

Fehérjék analitikája

11. előadás: Proteomika, peptidek és fehérjék tömegspektrometriája, bioinformatika

Előadó: Kecskeméti Ádám

Irodalom:

Pradip K. Ghosh – Introduction to Protein Mass Spectrometry, 4. és 6. fejezet

John R. Griffiths, Richard D. Unwin - Analysis of Protein Post-Translational Modifications by Mass Spectrometry - Introduction

John Cottrell - Database Searching for Protein Identification and Characterization (előadás) (http://www.matrixscience.com/pdf/asms_tutorial_2005.pdf)

Proteolízis

- Tripszin: Arg és Lys
 - Kimotripszin: Tyr, Phe és Trp, de Leu, Met, Ala, Asp, Glu is (lassabban)
 - Endoproteináz Arg-C, Lys-C, Asp-N, Glu-C, K (alifás/aromás)
 - Karboxipeptidáz A, B, Y, M
 - Papain (nemspecifikus)
 - Pepszin (nemspecifikus)
-
- Az emésztés jellege a keletkező peptidek méretét befolyásolja
 - Top-down; middle-down; bottom-up proteomika

Fehérjék szerkezete

Nonpolar

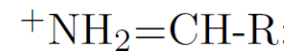
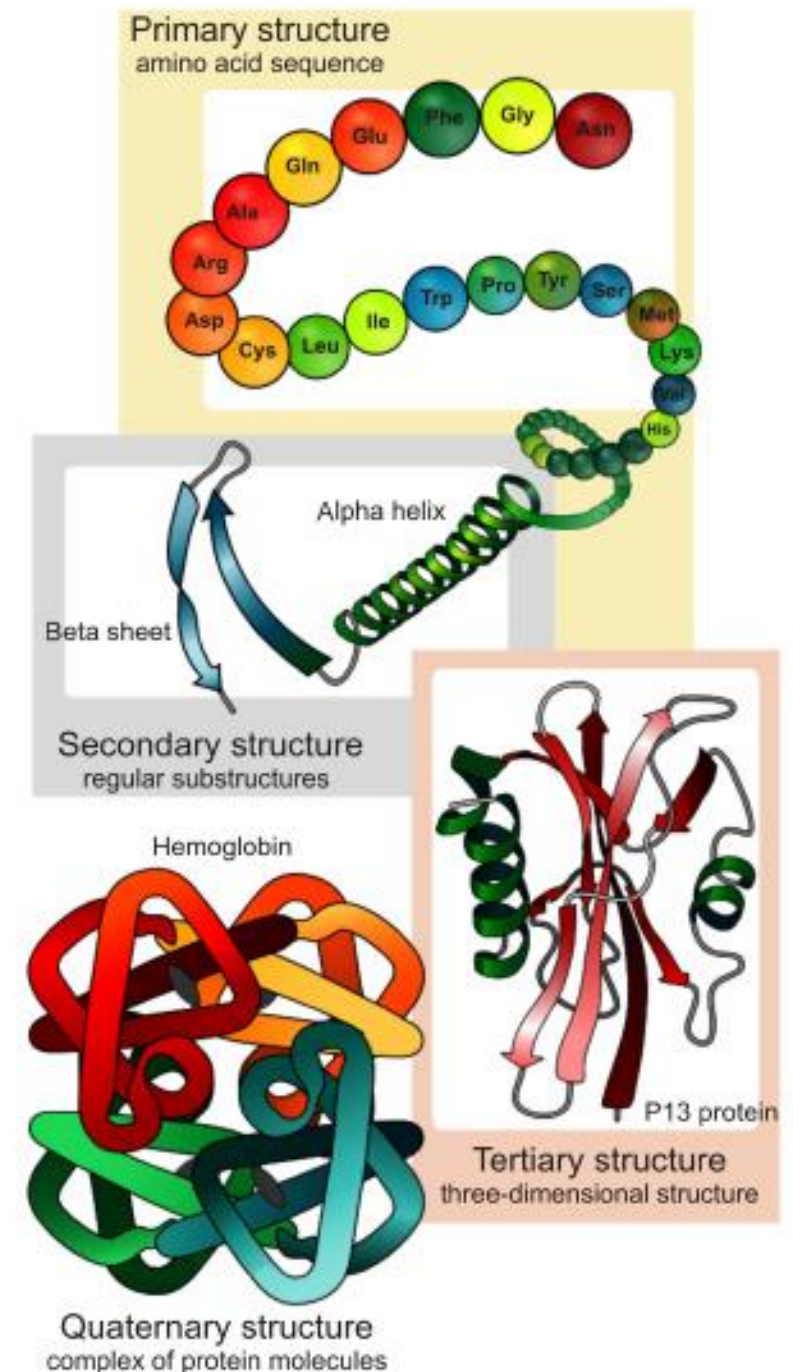


Table 4.1: Amino Acid Residue Names, Masses, and Immonium Ions.

Residue	1-Letter Code	3-Letter Code	Residue Mass	Immonium Ion (m/z)
Alanine	A	Ala	71.04	44.05
Arginine	R	Arg	156.10	129
Asparagine	N	Asn	114.04	87.09
Aspartic Acid	D	Asp	115.03	88.04
Cysteine	C	Cys	103.01	76
Glutamic Acid	E	Glu	129.04	102.06
Glutamine	Q	Gln	128.06	101.11
Glycine	G	Gly	57.02	30
Histidine	H	His	137.06	110.07
Isoleucine	I	Ile	113.08	86.1
Leucine	L	Leu	113.08	86.1
Lysine	K	Lys	128.09	101.11
Methionine	M	Met	131.04	104.05
Phenylalanine	F	Phe	147.07	120.08
Proline	P	Pro	97.05	70.07
Serine	S	Ser	87.03	60.04
Threonine	T	Thr	101.05	74.06
Tryptophan	W	Trp	186.08	159.09
Tyrosine	Y	Tyr	163.06	136.08
Valine	V	Val	99.07	72.08



Peptidhasadás MS/MS-ben

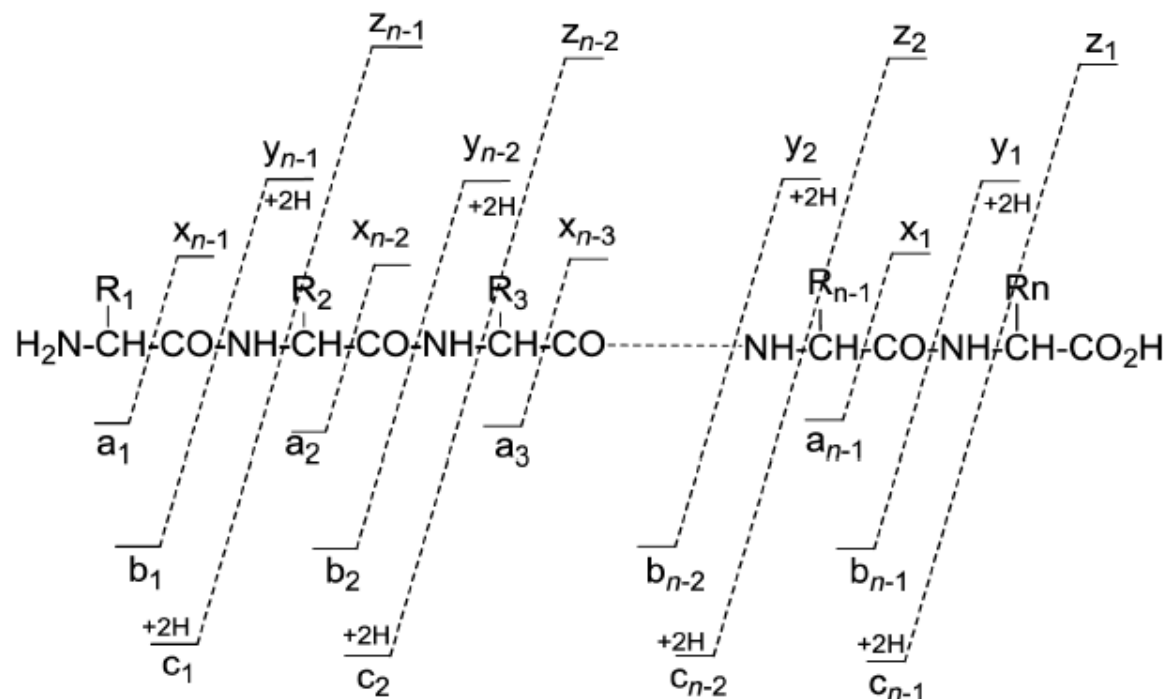
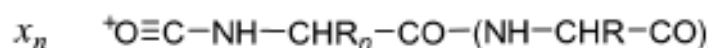
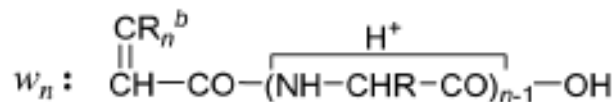
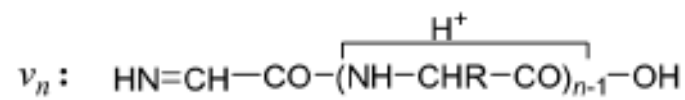
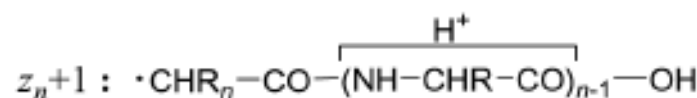
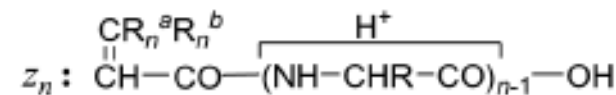
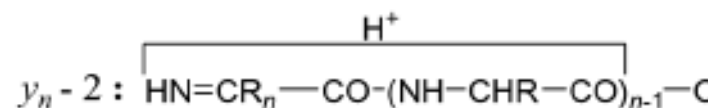
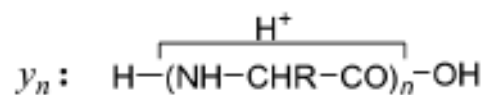
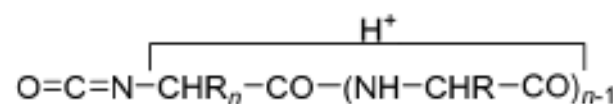


Figure 4.6: Nomenclature for peptide CID fragment ions (first proposed by Roepstorff and Fohlman,¹⁹⁶ later modified by Biemann¹⁹⁷). Typically, all of the above fragment ion types are observed in high-energy CID tandem mass spectra (although usually not in the same spectrum), but only *b*, *y*, and, less frequently, *a* and *z* fragments are observed in low-energy CID spectra.¹⁹⁹ Reproduced by permission of John Wiley & Sons, Inc. from *Mass Spectrom. Rev.*, Vol. 14, 49–73 (1995).

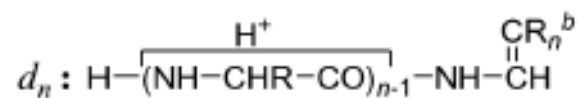
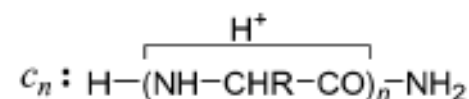
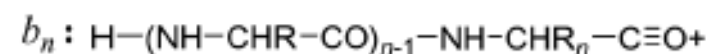
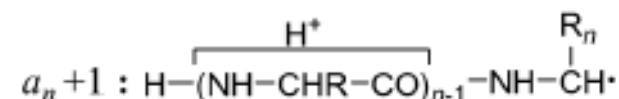
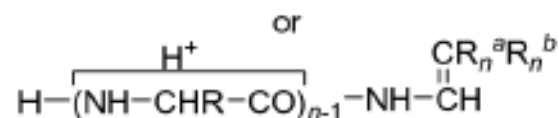
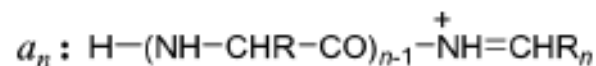
C-terminal ion types



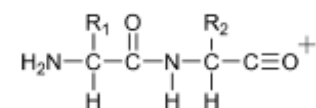
or



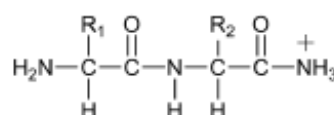
N-terminal ions



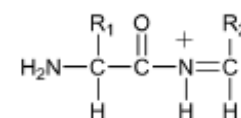
Sequence ions



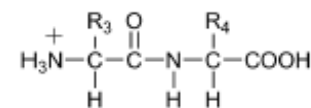
b₂



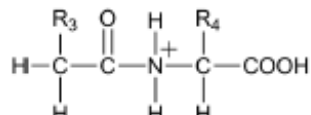
c₂



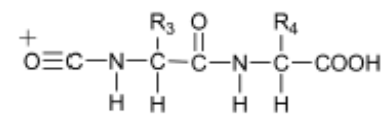
a₂



y₂

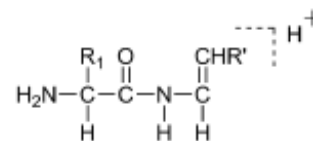


z₂

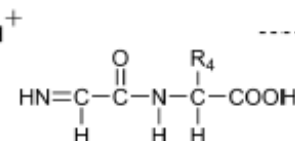


x₂

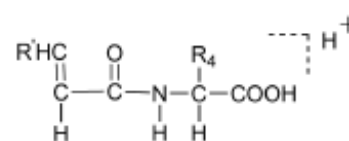
Satellite ions



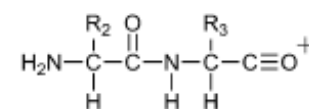
d₂



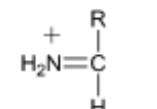
v₂



w₂



Internal ion



Immonium ion

Figure 4.7: Left: N-terminus ions; right: C-terminus ions. (R represents the chains of the amino acids; R_n^a and R_n^b are the beta substituents of the n th acid.) Reproduced by permission of Elsevier Limited from *Methods in Enzymology*, Vol. 193, Appendix 5. 1990.

Figure 4.9: Some fragment ions from a tetrapeptide. Ion types: $a \rightarrow [\text{N}]+[\text{M}]-\text{CHO}$; $b \rightarrow [\text{N}]+[\text{M}]-\text{H}$; $c \rightarrow [\text{N}]+[\text{M}]+\text{NH}_2$; $x \rightarrow [\text{C}]+[\text{M}]+\text{CO}-\text{H}$; $y \rightarrow [\text{C}]+[\text{M}]+\text{H}$; $z \rightarrow [\text{C}]+[\text{M}]-\text{NH}_2$; $a^* \rightarrow a-\text{NH}_3$; $b^* \rightarrow b-\text{NH}_3$; $y^* \rightarrow y-\text{NH}_3$; $a^\circ \rightarrow a-\text{H}_2\text{O}$; $b^\circ \rightarrow b-\text{H}_2\text{O}$; $y^\circ \rightarrow y-\text{H}_2\text{O}$; $d \rightarrow a$ -partial side chain; $v \rightarrow y$ -complete side chain; $w \rightarrow z$ -partial side chain.

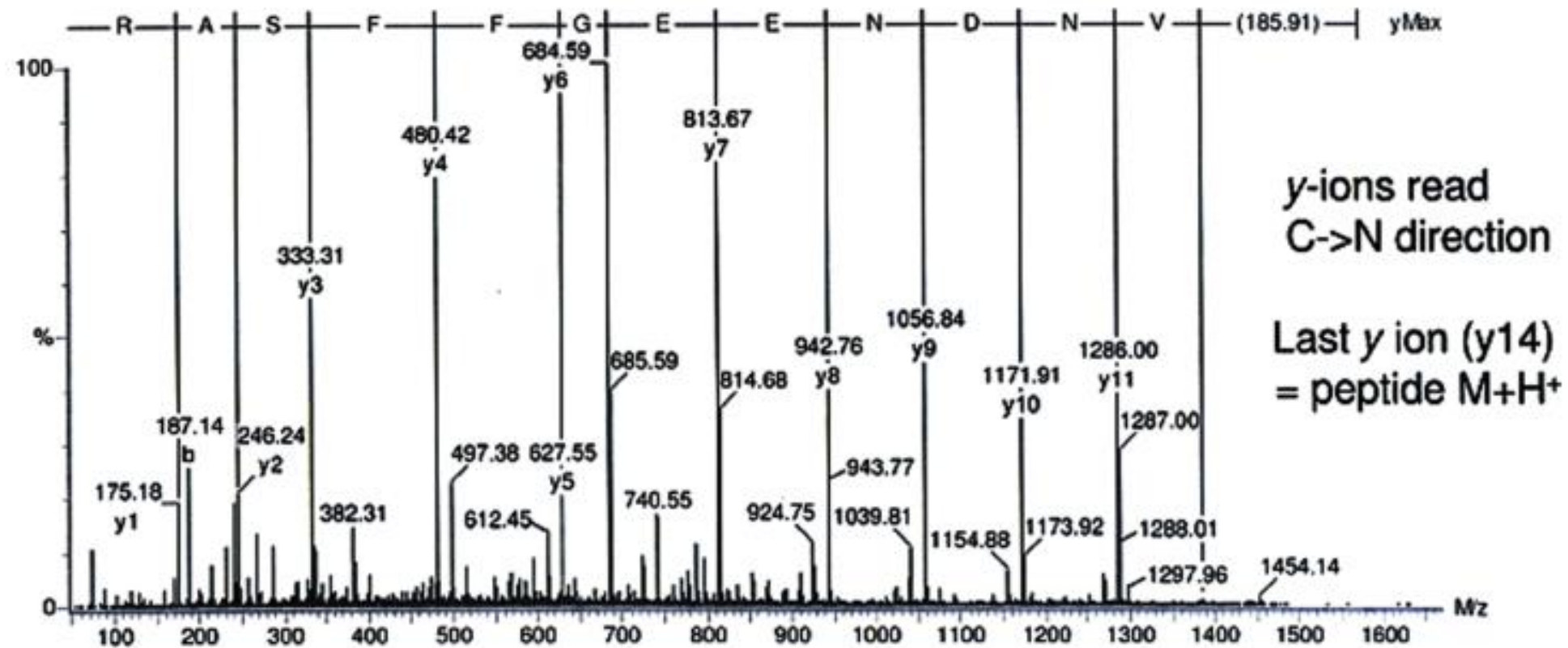
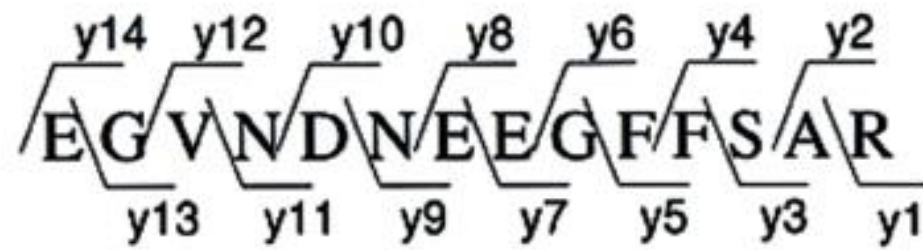


Figure 4.10: Low energy CID (Q-TOF) of doubly charged Glu-fibrinopeptide: *y*-series ions. Courtesy: Kevin Blackburn, Molecular and Structural Biochemistry, North Carolina State University. Reproduced by permission.

