

Proteomika

1.

Oprogramowanie do analiz proteomicznych
i praktyczna ocena wyników identyfikacji białek

Najczęstszy typ doświadczenia

- Preparat białek trawiony trypsyną
- Rozdział peptydów nanoHPLC odwróconej fazy
 - może być też wielowymiarowa chromatografia
- Spektrometr tandemowy/hybrydowy mas ze źródłem typu elektrorozpylacz
 - praca w trybie zależnym od danych 'data-dependent'
- Identyfikacja w oparciu bazę sekwencji białek

Identyfikacja

wstępne przetwarzanie danych

Dane surowe
(MS+MS/MS)

.RAW (Thermo)
.RAW (Waters)
fid, baf, yet (Bruker)
standardy mzXML, mzML

mzML „standard otwarty”

- bez kompresji $3,7\times$ Thermo RAW
- z kompresją $2,8\times$ RAW
- z dwukrotną kompresją $1,85\times$

ProteoWizard – przetwarzanie plików ze spektrometrów mas do różnych formatów

wstępne przetwarzanie
ocena m/z z widm MS
łączenie, selekcja, centrowanie
widm MS/MS

Oprogramowanie producenta
Mascot Distiller
MaxQuant

Pliki zawierające m/z prekursora,
m/z z widma MS/MS

.mgf (Mascot)
.pkl (Waters)
mzXML, mzML

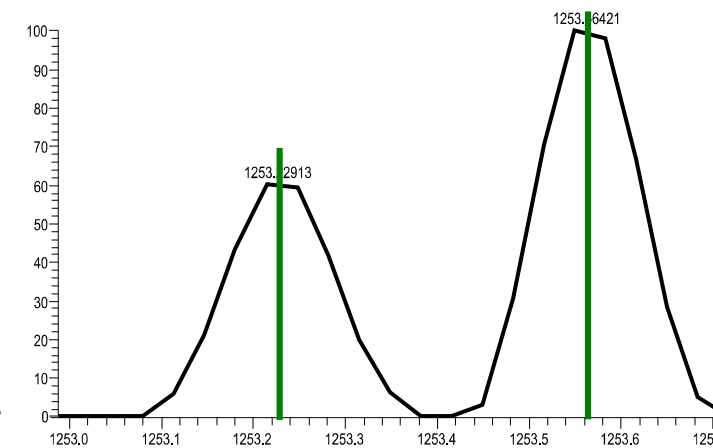
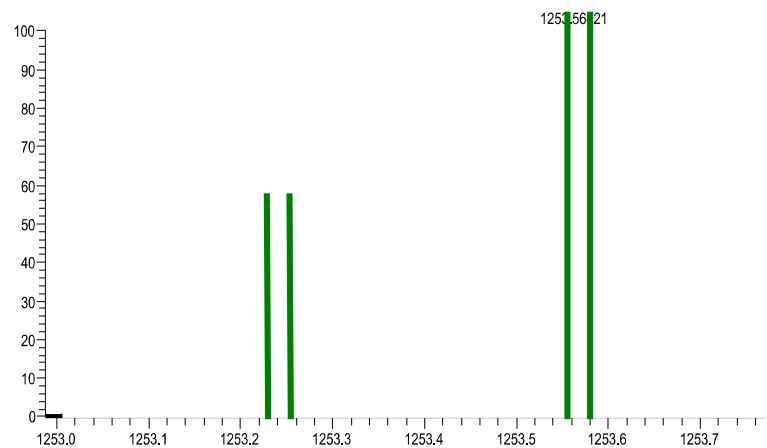
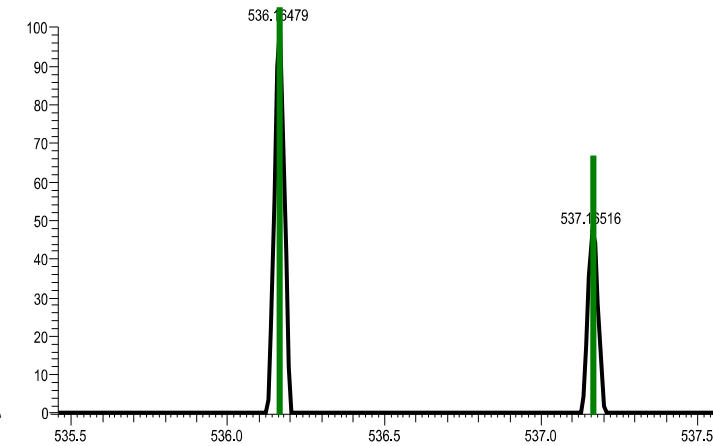
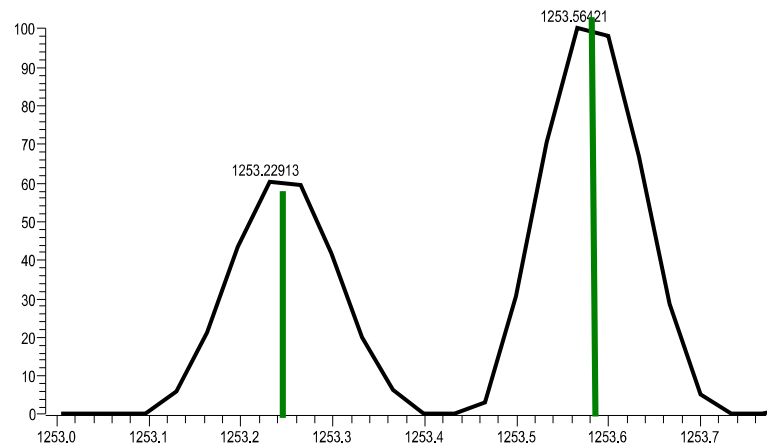
Oprogramowanie do identyfikacji
peptydów i białek

- **Mascot** (komercyjny)
- **Sequest** (komercyjny)
- **X!Tandem** (open-source)
- **Andromeda** (MaxQuant)
(darmowe źródła niedostępne)

Parametry

wstępne przetwarzanie

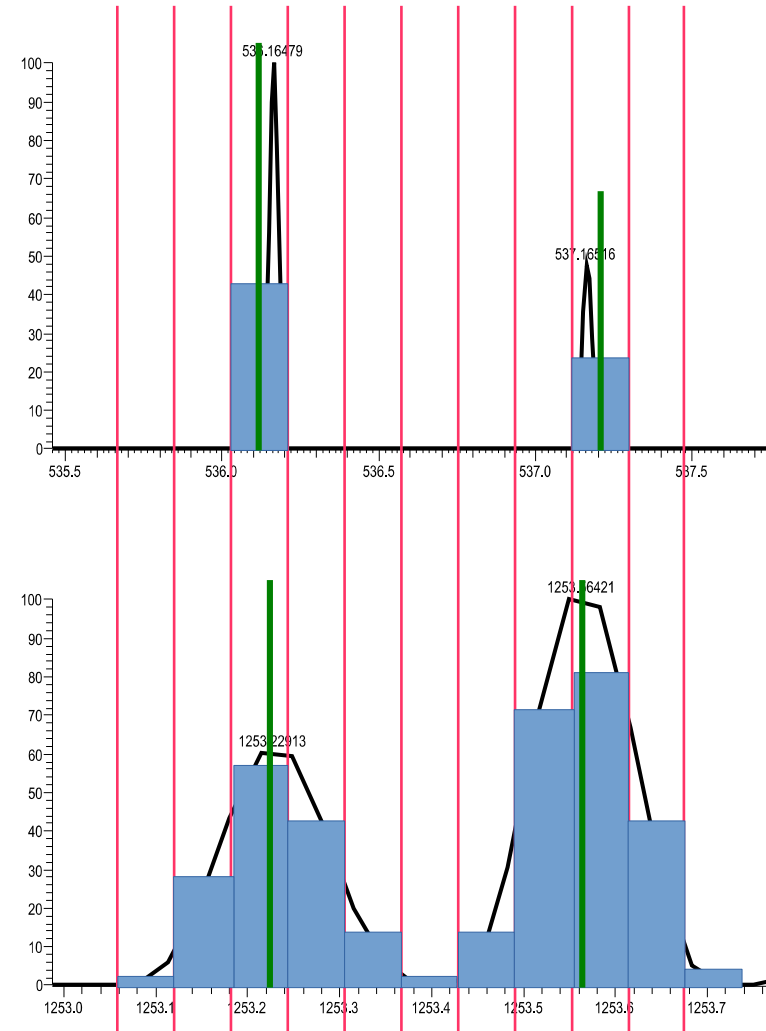
- Dane oryginalne (profil) i 'centroid'
- Łączenie widm MS/MS
- Decentroiding
 - brak danych o szerokości piku



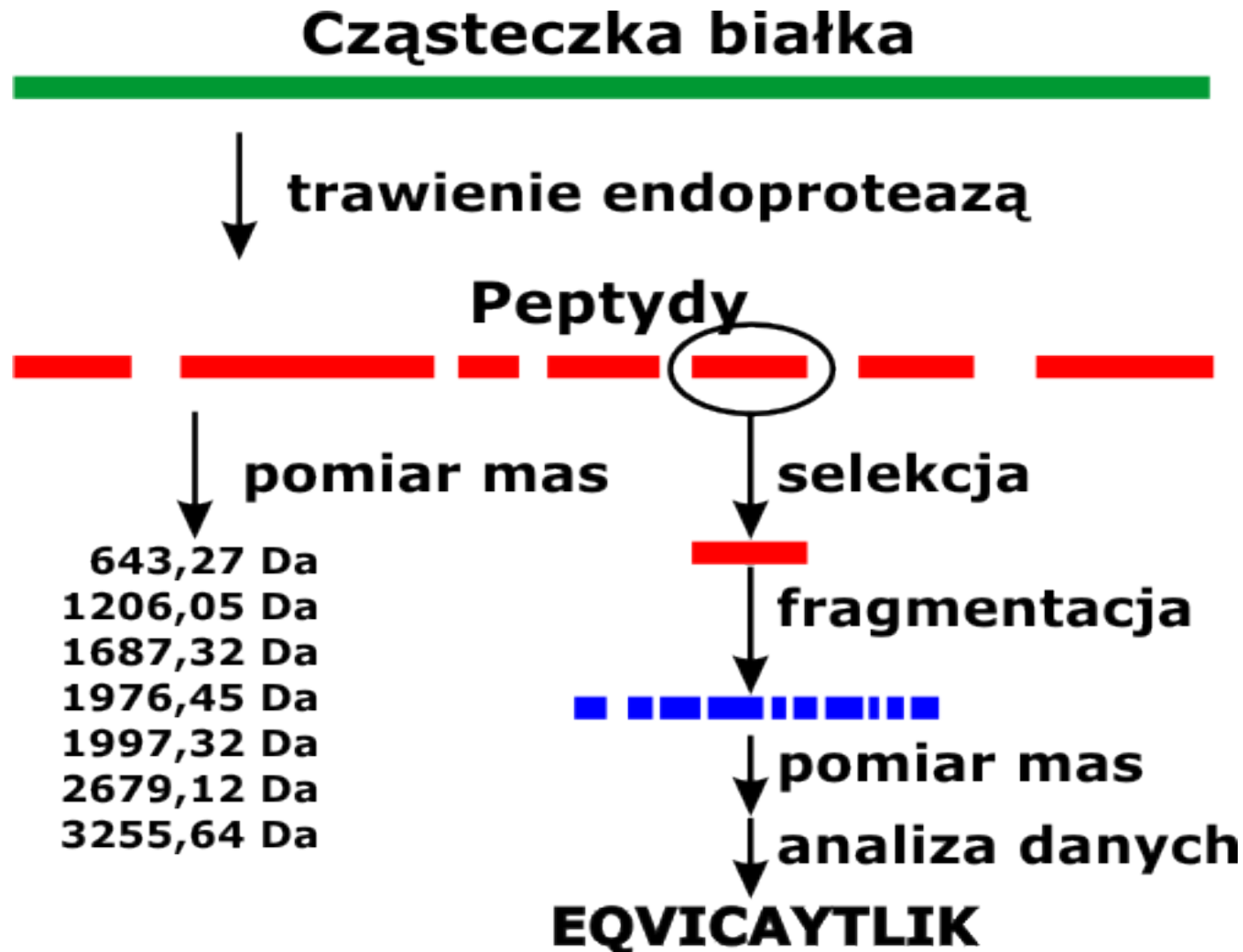
Parametry

wstępne przetwarzanie

- Dane oryginalne i centroidowane
- „Regriding”
 - liczba punktów na Dalton
 - błąd pomiaru wynikający ze zbyt małej rozdzielczości
- Przeliczanie mas prekursorów
- Centrowanie (centroid)
 - zapis danych do pliku MGF



