

**Humán vizeletből dúsított O-glikopeptidek tandem
tömegspektrometriás analízise
EThcD aktiváció alkalmazásával**

Doktori (Ph.D.) értekezés

Szerző: Pap Ádám

Témavezető: Dr. Darula Zsuzsanna



**Szegedi Biológiai Kutatóközpont
Proteomikai Kutatócsoport**

**Szegedi Tudományegyetem
Természettudományi és Informatikai Kar
Biológia Doktori Iskola**

2021.

Szeged

TARTALOMJEGYZÉK

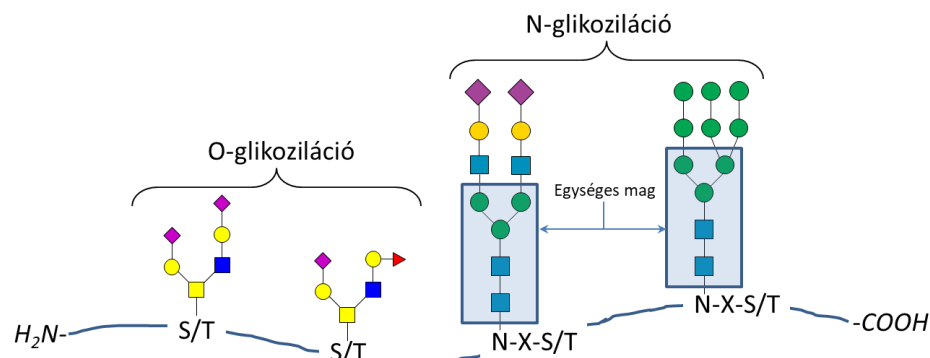
1. Bevezetés és Irodalmi Áttekintés	4
1.1 Az O-glikoziláció, mint poszttranszlációs módosítás.....	4
1.1.1 A mucin típusú O-glikoziláció.....	5
1.1.2 Makro- és mikroheterogenitás	8
1.2 Glikopeptidek analízise: Történeti áttekintés	8
1.3 Tömegspektrometriai alapok, fehérjeazonosítás LC-MS/MS technikával.....	10
1.4 O-glikopeptidek dúsítása komplex mintákból.....	20
1.5 O-glikopeptidek tandem tömegspektrometriája	21
1.5.1 Glikopeptidek fragmentációja ütközéses aktiválás esetén: nevezéktan, fő fragmension típusok.....	21
1.5.2 O-glikopeptidek analízise során alkalmazott fragmentációs technikák.....	23
1.5.2.1 Ütközéses aktiválás/rezonancia aktiválás ioncsapdában (ion trap CID, resonance CID).....	24
1.5.2.2 Ütközéses aktiválás ütközési cellában (beam-type CID, HCD).....	25
1.5.2.3 Gyökös fragmentáció: elektron transzfer disszociáció (ETD)	27
1.5.2.4 A gyökös és ütközéses aktiválás kombinációja: EThcD	28
1.6 O-glikopeptidek tömegspektrometriás adatainak elemzése	30
2. Célkitűzés.....	32
3. Anyagok és módszerek.....	33
3.1 Vizeletminták.....	33
3.2 Mintaelőkészítés	33
3.2.1 Emésztés cellulózmembrán szűrőn (FASP).....	33
3.2.2 Lektin affinitás kromatográfia	34
3.3 A tömegspektrometriás analízis és az adatkiértékelés során alkalmazott paraméterek....	35
3.3.1 nanoLC paraméterek	35
3.3.2 Glikopeptidek azonosítása automatizált adatkiértékeléssel (4.1 fejezet).....	36
3.3.3 Oxónium ion intenzitás-mintázat analízise HCD és EThcD spektrumokban (4.2 fejezet).....	37

3.3.4 Új O-glikán struktúrák azonosítása nem definiált módosítás kereséssel EThcD adatokból (4.3 fejezet).....	38
3.3.5 További új O-glikán struktúrák azonosítása GF-Hunterrel (4.4 fejezet)	39
4. Eredmények és megvitatásuk	40
4.1 Humán vizeletből dúsított glikopeptidek azonosítása automatizált adatkiértékeléssel	40
4.2 Oxónium ionok intenzitás-mintázata alacsony ütközési energiával (HCD NCE = 15%) felvett EThcD spektrumokban	43
4.3 Új O-glikán struktúrák azonosítása	47
4.3.1 Glikán fragmentációs szabályok alacsony ütközési energiával (NCE = 15%) felvett EThcD spektrumokban	48
4.3.1.1 Disziálsavat tartalmazó glikánok	48
4.3.1.2 O-acetil-neuraminsavat tartalmazó glikánok.....	49
4.4 További új O-glikán struktúrák azonosítása a HCD csúcslisták szűrésével.....	53
5. Konklúzió és kitekintés	68
6. Összefoglaló	70
6. Summary	75
7. Köszönetnyilvánítás	80
8. Irodalomjegyzék.....	81
9. Rövidítések jegyzéke.....	95
10. Mellékletek.....	97
11. Saját közlemények listája	102