Intakt fehérje tömegspektrometriája

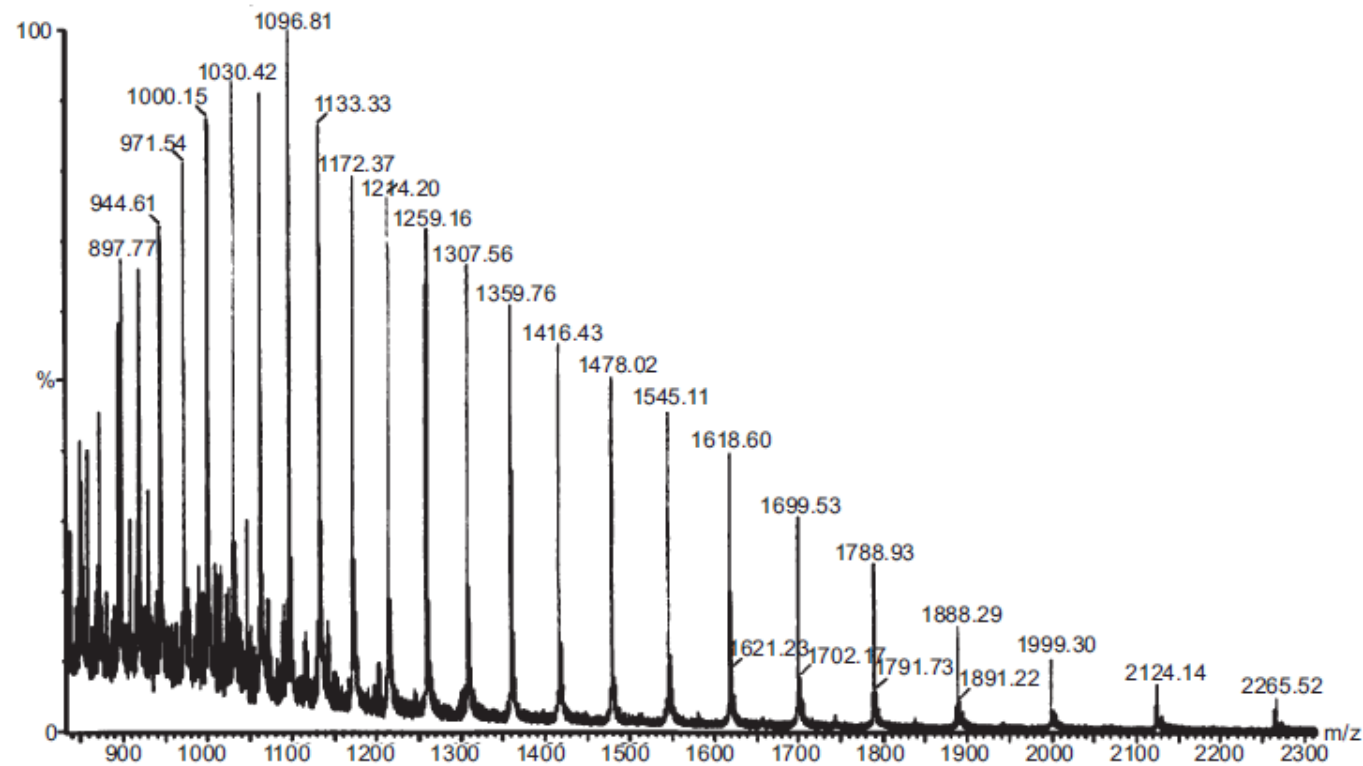


Figure 4.13: ESI spectrum of an intact protein. Courtesy: Kevin Blackburn, Molecular and Structural Biochemistry, North Carolina State University. Reproduced by permission.

Töltés és molekulatömeg kiszámítása

- Két szomszédos m/z értékére kell felírni az egyenletet (töltéskülönbségük 1), ami alapján töltés és molekulatömeg számítható

- $M(H^+) = 1,0079 \text{ Da}$

- Kisebb m/z -re: $\frac{m + (z + 1)M_{H^+}}{z + 1} = \left(\frac{m}{z}\right)_{z+1}$

- Nagyobb m/z -re: $\frac{m + zM_{H^+}}{z} = \left(\frac{m}{z}\right)_z$

$$z = \frac{\left(\frac{m}{z}\right)_{z+1} - M_{H^+}}{\left(\frac{m}{z}\right)_z - \left(\frac{m}{z}\right)_{z+1}}$$

$$m = \frac{\left[\left(\frac{m}{z}\right)_z - M_{H^+} \right] \cdot \left[\left(\frac{m}{z}\right)_{z+1} - M_{H^+} \right]}{\left(\frac{m}{z}\right)_z - \left(\frac{m}{z}\right)_{z+1}}$$

SRM, MRM

- SRM – selected reaction monitoring
 - Kiválasztott tömegnél fragmentálunk egy iont, majd következő lépésben a fragmentáció **egy** kiválasztott termékionját detektáljuk.
- MRM – multiple reaction monitoring
 - Kiválasztott tömegnél fragmentálunk egy iont, majd következő lépésben a fragmentáció **több** kiválasztott termékionját detektáljuk.
 - Nem szükséges teljes scan-t végezni, így gyorsabb analízis végezhető vele (csak a kiválasztott tömegeket követi)

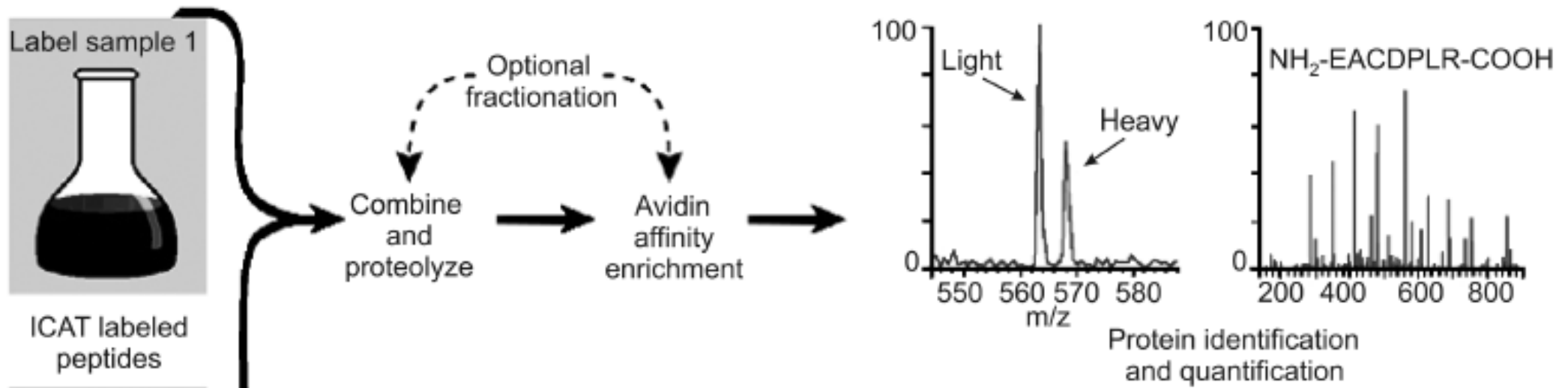
Kvantitatív proteomika – ICAT

- Isotope-coded affinity tags
- 2 elv:
 - Szomszédos aminosavak rövid szekvenciájának ismerete (5-25 aminosav) lehetővé teszi egyedi fehérjék azonosítását
 - Ha két kémiaiailag azonos, de eltérő izotópot tartalmazó jelölést alkalmazunk, akkor a jelölt molekulák elektroforetikus/kromatográfiás viselkedése megegyezik, míg az eltérő izotópok jelenléte miatt más tömegnél kaphatunk tömegcsúcsot. Ez pedig pontos kvantitálást tesz lehetővé.

ICAT lépései

- ICAT reagensek felépítése
 - Affinitás jelölő (pl. biotin) - affinitás dúsítást/izolálást tesz lehetővé jelölés után
 - Linker: stabil izotópot tartalmazza
 - Könnyű reagens (^1H -eket tartalmaz)
 - Nehéz reagens (8db ^2D -t tartalmaz)
 - Reaktív csoport –SH csoportok iránti specificitással

Könnyű
ICAT



Nehéz
ICAT

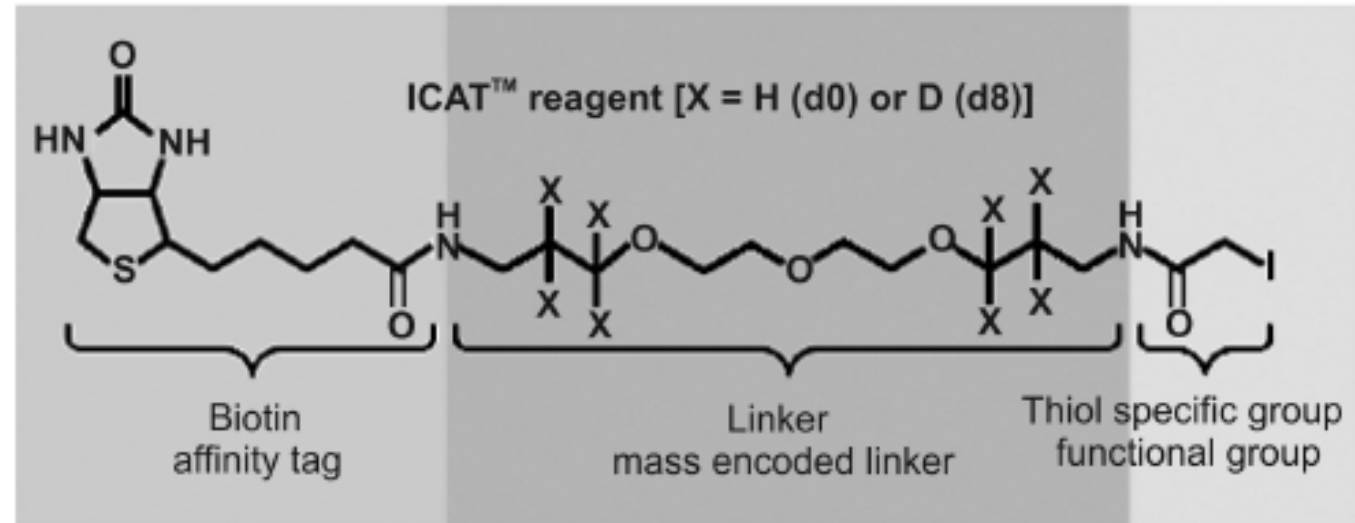
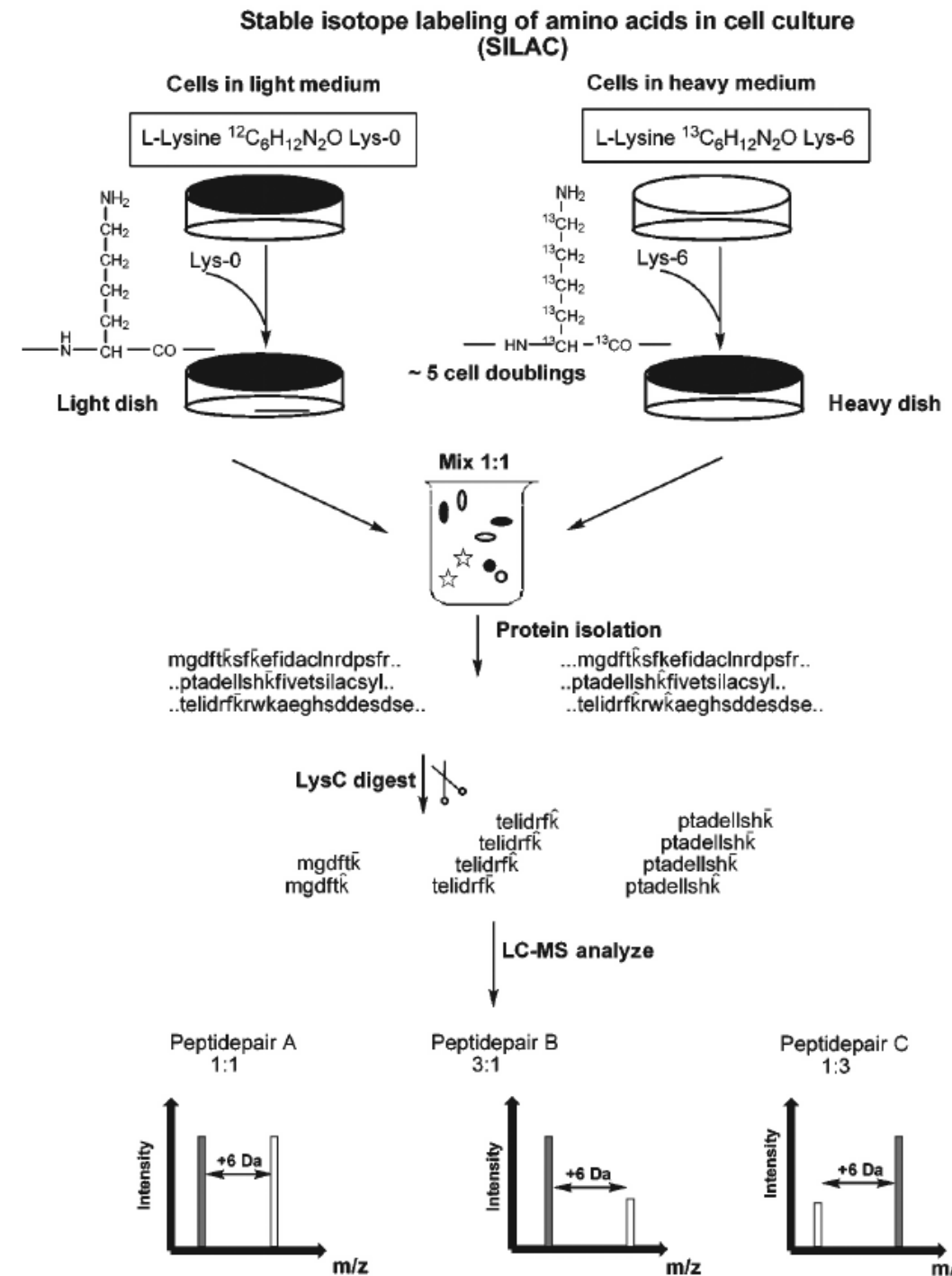


Figure 4.14: Quantitative proteomics with ICAT reagents. Source: www.imsb.ethz.ch/researchgroup/rudolfa/research. Courtesy: Ruedi Aebersold, Institute of Molecular Systems Biology, ETH Zürich. Reproduced by permission.

Kvantitatív proteomika – SILAC

- Stable isotope labeling by amino acids in cell culture
- Emlős sejtvonalak tenyésztése nem „hagyományos” aminosavakból zajlik, hanem stabil izotóp jelölésű (nem radioaktív) aminosavakból
- Pl. deuterált Leu, de ^{13}C és ^{15}N izotópok beépítése is lehetséges
- Az izotópjelölt aminosavakat tartalmazó közeg nem befolyásolja a sejtvonal növekedését
- Ezen fehérjepopulációkat kontrolhoz lehet hozzáadni, mert a jelölés „az aminosavakba van kódolva”, a jelölt/nem jelölt fehérjék aránya állandó marad
- A Leu-d0 – Leu-d3 tömegeltolódást kell követni, a csúcsok arányából lehet kvantizni