Identyfikacja peptydów



Parametry pracy algorytmu

- Baza danych
 - Genomowa
 - Większa (rodzaj, pełna baza, itp.)
 - Możliwość zawężenia
- Plik z danymi
 - format pliku (.mgf, .pkl, mzML)
- Modyfikacje potranslacyjne
 - występujące zawsze, najczęściej dotyczy modyfikacji wprowadzanych podczas przygotowania próby (np. karbamidometylacja cystein)
 - mogące wystąpić (Utlenianie metioniny
 - artefakt, Fosforylacja, Acetylacja, itp.)
- Enzym użyty do trawienia i parametry trawienia
 - trypsyna (Arg-C, Lys-C, CNBr, itp., kombinacje enzymów)
 - enzymy „półspecyficzne” np. semi-trypsyna
 - liczna niedotrawek
- Dopuszczalny błąd pomiaru masy
 - peptydu [ppm]
 - fragmentów [ppm, Da, mmu]
 - monoizotopowa / średnia
- Fragmentacja
 - CID w pułapce
 - CID w strumieniu np. HCD
 - ECD (Electron Capture)
 - ETD (Electron Transfer)

Parametry c.d.

często specyficzne dla programu

- Znakowanie peptydów (^{18}O , SILAC, iTRAQ)
- Użycie bazy losowej (Mascot – ‘decoy’)
 - Percolator scoring (Mascot)
 - Rodzaj bazy (MaxQuant: losowa, odwrócona)
- Wyszukanie typu ‘error tolerant’ (Mascot)
- MaxQuant
 - Maksymalna liczba modyfikacji na peptyd
 - Parametry wyszukiwania wstępnego i głównego
 - Ilościówka

MASCOT MS/MS Ions Search

Your name	mk		Email	mkotl@ibb.waw.pl	
Search title	Ble ble ble				
Database(s)	SprotHumanRM SprotRandHumanMerge Sumo Synaptosomy TAIR10		Enzyme	Trypsin	
			Allow up to	2 missed cleavages	
			Quantitation	None	
Taxonomy	All entries				
Fixed modifications	Carbamidomethyl (C)		> <	2-dimethylsuccinyl (C) 2-monomethylsuccinyl (C) 2-nitrobenzyl (Y) 2-succinyl (C) 2HPG (R) 3-deoxyglucosone (R) 3sulfo (N-term) 4-ONE (C) 4-ONE (H) 4-ONE (K) 4-ONE+Delta:H(-2)O(-1) (C)	
	Display all modifications ☒				
Variable modifications	Oxidation (M)		> <		
Peptide tol. ±	15	ppm	# ¹³ C	0	MS/MS tol. ± 50 mmu
Peptide charge	Mr		Monoisotopic	☐ Monoisotopic ☒ Average	
Data file	Browse... No file selected.				
Data format	Mascot generic		Precursor		
Instrument	HCD		Error tolerant	☐	
Decoy	☒		Report top	AUTO hits	
Start Search ...			Reset Form		

Co robi program?

- „Trawi” wszystkie sekwencje z bazy danych użytym enzymem (trypsyną) → **lista peptydów**
 - parametry: enzym, niedotrawki, modyfikacje
- Liczy masy peptydów, przewiduje wzory fragmentacji (liczy masy jonów potomnych)
 - parametry: typ fragmentacji
- Porównuje dostarczoną przez użytkownika listę mas peptydów z listą otrzymaną z analizy bazy danych
 - parametry: błąd pomiaru masy (widma MS)
- Znalezione peptydy o pasującej masie → porównanie danych MS/MS
 - często do jednej zmierzonej masy pasuje wiele peptydów z bazy danych

Co robi program c.d.

porównanie widma MS/MS z danymi z bazy danych

- Do listy fragmentów zmierzonej w skanie MS/MS porównywane są listy fragmentów wyliczone dla peptydów z bazy o pasującej masie
- Program liczy liczbę pików pasujących i nie pasujących → podstawa wyliczenia p-wartości
 - $\text{score} = -10\log(10)P$
 - score dla białka – kumulowany score peptydów
 - „istotność” zależy od wielkości bazy, warto liczyć FDR

MS/MS Fragmentation of EFAAEEISSMVLK
Found in AT5G02500.1 in TAIR10plus, HSC70-1, HSP70-1...
Match to Query 7695: 1565.824308 from(783.919430,2+) intensity(355477.4100) scans(16180)

Title: 9766: Scan 16180 (rt=7492.64) [Z:\DO\1407-lipiec\40704mk_jadra.raw]

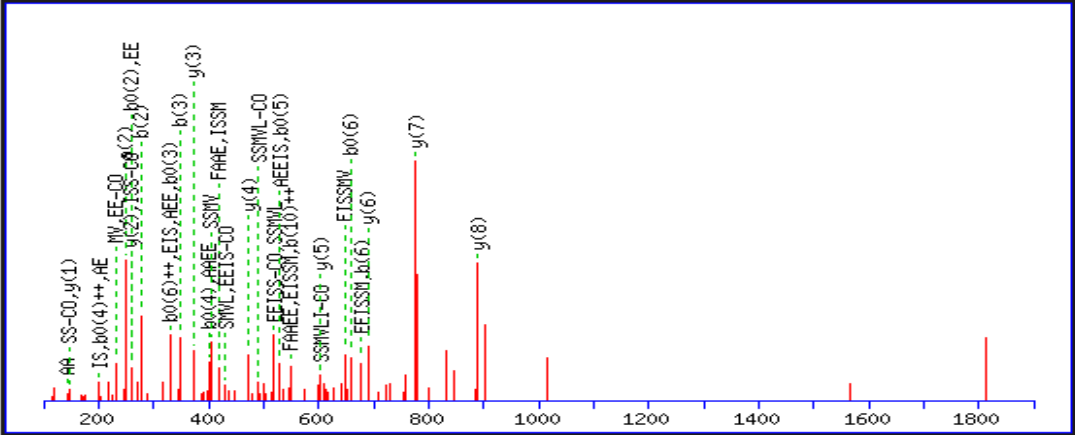
Monoisotopic mass of neutral peptide Mr(calc): 1565.8011

Fixed modifications: Carbamidomethyl (C)

Ions Score: 64 Expect: 8.7e-06

Matches : 50/238 fragment ions

using 41 most intense peaks



Protein Family Summary (scores adjusted by Percolator)

Significance threshold p< Max. number of families [\[help\]](#)
Ions score or expect cut-off Dendrograms cut at
Show Percolator scores ☒
Preferred taxonomy

▼Decoy search summary (reversed protein sequences)

Peptide matches	in TAIR10plus	in Decoy	FDR		
- above identity threshold	2053	9	0.44%	Adjust to	1%
- above identity or homology threshold	2053	9	0.44%	Adjust to	1%

Decoy results are available in [the decoy report](#).

Proteins (889)

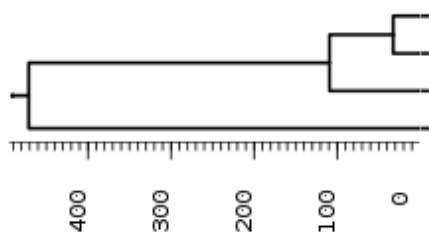
[Report Builder](#)

[Unassigned \(10882\)](#)

Protein families 1–10 (out of 889)

per page 1 [2](#) [3](#) [4](#) [5](#) [6](#) ... [89](#) [Next](#)

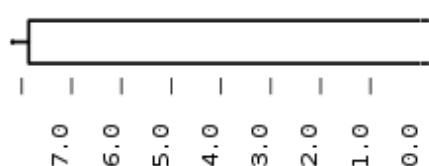
►1



1 AT5G02500.1
4 AT1G16030.1
2 AT3G09440.1
3 AT5G28540.1

1244 HSC70-1, HSP70-1, AT-HSC70-1, HSC70 | heat shock cognate protein 70-1 | chr5:5540
238 Hsp70b | heat shock protein 70B | chr1:5502386-5504326 REVERSE LENGTH=646
737 | Heat shock protein 70 (Hsp 70) family protein | chr3:2903434-2905632 REVERSE LENG
503 BIP1 | heat shock protein 70 (Hsp 70) family protein | chr5:10540665-10543274 REVER

►2



1 AT2G18960.1
2 AT2G07560.1

955 AHA1, PMA, OST2, HA1 | H(+)-ATPase 1 | chr2:8221858-8227268 FORWARD LENGTH=9
149 AHA6, HA6 | H(+)-ATPase 6 | chr2:3170394-3173952 REVERSE LENGTH=949

►3

ATCG00490.1

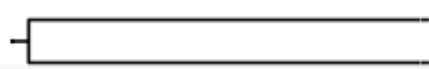
898 RBCL | ribulose-bisphosphate carboxylases | chrC:54958-56397 FORWARD LENGTH=47

►4

AT4G37930.1

746 SHM1, STM, SHMT1 | serine transhydroxymethyltransferase 1 | chr4:17831891-178347

►5



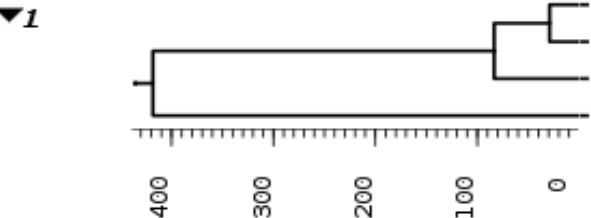
1 AT1G20620.1
2 AT4G35090.1

734 CAT3, SEN2, ATCAT3 | catalase 3 | chr1:7143142-7146193 FORWARD LENGTH=492
265 CAT2 | catalase 2 | chr4:16700937-16703215 REVERSE LENGTH=492

Protein families 1–10 (out of 1090)

10 per page 1 2 3 4 5 6 ... 109 Next Expand all Collapse all

Accession contains Find



1 AT5G02500.1
4 AT1G16030.1
2 AT3G09440.1
3 AT5G28540.1

1098 HSC70-1, HSP70-1, AT-HSC70-1, HSC70 | heat shock cognate protein 70
217 Hsp70b | heat shock protein 70B | chr1:5502386-5504326 REVERSE LEM
701 | Heat shock protein 70 (Hsp 70) family protein | chr3:2903434-290563
468 BIP1 | heat shock protein 70 (Hsp 70) family protein | chr5:10540665-1

