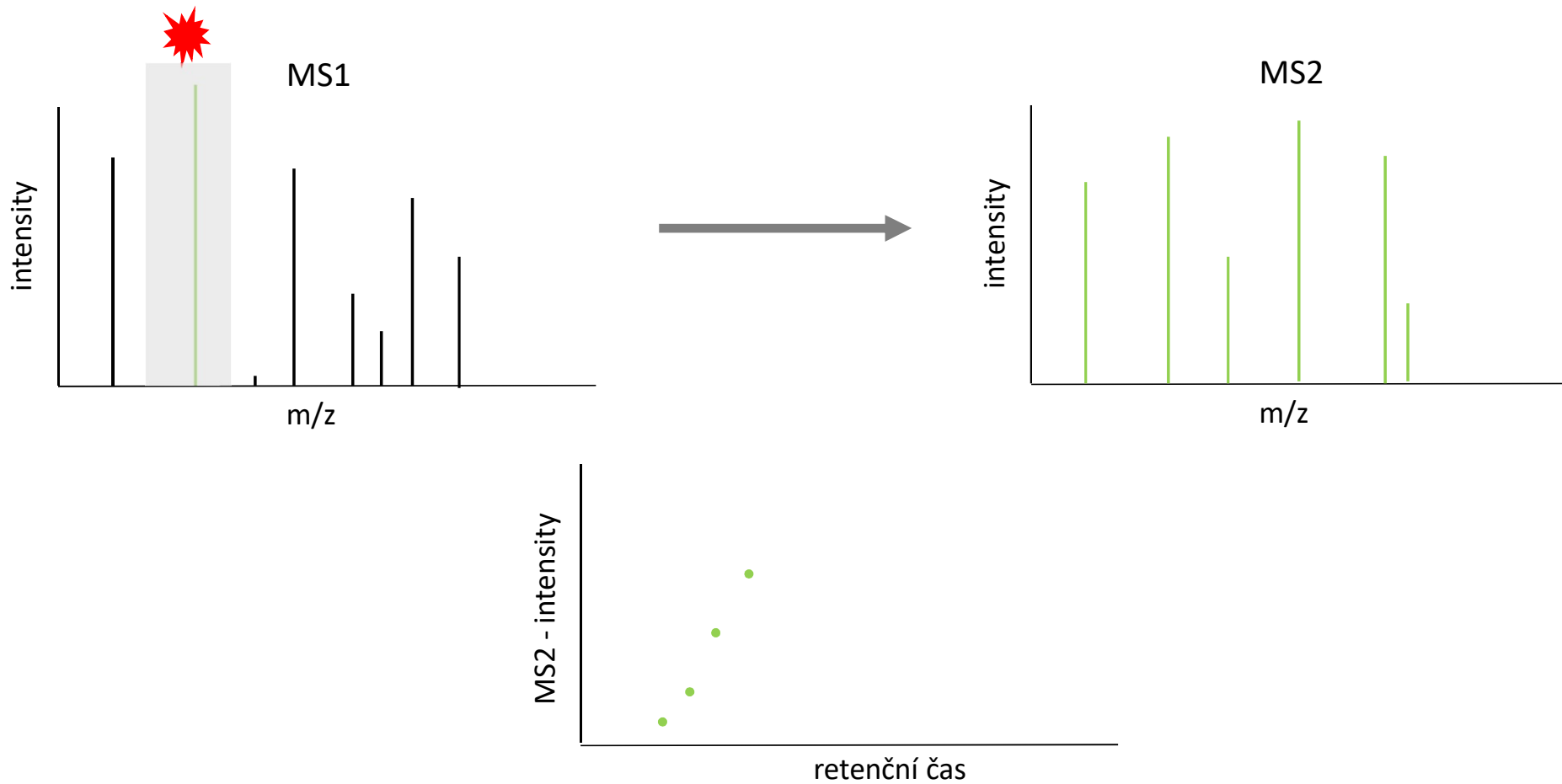
# Kde se u DIA bere kvantitativní informace?



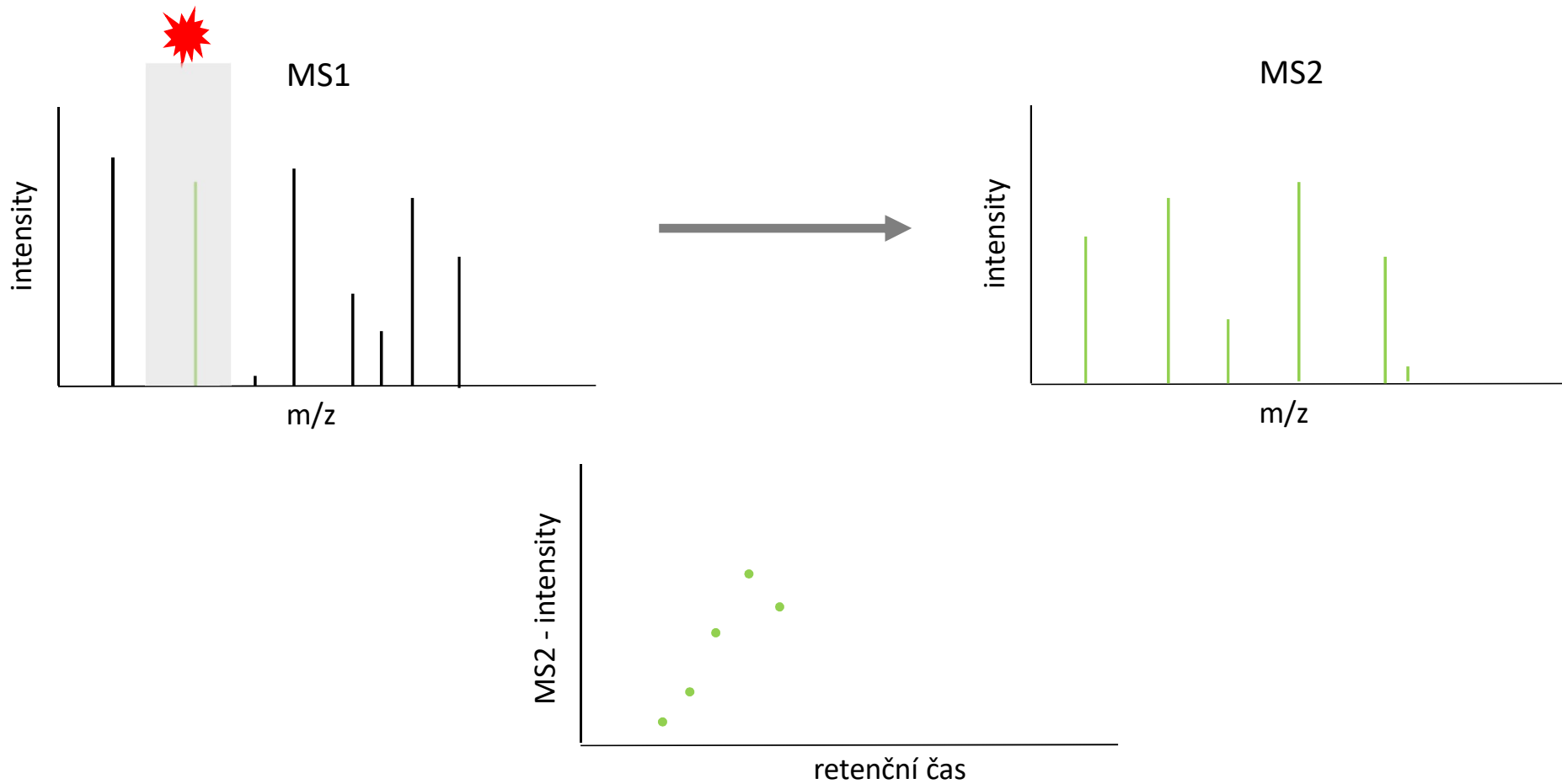
# Kde se u DIA bere kvantitativní informace?



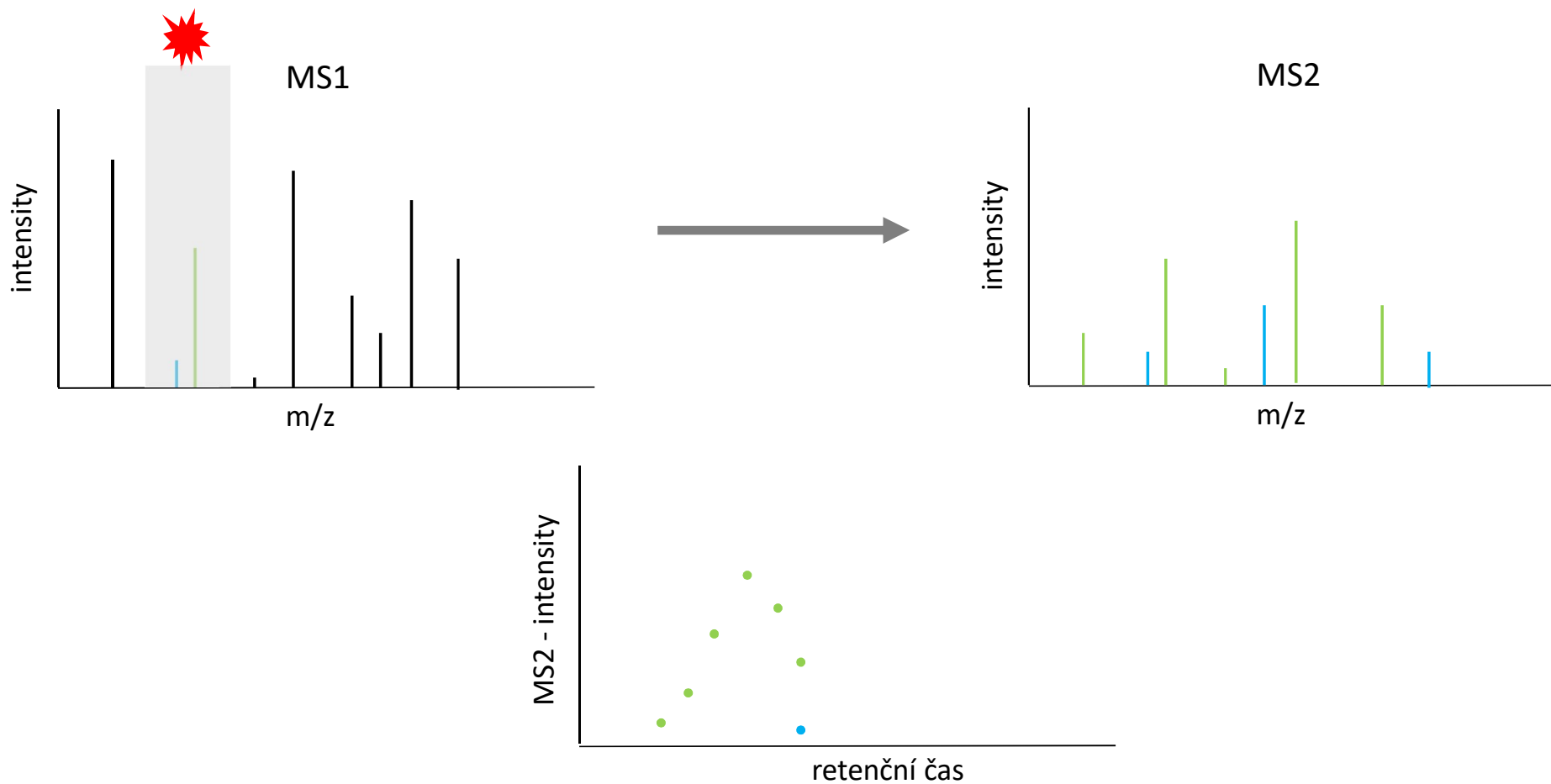
# Kde se u DIA bere kvantitativní informace?



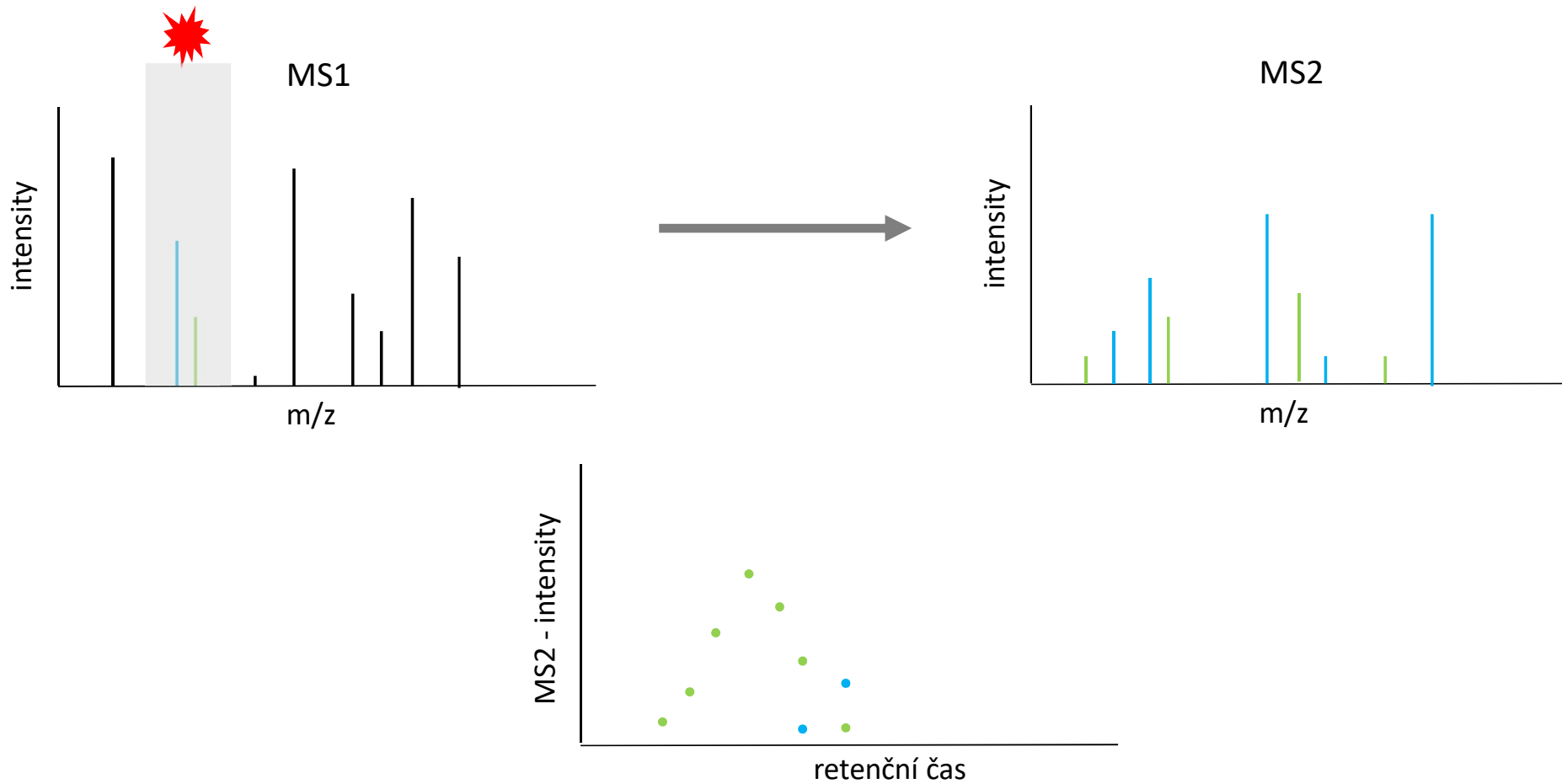
# Kde se u DIA bere kvantitativní informace?