Munkafolyamat - SILAC

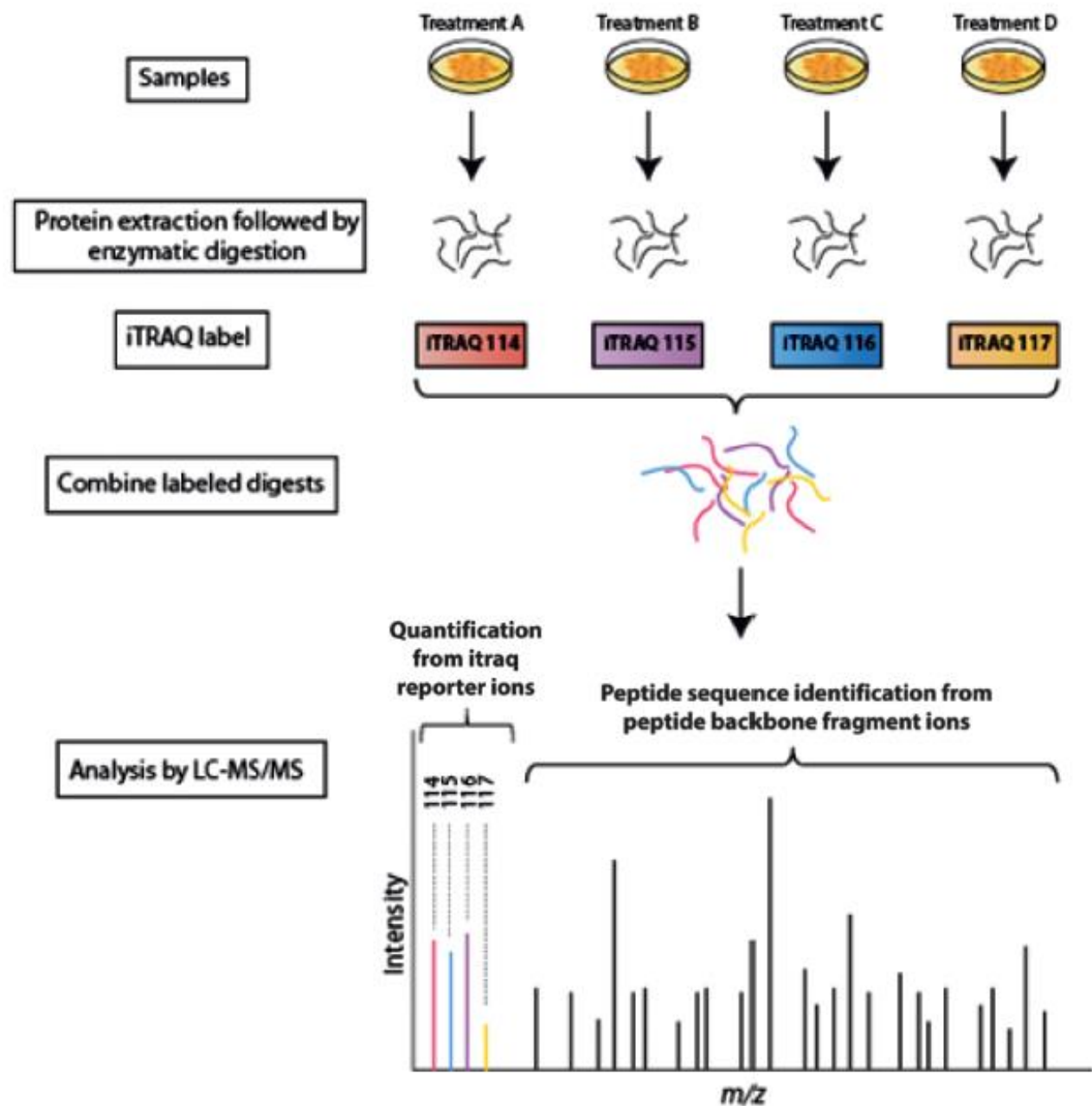
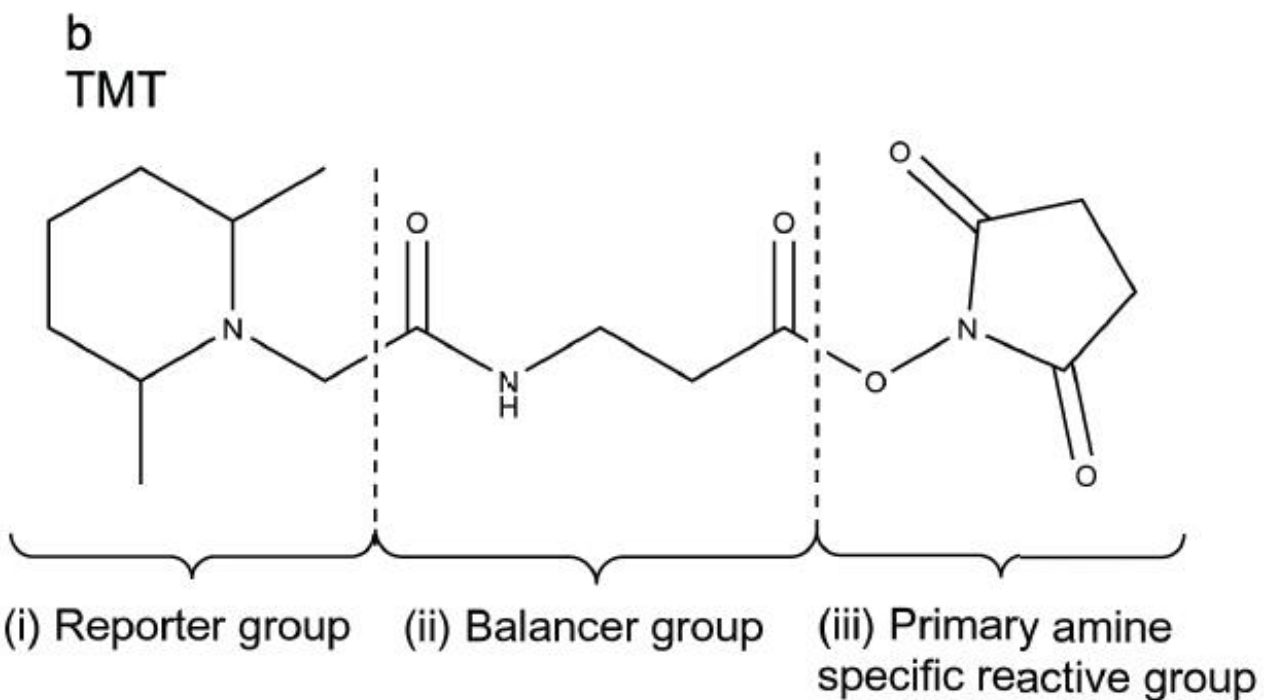
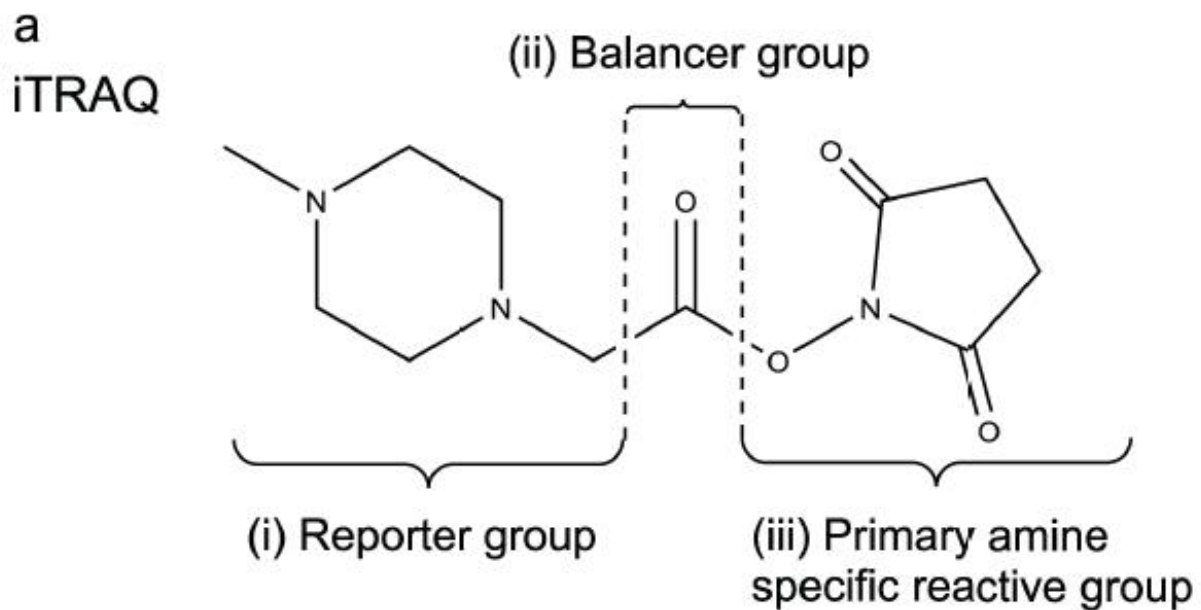


Kvantitatív proteomika – iTRAQ

- Isobaric tag for relative and absolute quantitation
- Nagyszámú mintához jobb
- Munkafolyamat
 - Sejt lizálás + összfehérje meghatározás
 - redukálás – alkilálás – enzimes emésztés (puffer aminmentes – primer és szekunder aminok kvencselik az iTRAQ jelölési reakciót)
 - 70% szerves oldószerre (alkohol) való beállítás
 - **Jelölés izobár iTRAQ reagenssel** (minden mintát külön)
- CID fragmentálás: ajánlott 10-15%-kal növelni a fragmentációs energiákon a natív peptideknél alkalmazotthoz képest

iTRAQ

- Reagens felépítése
 - Reporter csoport (riporter) (pl. N-Me-piperazin, M=114, 115, 116 és 117)
 - Balance csoport (kiegyensúlyozó) (pl. M=31, 30, 29, 28)
 - Reaktív csoport (pl. N-hidroxiszukcinimid)
- Minták jelölése után összekeverik őket, majd (LC)-MS/MS analízisnek teszik ki őket
- Relatív intenzitások megkaphatók a riporter ion relatív intenzitásaiból, az MS/MS spektrumban
- Max 8 minta kombinálható a megfelelő reagenssel (8-plex iTRAQ)



Kvantitatív proteomika – TMT

- Tandem mass tag
- Hasonló az iTRAQ-hez
- Tandem: a jelölő tag-ek vizsgálata kizárólag MS/MS-sel történik
- Nincs a mintában jelöletlen peptid
- Jelölő felépítése:
 - Érzékenységnövelő csoport (sensitization group)
 - Tömegkülönbséget okozó csoport (mass differentiation group)
 - Tömeg normalizációs csoport (mass normalization group)
 - Reaktív csoport

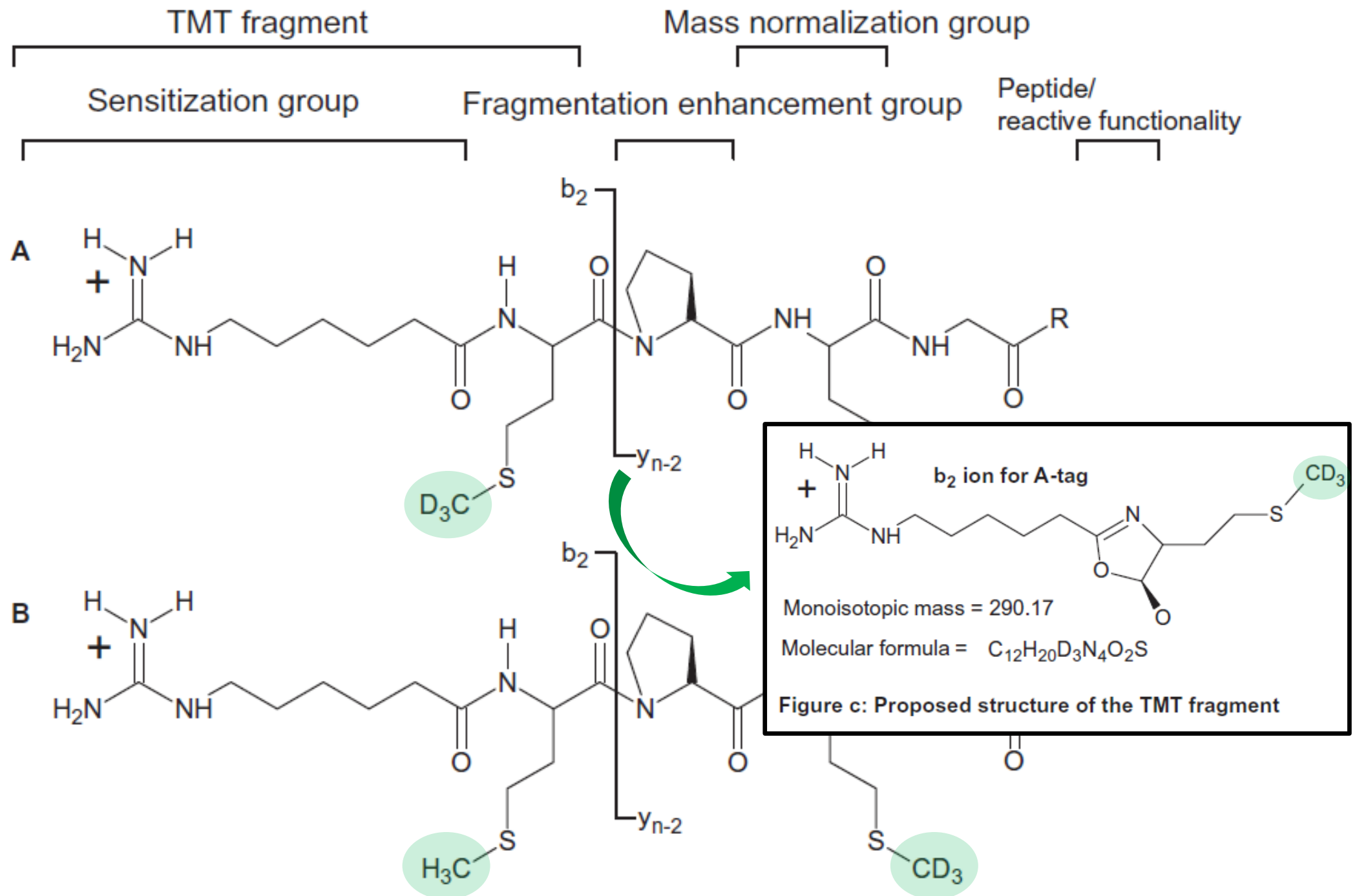


Figure b: Second generation TMT

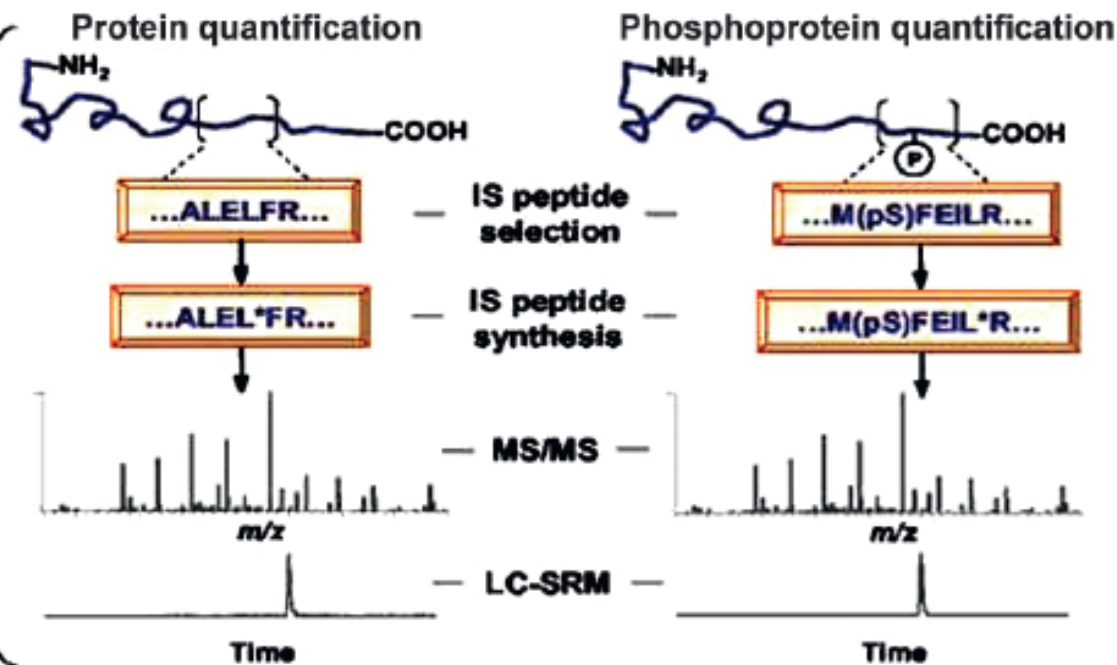
Kvantitatív proteomika – AQUA

- Abszolút kvantifikálás SRM vagy MRM-en alapszik
- MRM (maxis II)
 - Max. 100 kiválasztott tömeget tud követni, MS³ is lehetséges isCID energia segítségével
 - isCID-ként beállított energia jóval magasabb, mint a finomhangolási paraméterek között
 - Ütközési energia megadása
 - X Acq.: érzékenyítés, hányszor legyen nagyobb az MS/MS spektrumok felvételénél a mintaidő
- AQUA peptid: szintetikus peptid, izotóppal jelölve, néhány Da-nal nehezebb, mint az a peptid, amit vizsgálunk – tkp. Belső standard

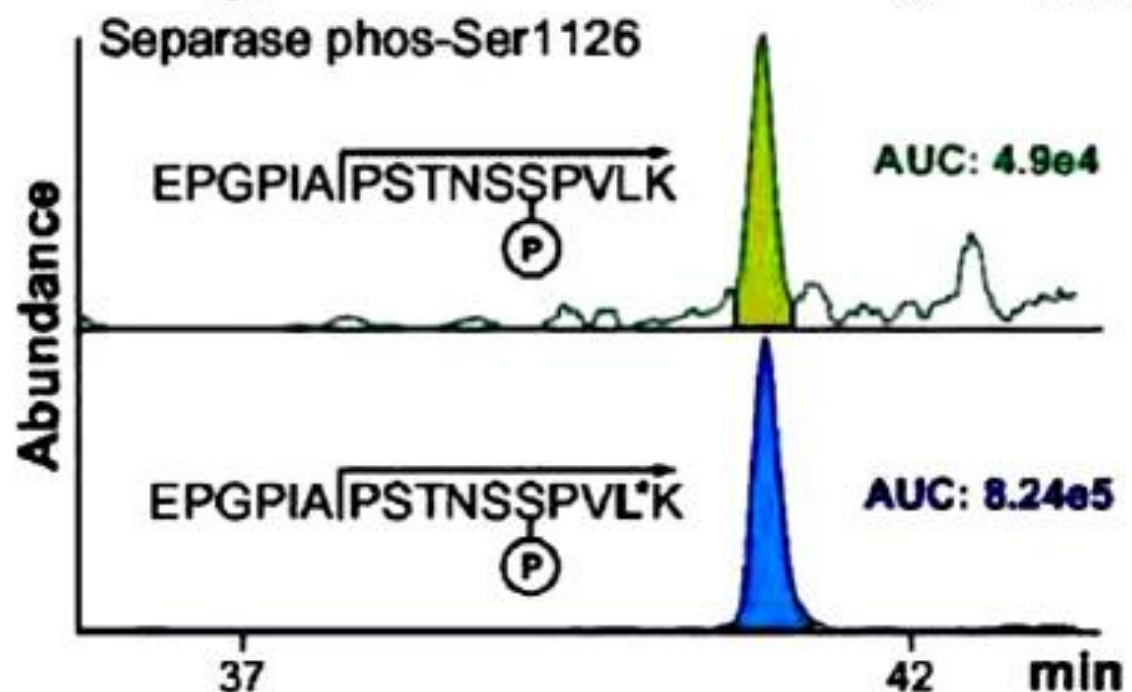
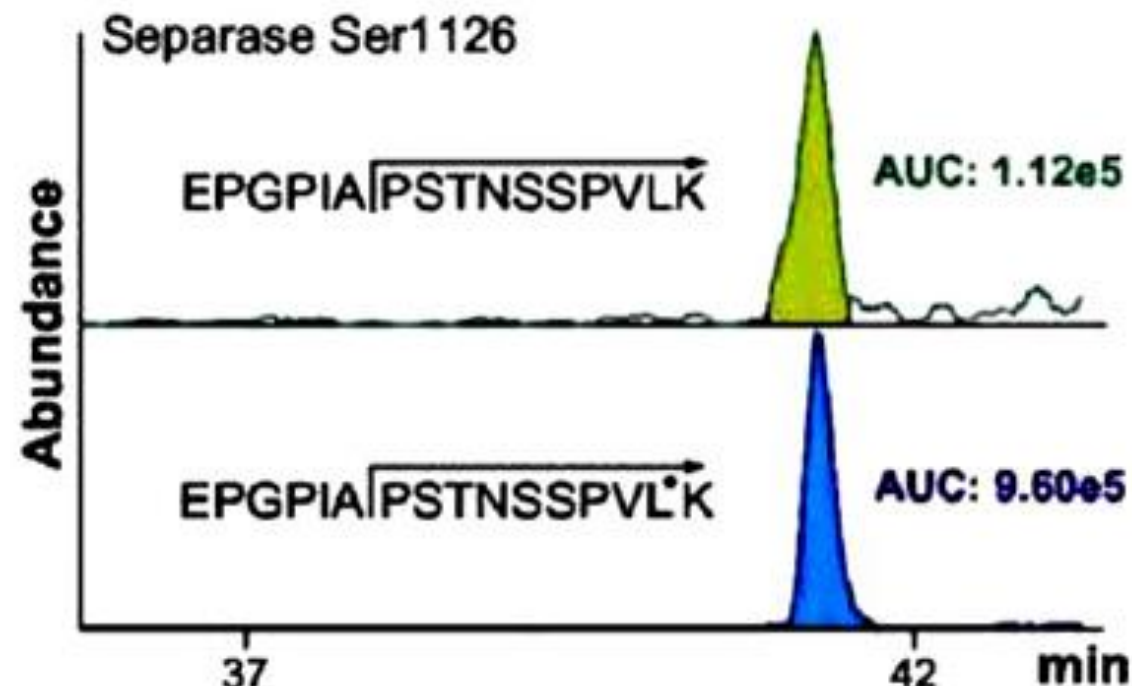
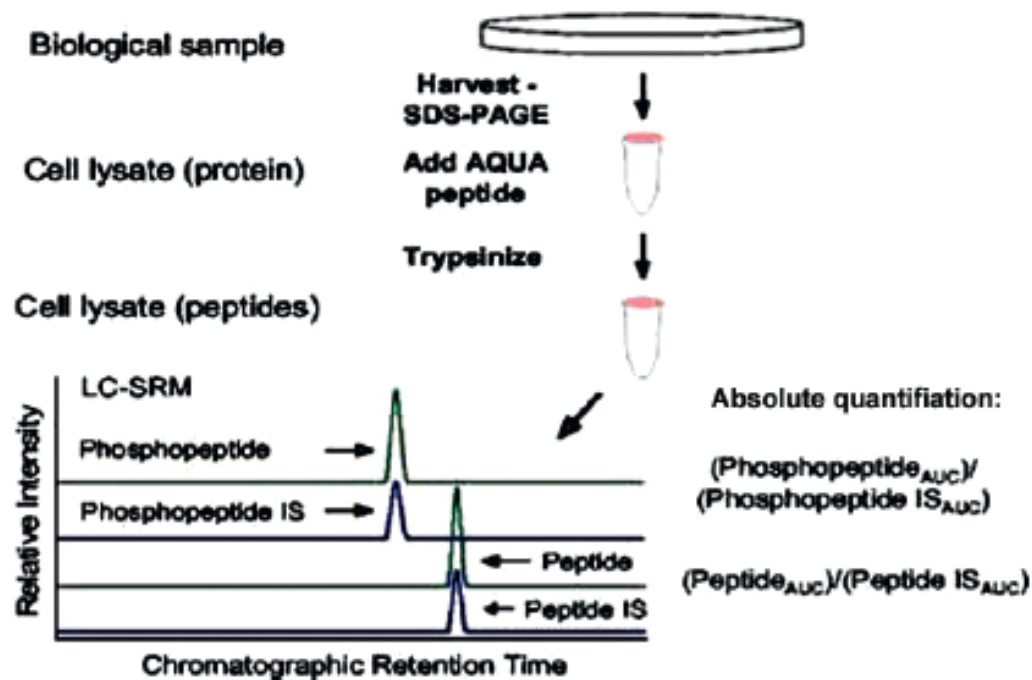
AQUA peptid szintézis