Threshold (0): 0 Cut

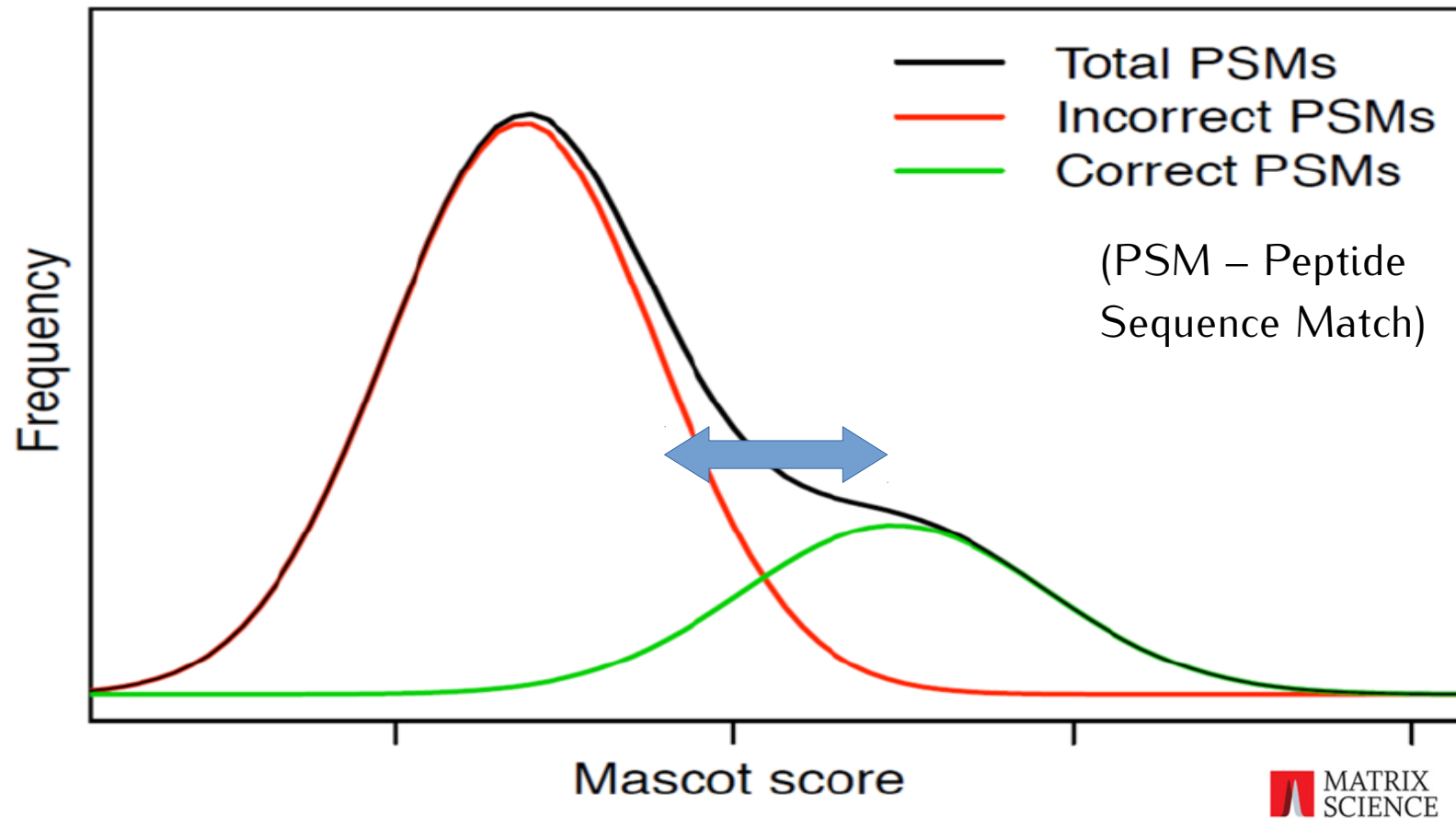
		Score	Mass	Matches	Sequences	emPAI	
✓ 1.1	AT5G02500.1	1098	71712	27 (21)	20 (15)	1.45	HSC70-1, HSP70-1, AT-HSC70-1, HSC70 heat shock cogn
✓ 1.2	AT3G09440.1	701	71559	22 (15)	19 (12)	0.96	Heat shock protein 70 (Hsp 70) family protein chr3:290
	1 same set of AT3G09440.1						
✓ 1.3	AT5G28540.1	468	73869	14 (11)	14 (11)	0.61	BIP1 heat shock protein 70 (Hsp 70) family protein chr5
✓ 1.4	AT1G16030.1	217	71270	10 (4)	8 (4)	0.20	Hsp70b heat shock protein 70B chr1:5502386-5504326
Redisplay All None							

55 peptide matches (53 non-duplicate, 2 duplicate)

Auto-fit to window

Query Dupes	Observed	Mr (expt)	Mr (calc)	ppm	M	Score	Expect	Rank	U	1	2	3	4	Peptide
373	345.1970	688.3794	688.3755	5.57	1	12	2	7	U					K.DAKLDK.S
1440	428.2388	854.4630	854.4610	2.40	1	9	1.4	4	U					K.HQKDISK.D
2237	481.2706	960.5267	960.5127	14.5	1	53	0.00015	1						R.LSKDEIEK.M
2495	495.2556	988.4966	988.4825	14.2	0	51	0.00019	1	U					R.LSQEEIDR.M
2516	496.2709	990.5272	990.5134	14.0	0	16	0.75	2	U					K.EVAEAFGLGR.T
3256	537.3289	1072.6432	1072.6281	14.1	0	27	0.016	1	U					K.LVPYQIVNK.D
3750	564.8147	1127.6148	1127.6047	9.01	2	6	0.89	2	U					K.NKKDISGNPR.A

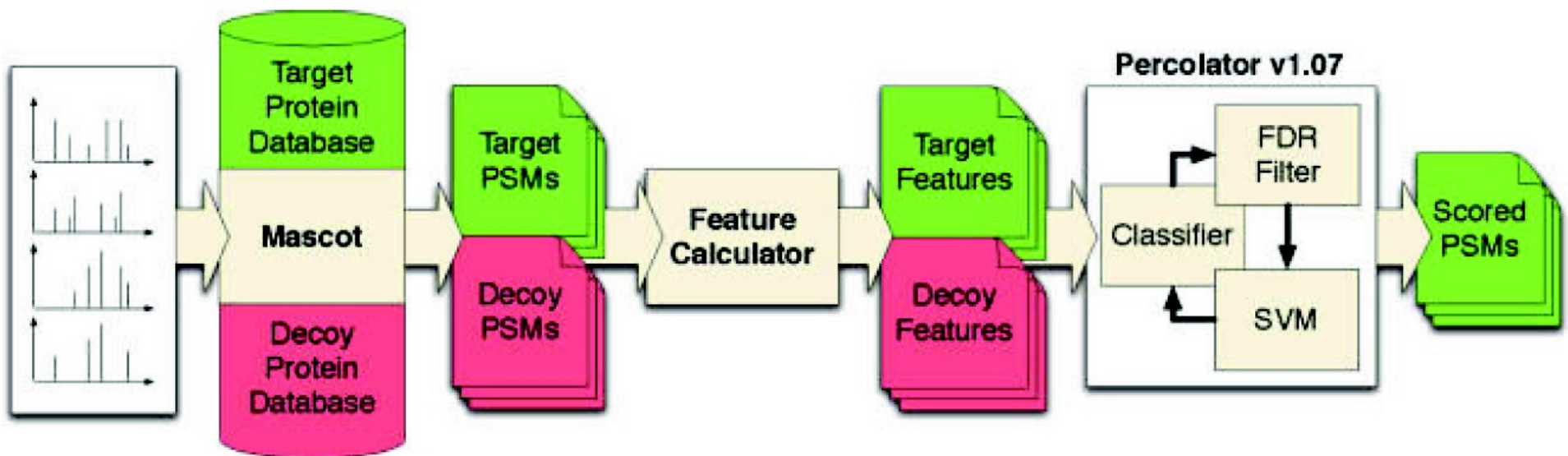
Poprawianie wyników



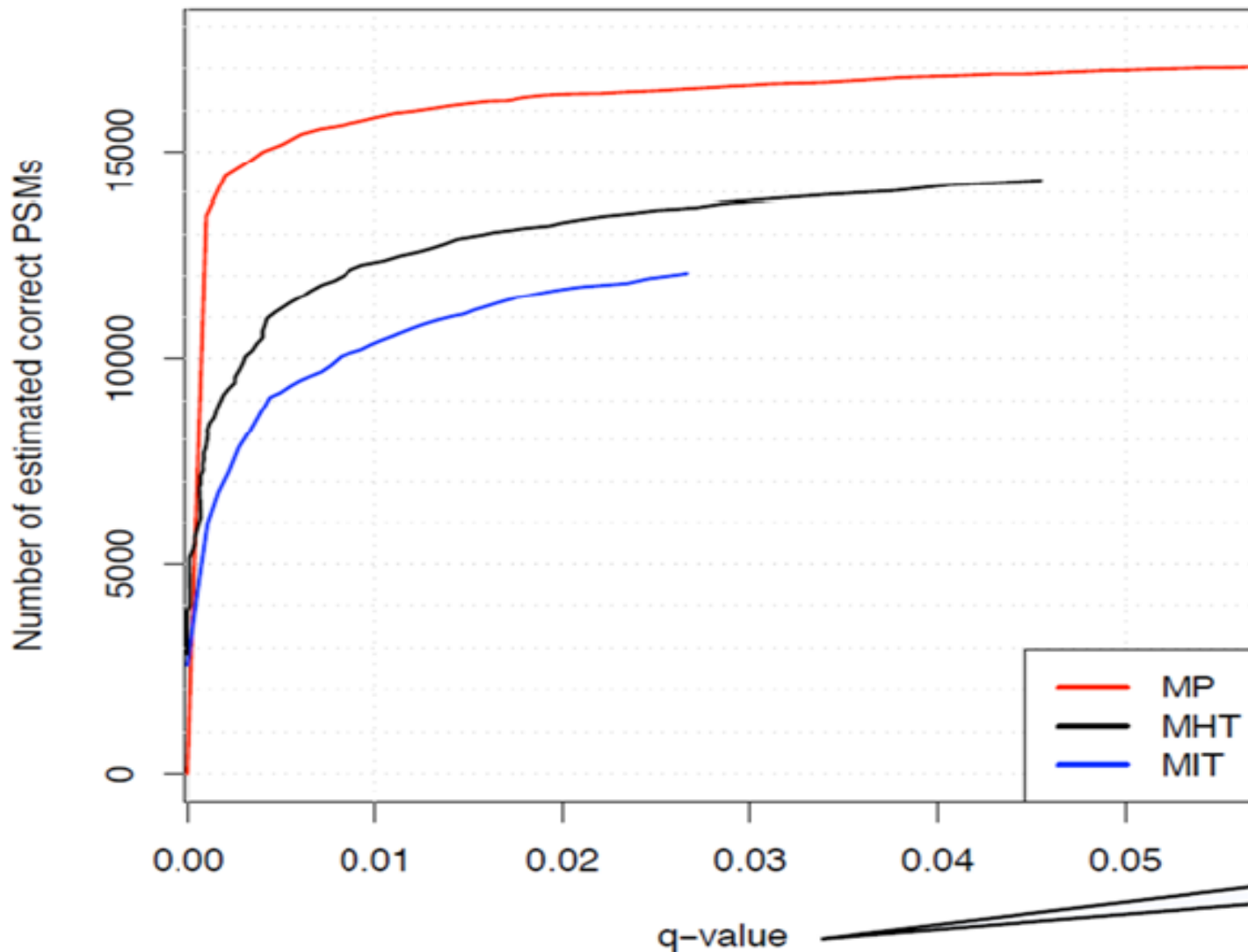
- Zmniejszanie błędu pomiaru masy
- Algorytmy, jak Percolator