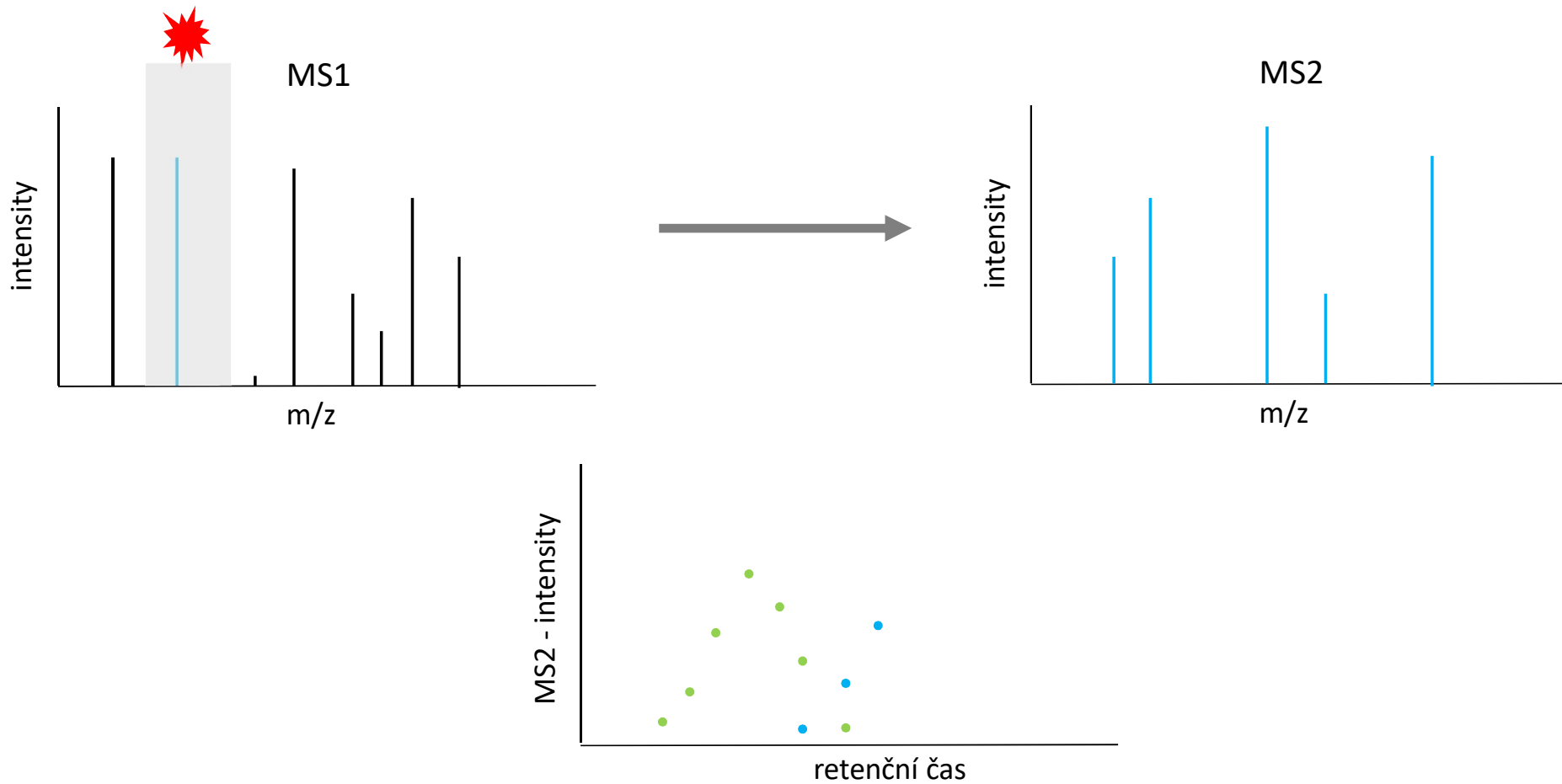
# Kde se u DIA bere kvantitativní informace?



# Kde se u DIA bere kvantitativní informace?

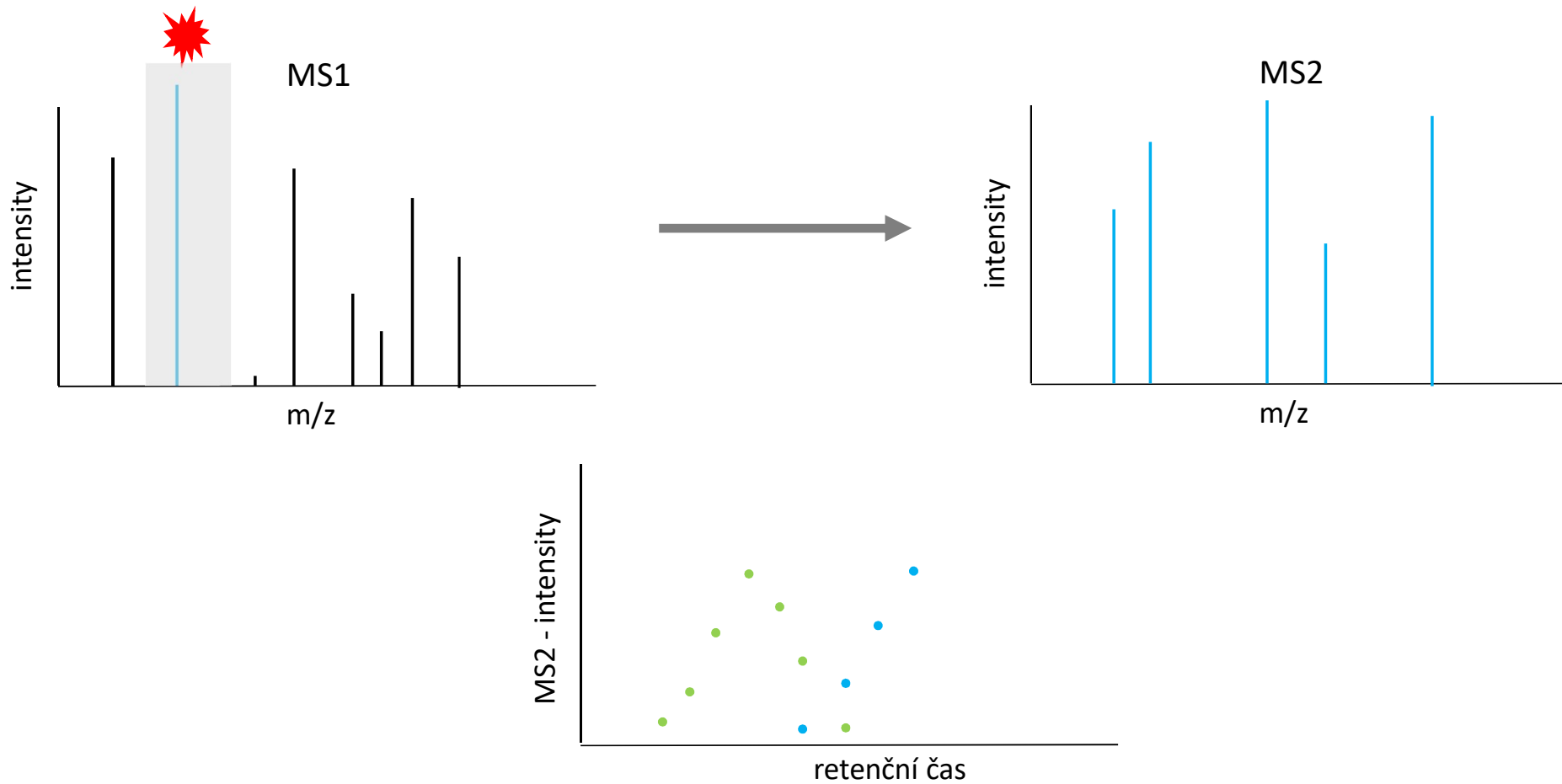


# Kde se u DIA bere kvantitativní informace?

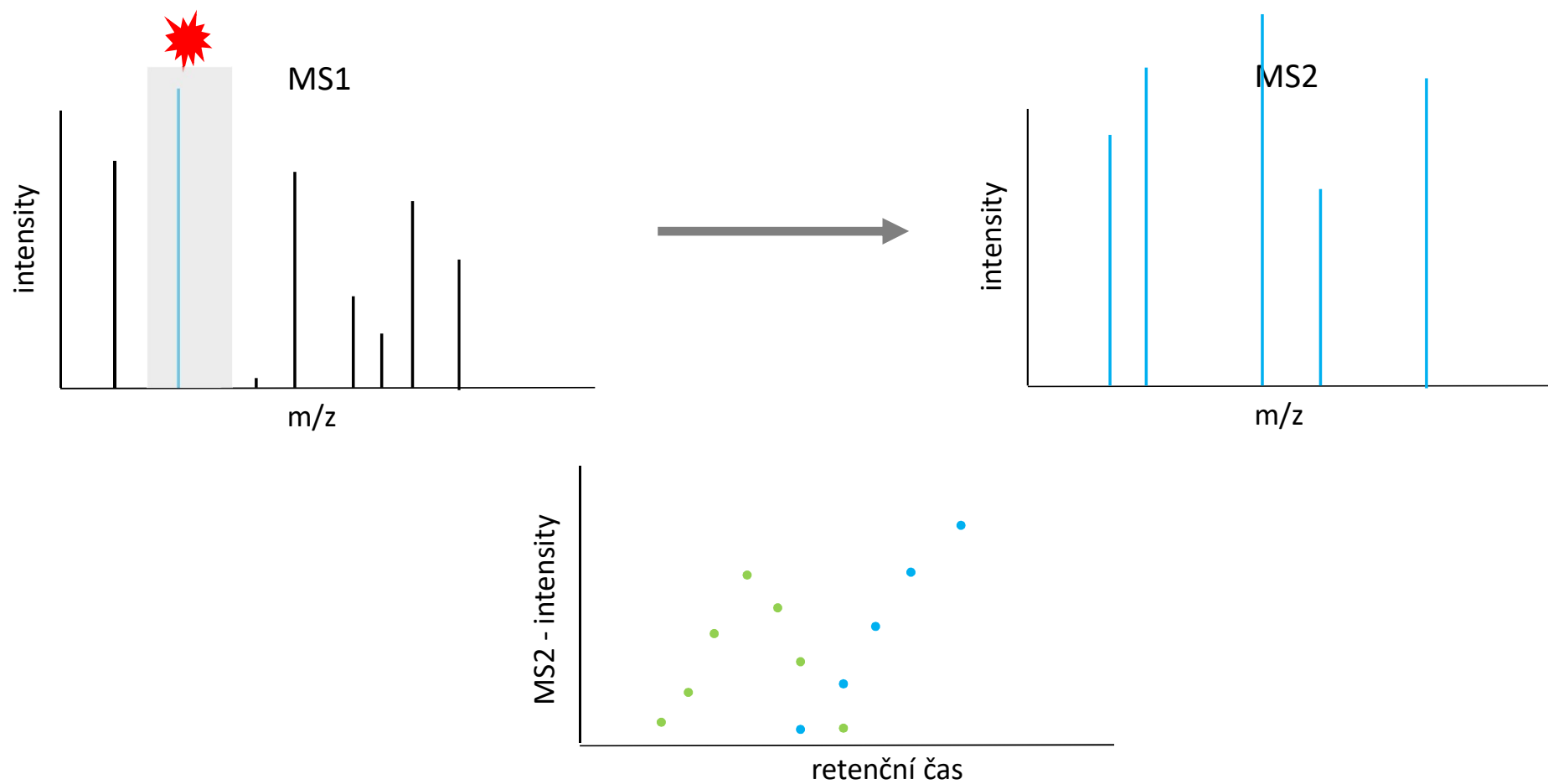




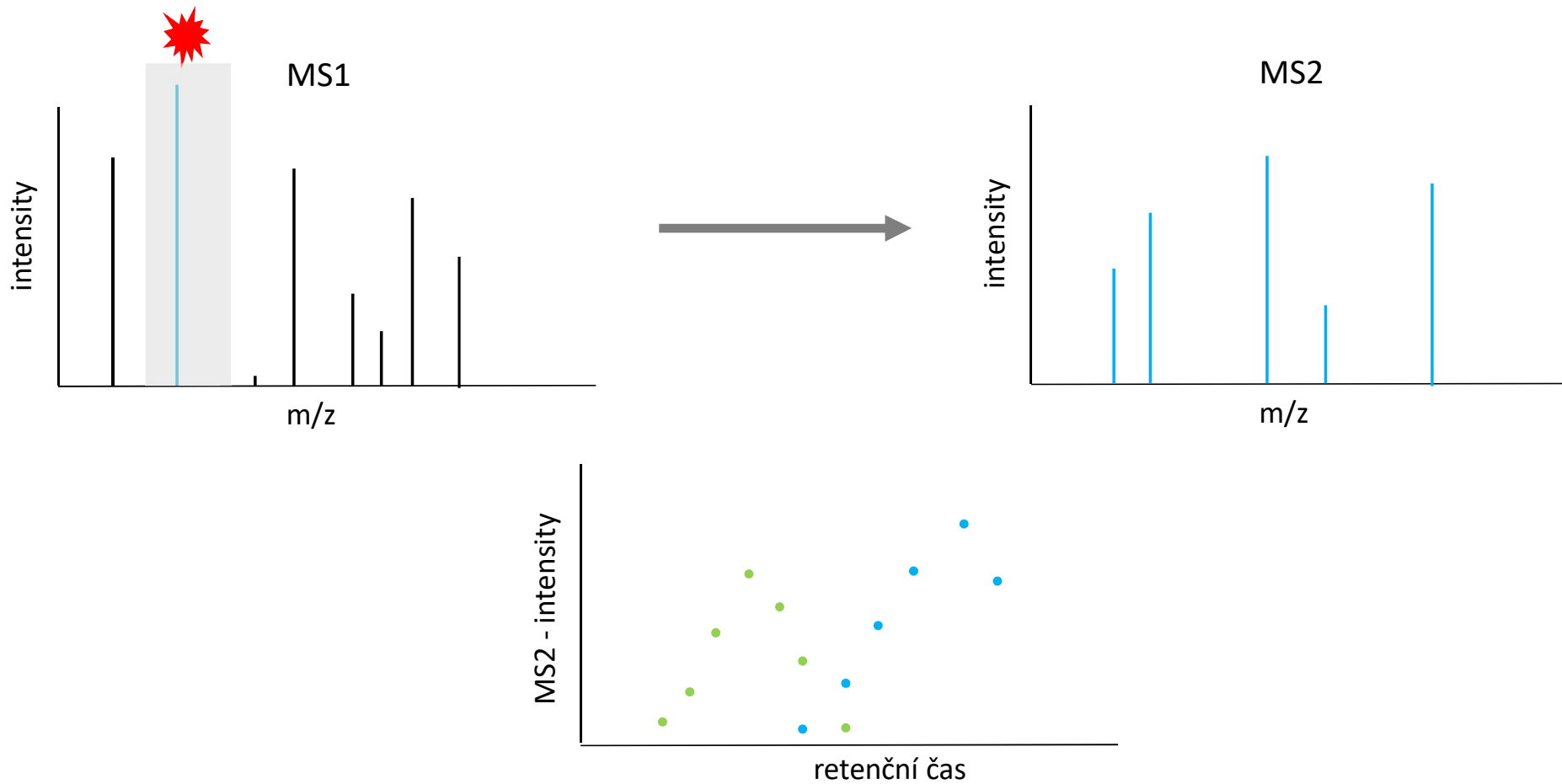
# Kde se u DIA bere kvantitativní informace?