- Egyedi peptidet kell választani, melynek minden formáját (PT módosított) kell egyidejűleg mérni
- Fehérjénként 2/3 peptid
- Jól ismert (fragmentáció, elválaszthatóság, SRM átmenetek) peptidet érdemes választani
- ^{13}C és ^{15}N izotópokkal történik a jelölés
- A szintetizált peptidet lemérik, a prekursor MS/MS spektrumában keresni lehet esetleg jelen lévő – prekurzornál intenzívebb – ionokat, amik alkalmasabbak lehetnek az analízishez (jobb S/N arány)

Stage 1: AQUA peptide development



Stage 2: Implementation



Peptid fragmentáció – MS/MS módszerek

- **CID**

- Ütközési cellában

- alacsony ütközésszám, N₂ vagy Ar
 - 20-100 eV energia

- Csapdában

- magas ütközésszám (100<), He
 - 20-100 eV energia

- ECD (electron capture dissociation – elektronbefogás dissz.)

- Szabad elektronokkal kölcsönhatás – fragmentáció

- EDD (electron detachment dissociation)

- A kölcsönható elektron lelök e⁻-okat az ionokról

Peptid fragmentáció – MS/MS módszerek

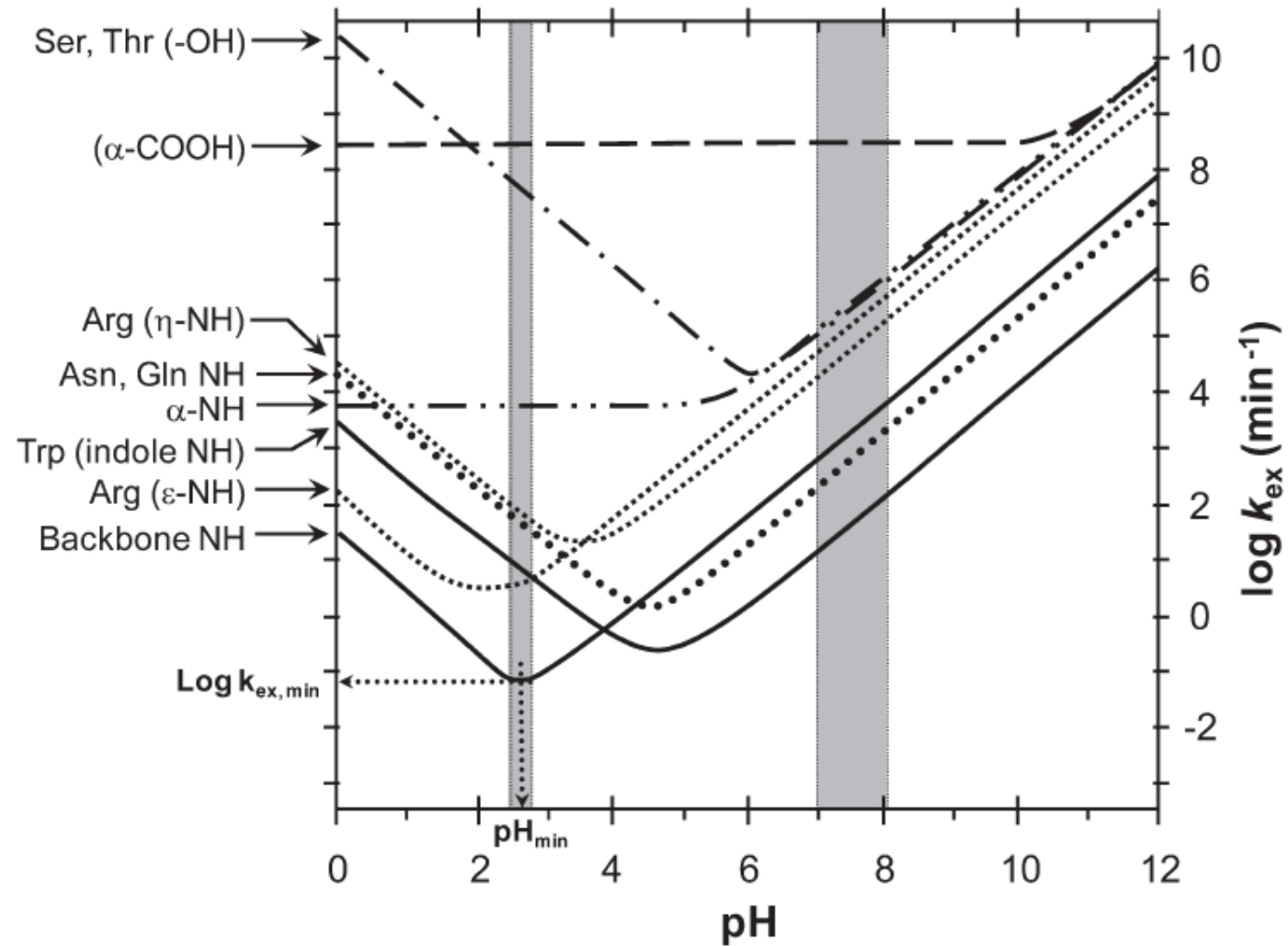
- **ETD (electron transfer dissociation)**
 - Reagens aniont abszorbeálja a prekursorion, amit fragmentáció követ
 - Negative ETD (NETD)

H-D csere

- Fehérjék konformációjára lehet ezáltal következtetni
- Minél kitékertebb a fehérje szerkezete, annál több labilis proton tud lecserélődni (partner pl. D_2O , CH_3COOD vagy CH_3OD)
- A cserefolyamat visszafelé is gyors – kvencselés szükséges (savanyítás és lehűtés $0\text{ }^\circ\text{C}$ -ra)
- Emésztés ilyen körülmények között
 - Tripszin nem jöhet szóba
 - Általában pepszinnel (nem specifikus)
 - Online reaktorok, redukáló és alkilezőszerek használata elterjedtebb a jó emésztési hatékonyság végett
 - Általában időben (és eltérő T-eken) vizsgálják a H-D csere lefolyását

H-D csere

- H-D csere sebessége a pH függvényében.
- Szürke régió: az a pH, ahol a H-D csere, ill. a kvencselés történik



H-D csere

- Fehérje-fehérje kölcsönhatás vizsgálata
 - Receptor fehérje hozzáférhető amid NH-kon H-D csere történik
 - Liganddal kölcsönhatásba lép (komplexet képeznek)
 - Nem deuterált oldószer hatására D-H (vissza)csere játszódik le ott, ahol az NH a komplexben is elérhető, míg a D a fehérjén marad ott, ahol a ligandfehérje meggátolja a cserereakciót
 - Akkor alkalmazható, ha a kölcsönhatás erős

Peptid fragmentációs mechanizmusok

- Kérdés: az energiát felvett prekursorion miként használja el az extra energiáját, azaz okoz helyenként kötéshasadást?
- A mechanizmussal megválaszolható a fragmentációs tömegspektrumon a csúcsok helyzete és intenzitása

Mobilis proton modell

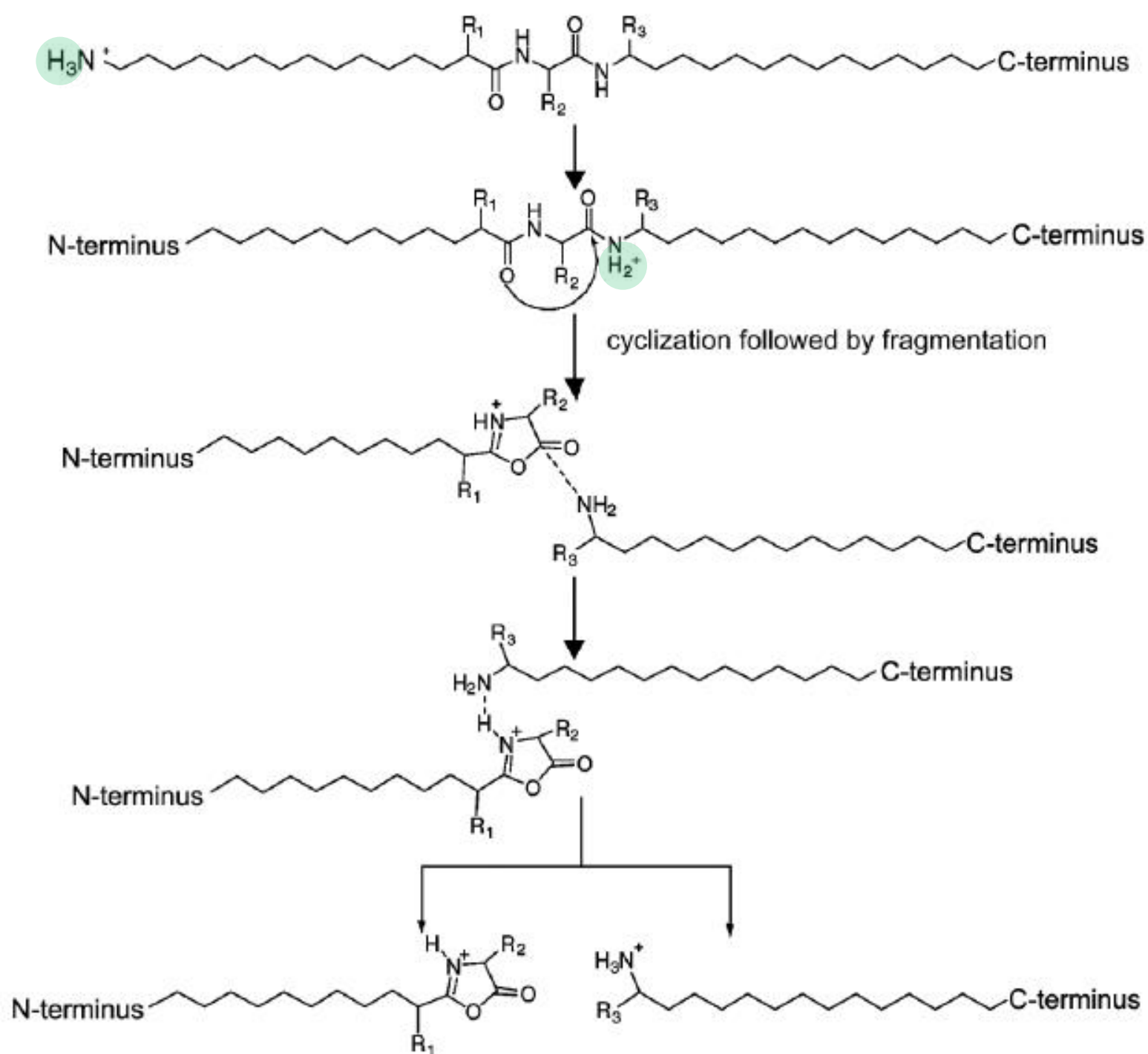
- ESI-ben a peptidek 1 vagy több H^+ -ra tesznek szert, amelyek valamely bázikus sajátságú helyen (láncon vagy oldalcsoporton) helyezkednek el (ált. N-atomon)
- Arg, Lys és Hys – nagy protonaffinitás; ha jelen vannak, azt feltételezzük, hogy a H^+ nem fog mozogni, hiányukban viszont a peptidlánc valamely N-jére kerülnek. Hasonló a helyzet, ha 2 H^+ van, de csak 1 Arg, Lys vagy Hys – a 2. H^+ a láncra kerül – mobilis lesz
- A H^+ jelenléte megbontja az amidkötés rezonanciastabilitását, és ez okozza az adott helyen a kötés felbomlását
- Hova kerül a H^+ ?
 - Bármely lánc N szóba jöhet
 - Terminális N

Peptid fragmentációs útvonalak

- 2 fő – versengő – kategóriába sorolhatók a peptid fragmentációs útvonalak
 - **Töltéssel irányított fragmentáció (charge directed)**
 - Nem szekvencia jellegű
 - Szekvencia jellegű fragmentáció
 - a_1-y_x
 - diketopiperazin y_x
 - oldalláncbéli fragmentáció
 - b_x-y_z
 - b_x-b_x-1
 - b_x-a_x
 - immónium fragmentáció
 - belső fragmentációs útvonal
 - **Töltéstől távoli fragmentáció (charge remote)**
 - Asp effekt
 - ox. met effekt

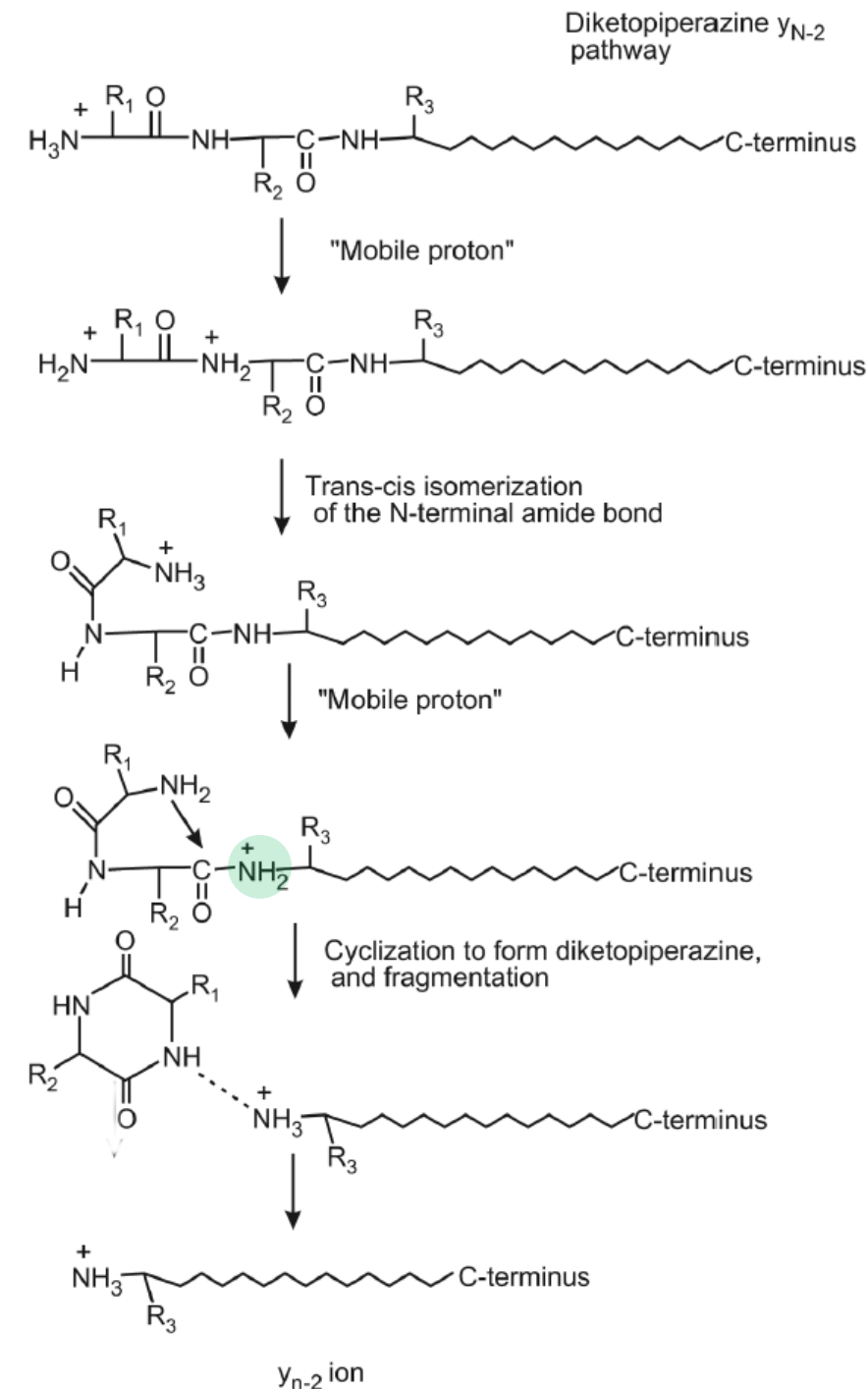
b_x - y_z útvonal

- A term. N H^+ -ja mobilis, átvándorol egy amid N-re
- Ezen amid C-t támadja egy nukleofil
- Protonált oxazolón származék képződik
- Átrendeződés után a C-terminális szekvencia b -ionként távozik
- alacsony energiájú CID fragmentációnál elérhető ez az útvonal



Diketopiperazin útvonala

- Diketopiperazinok 2 db cisz amidkötést tartalmaznak, amiatt ha a 2. és 3. aminosavak közti kötés hasad, egy izomerizáció szükséges
- Izomerizációt követően erre az amidkötésre mozog a mobilis H^+
- A terminális N támadja az amid C-t
- A diketopiperazin a C-terminális peptiddel „laza” komplexet képeznek, majd H^+ átmegy a peptidre (a ciklikus peptidnek kisebb a protonaffinitása)
- C-terminális peptid γ ionként távozik
- Ha távolabbi (γ_{N-n} ; $n > 2$) amidkötés hasad, a transz-cisz izomerizáció nem szükséges