Algorytm Percolator

- Używa wyszukiwań na bazie losowej
- Algorytm maszyny wektorów nośnych (Support Vector Machine)
- Nadaje się tylko do dużych zestawów danych
- Bierze pod uwagę szereg parametrów takich jak:
 - dokładność pomiaru MS i MS/MS
 - intensywność pików na widmie MS i MS/MS
 - itp. razem 18 cech



Wyniki Mascot Percolatora



Think FDR

Programy do dalszej analizy np. MScan

MScan - [magazyn/DaneMS/lpDaniel_analiza112014/DB_IP_scan_fdrPepF005.exp]

File Edit Data Tools Window Help

DB_IP X

Col-0

11216_Col0

110726_Col0

120223_Col0

111231_Col0

samples_merg

HD2

110726_HD2

11216_HD2

120223_HD2

111231_HD2

Proteins list: all groups

#	ID	Name	11216	110726	120223	111231	sample	110726	11216	120223	111231
1	ATCG0049...	ribulose-bisphosphate carboxylases	850.0	198.0	521.0	1330.0	935.0	31.0	252.0	427.0	1757.0
2	AT5G5991...	Histone superfamily protein	844.0	278.0	311.0	745.0	710.0	54.0	270.0	413.0	914.0
3	AT5G0250...	heat shock cognate protein 70-1	778.0	103.0	688.0	1480.0	1162.0	61.0	183.0	487.0	1395.0
4	AT3G2665...	glyceraldehyde 3-phosphate dehydrogenase ...	767.0	222.0	312.0	1097.0	901.0	25.0	137.0	602.0	1447.0
5	AT5G2854...	heat shock protein 70 (Hsp 70) family protein	746.0	26.0	713.0	932.0	1318.0	44.0	230.0	618.0	1654.0
6	AT2G2872...	Histone superfamily protein	739.0	183.0	362.0	898.0	732.0	34.0	194.0	382.0	1021.0
7	AT1G1290...	glyceraldehyde 3-phosphate dehydrogenase ...	734.0	199.0	297.0	1044.0	877.0	25.0	155.0	593.0	948.0
8	AT1G1290...	glyceraldehyde 3-phosphate dehydrogenase ...	734.0	199.0	297.0	1044.0	877.0	25.0	155.0	593.0	948.0

Protein info

Experiment info:

Name: heat shock cognate protein 70-1

ID: AT5G02500.2 Gene: HSC70-1, HSP70 Organism: Arabidopsis thaliana

Databases info:

DB: TAIR IDs: AT5G02500.2

Sequence:

MSGKGEGPAI GIDLGTYSY VGWQHDRV IANDQGNRT TPSYVAFTDS ERLIGDAAKN
QVAMNPVNTV FDAKRLIGRR FSDSSVQSDM KLNPFKIQAG PADKPMIYVE YKGEKEFAA
EETSSMLVIK HREIAEAYLG VTIKNAVVTY PAYFNDSQRQ ATKDAGVIAG LNVHRIINEP
TAAAIAYGLD KKATSVGKKN VLIFDLGGGT FDSVLLTIEE GIFEVKATAG DTHLGGEDFD
NRMVNHVQEK FKRSKKDIT GNPRALRLR TSCERAKRTL SSTAQTITIE DSYEGIDIFY
STITRARFEE LNMDFLRKCN EPVEKCLRDA KMDKSTVHDV VLVGGSTRIP KVVQLLDQFF
NGKELCKSIN PDEAVAYGAA VQGAILSGEG NEKMVQEAEX YKSEDEEHKK KVEAKNALEN
YAYNHRNTIQ DEKIGELPA ADKKKIEDST EQAIQMLEGN QLAEDAFED KMKLESTICN
PIIAKMYQGA GGEAGGPAS GMDDAPPAS GGAGPKIEEV D

Mass: 7542.79 Leng... 521 Coverage... 52.02%

Close

Peptides list (protein AT5G02500.2): all groups

#	Peptide	Mass	Proteins	11216 C...
1	DAGVIAGLNVMR	1214.64414	AT3G0944...	153.85
2	VEIANDQGNR	1227.62076	AT1G5641...	20.94
3	FSDSSVQSDMK	1229.52342	AT5G0250...	20.75
4	MVNHVQEFK + Oxidation (M)	1293.61759	AT1G5641...	36.74
5	FEELNMDLFR + Oxidation (M)	1328.60709	AT3G1258...	153.85
6	ELESICNPIAK	1385.72245	AT3G0944...	33.87
7	VQQLLDQFFNGK	1435.74596	AT1G5641...	23.03
8	ITPSYVAFTDSER	1472.67834	AT1G5641...	19.1
9	NQVAMNPVNTVFDK	1646.80864	AT1G5641...	59.65
10	IINEPTAAAIAYGLDK	1658.88794	AT1G1603...	153.85
11	IINEPTAAAIAYGLDKK	1786.9829	AT1G1603...	103.75
12	NAVVTYPAYFNDSQR	1679.82673	AT1G1603...	
13	DAGVIAGLNVMR + Oxidation (M)	1230.63906	AT3G0944...	
14	MVNHVQEFK	1277.62268	AT1G5641...	
15	EIAEAYLVGVIK	1305.71802	AT5G0250...	
16	NALENVAYNMR + Oxidation (M)	1373.6034	AT3G0944...	
17	STVHDVLVLVGSTR	1425.75759	AT5G0250...	
18	EFAAEEISSMLVIK	1565.8011	AT5G0250...	
19	NQVAMNPVNTVFDK + Oxidation (M)	1662.80356	AT1G5641...	
20	ATAGDTHLGGEDFNDR	1674.72339	AT1G1603...	
21	GEGPAIGDLGTYSYCVGWQHNR	2587.21254	AT1G5641...	
22	IQAGPADKPMIYVEYK	1821.93351	AT5G0250...	
23	IQAGPADKPMIYVEYK + Oxidation (M)	1837.92842	AT5G0250...	
24	SINPDEAVAYGAQVQGAILSGEGNEK	2559.24528	AT5G0250...	
25	FEELNMDLFR	1312.61217	AT3G1258...	

Common peptides list

#	Sequence	Mass	Ref. sam...	Sample
1	VPULGIWGGK	1151.70667	53.17	54.0
2	GLAYDTSDDQDITR	1696.75402	28.3	28.06
3	LPLQDVVK	974.54368	33.55	33.4
4	IGGIGTVPVGR	1024.60293	31.14	31.41
5	STTTGHLIYK	1119.59242	24.01	23.19
6	LVLPGLAK	938.58007	37.17	37.26
7	QVHPDIGISK	1179.62479	25.5	25.31
8	KPASEKPVEEK	1240.66632	15.33	15.23
9	DAVDITLHAR	1161.54145	20.44	20.01
10		1179.61355	34.12	33.53
11		1309.69517	46.44	46.56
12		1324.7463	28.32	28.16
13		1191.52302	28.14	27.45
14		1320.63099	36.17	36.02
15		1570.66818	17.07	16.3
16		797.50109	26.06	25.53
17		953.53345	41.56	42.03
18		1125.5666	19.12	18.42
19		1224.6714	16.13	15.52
20		788.44324	20.34	20.13
21		911.46472	21.53	21.3
22		961.47164	18.16	17.43
23		1020.52401	45.39	45.36
24		1038.4618	18.09	17.16
25		1406.66056	31.5	31.3
26		1547.80176	43.31	43.52
27		1167.56325	20.09	19.21
28		1227.62076	25.24	24.49
29		1135.59857	22.57	22.32
30		1614.75978	23.19	22.39
31		969.51311	27.47	27.19
32		972.48762	16.28	16.2
33		1313.70918	43.47	44.04
34		1498.61719	18.29	17.46
35		1375.6085	20.06	19.28

RT scatter plot

File View Data

RT [sec]

RT [sec]

a=0.9683, b=72.9676, R^2=0.9998

Col-0: 110726_Col0 vs Col-0: 11216_Col0

Score th.: scoreMMT 0 Peptide RT calc.: Max. inten...

Statistics:

Peptides in current sample: 741 Peptides in reference samp.: 186

Common peptides: 73 Samples similarity: 0.157

Queries with q<0.01: 0

Done

Sample: 111231_HD2

Total queries: 3831

Assigned queries: 3831 (0 not assigned to sequences, 0 not assigned to p

Target database queries: 3831

Decoy database queries: 0

Common queries: 0

Target/Decoy ratio: 9.223372036854777E13

Queries with q<0.01: 0

Done

47 GPQSPSGYSCK

48 AAAAGFEK

Proteomika

2.

Proteomika ilościowa

Oznaczanie ilości białka

- Bezwzględna ocena ilości konkretnego białka
 - oczyszczenie i pomiar ilości
 - trudne do oceny straty podczas oczyszczania
 - metody „pół ilościowe” np. western blot ze standardem dodanym
 - LC-MS ze standardem dodanym
- Bezwzględne oznaczenie ilości setek białek w jednym doświadczeniu?
 - 2DE + densytometria

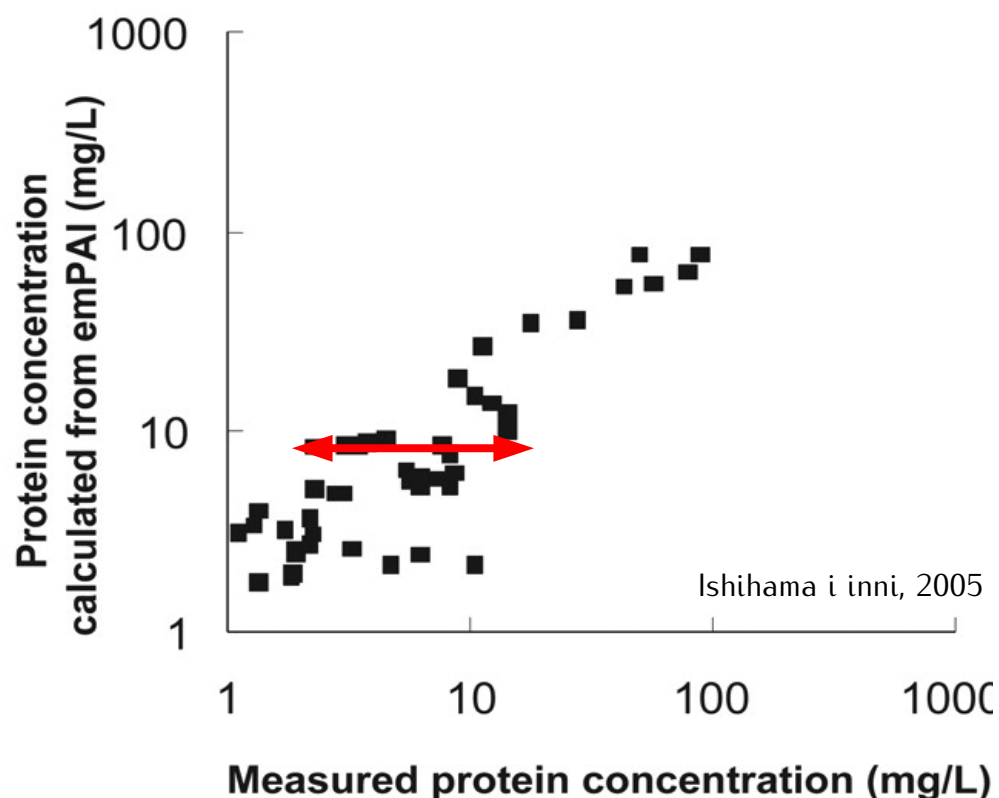
Liczenie widm

(spectral counting)