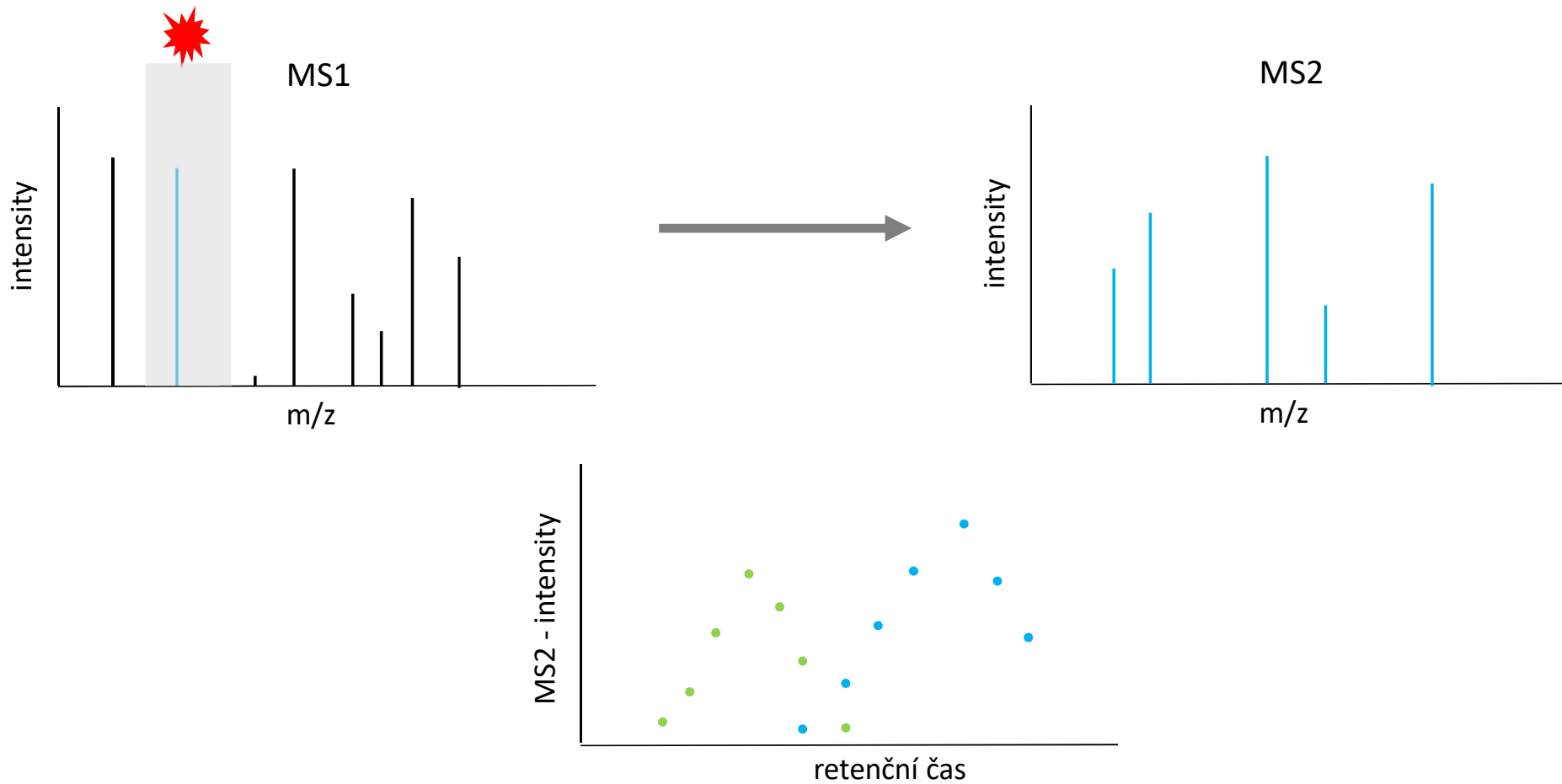
# Kde se u DIA bere kvantitativní informace?



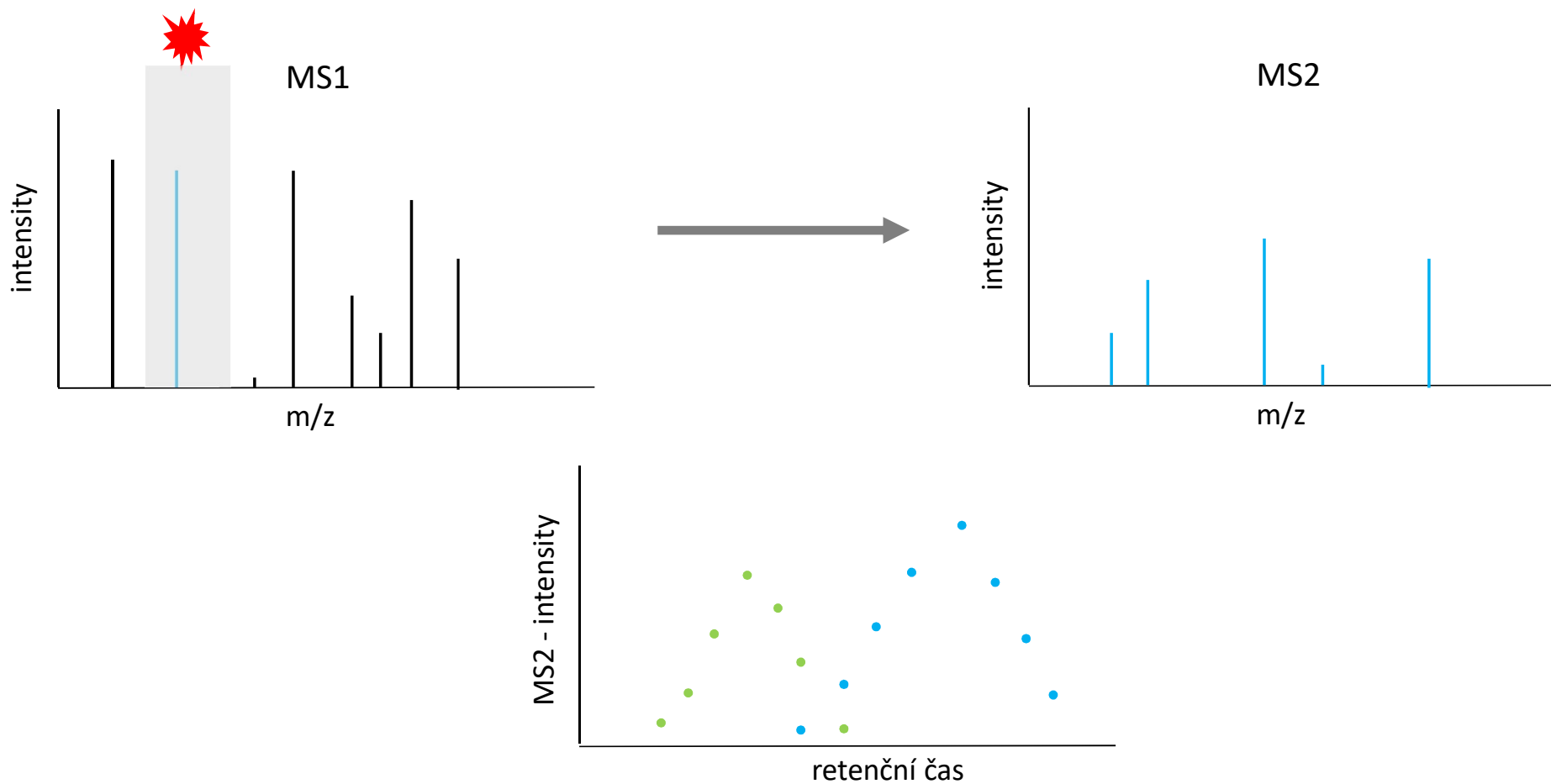
# Kde se u DIA bere kvantitativní informace?



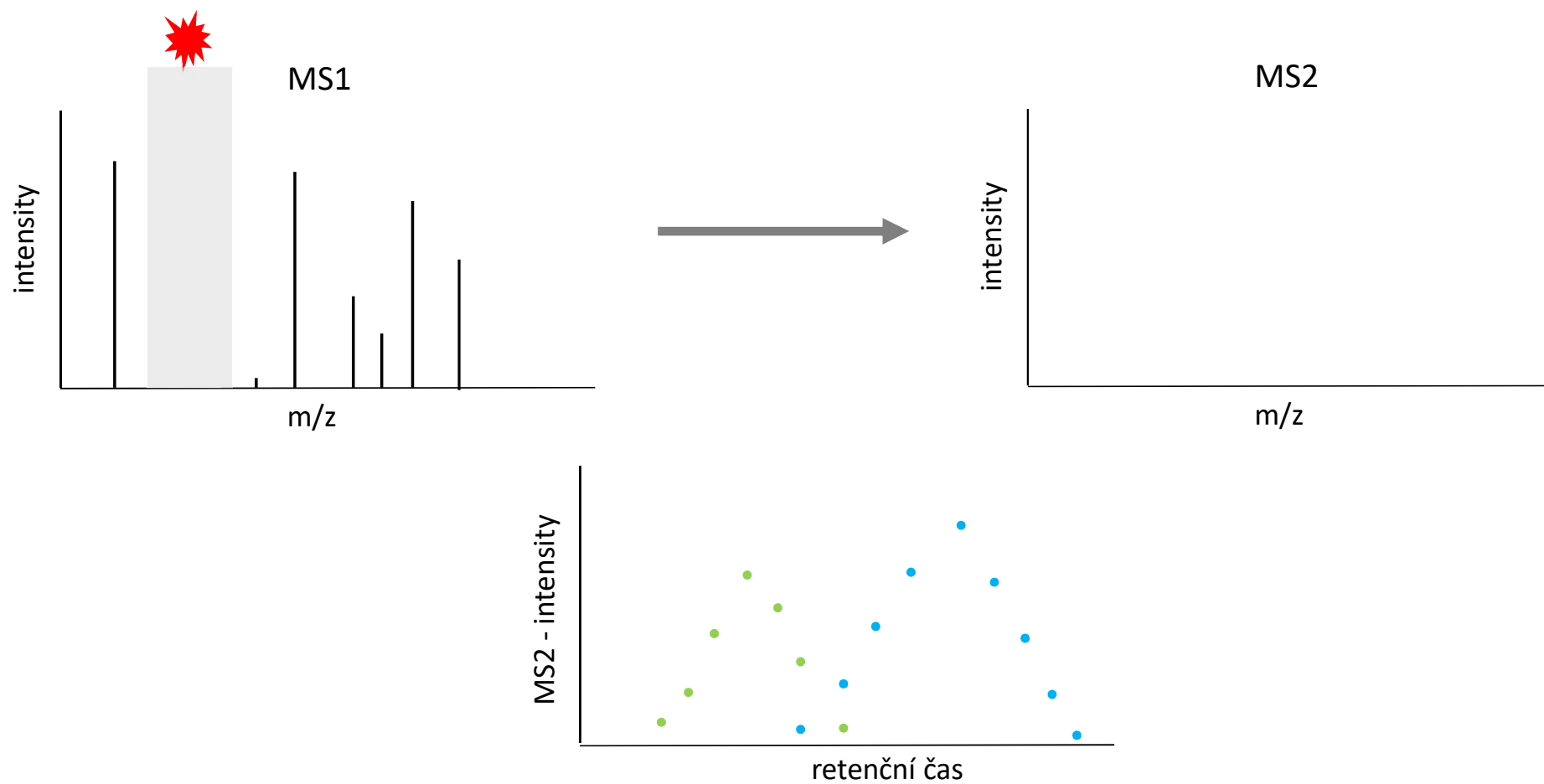
# Kde se u DIA bere kvantitativní informace?



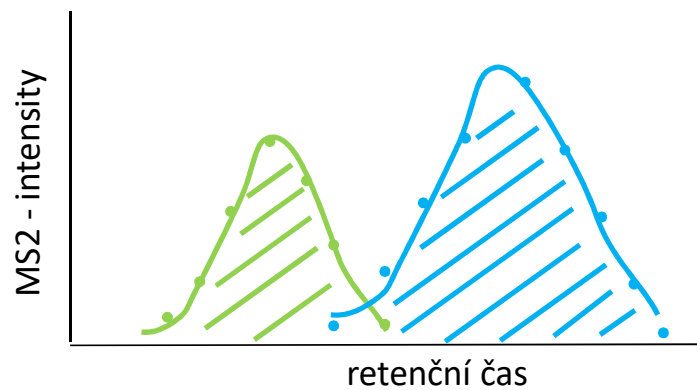
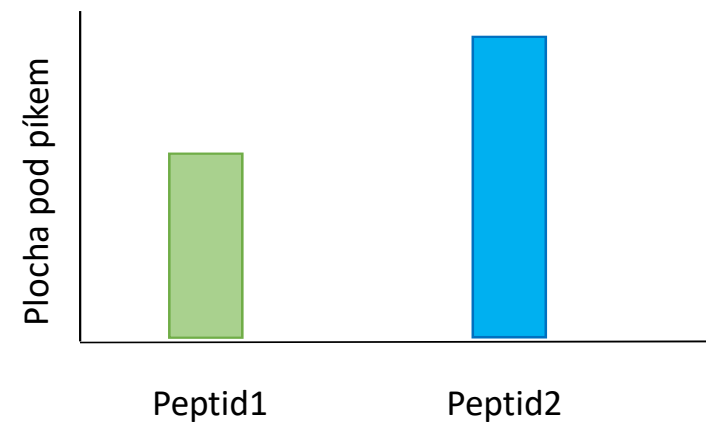
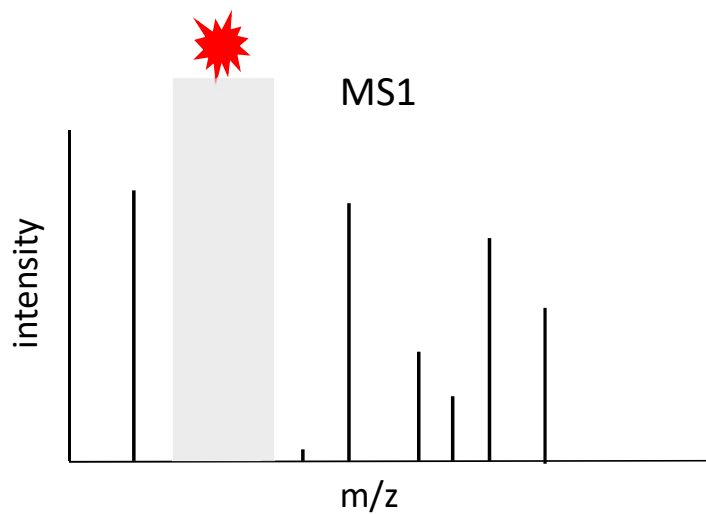
# Kde se u DIA bere kvantitativní informace?