Hys effektus

- Ha a fragmentált peptid tartalmaz Hys-t, a mobilis proton az imidazolgyűrű valamely N-jére kapcsolódik
- Egy másik H^+ mobilizálódik és a C-term. melletti amid N-re megy
- Az imidazol N-je nukleofilként támad a protonált amidkötés C-ére, ami ciklizációhoz vezet, ...
 - ami egy biciklikus b_x iont eredményez
 - vagy H^+ transzfer történik a gyűrűről az amid N-re, így az utóbbi szegmens y_z iont képez
- Azaz b_x vagy y_z ion is keletkezhet

Szekvencia átrendeződés

- Fragmentáció során ciklikus intermedierek keletkezhetnek
- Ha van rajtuk mobilis H^+ , akkor a gyűrű felnyílása okozhatja a szekvencia megváltozását
- *b* ionoknál lehetséges, *a* ionoknál viszont a gyűrűnyílás után az eredeti szekvenciasorrend kapható vissza
- Vitatott, hogy milyen mértékben fordul elő, shotgun proteomikai azonosításokat elvileg kevésbé zavarja

Poszttranszlációs módosítások

- Azonosításuk nehezebb az egyszerű peptidszekvencia megfejtésénél
- Míg egy-egy peptid nagy mennyiségben van jelen egy mintában, addig a módosított peptidek abundanciája alacsony. Fontos, hogy az MS/MS tömegspektrum minősége jó legyen
- Eltávolítandó a mintából azon enzimek, melyek PTM-t bontják vagy létrehoznak
- Legjobb a tripszin erre a célra
- Néhány PTM hidrofillé teszi a peptideket
 - C18 sómentesítés nem használható
 - Eltérő specificitású enzim használata megoldás (kimotripszin, Lys-N, endoproteináz Glu-C)

Poszttranszlációs módosítások

- PTM módosított peptidek dúsítása fontos
- De, általános dúsítási technikák nincsenek kifejlesztve. Kivétel a foszforiláció
- Nagy felbontású készülékek és minőségi fragmentáció (pl. HCD, ETD) fontos
- Kevés jól tanulmányozott proteom létezik (pl. hisztonok), de ez esetben is probléma a megfelelően hosszú peptidek generálása az emésztés során
- Legfontosabb - gyakran előforduló - PTM-ok: foszforilezés, ubikitilezés, glikozilezés, acetilezés

Poszttranszlációs módosítások

- Proteom mélyreható foszforiláltsági vizsgálata néhány napot vesz igénybe
- Single-shot megközelítés: egy injektálás/mérésből határozzák meg, robosztusabb és kevesebb anyagot igényel
- Frakcionálás: RP frakcionálás első lépésben, hogy csökkentsék a minta komplexitását. Több mérés, több meghatározott fehérje/PTM, de a frakcionálás bizonytalansága miatt kevésbé robosztus
- Célzott meghatározások: gyors, mert csak egy vagy néhány adott tömegű anyagot keresnek

Kémiai keresztkötések fehérjékben

- Fehérjék térszerkezetének vizsgálata
- Kovalens kötés két, térben egymáshoz közeli aminosav között, melyek lehetnek ugyanazon, vagy eltérő fehérjeláncon
- Pl.
 - adott egy fehérje
 - kovalens keresztkötéseket alakítanak ki a pl. Lys-ek között pl. bisz(szulfoszukcinimidil) szuberáttal
 - reakció kvencselés, termék tisztítás, tripszinnel vagy Lys-C-vel emésztés
 - peptidek frakcionálása (SEC)
 - detektálják azokat a peptideket, melyekben ezek a keresztkötések keletkeztek
 - a keresztkötések pedig csak akkor keletkezhetnek, ha a 2 Lys az eredeti fehérjekonformációban egymáshoz közel helyezkedett el
 - adatbáziskereséssel (xQuest, xProphet)

Kémiai keresztkötések fehérjékben

- Elvárások a reagensekkel szemben
 - stabilitás, reaktivitás és oldékonyság a fehérje natív konformációjának kedvező körülmények között
 - enzimes emésztés során ne történjen kötéshasadásuk
 - bizonyos távolságnál távolabbi aminosavakat ne tudjon összekötni, hogy értelmes térbeli adatokat szolgáltatasson
- Legtöbb reagens amin vagy tiol specificitással rendelkezik
 - Primer NH_2 - N-hidroxí- vagy szulfo-szukcinimidil észterek
 - A reaktív csoportokat valamilyen spacer/linker szakasz köti össze
 - Cys oldalláncok: pl. maleimidekkel, de a Cys gyakorisága kisebb a Lys-nél
 - Heterobifunkciós reagensek - Lys és Cys összekapcsolására

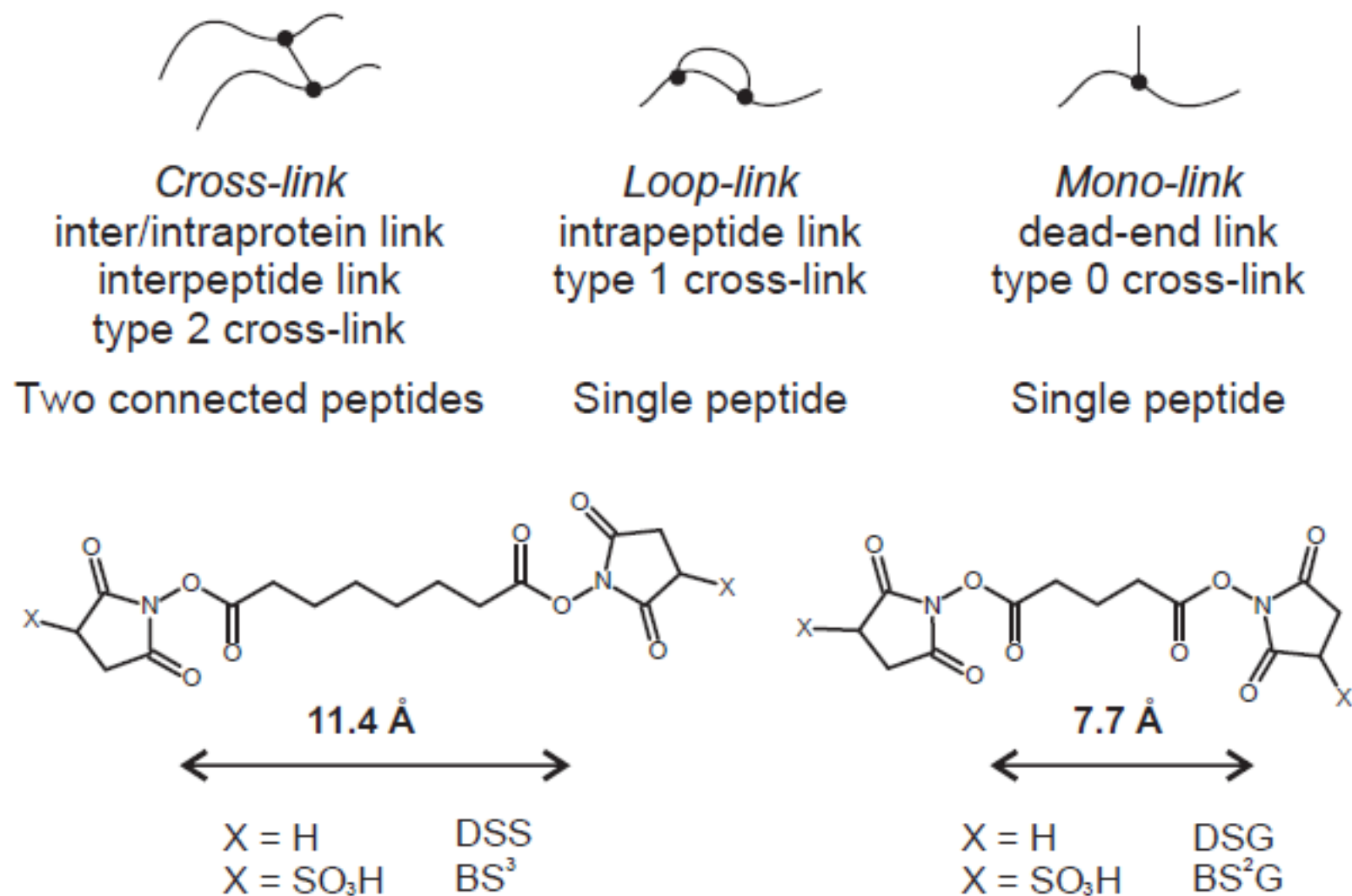


Figure 4.27: Top: Nomenclature of common products of chemical cross-linking reactions. Bottom: Structures of most commonly used amine-reactive cross-linking reagents. DSS, BS³, DSG, and bis(sulfosuccinimidyl) glutarate (BS²G). Reproduced by permission of the American Society for Biochemistry and Molecular Biology, Inc. from *Molecular & Cellular Proteomics*, **9**, 1634–1649, 2010.

Kémiai keresztkötések fehérjékben

- Formaldehid
 - speciális reagens
 - kétlépéses reakcióban tud keresztkötést kialakítani, egymáshoz nagyon közeli aminosavak között
 - Lys-t és Trp-t céloz elsősorban
 - Lys és Asp vagy Glu között alakít ki amidkötést
- Reagens jelölhető stabil izotóppal (egyedi izotópeloszlása lesz a reagensben lévő könnyű/nehézizotópok aránya miatt a keresztkötött peptideknek)
- Biotinilezett reagens - keresztkötött peptidek affinitáson alapuló dúsításához

Residue	1-letter code	3-letter code	Residue mass monoisotopic	Residue mass average	Immonium ion (m/z)
Alanine	A	Ala	71.03712	71.079	44.05 (<i>w</i>)
Arginine	R	Arg	156.10112	156.188	129 (<i>w</i>)
Asparagine	N	Asn	114.04293	114.104	87.09 (<i>m</i>)
Aspartic Acid	D	Asp	115.02695	115.089	88.04 (<i>m</i>)
Cysteine	C	Cys	103.00919	103.144	76 (<i>w</i>)
Glutamic Acid	E	Glu	129.04260	129.116	102.06 (<i>m</i>)
Glutamine	Q	Gln	128.05858	128.131	101.11 (<i>m</i>)
Glycine	G	Gly	57.02147	57.052	30 (<i>w</i>)
Histidine	H	His	137.05891	137.142	110.07 (<i>s</i>)
Isoleucine	I	Ile	113.08407	113.160	86.1 (<i>s</i>)
Leucine	L	Leu	113.08407	113.160	86.1 (<i>s</i>)
Lysine	K	Lys	128.09497	128.174	101.11 (<i>s</i>)
Methionine	M	Met	131.04049	131.198	104.05 (<i>m</i>)
Phenylalanine	F	Phe	147.06842	147.177	120.08 (<i>s</i>)
Proline	P	Pro	97.05277	97.117	70.07 (<i>s</i>)
Serine	S	Ser	87.03203	87.078	60.04 (<i>m</i>)
Threonine	T	Thr	101.04768	101.105	74.06 (<i>m</i>)
Tryptophan	W	Trp	186.07932	186.213	159.09 (<i>m</i>)
Tyrosine	Y	Tyr	163.06333	163.170	136.08 (<i>s</i>)
Valine	V	Val	99.06842	99.133	72.08 (<i>s</i>)

Amino acid	Immonium and related ion masses		Comments
Ala	44		
Arg	129	59, 70, 73, 87, 100, 112	129, 73 usually weak
Asn	87	70	87 often weak, 70 weak
Asp	88		Usually weak
Cys	76		Usually weak
Gly	30		
Gln	101	84, 129	129 weak
Glu	102		Often weak if C-terminal
His	110	82, 121, 123, 138, 166	110 very strong 82, 121, 123, 138 weak
Ile/Leu	86		
Lys	101	84, 112, 129	101 can be weak
Met	104	61	104 often weak
Phe	120	91	120 strong, 91 weak
Pro	70		Strong
Ser	60		
Thr	74		
Trp	159	130, 170, 171	Strong
Tyr	136	91, 107	136 strong, 107, 91 weak
Val	72		Fairly strong

Amino acid residue	Residue mass (Da)	Δ mass (Da)
Leucine	113.08407	
Isoleucine	113.08407	0
Hydroxyproline	113.04768	0.03639
Asparagine	114.04293	
Glycine + Glycine	114.04294 (2 57.02147)	0
Ornithine	114.07931	0.03638
Glutamine	128.05858	
Glycine + Alanine	128.05859 (57.02147 + 71.03712)	0
Lysine	128.09497	0.03639
Oxidized methionine	147.03540	
Phenylalanine	147.06842	0.03302
Glycine + Valine	156.08989 (57.02147 + 99.06842)	
Arginine	156.10112	0.01123
Glycine + Glutamic Acid	186.06407 (57.02147 + 129.04260)	
Alanine + Aspartic Acid	186.06407 (71.03712 + 115.02695)	0
Tryptophan	186.07932	0.01525
Valine + Serine	186.10045 (99.06842 + 87.03203)	0.03638
Proline + Threonine	198.10045 (97.05277 + 101.04768)	
Valine + Valine	198.13684 (2 99.06842)	0.03639

Amino acid	Characteristic losses (Da) from MH ⁺				Fragment	Mass calculation using residue weights	Mass calculation from other fragments
Ala	-				a _i	Σ residue weights 27	b _i 28
Arg	100				b _i	Σ residue weights + 1	MH ⁺ + 1 y _{n-i}
Asn	58				c _i	Σ residue weights + 18	b _i + 17
Asp	59				d _i	Σ residue weights 12 side chain	a _i (R _i 15)
Cys	47				for Ile	Σ residue weights 55 or 41	a _i 28 or 14
Gly	-				for Thr	Σ residue weights 43 or 41	a _i 16 or 14
Gln	59, 72				for Val	Σ residue weights 41	a _i 14
Glu	36, 60, 63, 73				v _i	Σ residue weights + 74	x _{i-1} + 29
His	81				w _i	Σ residue weights + 73	x _{i-1} + 28
Ile/Leu	57				for Ile	Σ residue weights + 87 or + 101	x _{i-1} + 42 or + 56
Lys	59, 72				for Thr	Σ residue weights + 87 or 89	x _{i-1} + 42 or + 44
Met	47, 48, 62, 75				for Val	Σ residue weights + 87	x _{i-1} + 42
Phe	91, 92				x _i	Σ residue weights + 45	y _i + 26
Pro	-				y _i	Σ residue weights + 19	MH ⁺ + 1 b _{n-1}
Ser	31				Y _i	Σ residue weights + 17	y _i 2
Thr	45				z _i	Σ residue weights + 2	y _i 17
Trp	130				Internal fragments		
Tyr	107, 108				b-type	Σ residue weights + 1	
Val	43				a-type	Σ residue weights 27	

Manuális *de novo* szekvenálás

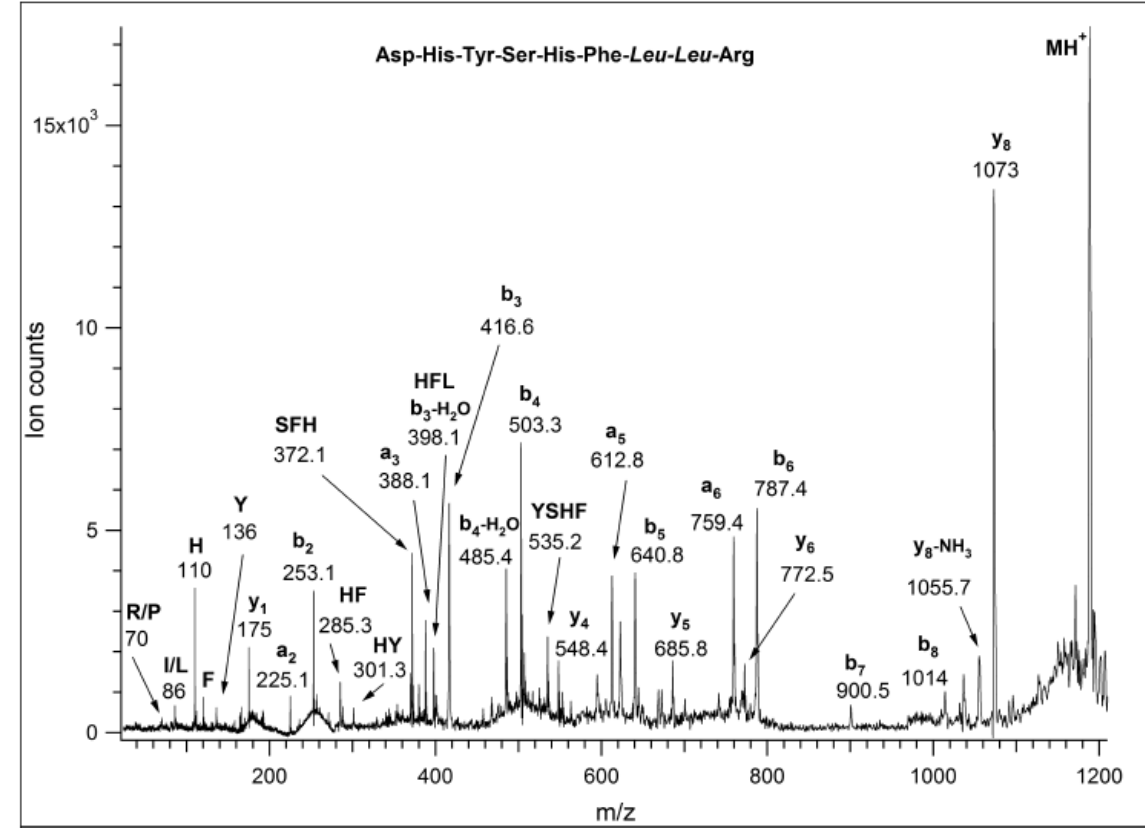
- A folyamat során a fragmentációs spektrumokból megfejtjük a fragmentált peptid aminosavsorrendjét
- CID, HCD: ált. b-, γ - és a-ionok
- ETD, ECD: ált. c-, z- és a-ionok

Manuális *de novo* szekvenálás

- Kiindulási lehetőségek:
 - Alacsony tömegnél immóniumionokat vizsgálni - milyen aminosavakból áll a peptid?
 - Tripszin - C-terminális mindig Lys vagy Arg
 - y_1 ion m/z (1^+ töltésnél) = maradékion + 19 - utalhat a C-terminális minőségére az alábbi ionok jelenléte
 - Lizin: $128,09 + 19 = 147,09$; Arginin: $156 + 19 = 175$
 - y_1 ismeretéből számítható a b_{n-1} ion tömege: $y_i + b_{n-i} = MH^+ + 1$
 - b_{n-1} ion tömege: prekursorion $(M+H)^+ - (128 \text{ vagy } 156) - 18 (H_2O)$
 - tripszines hasításnál a C-terminális bázikusabb, emiatt a b ionsorozat kevésbé abundáns, míg az y ionsorozat az
 - addig kell az y-ionok közti különbségeket keresni, míg megtaláljuk az összeset
 - végén ellenőrizendő, hogy a b-ionokkal egybevágnak-e az eredmény
 - PTM-ok befolyásolják a tömegeket (pl. karbamidometilezés Cys-eken)

Példa 1.