- Najprostszy sposób oceny ilości białka w badaniach LC-MS
- Liczenie współczynnika biorącego pod uwagę liczbę peptydów zidentyfikowanych i możliwych do zidentyfikowania
 - PAI – Protein Abundance Index
 - emPAI – Exponentially Modified PAI
- Wyniki skrajnie szacunkowe
- Współczynniki używające intensywności
 - iBAQ, TOP3, MeanInt

$$PAI = \frac{N_{obsd}}{N_{obsbl}}$$

$$emPAI = 10^{PAI} - 1$$

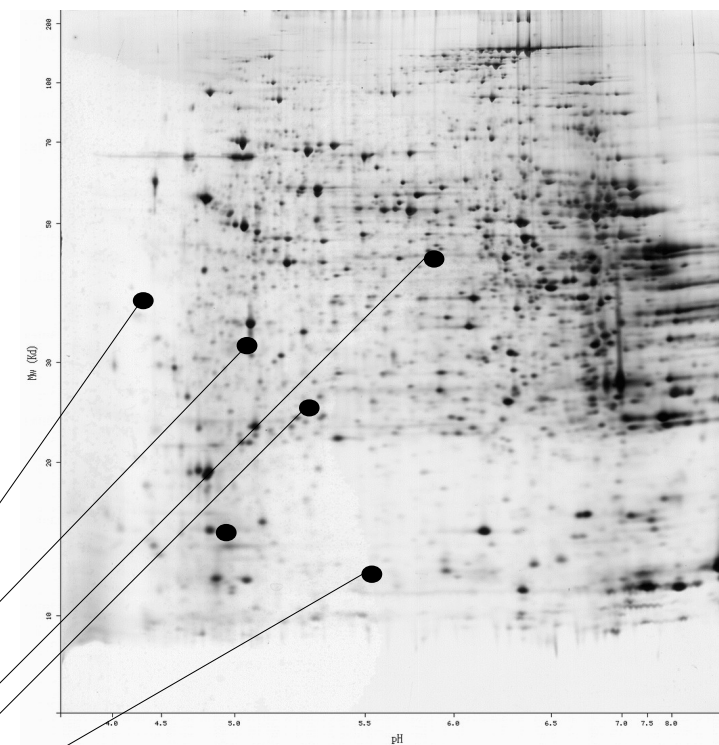
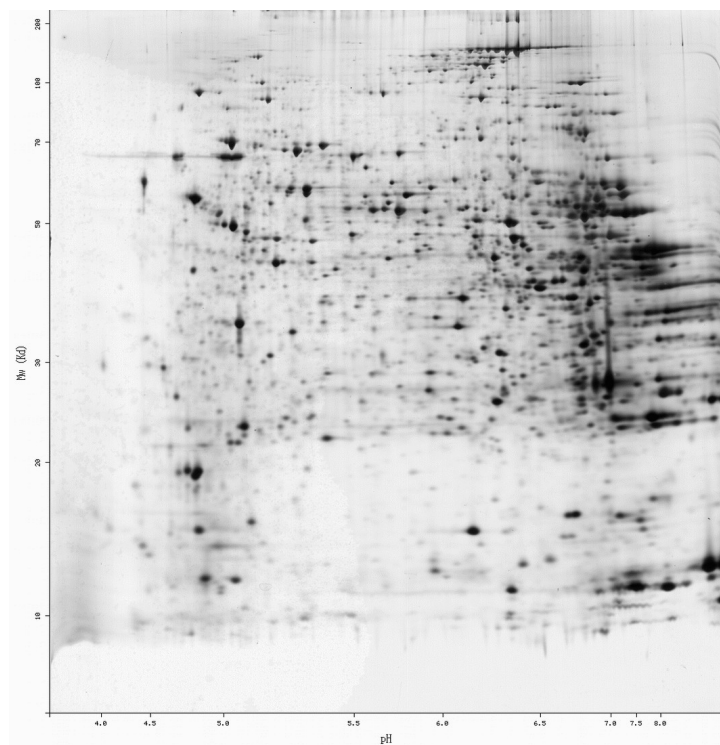


Proteomika różnicowa

- Biologów najbardziej interesują różnice w ilości konkretnych białek pomiędzy próbami
 - mutant – dziki
 - szczep chorobotwórczy – szczep niegroźny
 - reakcja na stres, itp.
 - Różnice odzwierciedlają zmiany zachodzące w komórce
 - Łatwiej badać różnice – kontrola jest „standardem”

Elektroforeza dwukierunkowa

metoda klasyczna



Plamki różnicowe

wycięte

białka strawione w żelu

peptydy poddane analizie spektrometrycznej i zidentyfikowane

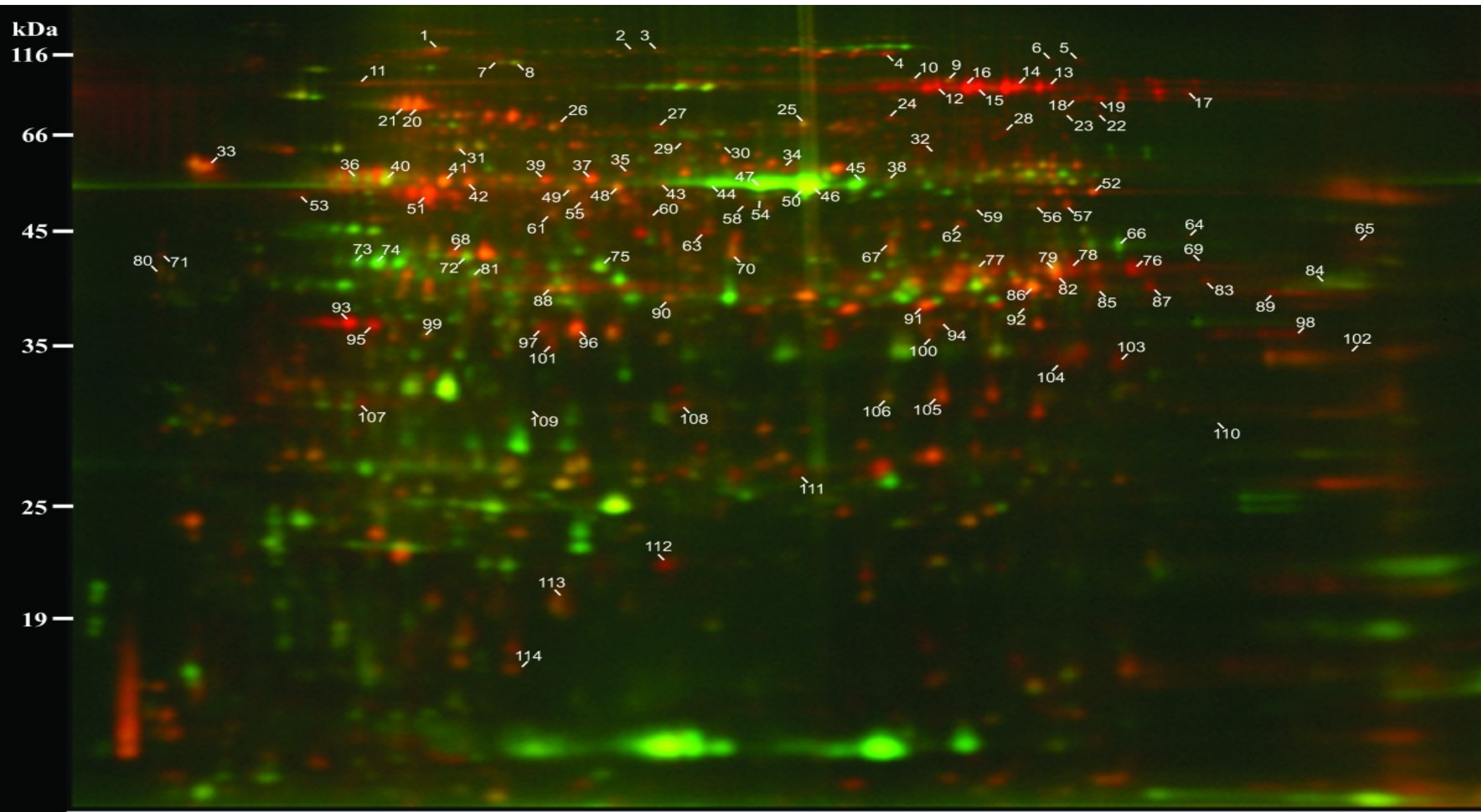
Elektroforeza 2D

- Prosta koncepcyjnie
 - różnice widoczne „na oko”
- Prosty sprzęt
- Łatwo oszacować ilość białka w preparacie
- Pracochłonna i czasochłonna
- Nie nadaje się do białek hydrofobowych i o skrajnym pI
- Mała czułość
 - zasłanianie jednych białek przez inne
- ograniczona liczba białek
- Kiepska powtarzalność
- Wysokie koszty (dodatkowo wiele analiz MS)

DIGE

Difference gel electrophoresis

Znakowanie białek z różnych prób rozróżnialnymi fluorofoforami, wspólny rozdział 2D-PAGE