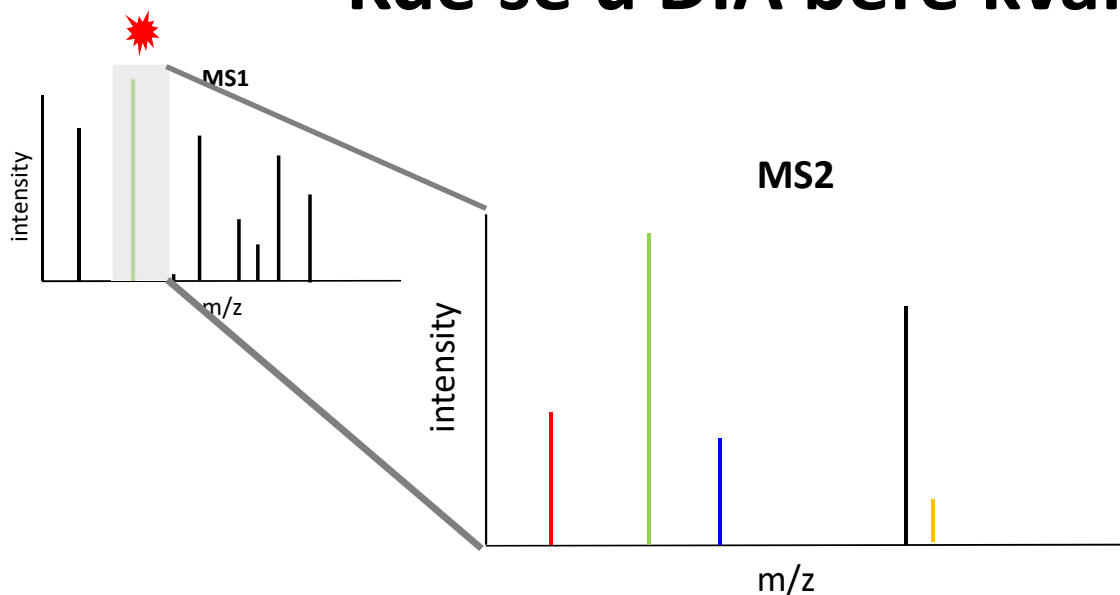
# Kde se u DIA bere kvantitativní informace?



# Kde se u DIA bere kvantitativní informace?

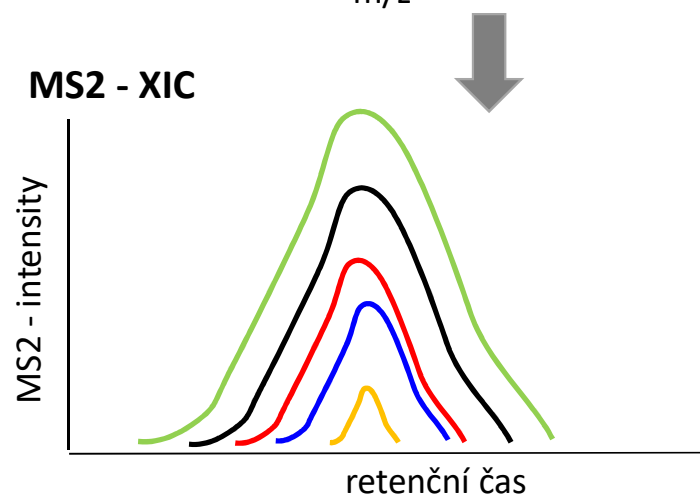


# Kde se u DIA bere kvantitativní informace?



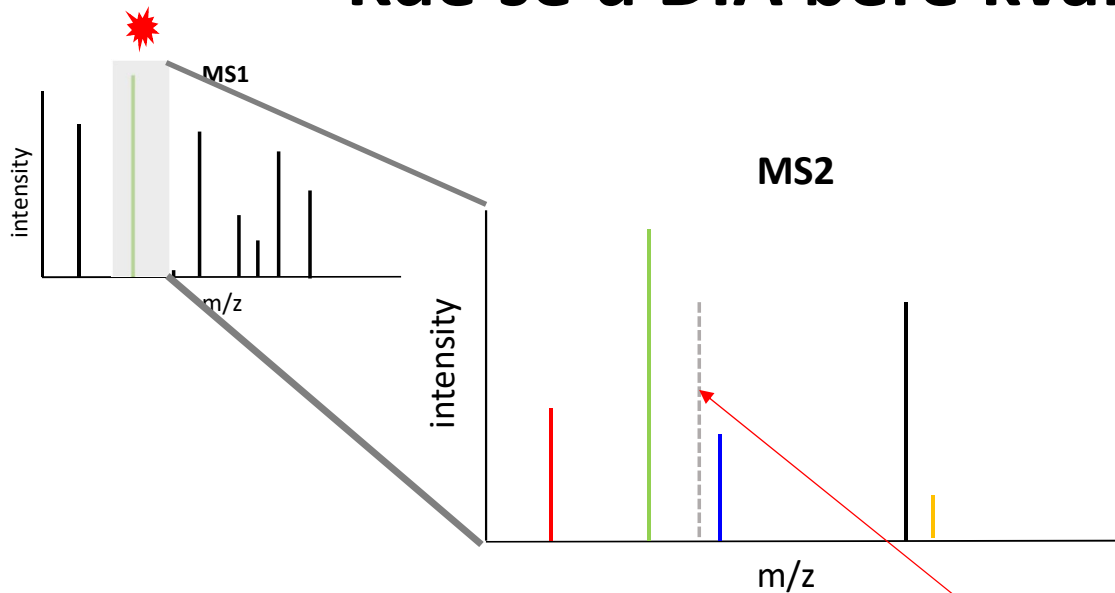
## Potvrzení identity prekurzoru:

- Poměry ploch extrahovaných chromatogramů jednotlivých fragmentů odpovídají poměrům intenzit píků ve fragmentačním spektru
- Překrývající se eluční profily



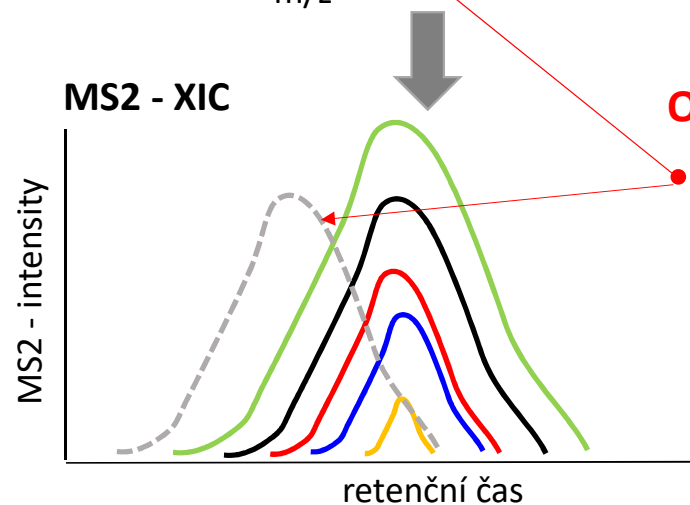


# Kde se u DIA bere kvantitativní informace?



## Potvrzení identity prekurzoru:

- Poměry ploch extrahovaných chromatogramů jednotlivých fragmentů odpovídají poměrům intenzit píků ve fragmentačním spektru
- Překrývající se eluční profily



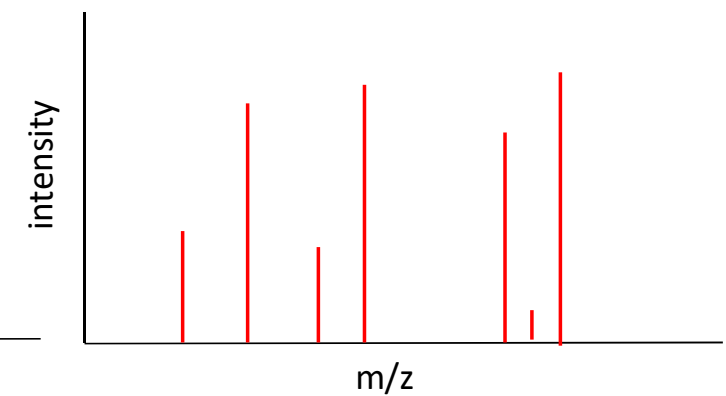
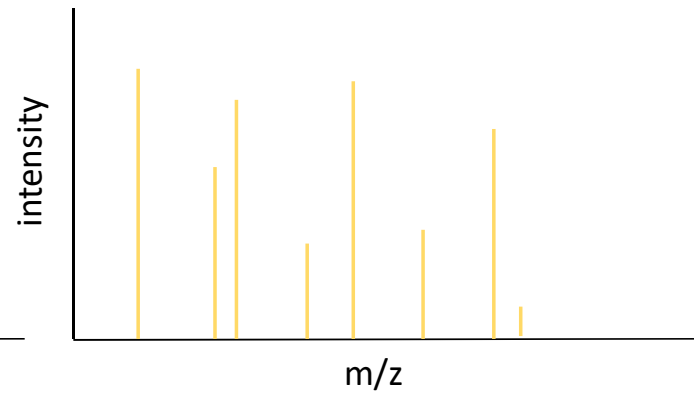
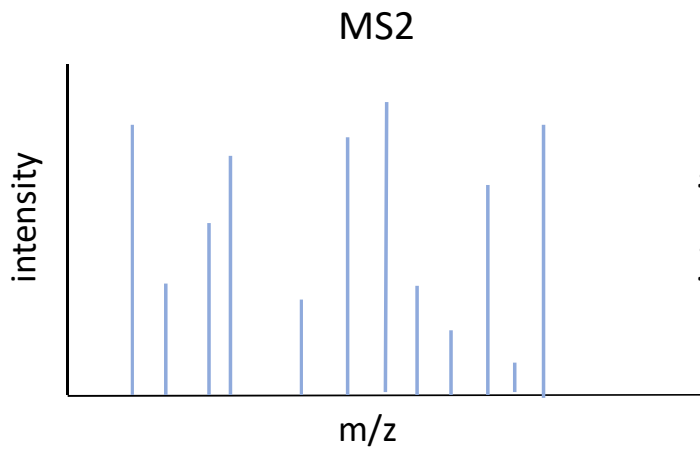
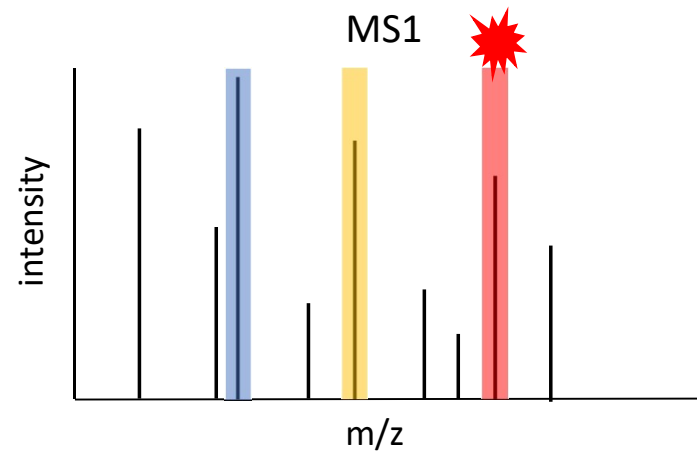
## Odhalení interferujících signálů

- Interferující signál nekopíruje přesně eluční profil

# Kde je hranice mezi DDA a DIA?

**DDA**

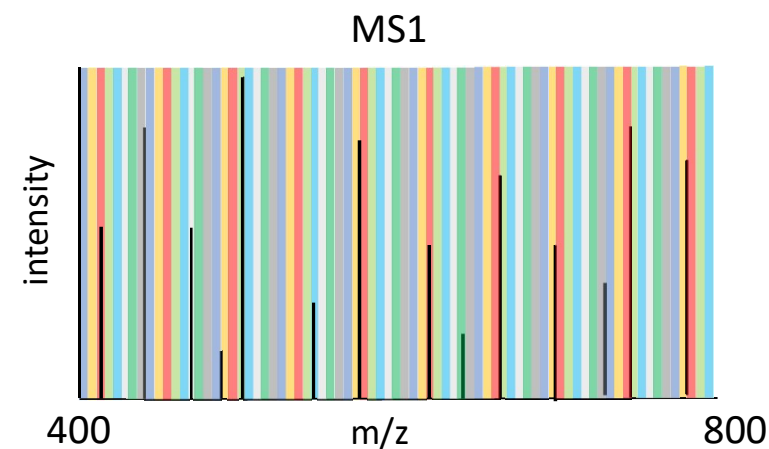
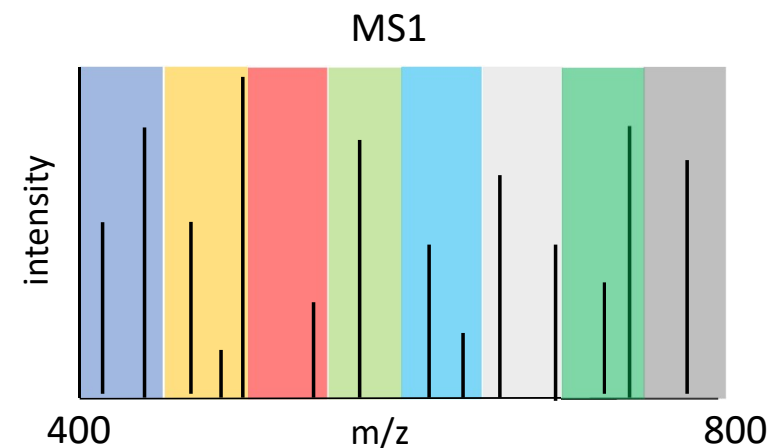
Izolační okno obvykle 0,7 – 2 Da



# DIA

## Analyzovaný hmotnostní rozsah 400 – 800 m/z

- celková šířka = 400
- **Šířka okna 50 Da** – potřeba 8 oken
  - Proměření 8 oken zabere relativně krátký čas
  - MS2 spektra budou extrémně komplexní, což bude snižovat množství identifikací
- **Šířka okna 2 Da** – potřeba 200 oken
  - Spektra budou málo komplexní – vyšší počet identifikací
  - Proměření 200 oken zabere hodně času – dlouhý cycle time



### Úzké izolační okno:

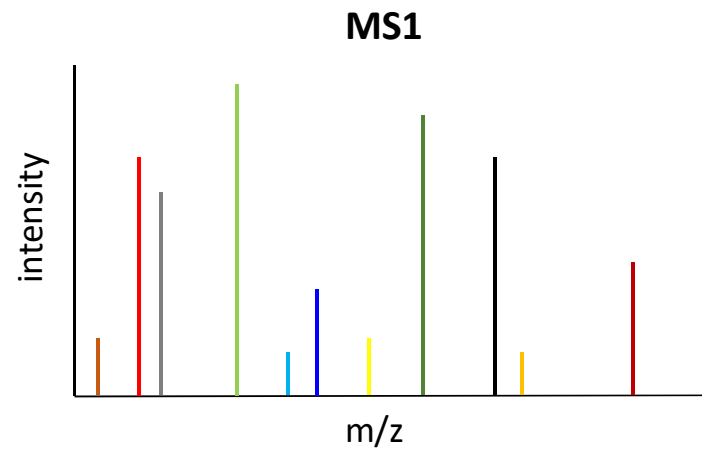
- Méně koizolovaných peptidů
- Jednodušší spektra
- Více oken pokrývajících daný hmotnostní rozsah