- $MH^+ = 1187,6$ Da (egyszeresen töltött)

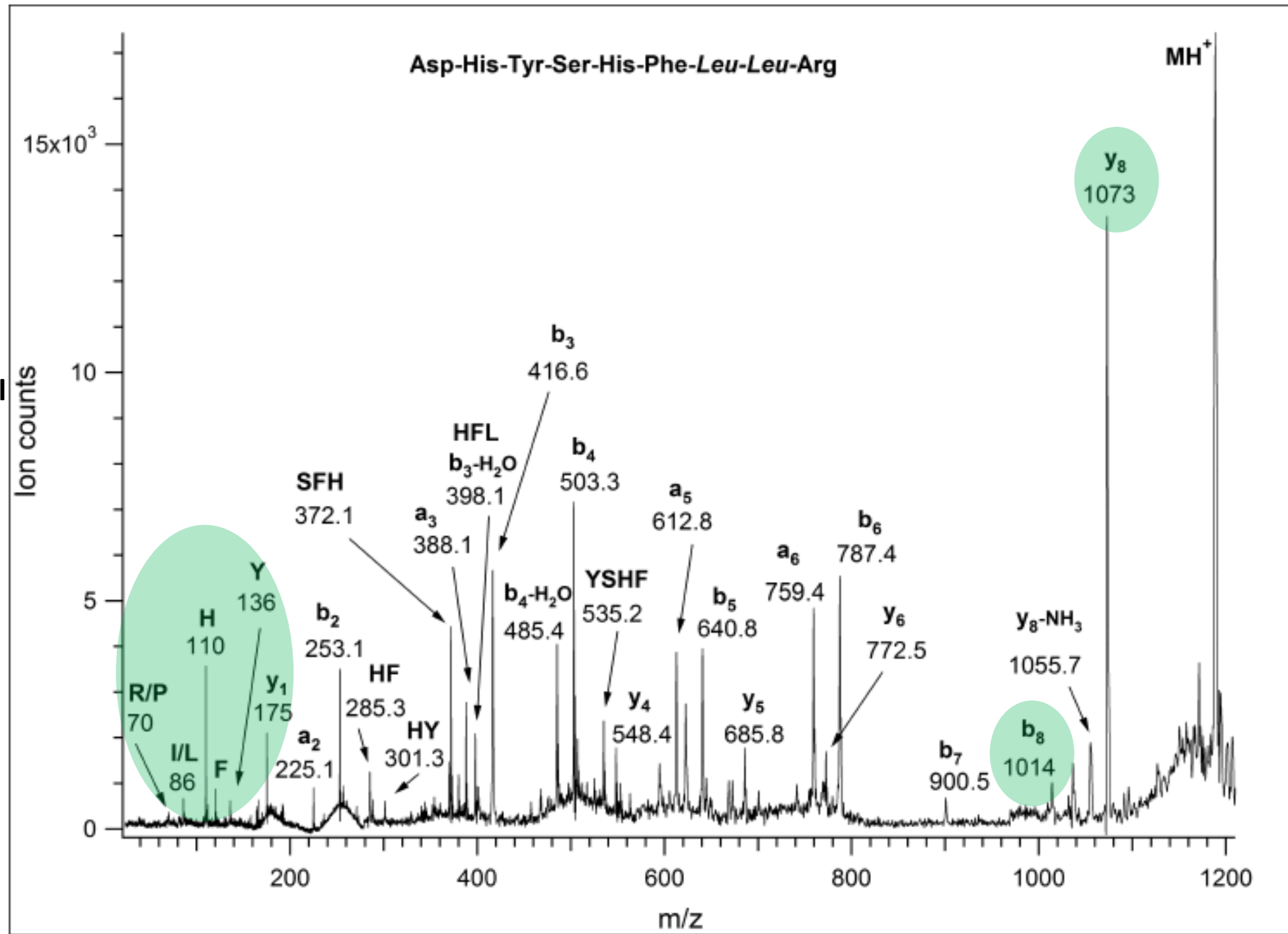


Immónium ionok
 y_1 C-terminális Arg

b_{n-1} ion: 1014

1073-nál y_{n-1} ion
115 Da vesztés a MH^+ -ból
(116 Da b_1 tömeg)
= Asp N-terminális

(Asp- -Arg)



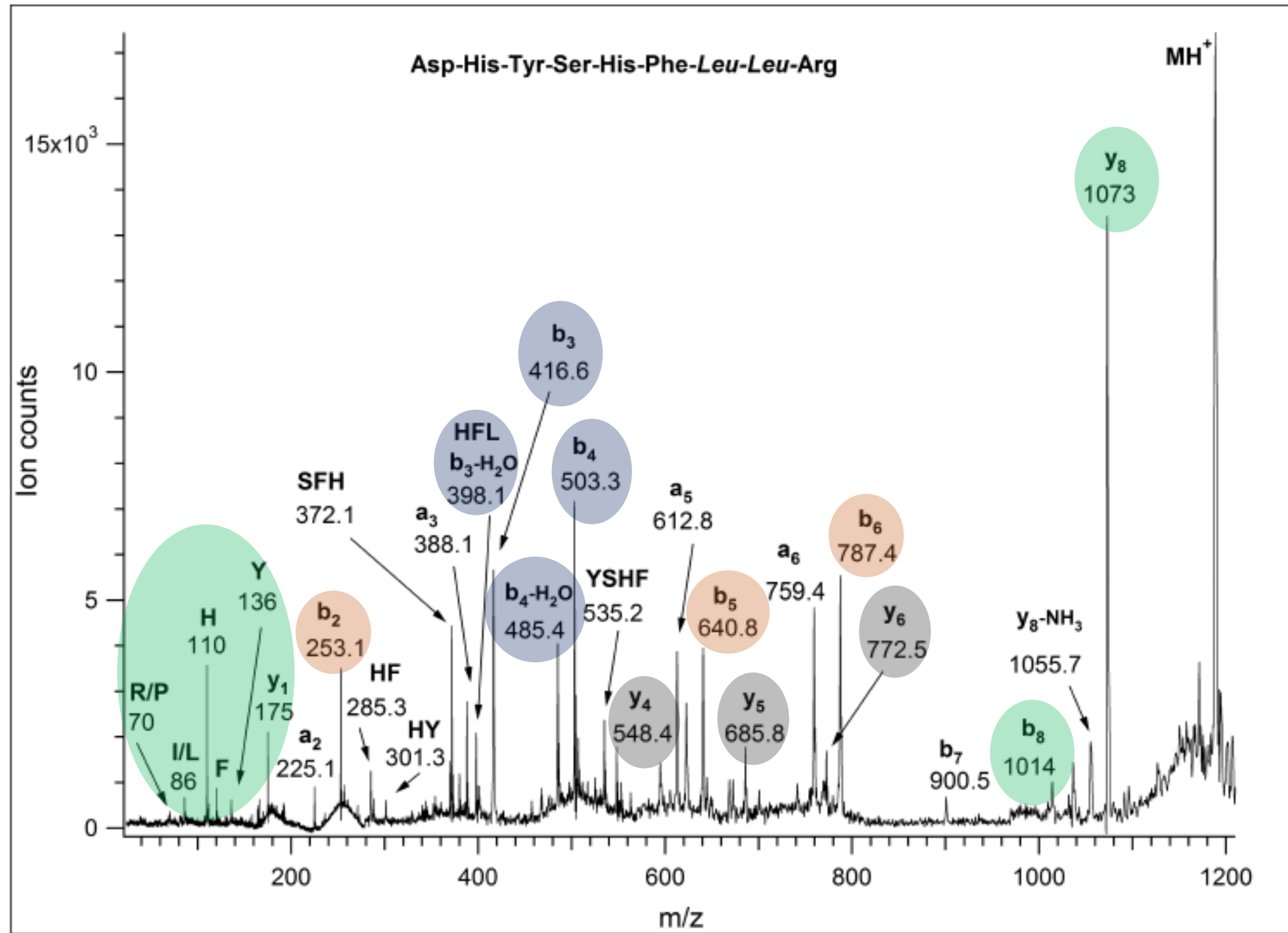
b-ionok azonosítása:
 ha H₂O-t (18) vagy CO-t (28) vesztenek

b-ionok azonosítása:
 H₂O-t (18): 503,3; 416,6
 CO-t (28): 787,4; 640,8; 253

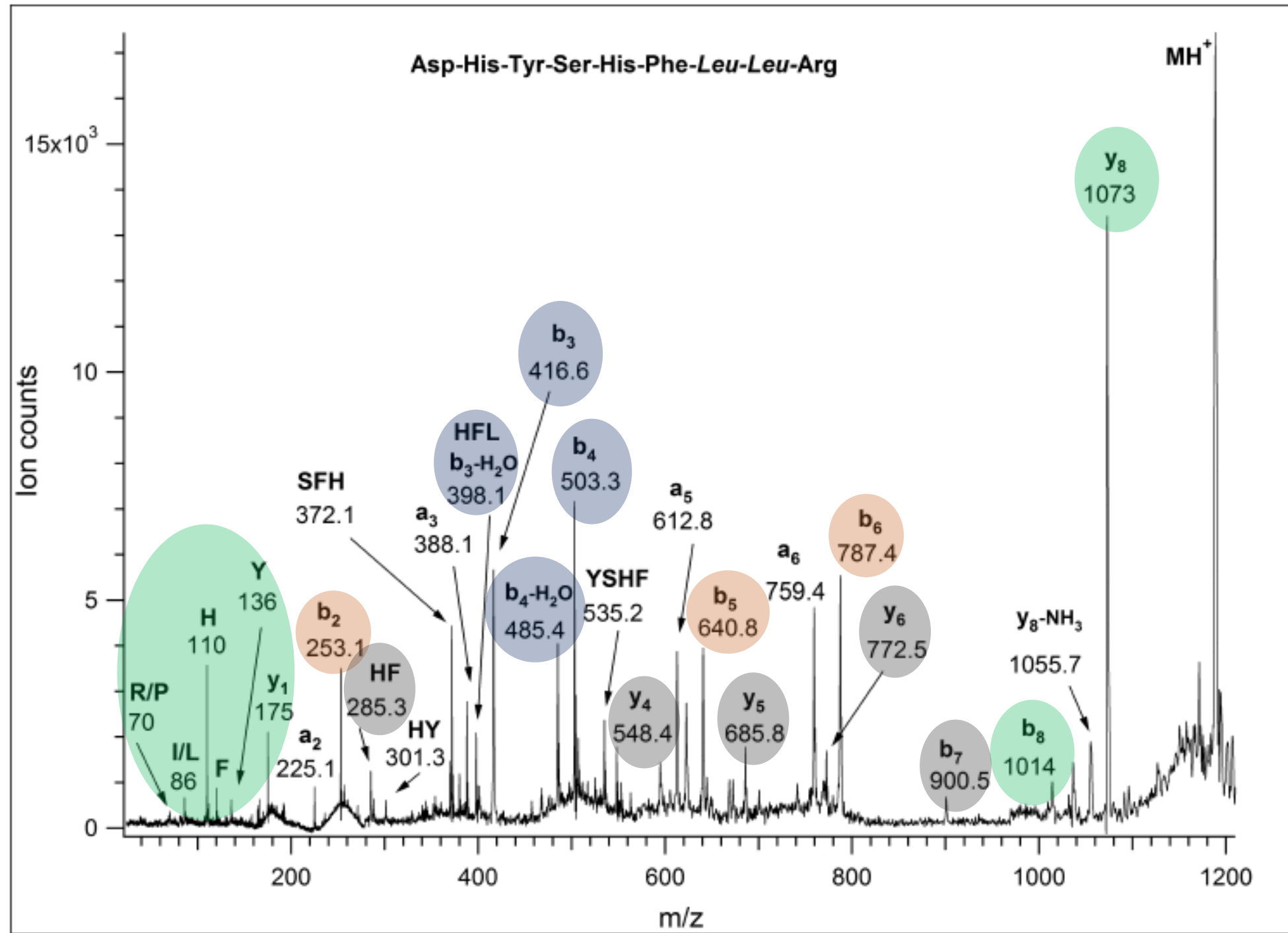
787 és 640 közti
 kül. 147 Da: Phe
 640 y párja (MH-ból) az 548-as
 (Asp - - Phe - - Arg)

640,8 és 503,3 közti
 kül. 137 Da: Hys
 503 y párja (MH-ból) az 685-ös
 (Asp - - Hys - Phe - - Arg)

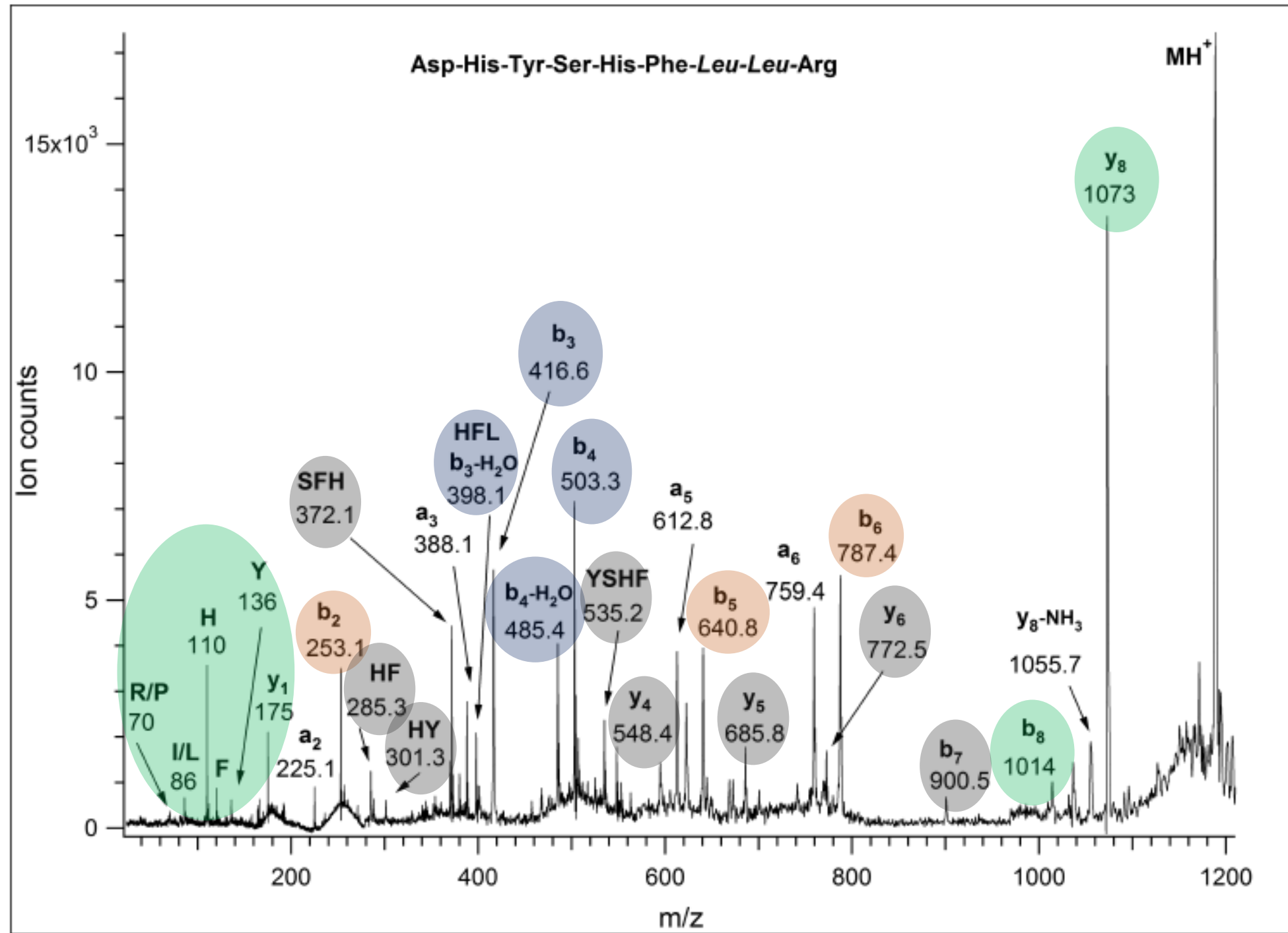
503,3 és 416,6 közti
 kül. 87 Da: Ser
 416,6 y párja (MH-ból) a 772,3-ös
 (Asp - - Ser - Hys - Phe - - Arg)



416,6-ból következő lépés:
 285,3 - 131 Da kül: Met, de ez nem jelenik meg immóniumionként, ill. nincs 18 vagy 28 vesztett „párja”
 285,3 belső fragmens: Hys-Phe
 416,6-ból 253:
 163 Da kül: Tyr (Asp - - Tyr - Ser - Hys - Phe - - Arg)
 N-term: Asp, így 253 - 116 (b_1) = 137 (Hys) (Asp - Hys - Tyr - Ser - Hys - Phe - - Arg)
 MH^+ : 1187,6 - megfejtett aminosavak: 226 Da Leu/Ile-t tartalmaz (immóniumion): 113. Innen adódik, hogy a másik is 113, azaz Leu/Ile.