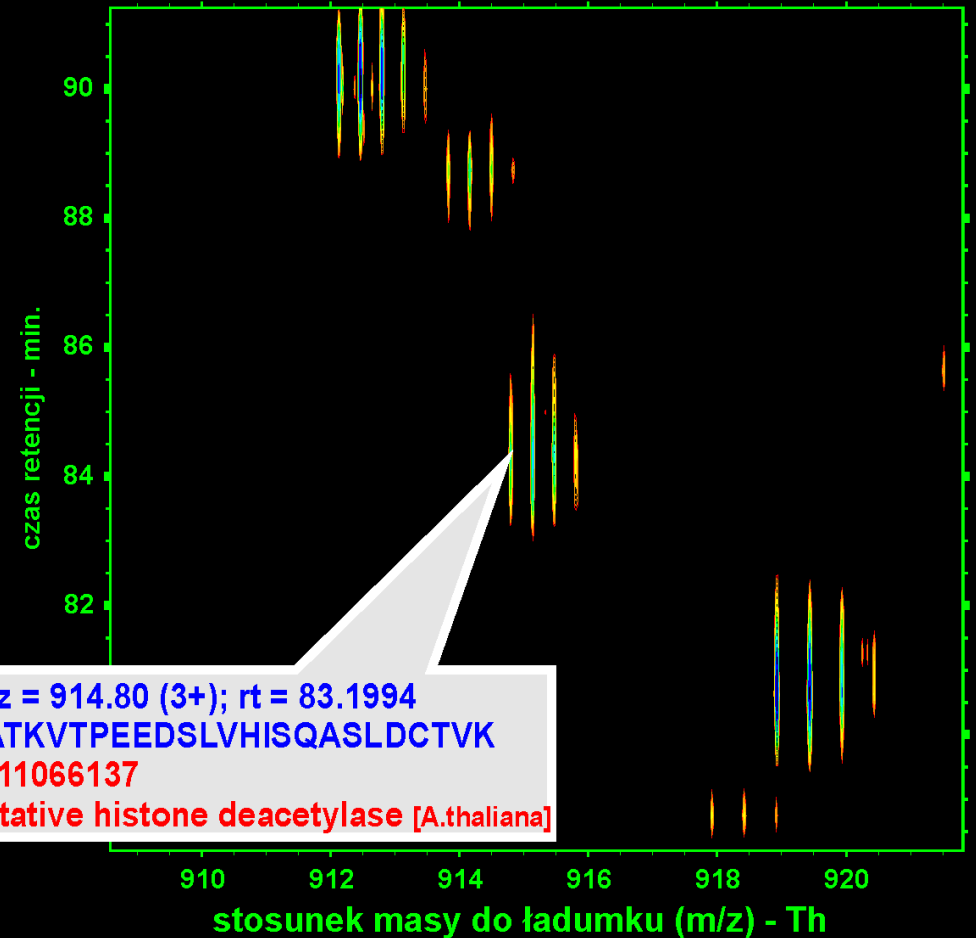
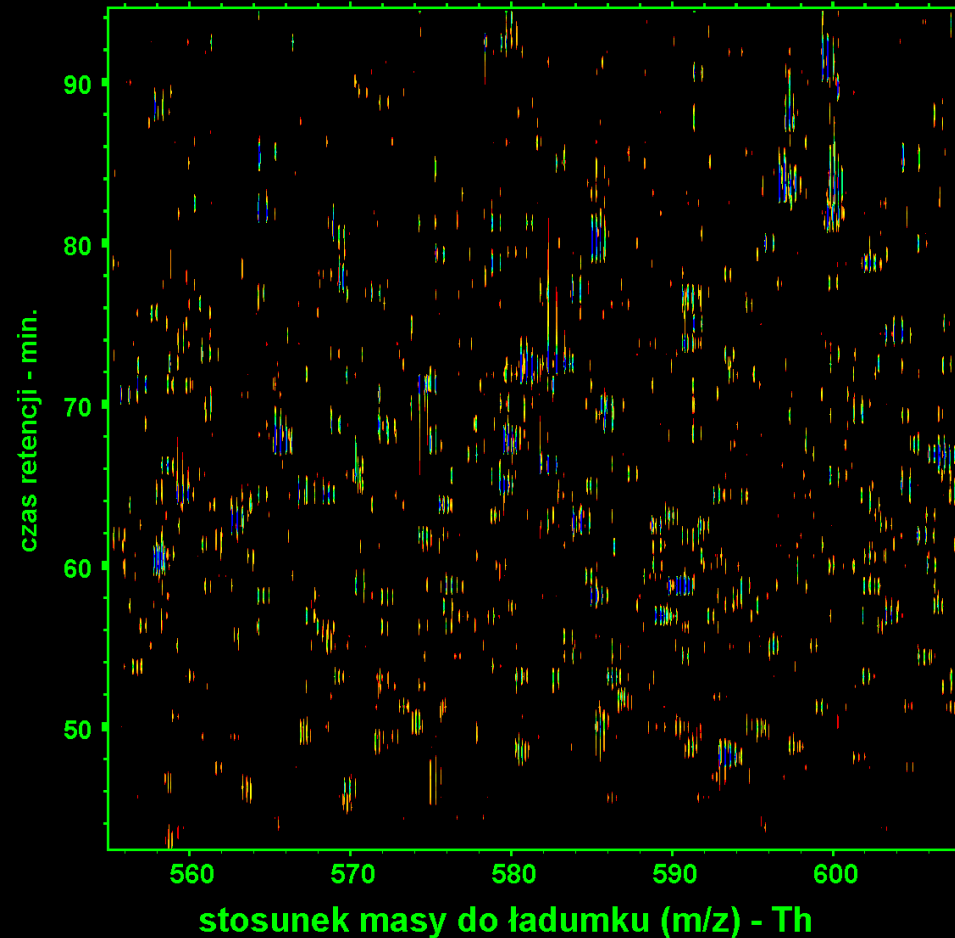
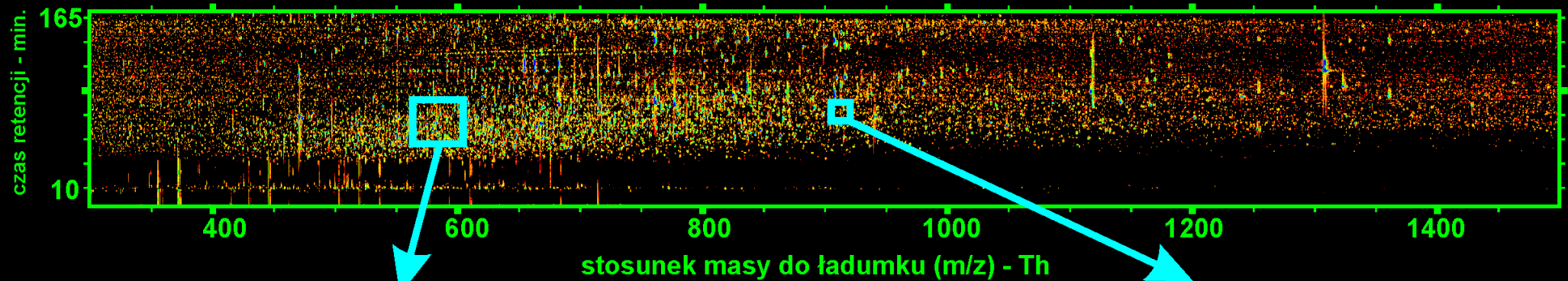


Użycie metod typu 'shotgun'

- Spektrometr mas rejestruje intensywność sygnału
- Intensywność zależy od stężenia peptydu ale
 - wydajność jonizacji ciężka do przewidzenia
 - właściwości peptydu, obecność innych substancji w próbce, parametry źródła jonów
 - można założyć, że w dwóch porównywalnych próbach wydajność jonizacji będzie zbliżona
 - porównanie sygnałów pozwoli na stwierdzenie różnic w stężeniu określonych peptydów (i białek)

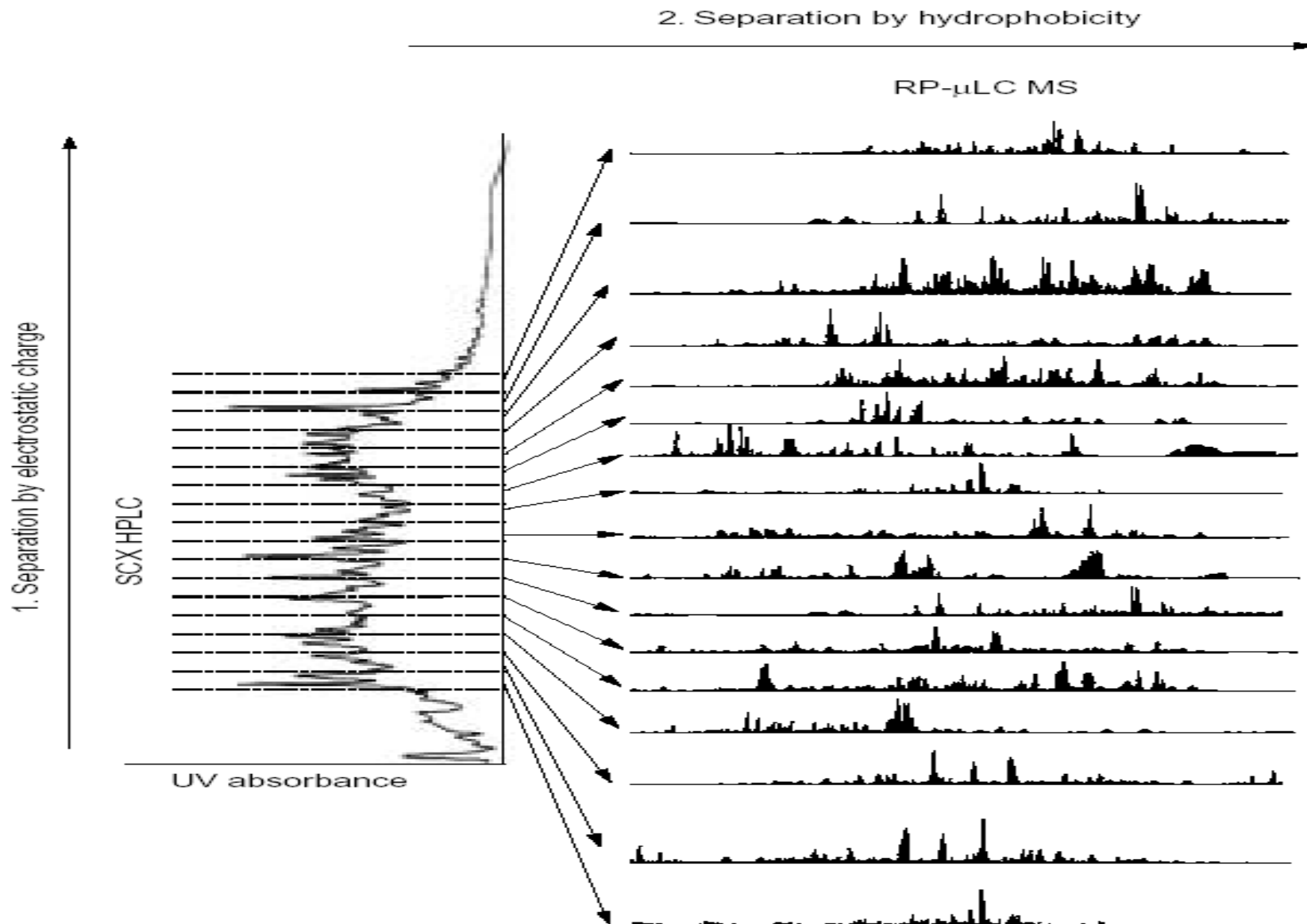
Mapa peptydowa jądra komórkowego *Arabidopsis*

HPLC - ESI - LTQ-FTICR



Rozdział 2D

typowa metoda MudPIT



Analizy bez znakowania peptydów

Label-free

- Kolejne porównywane próby peptydów są poddawane analizie w układzie LC-MS
 - lub 2D-LC-MS, IEF → LC-MS, itp.
- Dane o intensywności z widm MS są używane do porównania ilości peptydów/białek
 - widma MS/MS są używane do identyfikacji

Rozdzielenie rejestracji danych MS i MS/MS

Metody 'Accurate Mass Tags'

- Spektrometr rejestruje widma wielkiej liczby peptydów w czasie jednego przebiegu
 - tylko nieliczne są fragmentowane
- Czas retencji peptydów z kolumny HPLC jest powtarzalny
 - należy zapewnić powtarzalne warunki rozdziatu i skład badanej próby (porównywanie podobnych preparatów)
 - możliwa identyfikacja peptydów na podstawie dokładnego pomiaru masy i czasu retencji
- Jeśli peptyd został zsekwencjonowany tylko w jednym przebiegu, można zidentyfikować jego pik w innym przebiegu
 - porównanie ilości
 - można zwiększyć liczbę identyfikacji i porównań ilości
 - budowa bazy danych pozwalającej przypisać peptydy na podstawie m/z i czasu retencji

Metody 'label free'