### Široké izolační okno:

- Méně oken pro pokrytí daného hmotnostního rozsahu
- Více koizolovaných peptidů
- Komplexnější spektra

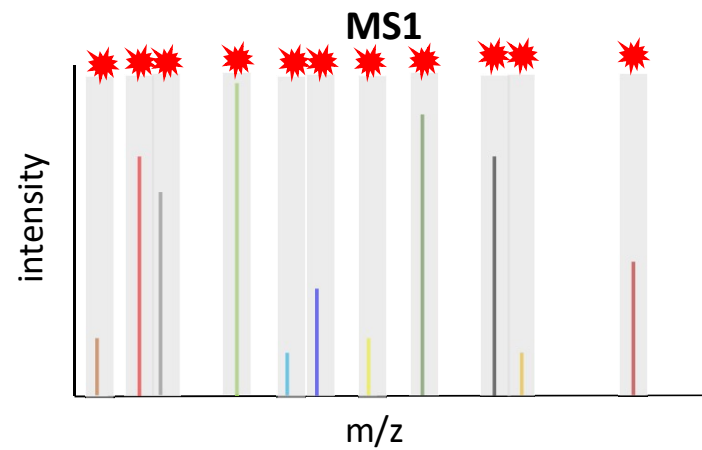
# Kde je hranice mezi DDA a DIA? Narrow window DIA

**DDA**



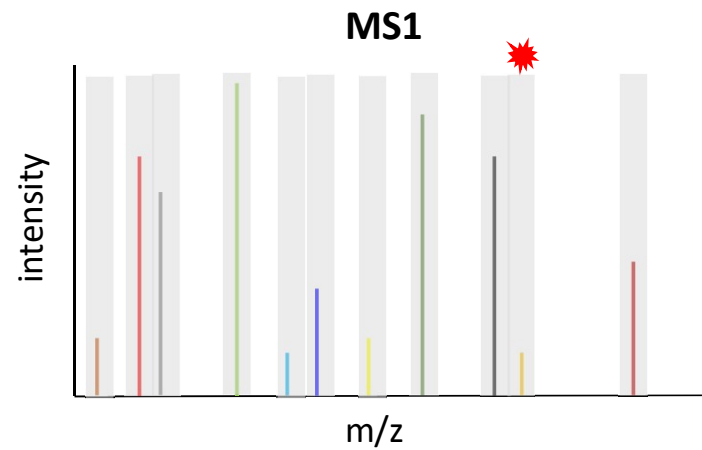
# Kde je hranice mezi DDA a DIA? Narrow window DIA

**DDA**

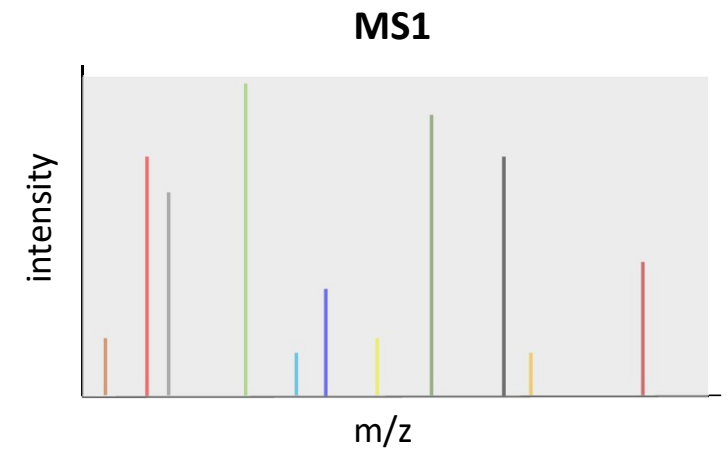


# Kde je hranice mezi DDA a DIA? Narrow window DIA

**DDA**



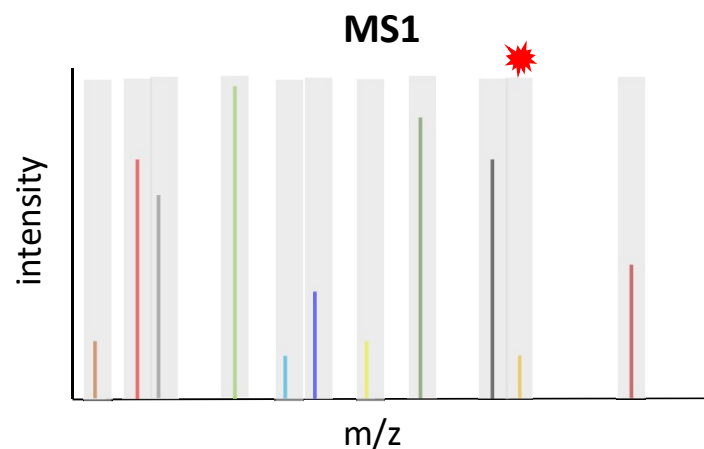
**DIA**



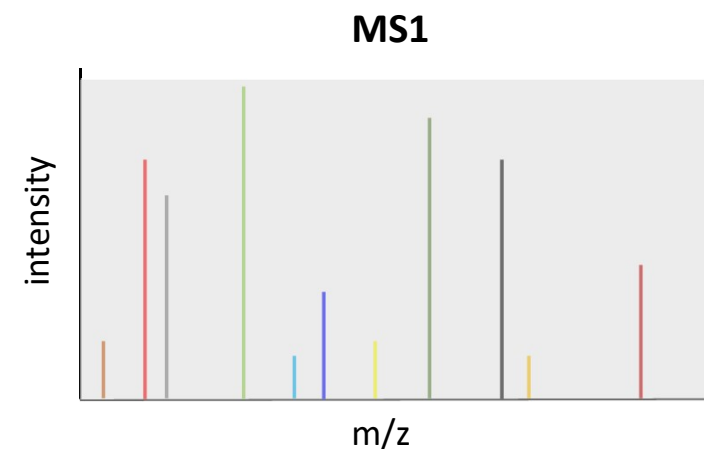
# Kde je hranice mezi DDA a DIA?

## Narrow window DIA

DDA



DIA

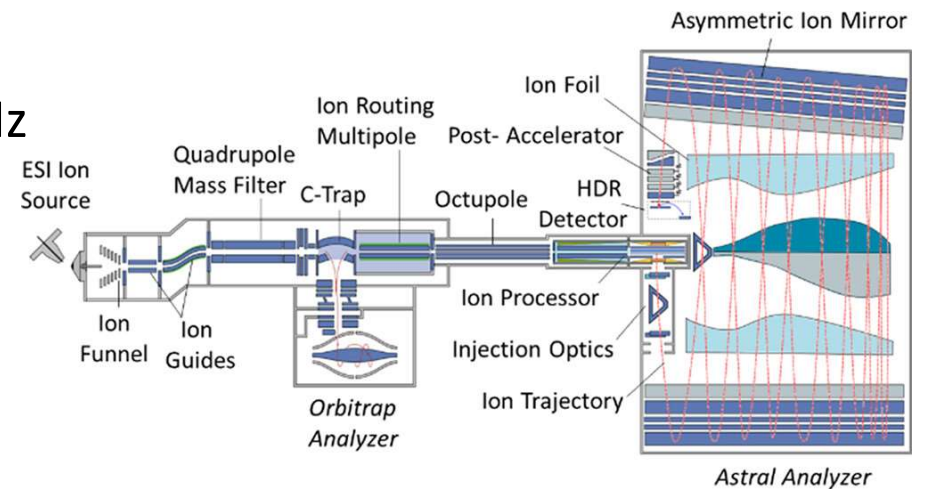
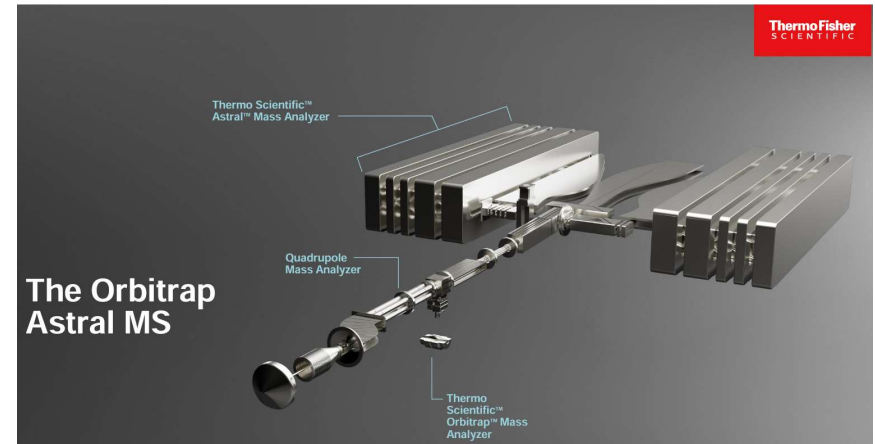


**DIA s dostatečně úzkým izolačním oknem (2-3 Da) se v principu podobá DDA = Narrow Window DIA**

# Narrow window DIA

## Hmotnostní analyzátor Astral

- Detekce MS/MS spekter
- Na principu TOF analyzátoru
- Dlouhá dráha letu – 30 m – vysoké rozlišení
- Vysoká skenovací frekvence – reálně kolem 200 Hz



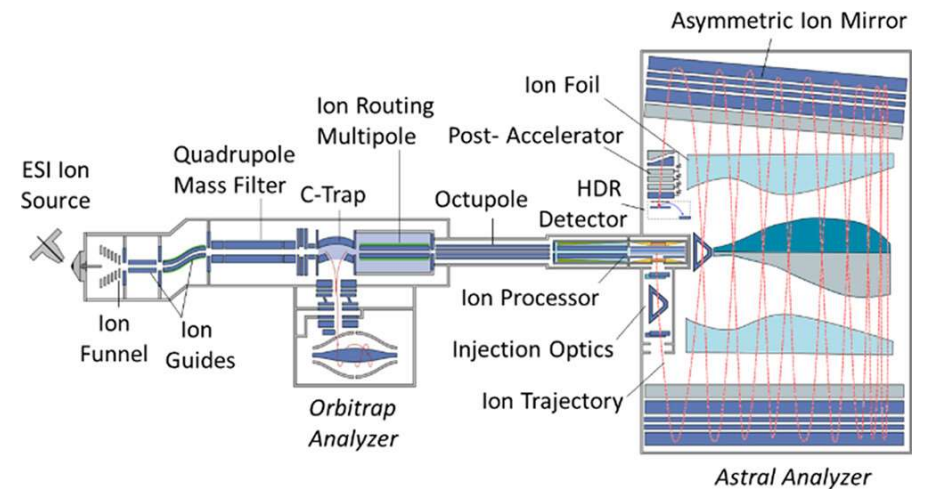
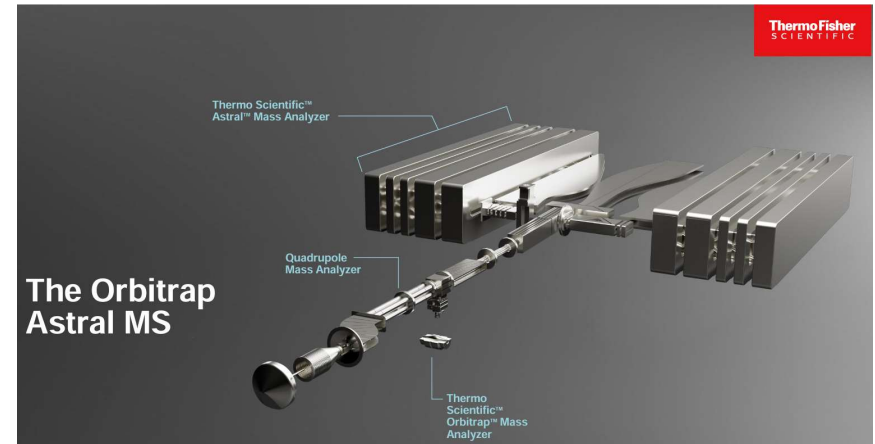
Stewart et al., Parallelized Acquisition of Orbitrap and Astral Analyzers Enables High-Throughput Quantitative Analysis, Analytical Chemistry Vol 95/Issue 42, 2023



# Narrow window DIA

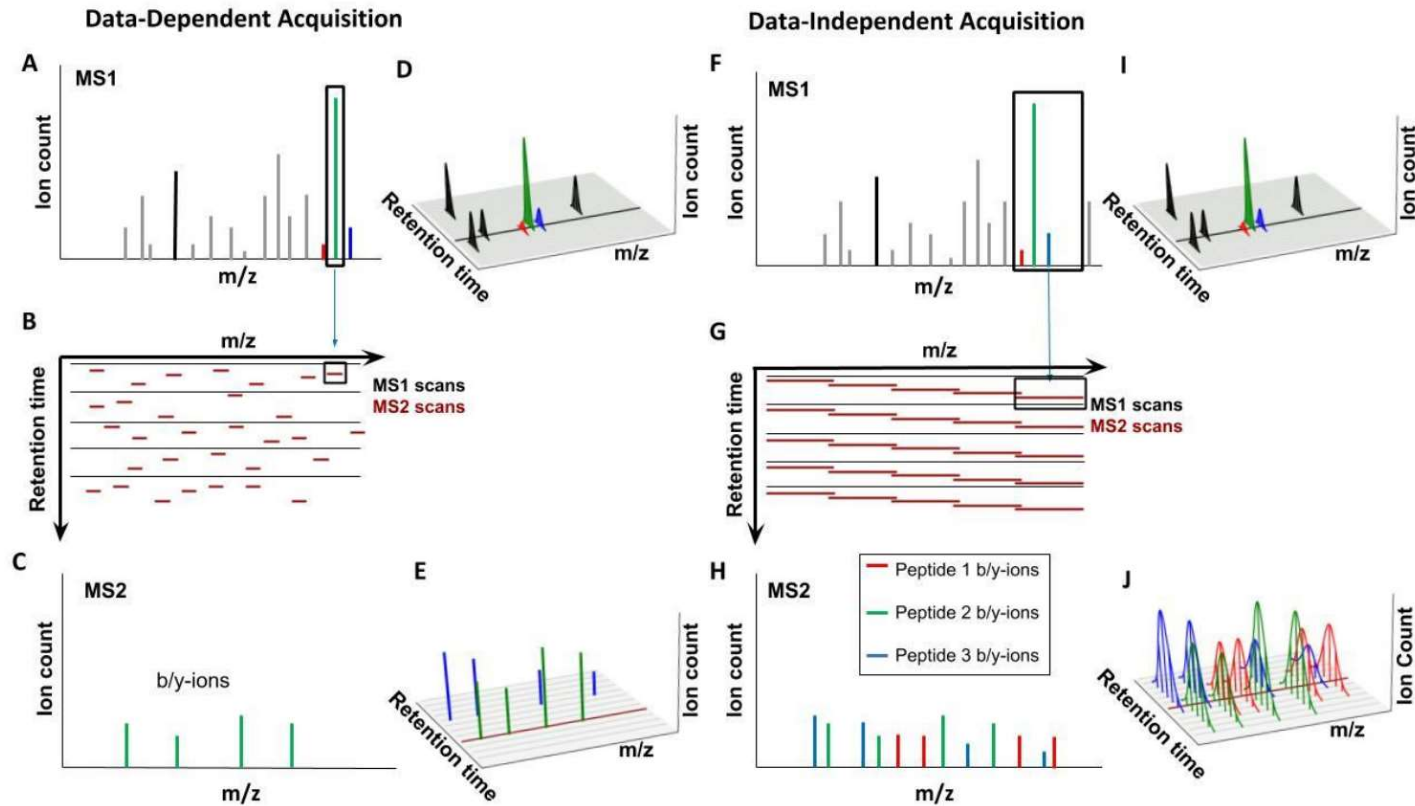
## Hmotnostní analyzátor Astral

- Vysoká skenovací frekvence – reálně kolem 200 Hz
- Umožňuje analyzovat více oken při zachování přijatelné délky cyklu