372,1; 301,3; 285; 535,2;
672: belső fragmensek



Manuális *de novo* szekvenálás

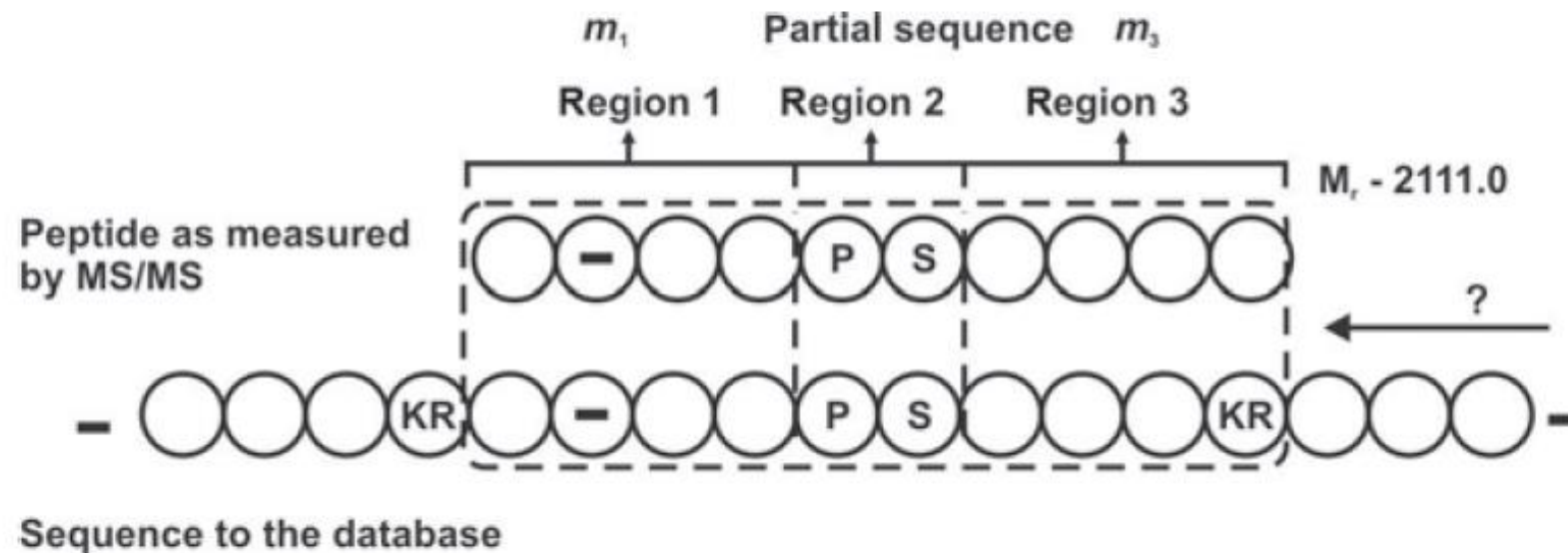
- CID: b-ionok nem jelennek meg ennyire (inkább csak kisebb tömegeknél), y-ionok abundánsabbak
- y-ionok különbségei a residue (maradék) tömegek, ezek sorban meg is adhatják a szekvencia egy részletét

MS-alapú bioinformatika - Peptid térkép vizsgálat

- Peptide mass fingerprinting - 1993 - 5 független csoport
 - nincs fragmentáció, csak a peptid tömegei kellene az azonosításhoz
 - összes tömeget meg kell adni hozzá - ált. MALDI
 - tiszta fehérjék vagy nagyon egyszerű keverékek vizsgálatára

MS-alapú bioinformatika- Peptide sequence tag

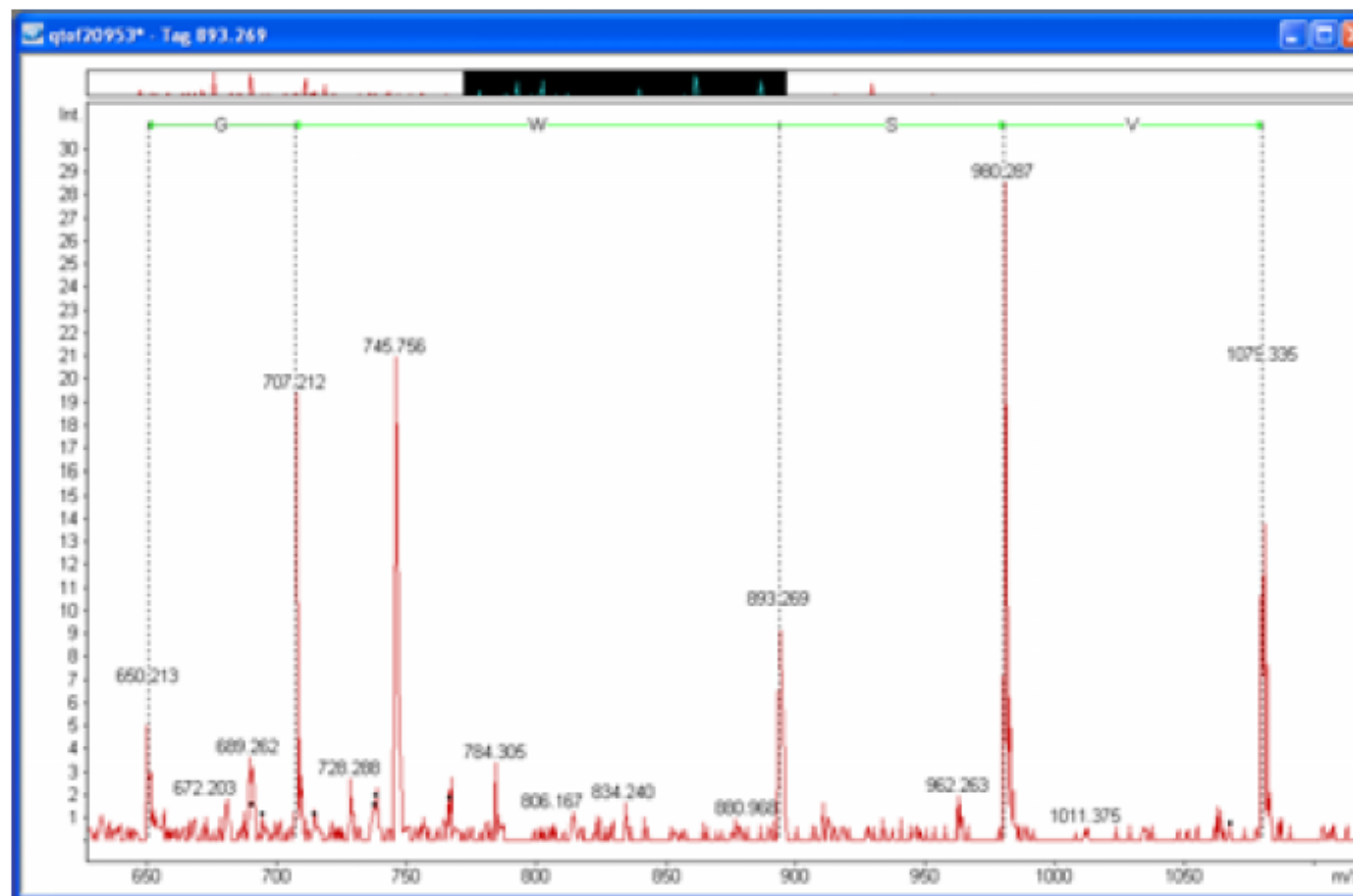
- Peptide sequence tag - Mann (EMBL Heidelberg)
 - néhány aminosavból álló szekvencia azonosítva a peptid közepén (m_2)
 - m_1 és m_3 : peptid maradéka, a szekvencia „tag” előtt és után
 - $M = m_1 + m_2 + m_3$



MS-alapú bioinformatika - Peptide sequence tag

- Peptide sequence tag
 - Feltéve, hogy b-ionokat detektálunk:
 - M-ből, ill. a mért „köztes” tömegekből kiszámítható egyszerűen m_1 és m_3
 - N- és C-terminális a tripszin hasítóhelyei legyenek
 - γ -ionok esetén hasonlóan járunk el. Az m_1 és m_3 tömegek 2 Da-nal nagyobbak, ill. kisebbek lesznek
 - PeptideSearch szoftverrel kereshető
 - InsPecT: poszttranszlációs módosítások keresése
 - Minél hosszabb a részleges szekvencia, annál specifikusabb a keresés
 - Általában a spektrumok értékelése (tag-keresés) manuálisan megy, nem alkalmas sok peptid azonosítására nagyszámú fehérjékből
 - Gyors keresés (tag egyfajta szűrő), jó tag esetén elég valószínű a jó találat

MS-alapú bioinformatika



1489.430 tag(650.213,GWSV,1079.335)

MS-alapú bioinformatika - MS/MS ion keresés

- MS/MS tömegspektrumokat kell hozzárendelni egy peptidszekvenciához a lehető legrövidebb idő alatt
- Lassú lehet - pl. no enzyme, több variábilis módosítás, nagyméretű adatbázis/adatfájl
- Valójában peptideket azonosít, nem fehérjéket
- Ez a módszer alkalmas PTM keresésre

MS-alapú bioinformatika - MS/MS ion keresés

- SEQUEST - Yates group
 - az algoritmus egy MS/MS tömegspektrumhoz rendel egy peptidszekvenciát
- MASCOT
 - online server: ingyenes keresés
 - MASCOT Server: licenzhez kötött
 - Distiller: egyszerűsített felületen kereshetők a nyers (LC)-MS/MS adatfájlok
- PEAKS DB
 - de novo szekvenálást alkalmaz

MS-alapú bioinformatika - MS/MS ion keresés

- X! Tandem
 - automatikus PTM keresés, de csak olyan peptideken, amiket már korábban azonosított
- OMSSA, OMSSAPercolator
 - az MS/MS csúcsokat összeveti egy adatbázis *in silico* emésztett csúcsaival
- Scaffold
 - találati valószínűségek pontosabb megadása
- Fizetős: Mascot, Sequest, Phenyx, SpectrumMill, IdentityE, Byonic
- Ingyenes: X!Tandem, OMSSA, MS-Amanda, MS-GF+, MaxQuant

MS-alapú bioinformatika - DDA vs DIA

- Data dependent acquisition (DDA) vs. Data independent acquisition (DIA)
 - DIA: adott m/z tartomány összes ionja fragmentálódik és detektálódik egy második fázisban
 - nem kell tudni tömegeket, nincs előre kiválasztás
 - pl. MS^E , ARM (all reaction monitoring), AIF (all ion fragmentation), SWATH-MS
 - DDA: egy rögzített számú prekursorion fragmentálódik csak és detektálódik a második fázisban
 - pl. AutoMS, SRM, MRM
 - Az MS/MS spektrumokat a mérés után *de novo* szekvenálással, vagy újabban algoritmusokkal (pl. MASCOT, SEQUEST, Andromeda) végzik
 - Kis koncentrációjú (pl. PT módosított) peptidek detektálása nem feltétlenül valósul meg

MS-alapú bioinformatika

- Targeted (célzott) analízis
 - specifikus ion átmenetek monitorozása
 - Pl. precursor ion scanning (PIS)
 - QqQ készülék 3. kvadrupól egy adott semleges vesztesre van állítva - 79 Da foszforiláció vizsgálatánál
 - Ha ez a fragmens megjelenik, akkor tuti, hogy a fragmentált peptid amiből származott foszfopeptid
- MRM/SRM
 - MRMAid vagy Skyline szoftverek segíthetnek kiválasztani a mérésekhez használható MRM átmeneteket adott fehérjék vizsgálatához

PTM vizsgálat

- Leggyakoribbak: foszforilezés, metilezés, ubikvitinálás, glikozilálás
- összesen több mint 200 fajta PTM
- bottom-up megközelítéssel szokás vizsgálni
- Számítási nehézségek: legtöbb program (kivéve pl. Byonic) maximálja a variábilis PTM-ok számát 3-ban
- UNIMOD: PTM-ok adatbázisa

PTM vizsgálat

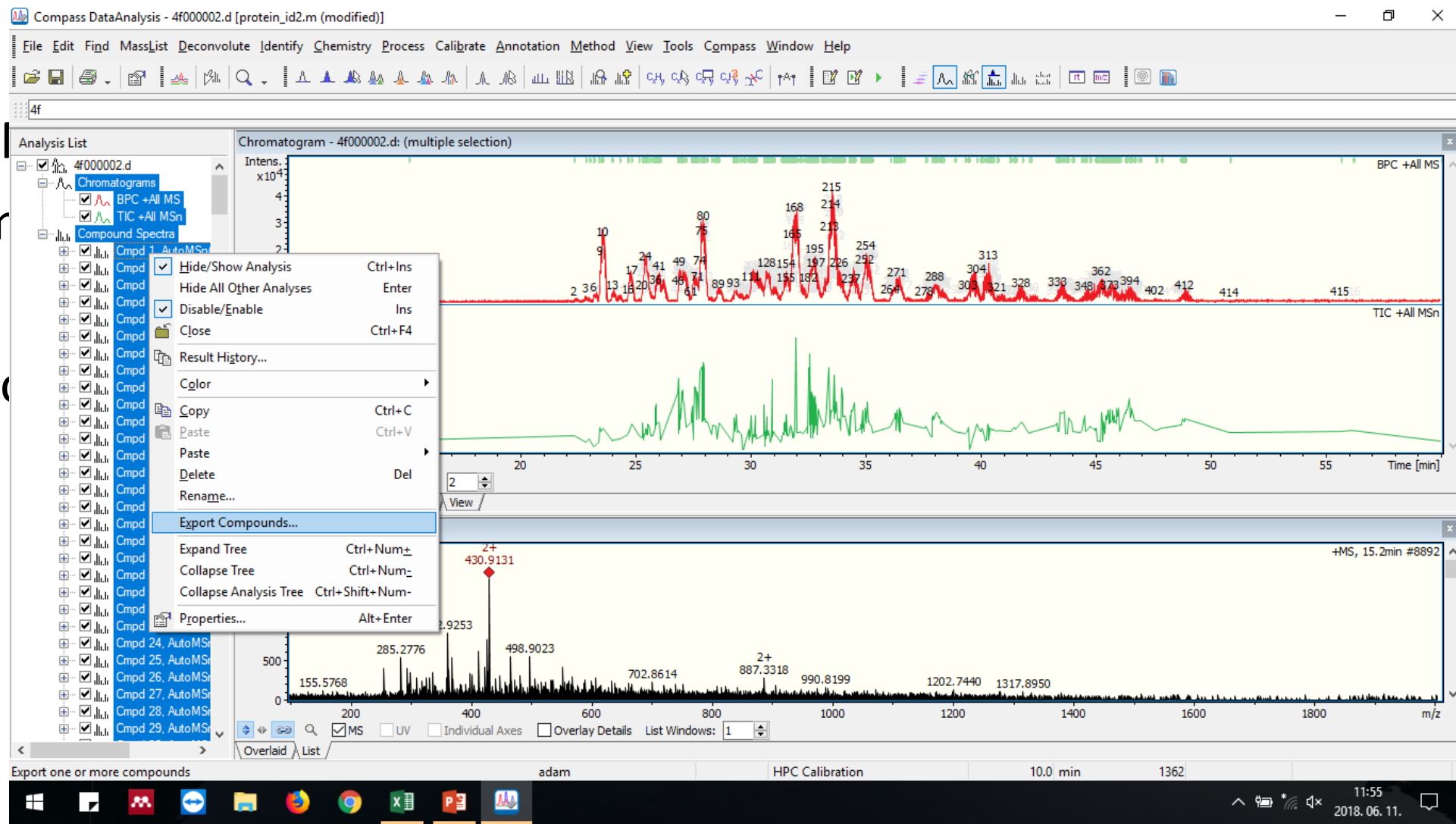
- MS detektálásuk:
 - globális módszer vagy célzott (targeted) módszer
 - első körben tisztázni kell milyen PTM-t akarunk vizsgálni
 - ez egy tömegkülönbséget/eltolódást okozhat a tömegspektrumokban
 - Pl. ubikvitinált fehérjék - tripszines emésztés - ubikvitinezett lizinhez 2 Gly fog kapcsolódni (K-GG), ami MS/MS-sel azonosítható
 - Pl. foszforilezett fehérjék - dúsítás SCX/IMAC oszlopokon (immobilizált fém-affinitás oszlop)
- Mivel a sok variábilis PTM a keresés időtartamát növeli jelentősen, emiatt egy általános - gyors - keresés után érdemes az adott fehérjetalálattal lefuttatni újra a keresést sok PTM-mel.

Gyakorlati megvalósítás - Peak picking

- Csúcslista fájl generálása (100 kB - néhány MB) akár több GB méretű LC- vagy CE-MS/MS analízisből
- Adatok hatékony redukálása fontos
 - részben mérési paraméterekkel szabályozható,
 - ill. peak picking paraméterekkel
 - Pl. kidobhatóak a 700 Da (nem m/z) alatti prekurzorokból származó, vagy kevés csúcsot (<10) tartalmazó MS/MS spektrumok
 - Csúcsok időtartományában érdemes a keresést futtatni

Peak picking - DataAnalysis

- Calibration
- Finding Auto
- Generating m
- Deconvolute
- Export compo



M00111

Too many peptide masses in your data file. Mascot has no limit, but this system has been configured to have a limit of [number] for [number] searches

Further help:

There is a limit to the number of peptide masses allowed. Are you entering ms-ms data into a pmf form? The maximum number of queries is set by adding a value 'MaxQueries' into the options section of the mascot.dat file. If no value is specified, then the default of 10,000 is used. On the Matrix Science public web site, the limit is **1200**

Actions:

- Show message to end user
- Terminate search
- Message put into errorlog.txt file
- Message not put into monitor.log file
- Message not emailed to the Mascot administrator
- Message not put into the search results file

Kizárások alkalmazása a

tal
kol
cső

a: A
rec
b: A
per
45
c: S
d: A
e: S
exc
idő
kizá
f, u
spe
g, u
frel
hel
h, u
csa

Intens.
x10⁵
1.0
0.5

a

		a)	b)	c)	d)	e)	f)	g)	h)
	MS/MS komponensek szama	1601	2578	1405	1089	96	372	115	177
SC%	BSA	40	23	44	44	12	40	47	40
	Lizozim	52	36	52	52	8	48	52	37
	HB α	26	8	19	30	-	11	26	19
	HB β	-	-	8	17	8	8	17	17
	HRP	-	-	-	-	-	-	-	-
	Trypsin	-	-	8	4	-	13	13	4

0

5

10

15

20

25

30

35

40

Time [min]

04.d: TIC +All MS

02.d: TIC +All MS

05.d: TIC +All MS

06.d: TIC +All MS

07.d: TIC +All MS

08.d: TIC +All MS

09.d: TIC +All MS

Keresés - MASCOT

- Fix módosítások - nem növelik a számítási időt
- Variábilis (lehetséges) módosítások - drasztikusan növelik a számítási időt. PTM-ok keresésének elterjedt módja
- Tömeghiba megadása pontosan
- Enzim megadása: ált. tripszin - no enzim növeli a keresés idejét
- Adatbázisok:
 - legjobbak: NCBI nr, UniRef100, MSDB
 - legjobban annotált: SwissProt - de ez kevésbé alkalmas adatbáziskeresésekre (non-redundant - pontosan a benne szereplő szekvenciának kell megjelenie)
 - legnagyobb: EST



Access Mascot Server

You are welcome to submit searches to this free Mascot Server. Searches of MS/MS data are limited to 1200 spectra and some functions, such as no enzyme searches, are unavailable. Automated searching of batches of files is not permitted. If you want to automate search submission, perform large searches, search additional sequence databases, or customise the modifications, quantitation methods, etc., you'll need to license your own, in-house copy of Mascot Server.

Peptide Mass Fingerprint

The experimental data are a list of peptide mass values from the digestion of a protein by a specific enzyme such as trypsin.

[Perform search](#) | [Example of results report](#) | [Tutorial](#)

Sequence Query

One or more peptide mass values associated with information such as partial or ambiguous sequence strings, amino acid composition information, MS/MS fragment ion masses, etc. A super-set of a sequence tag query.

[Perform search](#) | [Example of results report](#) | [More information](#)

MS/MS Ions Search

Identification based on raw MS/MS data from one or more peptides.