- Możliwe badanie wielkiej liczby białek
- Duża czułość
- Białka hydrofobowe i o skrajnym pI
- Łatwa i stosunkowo tania obróbka prób
 - mało etapów – małe straty
 - mało materiału
- Badanie podobnych próbek
 - wpływ mutacji, warunków, choroby, itp.
 - słabiej spisuje się przy badaniu różnych proteomów
- Skomplikowane przetwarzanie danych
- Wrażliwe na zakłócenia rozdziatu i zmiany czułości spektrometru mas

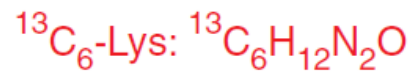
Użycie standardów znakowanych izotopowo

- Możliwe rozróżnienie sygnałów od różnych izotopów przy pomocy spektrometrii mas
 - porównywane próby można wyznakować różnymi izotopami i zmieszać przed analizą
 - różnice w składzie izotopowym nie zmieniają istotnie właściwości peptydów – czas retencji ten sam
- Znakowanie metaboliczne
- Znakowanie *in-vitro*

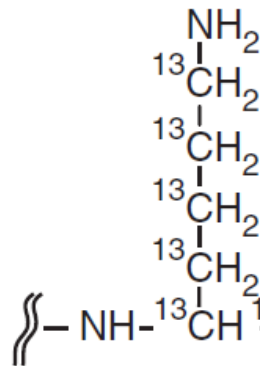
Znakowanie metaboliczne

- Cięższe izotopy dodawane do składników pożywki
- SILAC (Stable Isotope Labeling by Amino acids in cell Culture)
 - Aminokwasy najczęściej Lizyna i Arginina, rzadziej tyrozyna i metionina
 - Izotopy ^{13}C , ^{15}N rzadziej ^2H
 - Po kilka atomów w aminokwasie zmienionych (+masa kilka Da)
 - Możliwe kombinacje atomów o różnej masie
 - Kultury komórkowe
 - Organizmy modelowe np. Muszka owocowa, Mysz
 - Potrzeba kilku pokoleń

SILAC

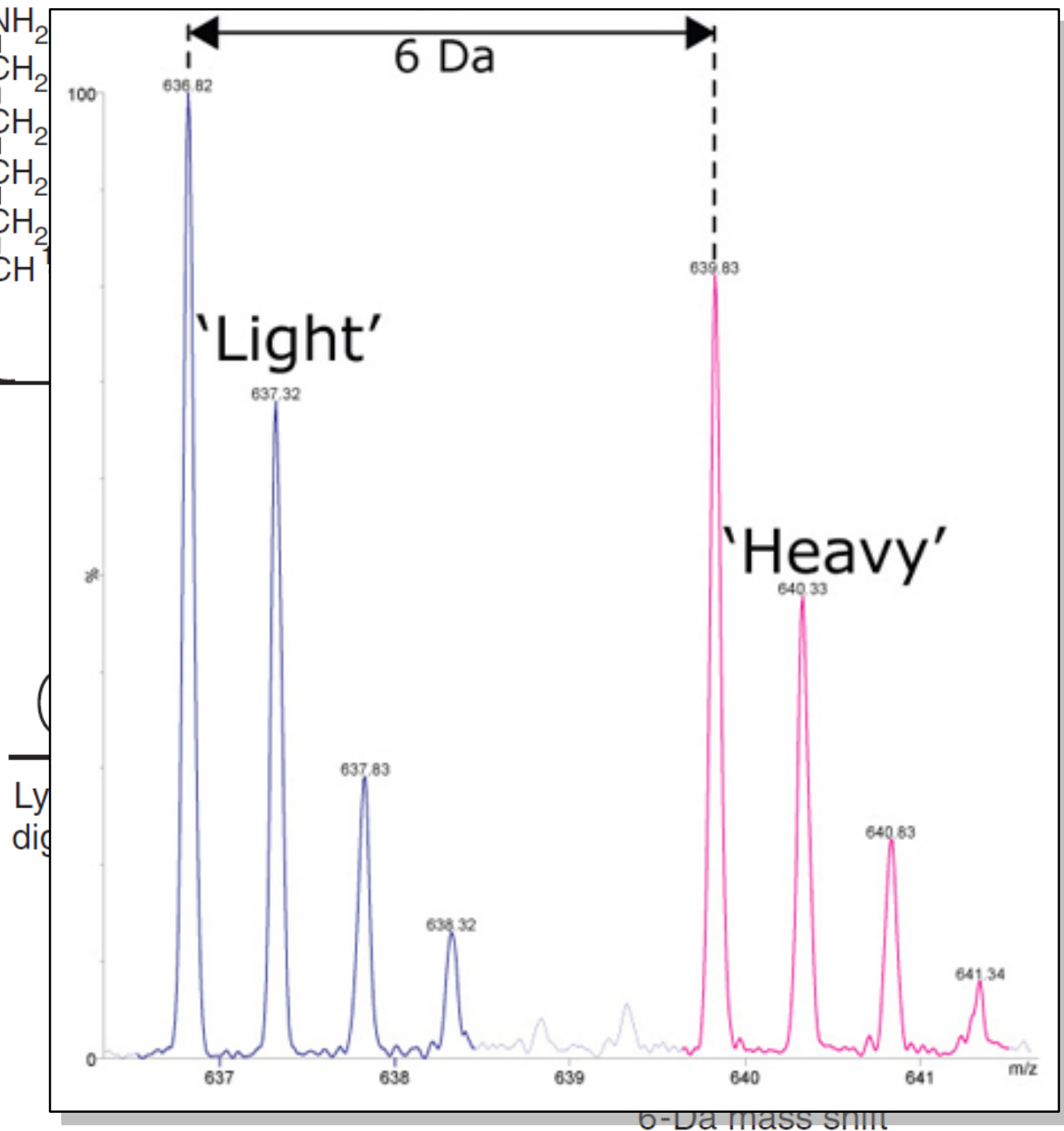


Cells in light,
 $^{12}\text{C}_6\text{-Lys}$ medium



mahspvavqvpqgmqnniadv
eelft^klerig^kgsfgev^k
ginrtqqvvaikⁱidleae
deiesilacdiqqeivlsq
cdssyvt^khask^kstyleyyg
syl^kgsklw...

Heavy proteins



SILAC

- dobra jakość wyników
- eliminuje błędy na wszystkich etapach izolacji białek
- większość peptydów znakowana
- działa na wszystkich spektrometrach
- ograniczone zastosowanie (zwykle *in-vitro*)
- dość duże zużycie znakowanych aminokwasów
→ koszty

Znakowanie metaboliczne i organizmy samożywne

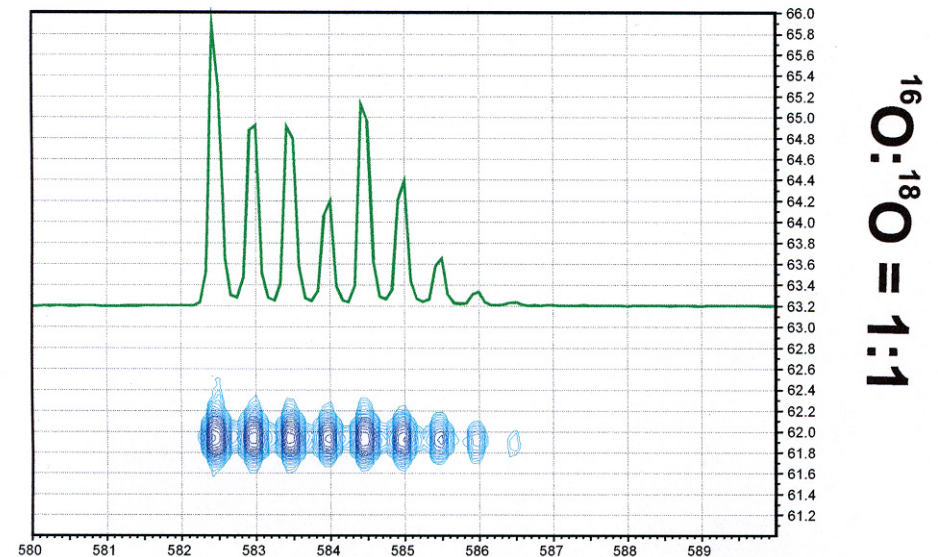
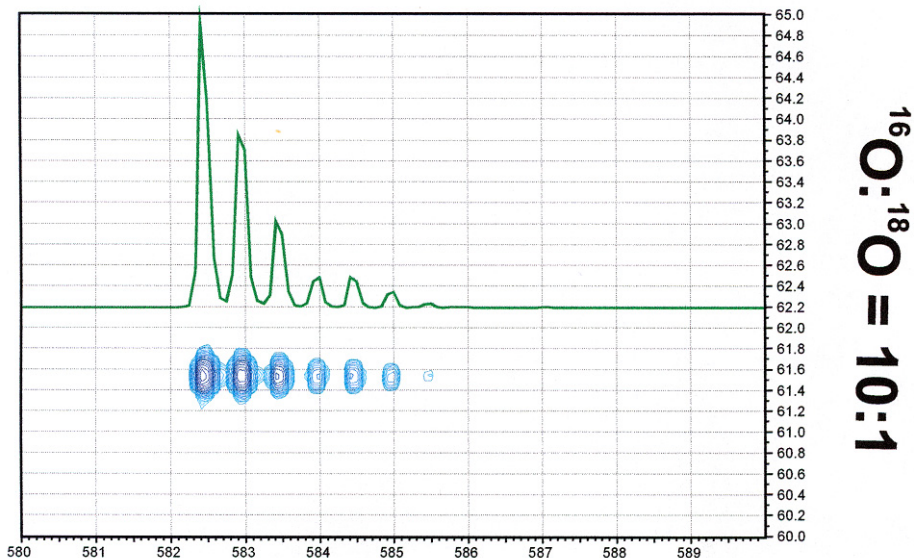
- Rośliny syntetyzują wszystkie aminokwasy
 - można dostarczyć nadmiaru aminokwasów do pożywki
 - mała wydajność inkorporacji (~80%)
- Zastąpienie ^{14}N przez ^{15}N w pożywce
 - możliwe *in-vitro* i np. w szklarni (hydroponika)

Znakowanie *in-vitro*

- Znakowane są wyizolowane białka lub peptydy
- Dodanie ^{18}O podczas trawienia trypsyną
- ICAT (Isotopically coded affinity tags)
- iTRAQ — znaczniki izobaryczne

Znakowanie ^{18}O

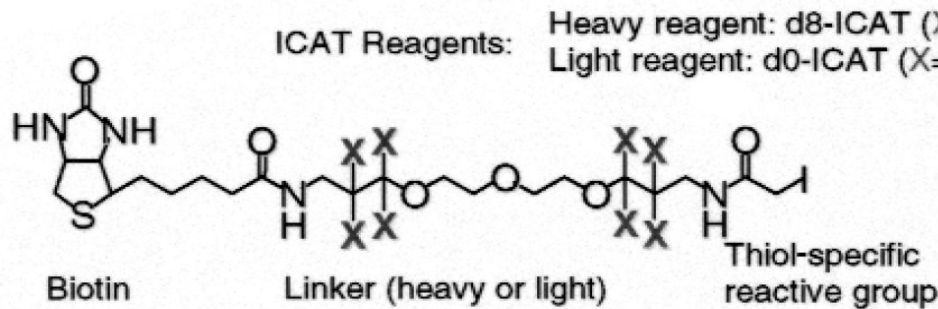
- Trypsyna podczas trawienia przyłącza dwa atomy tlenu do odtwarzanej grupy $-\text{COOH}$
 - Trypsyna wymienia atomy tlenu grup COOH lizyn i arginin na C-końcach peptydów.