Stewart et al., Parallelized Acquisition of Orbitrap and Astral Analyzers Enables High-Throughput Quantitative Analysis, Analytical Chemistry Vol 95/Issue 42, 2023

# DDA vs. DIA - shrnutí



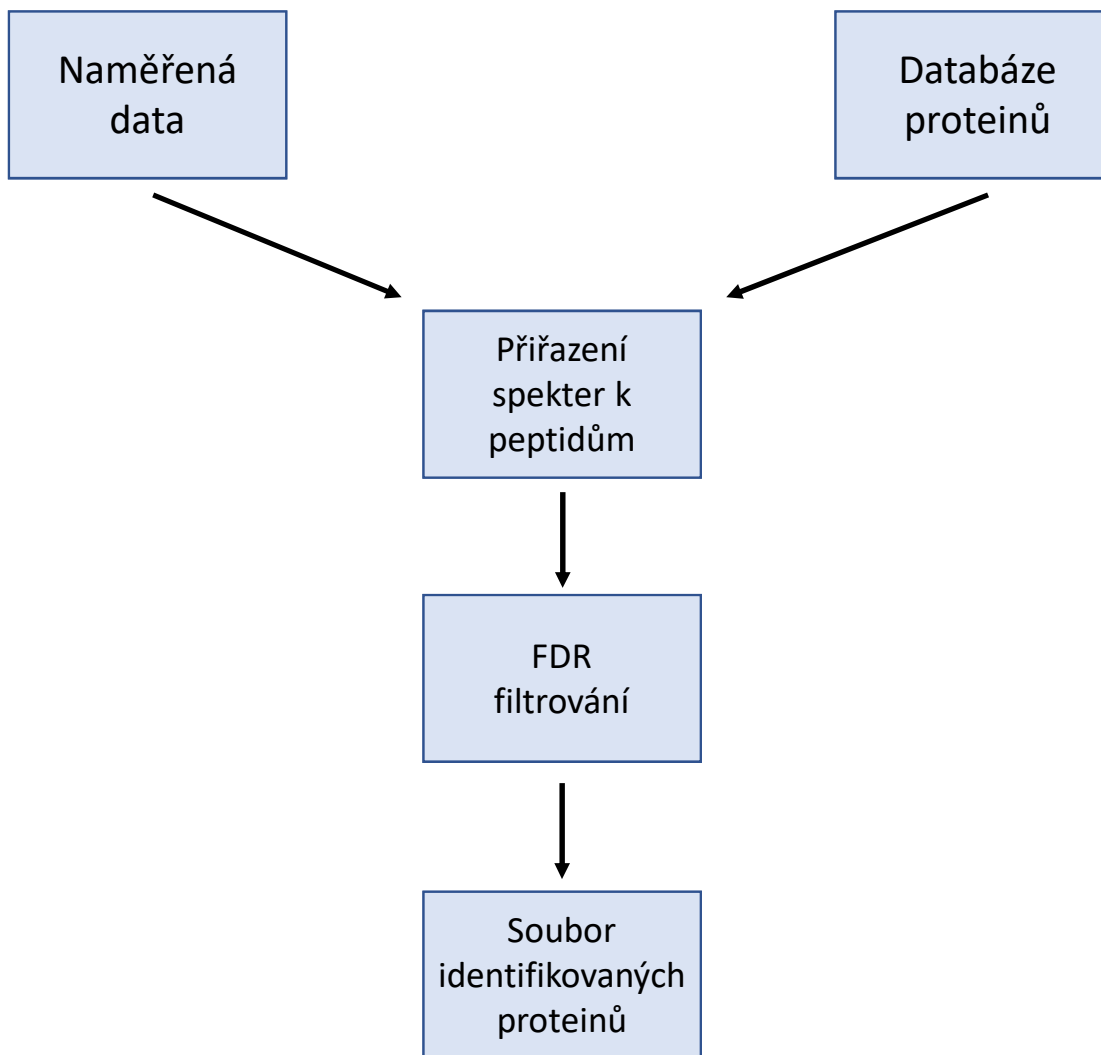
Pappireddi et al, A Review on Quantitative Multiplexed Proteomics, ChemBioChem 2019

## DDA vs. DIA - shrnutí

|                                       | DDA     | DIA      |
|---------------------------------------|---------|----------|
| Počet identifikací z lidského lyzátu: | 5 000   | 8 000    |
| Kvantifikace na základě:              | MS1     | MS2      |
| Potřeba spektrální knihovny:          | Ne      | Ano/Ne   |
| Počet missing values:                 | 30-40 % | 15 %     |
| Přesnost kvantifikace (CV):           | 25 %    | 5 – 10 % |

# **Od spekter k souboru proteinů**

**Jak z 250 000 spekter získat 5000 proteinů?**



# Databáze

- Vedle naměřených dat **potřebujeme** k vyhodnocení databázi proteinů – **neprobíhá de novo sekvenace**
- FASTA formát: hlavička strukturovaná tak, aby se z ní daly automaticky extrahovat informace

```
>sp|P61206|ARF3_RAT ADP-ribosylation factor 3 OS=Rattus norvegicus OX=10116 GN=Arf3 PE=2 SV=2
MGNIFGNLLKSLIGKKEMRILMVGLDAAGKTTILYKLLKLGIVTTIPTIGFNVETVEYKNISFTVWDVGGQDKIRPLWRHYFQ
NTQGLIFVVDSNDRERVNEAREELMRMLAEDELDAVLLVFANKQDLPNAMNAAEITDKLGLHSLRHRNWYIQATCATSG
DGLYEGLDWLANQLKNKK
```

# Databáze

- Vedle naměřených dat **potřebujeme** k vyhodnocení databázi proteinů – **neprobíhá de novo sekvenace**
- FASTA formát: hlavička strukturovaná tak, aby se z ní daly automaticky extrahovat informace

```
>sp|P61206|ARF3_RAT ADP-ribosylation factor 3 OS=Rattus norvegicus OX=10116 GN=Arf3 PE=2 SV=2
MGNIFGNLLKSLIGKKEMRILMVGLDAAGKTTILYKLLKLGIVTTIPTIGFNVETVEYKNISFTVWDVGGQDKIRPLWRHYFQ
NTQGLIFVVDSNDRERVNEAREELMRMLAEDELDAVLLVFANKQDLPNAMNAAEITDKLGLHSLRHRNWYIQATCATSG
DGLYEGLDWLANQLKNNK
```

# Databáze

- Vedle naměřených dat **potřebujeme** k vyhodnocení databázi proteinů – **neprobíhá de novo sekvenace**
- FASTA formát: hlavička strukturovaná tak, aby se z ní daly automaticky extrahovat informace

```
>sp|P61206|ARF3_RAT ADP-ribosylation factor 3 OS=Rattus norvegicus OX=10116 GN=Arf3 PE=2 SV=2
```

- Obsah databáze by vždy měl odpovídat pravděpodobnému obsahu vzorku = **vzorky lidských buněčných linií prohledávat proti databázi lidských proteinů, ne proti databázi všech savců nebo eukaryot**
- Použití příliš velké databáze znehodnocuje výsledky
- **I špatná databáze dá vždy nějaký výsledek**



# Výsledek vždy závisí na použité databázi!

Vyhodnocení vzorku z lidských HeLa buněk proti dvěma různým databázím:

| Databáze                        | Human | Mus musculus |
|---------------------------------|-------|--------------|
| Počet identifikovaných proteinů | 5 352 | 3 634        |

**Když použijeme databázi myši k vyhodnocení dat z lidského vzorku, dostaneme přes 3600 identifikací!**

# Databáze

- Zdroje databází:
  - Uniprot.org

**UniProt** BLAST Align Peptide search ID mapping SPARQL **Proteomes** Advanced | List Search

## Proteomes · Homo sapiens (Human)

### Overview

|   |  |
|---|--|
| <b>Status</b> Reference proteome  | <b>Genome representation</b> Full                |
| <b>Number of entries</b> : 82,678   | <b>Completeness (CPD)</b> : Outlier (high value) |
| <b>Gene count</b> : 20,594 <a href="#">Download one protein sequence per gene (FASTA)</a> | <b>BUSCO</b> :                                   |
| <b>Proteome ID</b> : UP000005640  | n:13780 · primates_odb10                         |
| <b>Taxonomy</b> Homo sapiens (Human)  | C:99.5% (S:37.4% D:62.1%) F:0% M:0.5%            |
| <b>Genome assembly and annotation</b> : GCA_000001405.28 from Ensembl                     |  |

Homo sapiens (Homo sapiens sapiens) or modern humans are the only living species of the evolutionary branch of great apes known as hominids. Divergence of early humans from chimpanzees and gorillas is estimated to have occurred between 4 and 8 million years ago. The genus Homo (Homo habilis) appeared in Africa around 2.3 million years ago and shows the first signs of stone tool usage. The exact lineage of Homo species ie. H. habilis/H. ergaster to H. erectus to H. rhodesiensis/H.heidelbergensis to H. sapiens is still hotly disputed. However, continuing evolution and in particular larger brain size and complexity culminates in Homo sapiens. The first anatomically modern humans appear in the fossil record around 200,000 years ago. Modern humans migrated across the globe essentially as hunter-gatherers until around 12,000 years ago when the practice of agriculture and animal domestication enabled large populations to grow leading to the development of civilizations.

- NCBI - National Center for Biotechnology Information

**NIH** National Library of Medicine  
National Center for Biotechnology Information [Log in](#)

Taxonomy   [Search](#)  
[Create alert](#) [Limits](#) [Advanced](#) [Help](#)

Display Settings:

[Homo sapiens](#)  
[\(human\)](#), [species](#), [primates](#)  
[Nucleotide](#) [Protein](#)

Send to:   
 Nucleotide  
 Protein  
 Assembly  
 BioProject

# Databáze

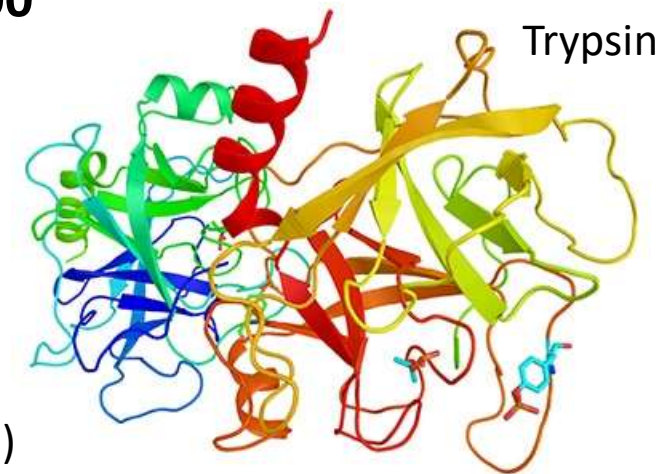
- Parametry, které je nutné definovat spolu s použitou databází:
  - Použitá proteáza (trypsin)
  - Modifikace

# Databáze – proteasy

- Proč je nutné uvést proteázu použitou k digesci?
  - Počet všech teoretický peptidů o délce 8 – 14 AMK generovaných na základě databáze lidských proteinů je asi **70 000 000**
  - Počet tryptických peptidů o stejných parametrech je asi **900 000**

- Miscleavage

- Proteasa dané štěpné místo někdy vynechá (sterická blokáce, přítomnost modifikací)
- Specifikujeme kolik vynechání (miscleavage) chceme tolerovat (obvykle 1-2)



# Databáze - modifikace

- Společně s databází proteinů a použitou proteasou musíme specifikovat modifikace, které lze ve vzorku očekávat
- **Modifikace obecně se vyskytující ve všech vzorcích:**
  - Modifikace cysteinu – vznik modifikací volných –SH během přípravy vzorku. Záleží na alkylačním činidle (IAA, CAA – carbamidomethyl)
  - Oxidace methioninu
  - Acetylace N-konce
- **Modifikace specifické pro vzorek:**
  - Fosforylace, acetylace, methylace, hydroxylace, ubikvitinylace (GlyGly)
- **Fixní** – počítá se, že modifikace je na dané aminokyselině přítomna vždy (alkylace Cys)
- **Variabilní** – modifikace se na dané aminokyselině vyskytovat může a nemusí
- Software následně generuje varianty peptidů obsahující všechny možné kombinace zadaných modifikací

# Databáze - modifikace

## Peptid LVSCAGTFK:

1 fixní

LVSC(cam)AGTFK

**1 sekvence**

1 fixní + 1 variabilní

LVSC(cam)AGTFK

LVS(ph) C(cam)AGTFK

LVSC(cam)AGT(ph)FK

LVS(ph) C(cam)AGT(ph)FK

**4 sekvence**

1 fixní + 2 variabilní

LVSC(cam)AGTFK

LVS(ph) C(cam)AGTFK

LVSC(cam)AGT(ph)FK

LVS(ph) C(cam)AGT(ph)FK

LVS(ph) C(cam)AGTFK(me)

LVSC(cam)AGT(ph)FK(me)

LVS(ph) C(cam)AGT(ph)FK(me)

LVSC(cam)AGTFK(me)

**8 sekvencí**

# Databáze - modifikace

## Peptid LVSCAGTFK:

1 fixní

LVSC(cam)AGTFK

1 fixní + 1 variabilní

LVSC(cam)AGTFK

LVS(ph) C(cam)AGTFK

LVSC(cam)AGT(ph)FK

LVS(ph) C(cam)AGT(ph)FK

1 fixní + 2 variabilní

LVSC(cam)AGTFK

LVS(ph) C(cam)AGTFK

LVSC(cam)AGT(ph)FK

LVS(ph) C(cam)AGT(ph)FK

LVS(ph) C(cam)AGTFK(me)

LVSC(cam)AGT(ph)FK(me)