[Perform search](#) | [Example of results report](#) | [Tutorial](#)

More info

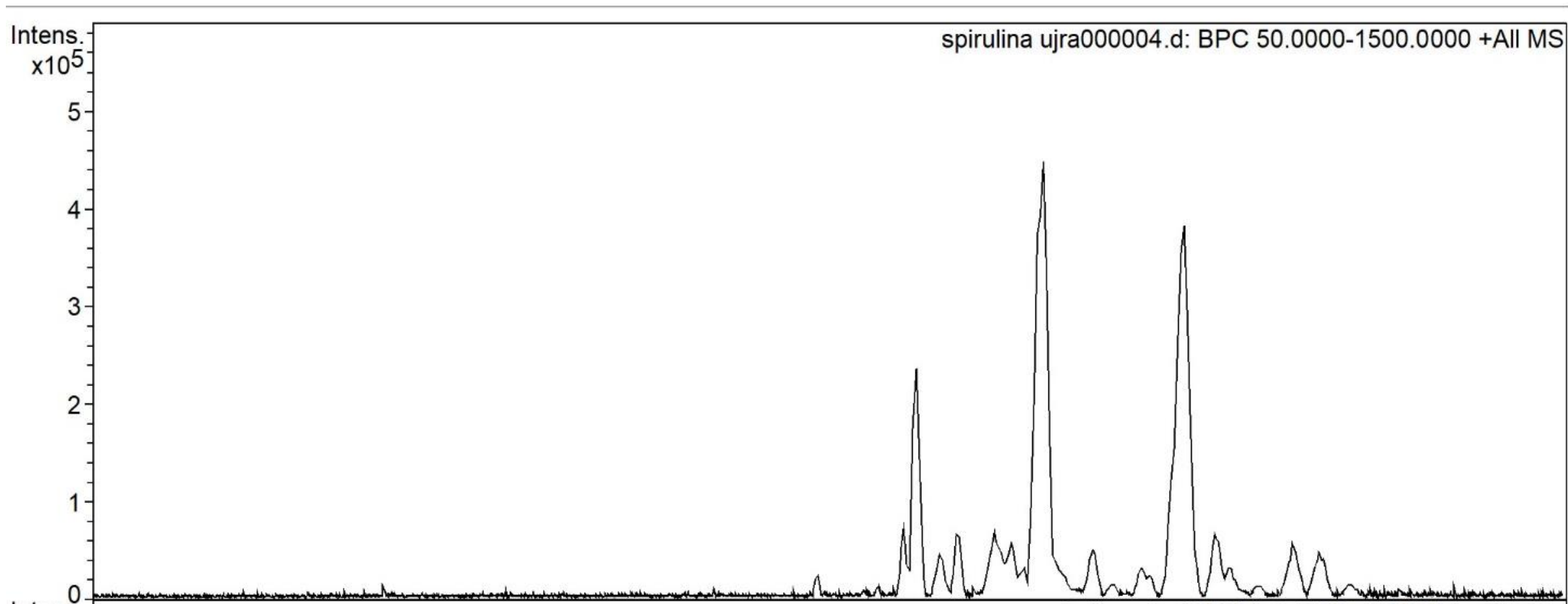
- > [Mascot overview](#)
- > [Search parameter reference](#)
- > [Data file format](#)
- > [Results report overview](#)



MASCOT MS/MS Ions Search

Your name	<input type="text" value="Adam Kecskemeti"/>	Email	<input type="text" value="toprongyos@gmail.com"/>
Search title	<input type="text"/>		
Database(s)	<input type="text" value="SwissProt (AA)"/>	<input type="button" value="v"/>	<input type="button" value="^"/>
			<ul style="list-style-type: none">Amino acid (AA)contaminantscRAPNCBIprotNucleic acid (NA)Environmental_ESTFungi_ESTHuman_ESTInvertebrates_ESTMammals_EST
Taxonomy	<input type="text" value="All entries"/>		
Enzyme	<input type="text" value="Trypsin"/>	Allow up to	<input type="text" value="1"/> missed cleavages
Quantitation	<input type="text" value="None"/>		
Fixed modifications	<input type="text" value="-- none selected --"/>	<input type="button" value="v"/>	<input type="button" value="^"/>
			<ul style="list-style-type: none">Acetyl (K)Acetyl (N-term)Acetyl (Protein N-term)Amidated (C-term)Amidated (Protein C-term)Ammonia-loss (N-term C)Carbamidomethyl (N-term)Carbamyl (K)Carbamyl (N-term)Carboxymethyl (C)Cation:Na (C-term)
	<input type="checkbox"/> Display all modifications		
Variable modifications	<input type="text" value="Carbamidomethyl (C)"/>	<input type="button" value="v"/>	<input type="button" value="^"/>
Peptide tol.	<input type="text" value="0.3"/> Da	<input type="text" value="# <sup>13</sup>C 0"/>	MS/MS tol. <input type="text" value="0.3"/> Da
Peptide charge	<input type="text" value="1+, 2+ and 3+"/>		Monoisotopic <input checked="" type="radio"/> Average <input type="radio"/>
Data file	<input type="text" value="Fájl kiválasztása"/> p4.mgf		
Data format	<input type="text" value="Mascot generic"/>		
Instrument	<input type="text" value="ESI-QUAD-TOF"/>		
Decoy	<input type="checkbox"/>		
	<input type="button" value="Start Search ..."/>		<input type="button" value="Reset Form"/>

Mascot keresési eredmény vs mérési fájl



1. [TRY1 BOVIN](#) Mass: 25769 Score: 474 Matches: 9(7) Sequences: 6(5)

Cationic trypsin OS=Bos taurus PE=1 SV=3

Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Unique	Peptide
<input checked="" type="checkbox"/> 49	510.7489	1019.4833	1019.4957	-0.0124	0	74	6.1e-05	1	U	K.APILSDSSCK.S
<input checked="" type="checkbox"/> 53	539.2616	1076.5086	1076.5172	-0.0085	0	(64)	0.00053	1	U	K.APILSDSSCK.S + Carbamidomethyl (C)
<input checked="" type="checkbox"/> 55	577.2885	1152.5624	1152.5663	-0.0038	0	62	0.00073	1	U	K.SSGTSYPDVLK.C
<input checked="" type="checkbox"/> 57	584.7926	1167.5707	1167.5747	-0.0039	0	19	13	1	U	K.VCNYVSWIK.Q + Carbamidomethyl (C)
<input checked="" type="checkbox"/> 69	717.3618	1432.7091	1432.7133	-0.0042	0	(18)	17	1	U	K.LQGIVSWGSGCAQK.N
<input checked="" type="checkbox"/> 71	745.8750	1489.7355	1489.7348	0.0007	0	75	4e-05	1	U	K.LQGIVSWGSGCAQK.N + Carbamidomethyl (C)
<input checked="" type="checkbox"/> 75	1082.0290	2162.0434	2162.0491	-0.0057	0	130	8.6e-11	1	U	R.LGEDNINVVEGNEQFISASK.S
<input checked="" type="checkbox"/> 76	721.6911	2162.0516	2162.0491	0.0024	0	(78)	1.3e-05	1	U	R.LGEDNINVVEGNEQFISASK.S
<input checked="" type="checkbox"/> 82	758.3925	2272.1557	2272.1521	0.0036	0	114	3.4e-09	1	U	K.SIVHPSYNSNTLNNDIMLIK.L

2. [PHCA ARTFS](#) Mass: 4717 Score: 126 Matches: 1(1) Sequences: 1(1)

C-phycoyanin alpha chain (Fragment) OS=Arthrospira fusiformis GN=cpcA PE=1 SV=1

Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Unique	Peptide
<input checked="" type="checkbox"/> 73	772.9000	1543.7855	1543.7842	0.0013	0	126	3.3e-10	1	U	K.TPLTEAVSIADSQGR.F

Proteins matching the same set of peptides:

[PHCA ARTPT](#) Mass: 17591 Score: 126 Matches: 1(1) Sequences: 1(1)

C-phycoyanin alpha chain OS=Arthrospira platensis GN=cpcA PE=1 SV=2

Protein sequence coverage: 34%

Matched peptides shown in **bold red**.

1 MK**TPLTEAVS IADSQGR**FLS STQIQVLFGR FRQAKAGLXA AKAL

Unformatted sequence string: [44 residues](#) (for pasting into other applications).

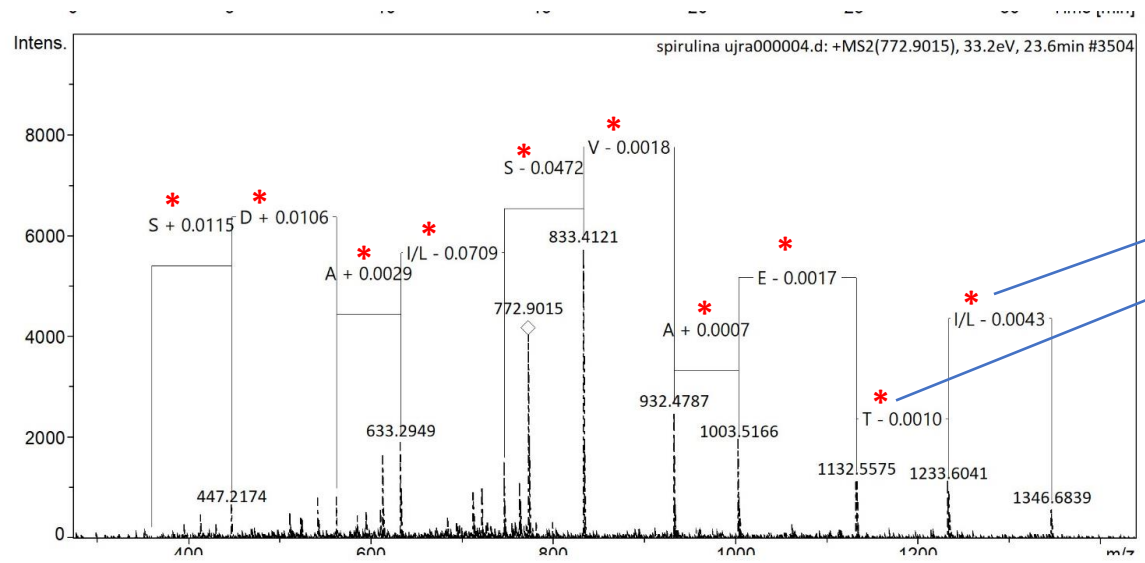
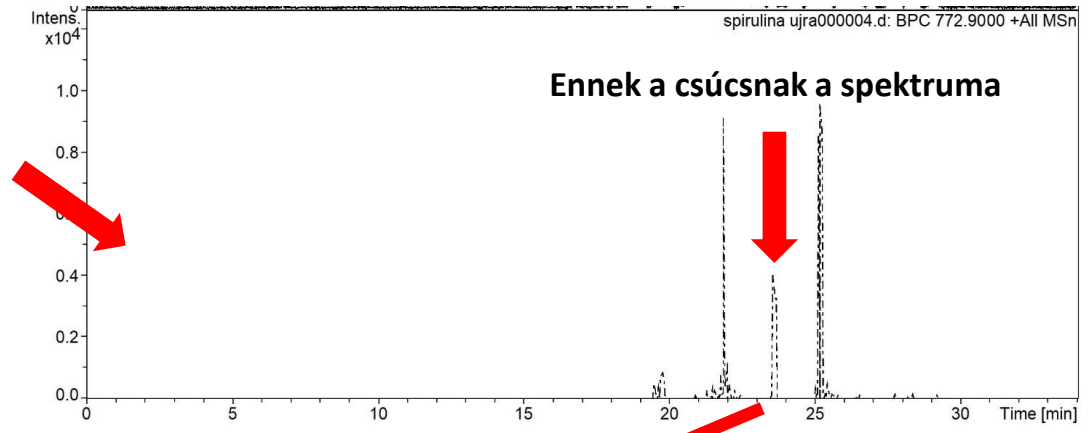
Sort by residue number increasing mass decreasing mass

Show matched peptides only predicted peptides also

Query	Start - End	Observed	Mr (expt)	Mr (calc)	Delta M	Score	Expect	Rank	U	Peptide
73	3 - 17	772.9000	1543.7855	1543.7842	0.0013	0 126	3.3e-10	1	U	K.TPLTEAVSIADSQGR.F

#	b	b ⁺⁺	b [*]	b ^{*++}	b ⁰	b ⁰⁺⁺	Seq.	y	y ⁺⁺	y [*]	y ^{*++}	y ⁰	y ⁰⁺⁺	#
1	102.0550	51.5311			84.0444	42.5258	T							15
2	199.1077	100.0575			181.0972	91.0522	P	1443.7438	722.3755	1426.7172	713.8623	1425.7332	713.3703	14
3	312.1918	156.5995			294.1812	147.5942	L	1346.6910	673.8492	1329.6645	665.3359	1328.6805	664.8439	13
4	413.2395	207.1234			395.2289	198.1181	T	1233.6070	617.3071	1216.5804	608.7938	1215.5964	608.3018	12
5	542.2821	271.6447			524.2715	262.6394	E	1132.5593	566.7833	1115.5327	558.2700	1114.5487	557.7780	11
6	613.3192	307.1632			595.3086	298.1579	A	1003.5167	502.2620	986.4901	493.7487	985.5061	493.2567	10
7	712.3876	356.6974			694.3770	347.6921	V	932.4796	466.7434	915.4530	458.2302	914.4690	457.7381	9
8	799.4196	400.2134			781.4090	391.2082	S	833.4112	417.2092	816.3846	408.6959	815.4006	408.2039	8
9	912.5037	456.7555			894.4931	447.7502	I	746.3791	373.6932	729.3526	365.1799	728.3686	364.6879	7
10	983.5408	492.2740			965.5302	483.2688	A	633.2951	317.1512	616.2685	308.6379	615.2845	308.1459	6
11	1098.5677	549.7875			1080.5572	540.7822	D	562.2580	281.6326	545.2314	273.1193	544.2474	272.6273	5
12	1185.5998	593.3035			1167.5892	584.2982	S	447.2310	224.1191	430.2045	215.6059	429.2205	215.1139	4
13	1313.6583	657.3328	1296.6318	648.8195	1295.6478	648.3275	Q	360.1990	180.6031	343.1724	172.0899			3
14	1370.6798	685.8435	1353.6533	677.3303	1352.6692	676.8383	G	232.1404	116.5738	215.1139	108.0606			2
15							R	175.1190	88.0631	158.0924	79.5498			1

722.900 extracted



P	1443.7438
L	1346.6910
T	1233.6070
E	1132.5593
A	1003.5167
V	932.4796
S	833.4112
I	746.3791
A	633.2951
D	562.2580
S	447.2310

Egyezés megtalálása/elfogadása

- Találatok szűrése - pl. Sequence Tag
- Relative/arbitrary score értékek
 - PMF - hány db peptidtömeget sikerült párosítania a mértékkel
 - szinte bármilyen adatot felhasználva lehet generálni találatot, nagy (1-3 MDa) fehérjékre
 - MOWSE pontozási algoritmussal ezek kiszűrhetők - Pappin group
 - talált peptideket a fehérje méretéhez viszonyítja
 - MS/MS - Sequest-ben a kereszt-korrelációs koefficiens

Egyezés megtalálása/elfogadása

- Abszolút score - Valószínűség-alapú
 - valószínűségszámítás alapján pontoz különféle algoritmusokkal
 - találatok statisztikailag szignifikánsak-e
 - kiszámítja, hogy a kísérleti adat és egy adott peptidből származó - elméleti - spektrum egyezése mekkora a valószínűséggel véletlen. Valós találat esetén ez a valószínűség kicsi
 - a score értékek alávethetők klasszikus statisztikai teszteknek
 - adott valószínűségi érték (leggyakrabban FDR) fölött tekintjük elfogadhatónak a találatot
 - így ezzel rendelkezésünkre áll egy olyan pontozási módszer, amellyel elvethetjük biztosan a random találatokat

Egyezés megtalálása/elfogadása

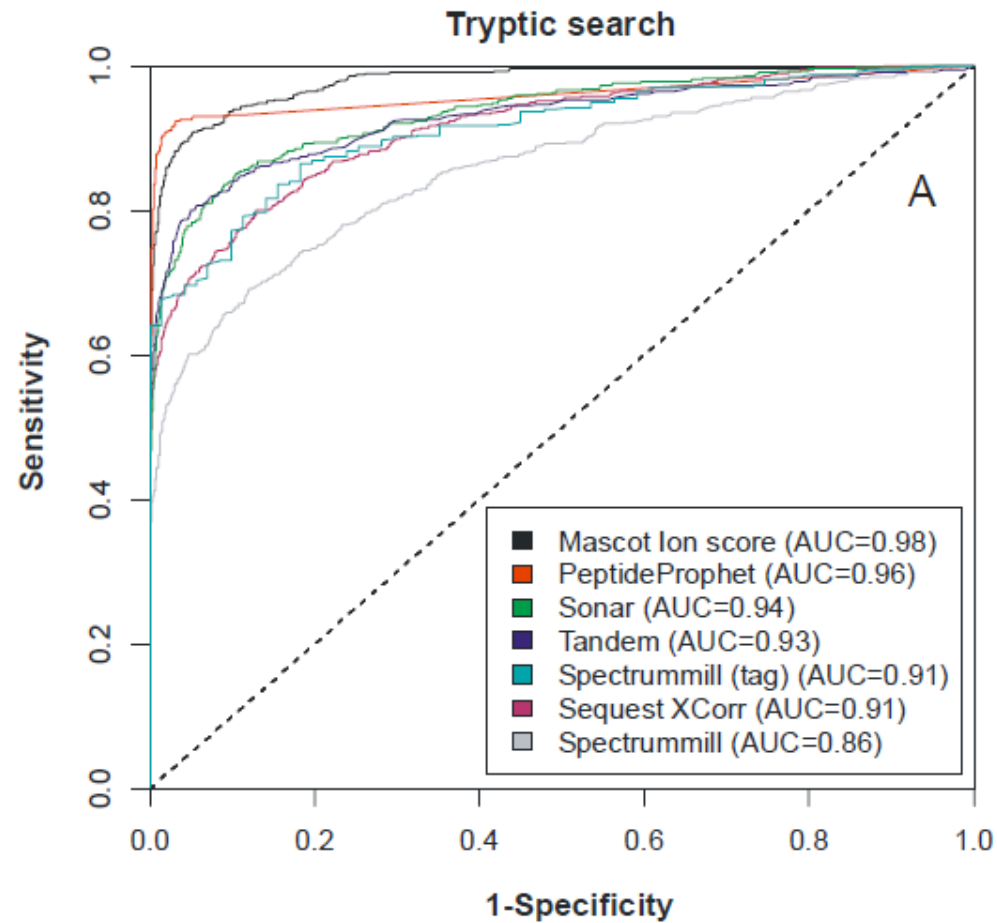
- Arbitrary score konvertálása probability score-rá: sok arbitrary score eloszlása normális. A kiugró érték nagyobb valószínűséggel való találat
- Normalizálás: pl. PeptideProphet-tel

Validálás

- Decoy adatbázis: minden más keresési paraméter megegyezik, a szekvenciák meg vannak fordítva vagy össze vannak keverve benne
- Decoy adatbázisból arra számítasz, hogy nem lesz találat

Validálás

ROC plot - trypsin (IPI db)



Kapp E. A., *et al.*,
Proteomics (HUPO-
PPP special issue),
August 2005

Reportálás

- Pl. ProteinProphet, DBParser
- ROC görbék

Keresés megghiúsulásának okai

- Gyenge minőségű adatok
- Prekurzor töltésének rossz meghatározása
- Proton helyett Na-mal repül
- Alulbecsült tömeghiba
- Szekvencia nincs benne az adatbázisban
- Meg nem adott PTM
- Enzim specificitás nem stimmel - szennyező proteázok jelenléte

Gyakorlat

- Bottom-up proteomika:
 - Peaklist generálás, paraméterek beállítása
 - Sequence tag keresés
 - MS/MS ion keresés
 - Mascot
 - MaxQuant
- Néhány hasznos eszköz bemutatása:
 - pI/MW tool
 - Peptide cutter
 - Uniprot