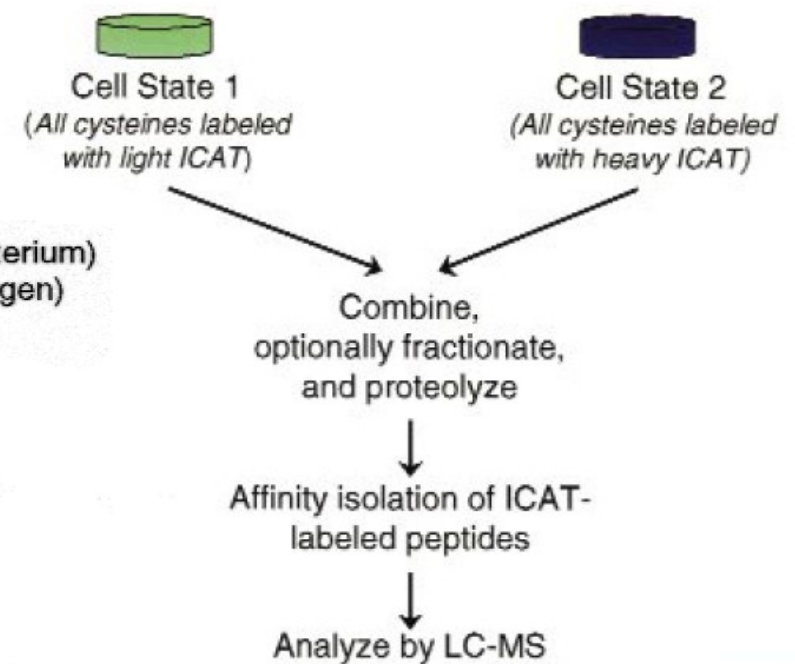
^{18}O

- znakowane prawie wszystkie peptydy
- znakowanie podczas trawienia białek lub znakowanie peptydów po trawieniu
- dobre porównanie ilości
- wymiana ^{18}O na ^{16}O
 - tripsyna w roztworze
- należy wymienić całość wody w próbce przed znakowaniem
 - np. liofilizacja
- H_2^{18}O dość drogie

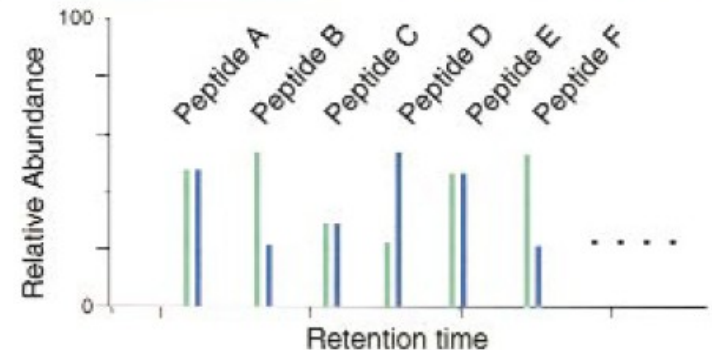
ICAT



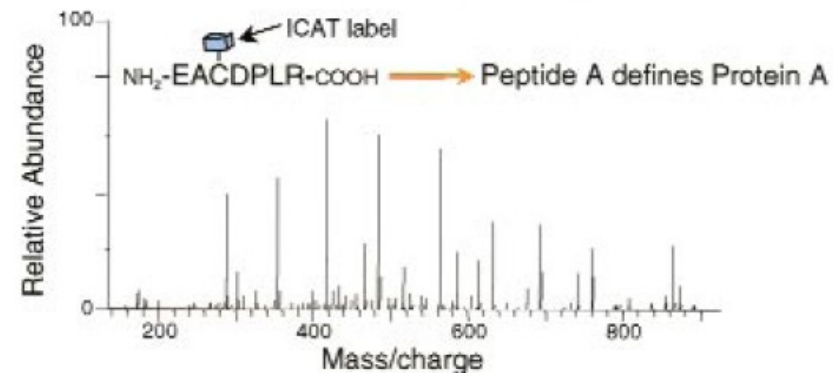
- Isotopically coded affinity tag
- Przyłącza się do zredukowanej cysteiny
 - Oczyszczanie
 - Niewiele peptydów wyznakowanych i wzbogaconych
- Zmienia się m/z i czas retencji



Quantitate relative protein levels by measuring peak ratios



Identify proteins by sequence information (MS/MS scan)

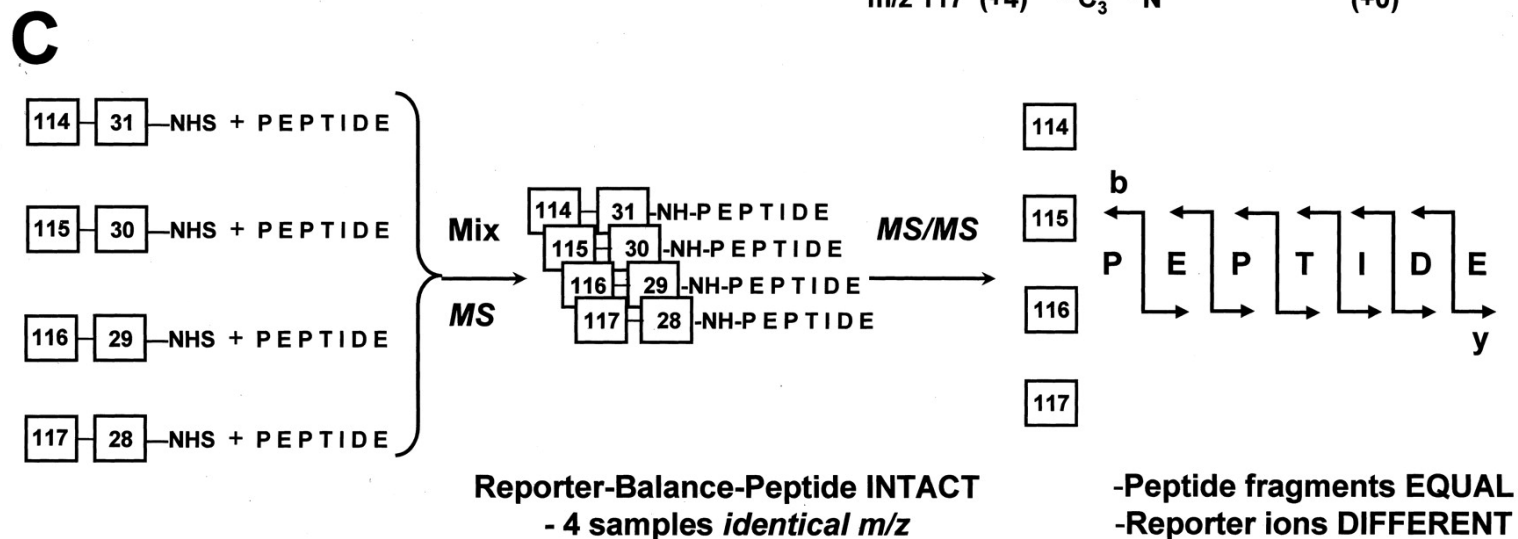
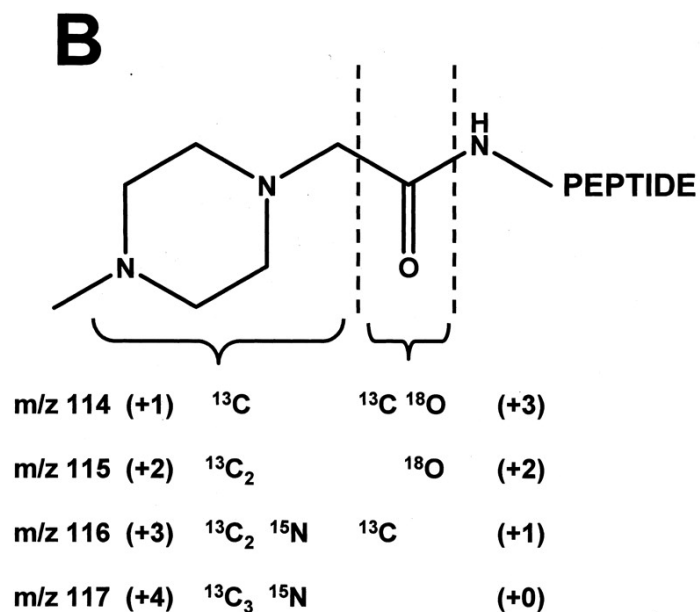
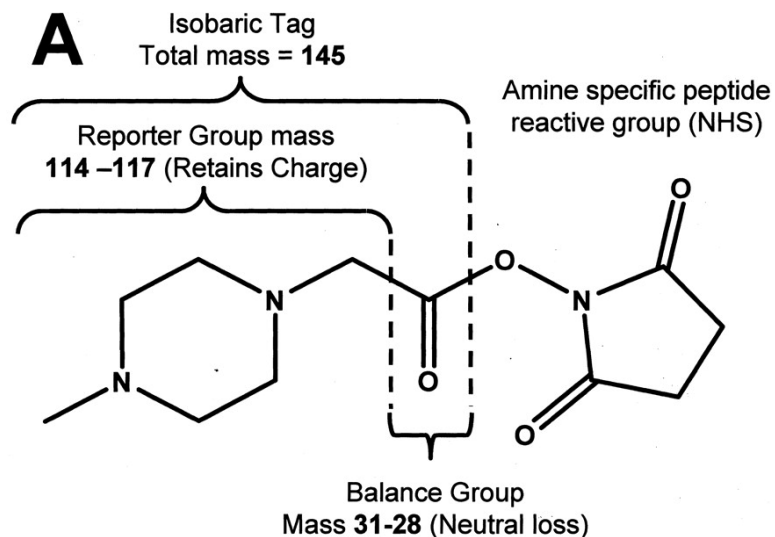


Znaczniki izobaryczne iTRAQ

Isobaric tag for relative and absolute quantitation

- Znacznik przyłącza się do grup aminowych
- Wiele wersji znacznika np. 4 lub 8
- masy poszczególnych wersji **równe!**
- przy fragmentacji powstaje jon reporterowy
 - z każdego znacznika powstaje jon o innej masie
- porównanie intensywności sygnałów jonów reporterowych na widmach MS/MS

A, diagram showing the components of the multiplexed isobaric tagging chemistry.



Ross P L et al. Mol Cell Proteomics 2004;3:1154-1169



iTRAQ

- dobra ocena ilości
- wiele prób na raz
- prostsze przetwarzanie danych niż w ^{18}O i SILAC
- znakowanie dość proste
- nie działa z fragmentacją w pułapce jonowej
- możliwa mniejsza czułość
 - dużo prób na raz
- wysoka cena

Oprogramowanie

- Przy metodach 'label-free' i znakowaniu izotopowym
 - Potrzebna analiza widm MS
 - Znajdowanie pików i ocena objętości
 - Nakładanie i porównywanie map z HPLC-MS w metodach typu AMT (eliminacja zakłóceń rozdziatu HPLC)
 - Większa objętość danych (np. ~1GB na przebieg)
 - przy identyfikacji używane tylko na początku analiz
- iTRAQ – podobnie, jak przy identyfikacji dokładniejsza analiza tylko widm MS/MS

Podsumowanie

- Metody klasyczne
 - 2D–PAGE
- Metody opierające się na LC–MS
 - bez znakowania ('label–free')
 - znakowanie metaboliczne
 - znakowanie wyizolowanych białek lub strawionych peptydów
- Optymalny wybór metody