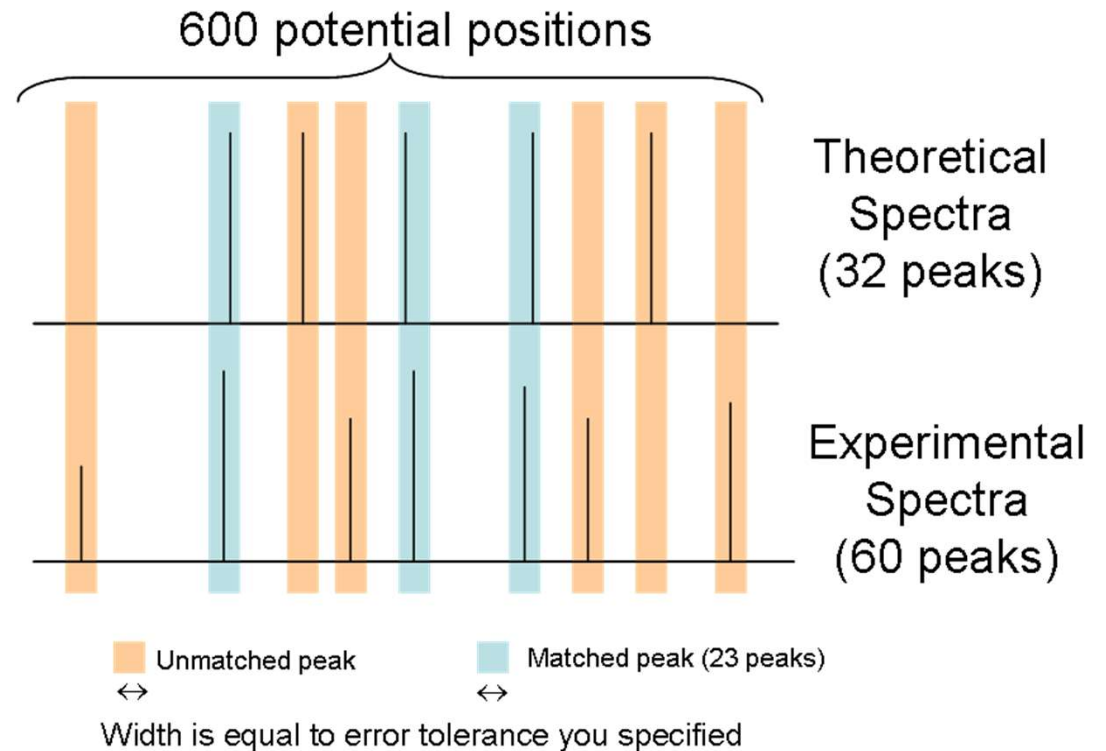
LVS(ph) C(cam)AGT(ph)FK(me)

LVSC(cam)AGTFK(me)

- Variabilní modifikace výrazně expandují databázi a navyšují tzv. search space a tím i výpočetní náročnost. Použití modifikací, které se ve vzorku nevyskytují nebo se v něm vyskytují jen sporadicky, znehodnocuje výsledek.
- **Vždy by měly být specifikovány jen ty modifikace, které se ve vzorku opravdu a ve výrazné míře vyskytují.**

# Vyhodnocení spekter a skórování

1. Na základě zadané databáze vygenerujeme všechny teoretické tryptické peptidy
2. Generování teoretických spekter
3. Hledáme počet shod mezi teoretickým a naměřeným MS2 spektrem
4. Jaká je pravděpodobnost, že nalezená shoda je náhodná?
5. Vyčíslíme pomocí **skóre** (obvykle  $-\log$ )

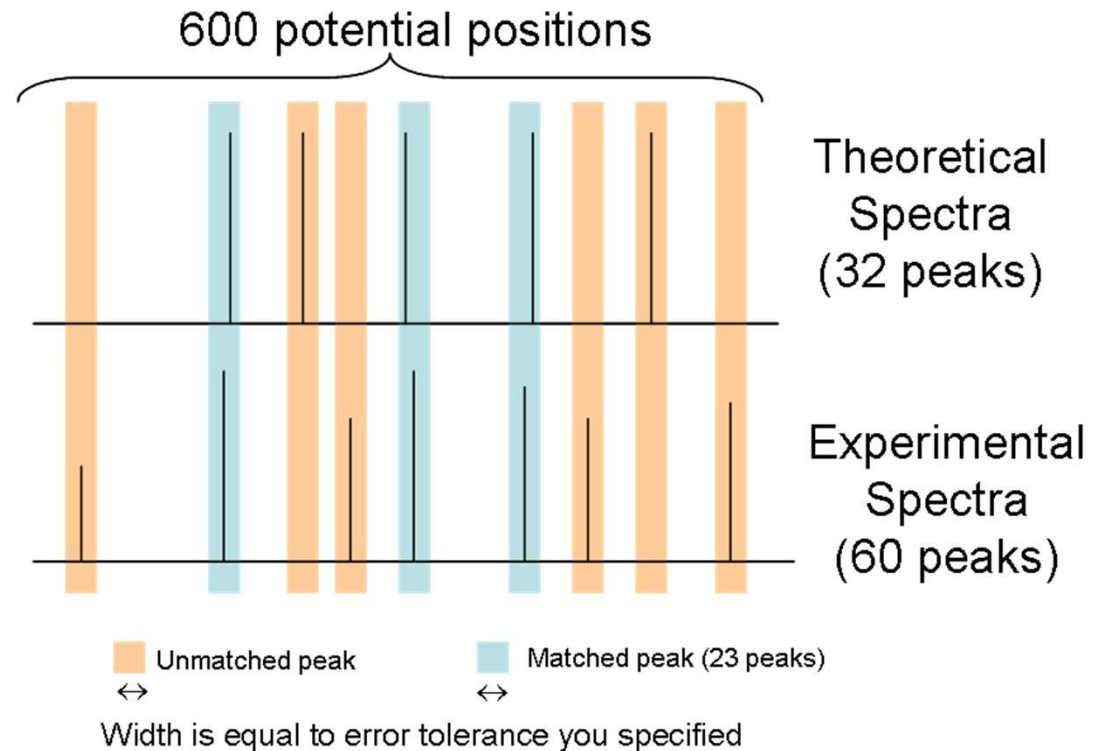


$$s(q, \text{loss}) = -10 \log_{10} \sum_{j=k}^n \left[ \binom{n}{j} \left( \frac{q}{100} \right)^j \left( 1 - \frac{q}{100} \right)^{n-j} \right]$$



# Vyhodnocení spekter a skórování

Skóre = pravděpodobnost, že shoda naměřených a teoretických fragmentů je náhodná

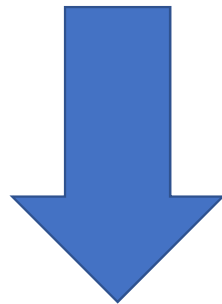


$$s(q, \text{loss}) = -10 \log_{10} \sum_{j=k}^n \left[ \binom{n}{j} \left( \frac{q}{100} \right)^j \left( 1 - \frac{q}{100} \right)^{n-j} \right]$$

# False Discovery Rate

Samotné skóre nestačí pro zajištění dostatečné spolehlivosti identifikací.

Při porovnávání vysokého množství spekter v rádech statisíců dostáváme určitou frakci falešně pozitivních přiřazení.



**Filtrování pomocí odhadu False Discovery Rate (FDR)**

# False Discovery Rate

$$\text{FDR} = \frac{D}{D+T}$$

D – počet hitů z decoy databáze

T – počet hitů z target databáze

Pokud nastavíme treshold FDR na 1%, dostaneme dataset obsahující 99% správných identifikací a 1% falešně pozitivních.

# FDR – target-decoy přístup

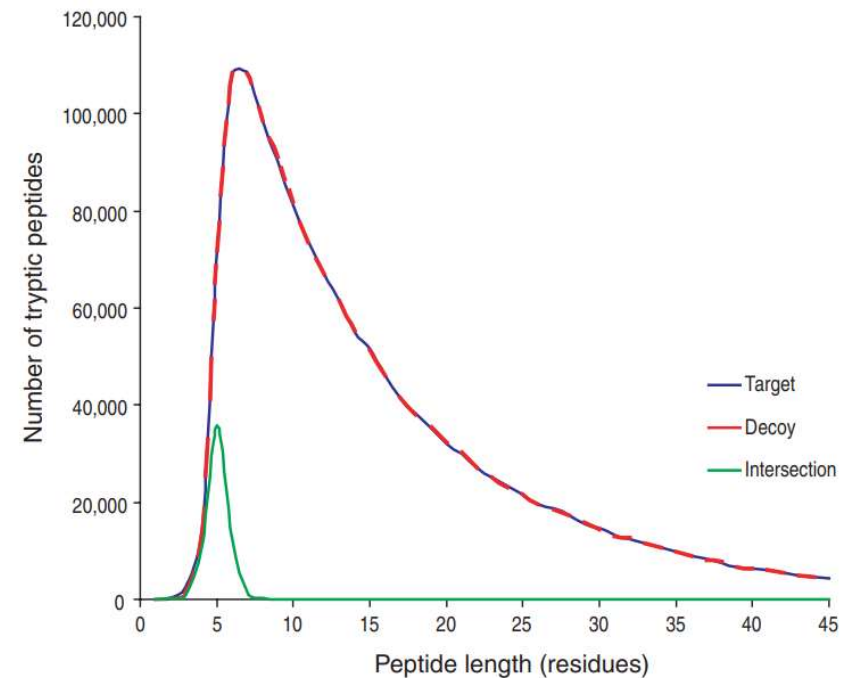
1. Vygenerujeme databázi neexistujících proteinů (návnad) otočením všech sekvencí v originální (target) databázi = **decoy** databáze

```
>sp|P01308|INS_HUMAN Insulin OS=Homo sapiens OX=9606 GN=INS PE=1 SV=1  
MALWMRLLPLLALLALWGPDPAAAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGGQVELGGGPGAGSLQPL  
ALEGSLQKRGIVEQCCTSICSLYQLENYCN
```

```
>REV:sp|P01308|INS_HUMAN Insulin OS=Homo sapiens OX=9606 GN=INS PE=1 SV=1  
NCYNELQYLSCISTCCQEVIGRKQLSGELALPQLSGAGPGGGLEVGQVQLDEAERRTKPTYFFGREGCVLYLAEVLHSGCL  
HQNVFAAAPDPGWLALLALLPLLRMWLAM
```

## Předpoklady decoy databáze:

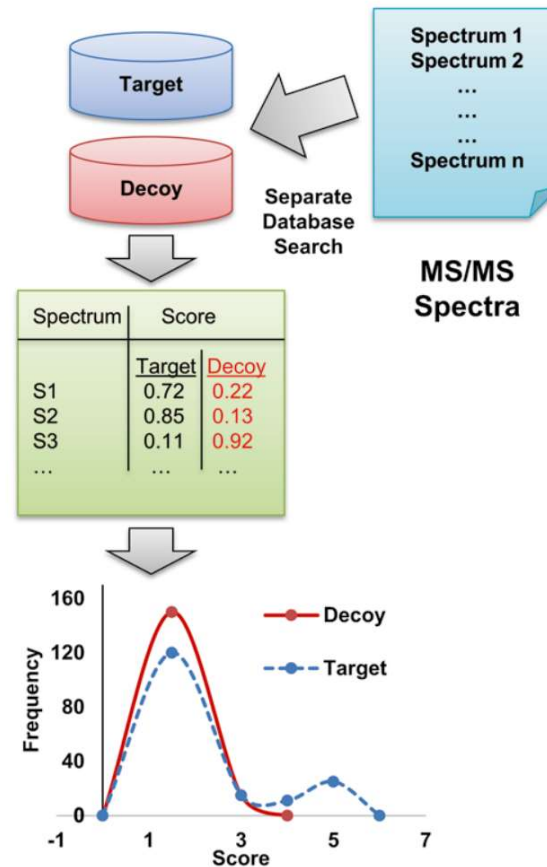
- Decoy databáze obsahuje identické množství proteinů jako target databáze
- Distribuce délek peptidů je identická u obou databází
- Sekvence peptidů z target a decoy databáze se nepřekrývají (platí od zhruba 7 - 8 AMK)



Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry, Nature methods 2007

# FDR – target-decoy přístup

1. Vygenerujeme databázi neexistujících proteinů (návnad) otočením všech sekvencí v databázi = **decoy databáze**
2. Naměřená spektra porovnáme jak proti smysluplné databázi (target), tak proti decoy databázi
3. Každé shodě je přiřazeno skóre.
4. V datasetu ponecháme jen identifikace odpovídající určité FDR (obvykle 1%)

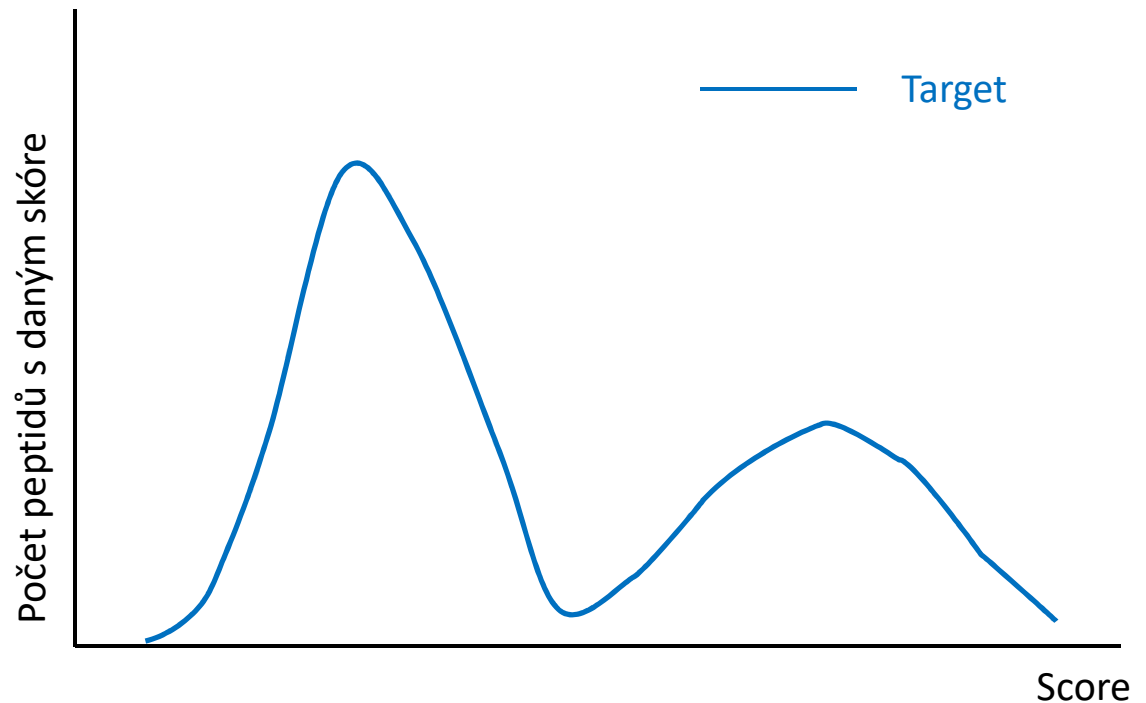


$$FDR = \frac{D}{D+T}$$

D – počet hitů z decoy databáze nad tresholdem  
T – počet hitů u target databáze

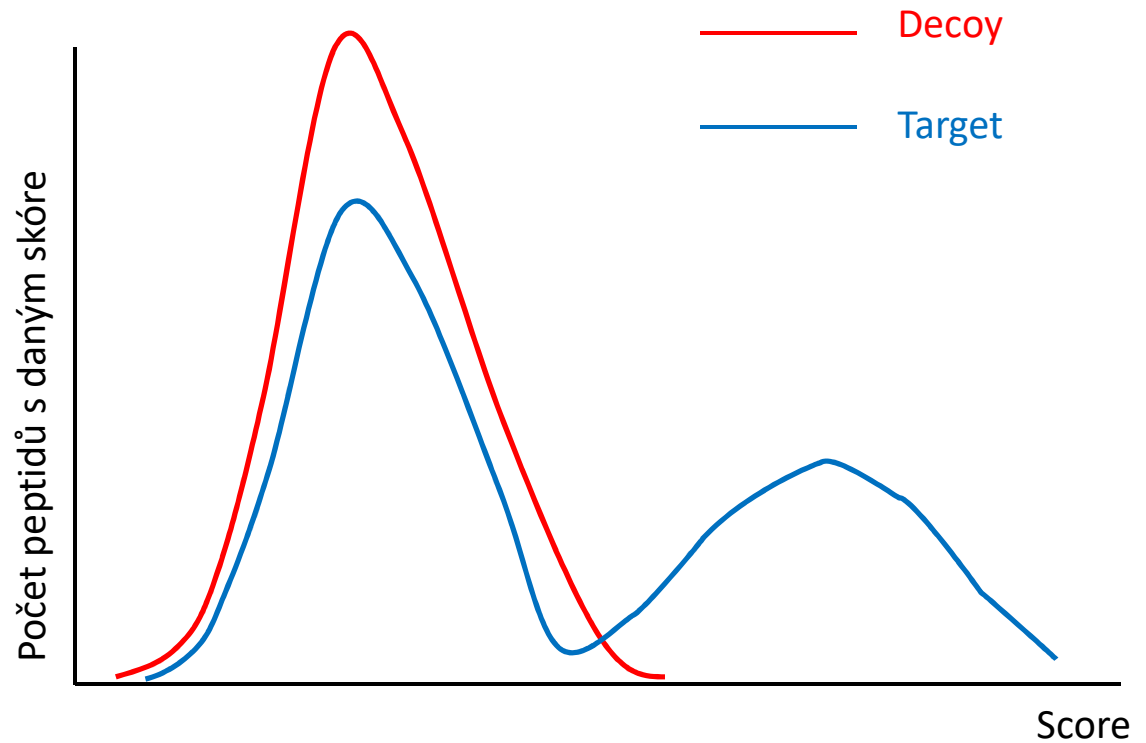
# FDR – target-decoy přístup

1. Vygenerujeme databázi neexistujících proteinů (návnad) otočením všech sekvencí v databázi = **decoy databáze**
2. Naměřená spektra porovnáme jak proti smysluplné databázi (target), tak proti decoy databázi
3. Každé shodě je přiřazeno skóre.
4. V datasetu ponecháme jen identifikace odpovídající určité FDR (obvykle 1%)



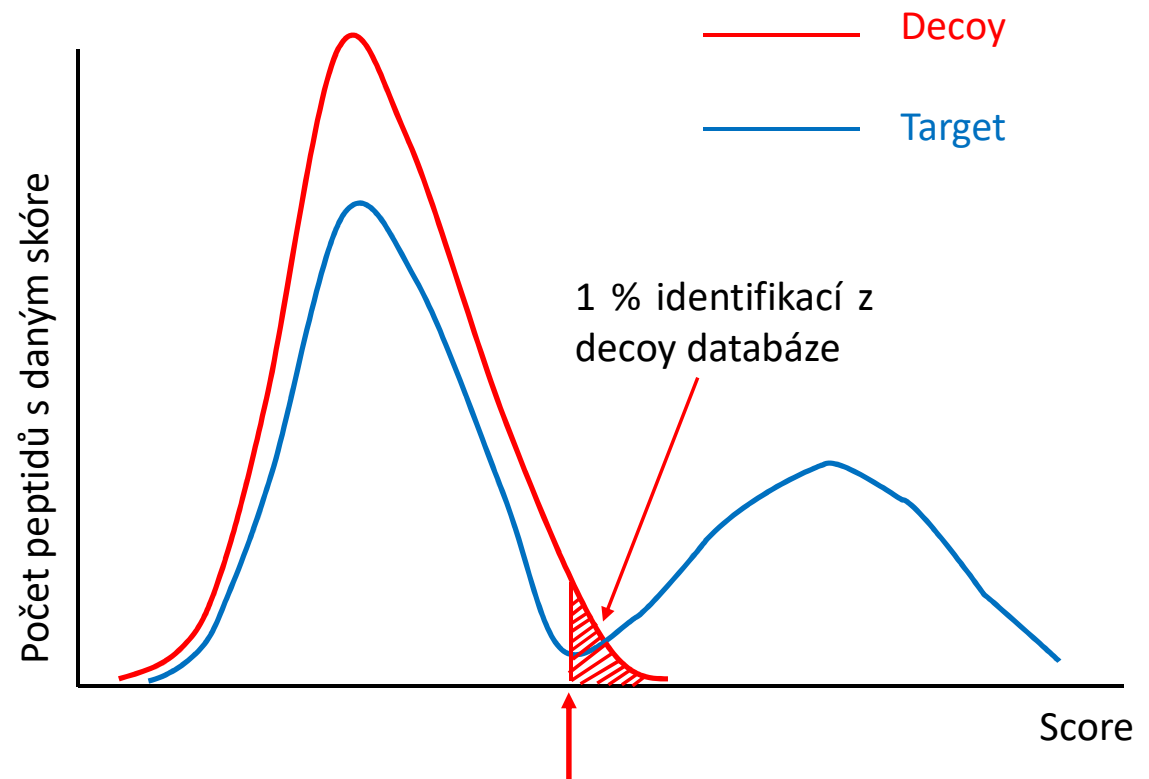
# FDR – target-decoy přístup

1. Vygenerujeme databázi neexistujících proteinů (návnad) otočením všech sekvencí v databázi = **decoy databáze**
2. Naměřená spektra porovnáme jak proti smysluplné databázi (target), tak proti decoy databázi
3. Každé shodě je přiřazeno skóre.
4. V datasetu ponecháme jen identifikace odpovídající určité FDR (obvykle 1%)



# FDR – target-decoy přístup

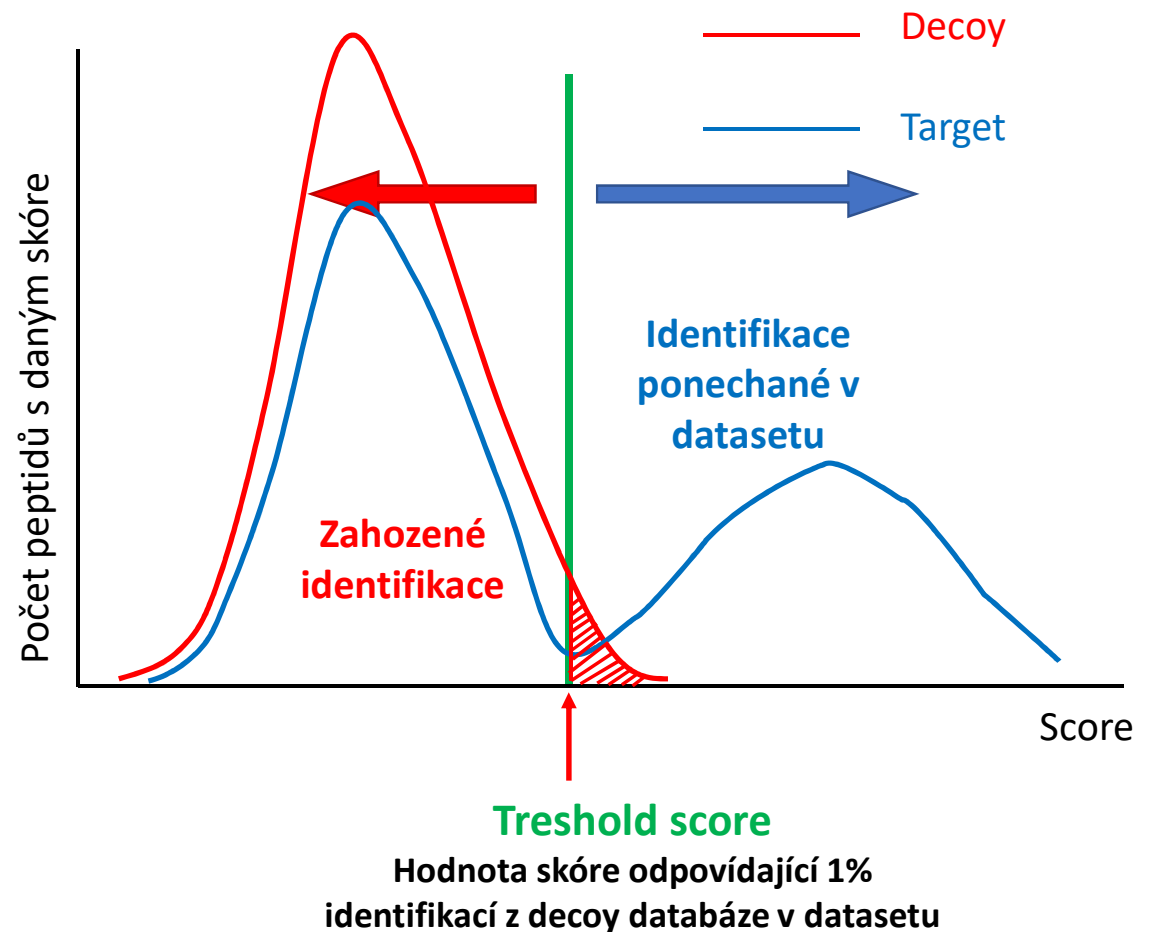
1. Vygenerujeme databázi neexistujících proteinů (návnad) otočením všech sekvencí v databázi = **decoy databáze**
2. Naměřená spektra porovnáme jak proti smysluplné databázi (target), tak proti decoy databázi
3. Každé shodě je přiřazeno skóre.
4. V datasetu ponecháme jen identifikace odpovídající určité FDR (obvykle 1%)





# FDR – target-decoy přístup

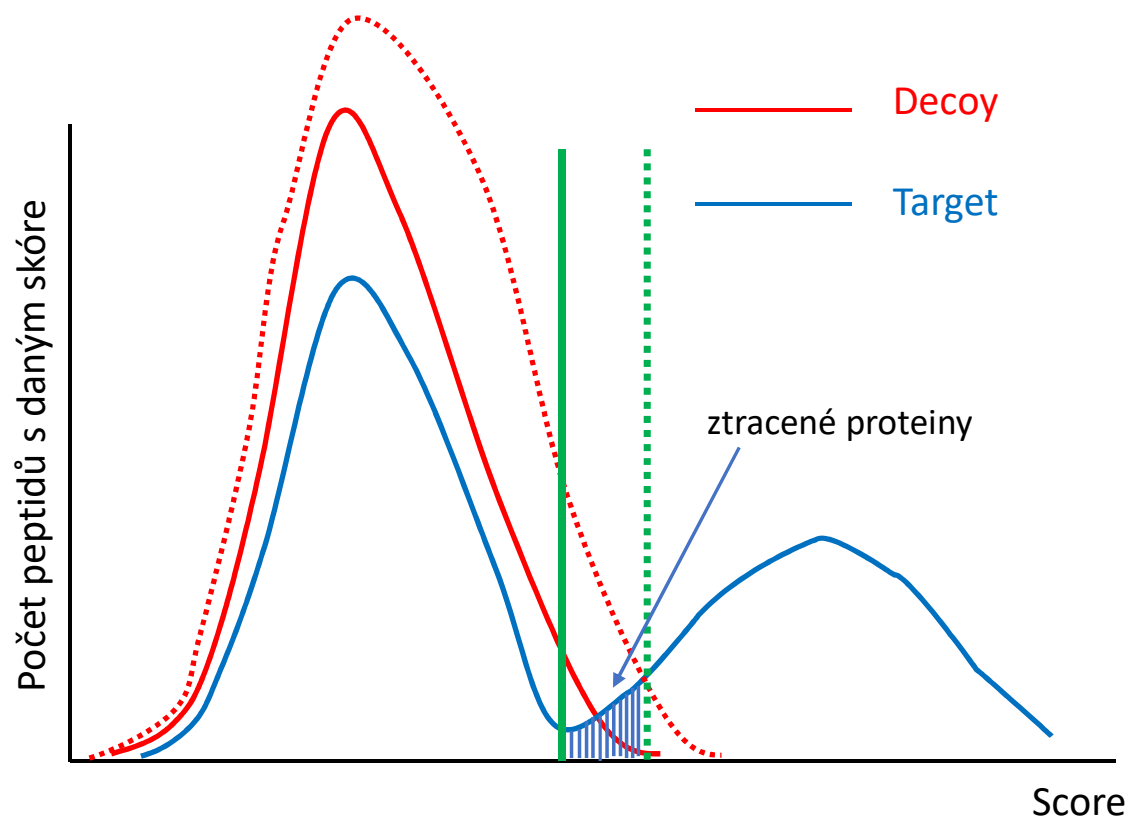
1. Vygenerujeme databázi neexistujících proteinů (návnad) otočením všech sekvencí v databázi = **decoy databáze**
2. Naměřená spektra porovnáme jak proti smysluplné databázi (target), tak proti decoy databázi
3. Každé shodě je přiřazeno skóre.
4. V datasetu ponecháme jen identifikace odpovídající určité FDR (obvykle 1%)



# FDR – target-decoy přístup

Ovlivnění výsledků :

- použitím nesespecifické nebo redundantní databáze
- definováním neúměrného počtu variabilních modifikací se zanedbatelným výskytem ve vzorku



# Q-value

- asociovaná vždy s konkrétním peptidem
- **minimální hodnota FDR, při které daný peptid projde FDR filtrováním**
- pokud má peptid  $q$ -value 0.01, znamená to, že filtrováním projde, pokud bude nastavena FDR alespoň na 1% (při FDR 0.5% by byl vyřazen)
- **vždy závisí na konkrétním vyhodnocení – stejný peptid bude mít různou  $q$ -value ve dvou různých experimentech** (i když skóre bude stejné)

# Q-value – Proč záleží na použité databázi?

- **Příklad 1:**

1. Data z lidského vzorku prohledáme proti **NESPECIFICKÉ DATABÁZI VŠECH EUKARYOT**
2. Seřadíme identifikované peptidy podle skóre od nejvyššího
3. Peptid EAMRPK je na pozici 100
4. Mezi těmito 100 peptidy je 10 falešně pozitivních
5. ***q*-value EAMRPK je  $10/100 = 0.1$**

- **Příklad 2:**

1. Stejná data z lidského vzorku nyní prohledáme proti **SPECIFICKÉ LIDSKÉ DATABÁZI**
2. Seřadíme identifikované peptidy podle skóre od nejvyššího
3. Z původních 100 vypadlo 20 identifikací z důvodu menší databáze. Z toho 6 bylo falešně pozitivních.
4. EAMRPK je na pozici 80.
5. Mezi těmito 80 peptidy jsou 4 falešně pozitivní.
6. ***q*-value EAMRPK je  $4/80 = 0.05$**

# Q-value

Příklad 1 vs příklad 2:

stejné spektrum

dvě různé databáze

stejný peptid

dvě různé q-value

stejné skóre



**Dva různé výsledky!**

- **Zda peptid projde filtrováním je ovlivněno:**
  - použitou databází – čím specifitější pro daný organismus, tím lépe
  - množstvím zadaných variabilních modifikací – expandují search space podobně jako nespecifická databáze

# Příklady software pro necílenou proteomiku

## DDA



MaxQuant – [www.maxquant.org](http://www.maxquant.org)  
FREE  
[www.youtube.com/@MaxQuantChannel](http://www.youtube.com/@MaxQuantChannel)

## DIA



DiaNN - <https://github.com/vdemichev/DiaNN>  
FREE



Proteome Discoverer  
ThermoFisher Scientific



Spectronaut  
Biognosys

pavel.talacko@natur.cuni.cz