1. BEVEZETÉS ÉS IRODALMI ÁTTEKINTÉS

1.1 Az O-glikoziláció, mint poszttranszlációs módosítás

Az életfolyamatok komplexitását többek között a fehérjék poszttranszlációs módosításai (PTM) teszik lehetővé olyan módon, hogy befolyásolják és/vagy kontrollálják a fehérjék biológiai funkcióit, lokalizációját, kölcsönhatásait, élettartamát. Az egyik leggyakoribb PTM a glikoziláció, ami különböző módosítások gyűjtőneve. Attól függően, hogy milyen atomon keresztül kapcsolódik a glikán a fehérjéhez, különböző típusokat különböztetünk meg. Az N-, illetve az O-glikoziláció a leggyakoribb (1. ábra), míg a C- és az S-glikoziláció a ritkább PTM-ek közé tartozik [Varki, 2015; Vliegenthart, 1998; Maynard, 2016]. A sejten belüli irányító funkciót betöltő GlcNAc módosításon kívül a glikozilációs folyamatok az endoplazmatikus retikulumban és a Golgi-készülékben játszódnak le, és így csak a sejten kívülre kerülő fehérjéket vagy fehérje-részeket érintik. Ebből a szempontból „sejten kívülinek” tekintjük bizonyos sejtstruktúrákat (ER, Golgi-készülék, lizoszómák) belső kompartmentumát is. Az N-glikánok esetében a cukor a fehérjéhez egy Asn oldalláncon keresztül kapcsolódik. A transláció folyamán egy $\text{GlcNAc}_2\text{Man}_9\text{Glc}_3$ szerkezet kerül a preferált, azaz konszenzus szekvenciában lévő aminosavra: $\text{Asn-Xxx-Ser/Thr(/Cys)}$, ahol Xxx nem lehet Pro. Ez az oligoszacharid a Golgi-készüléken áthaladva mindig komplex átalakuláson megy keresztül, de a $\text{GlcNAc}_2\text{Man}_3$ alapszerkezet (1. ábra) egységes az élővilágban. Az O-glikánok a fehérjékhez leggyakrabban Ser és Thr aminosavak oldalláncain keresztül kapcsolódnak, de Tyr módosítását is detektálták már [Varki, 2015; Halim, 2011; Steentoft, 2011; Trinidad, 2013], és ismert a hidroxilizin (Hyl) glikozilációja is [Sricholpech, 2012]. A fehérjéhez közvetlenül kapcsolódó monoszacharid többféle lehet (Fuc, Gal (kollagén Hyl glikoziláció), Glc, GalNAc, GlcNAc, Man és Xyl), erre épülhetnek a Golgi-készülékben különböző méretű és komplexitású cukorszerkezetek [Varki, 2015].



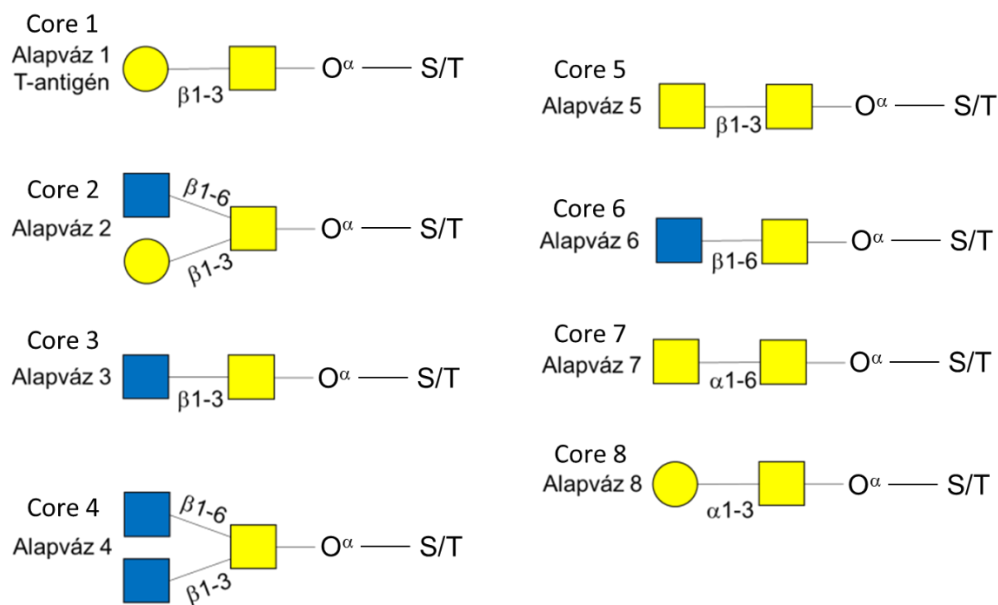
1. ábra A két leggyakoribb glikoziláció típus: az N-, illetve az O-glikoziláció.

1.1.1 A mucin típusú O-glikoziláció

A mucin típusú O-glikoziláció a leggyakoribb O-glikozilációs altípus emlősökben, így az emberben is. A szintézis kezdeti lépéséért felelős GalNAc-transzferáznak több mint 20 különböző izoformája ismert, amelyek feladata a GalNAc hozzákapcsolása a fehérje Ser, Thr esetleg Tyr oldalláncaihoz [Varki, 2015]. A GalNAc-transzferázok nagy számából és bizonyos esetekben átfedő, de eléggé definiálatlan szubsztrát-specifitásából [Gill, 2011] következik, hogy a mucin típusú O-glikozilációnak nincsen szorosabb értelemben vett konszenzus szekvenciája, habár bizonyos tendenciák megfigyelhetők. Például a Thr módosítás gyakoribb, mint a Ser glikozilációja [Darula, 2012], és csak néhány Tyr módosítást jelentettek idáig [Halim, 2011; Steentoft, 2011]. Az O-glikozilált szekvencia régiókban a Pro és kisebb mértékben a Ser, Thr, Ala és Gly aminosavak felülreprezentáltak [Wilson, 1991; Thanka Christlet, 2001].

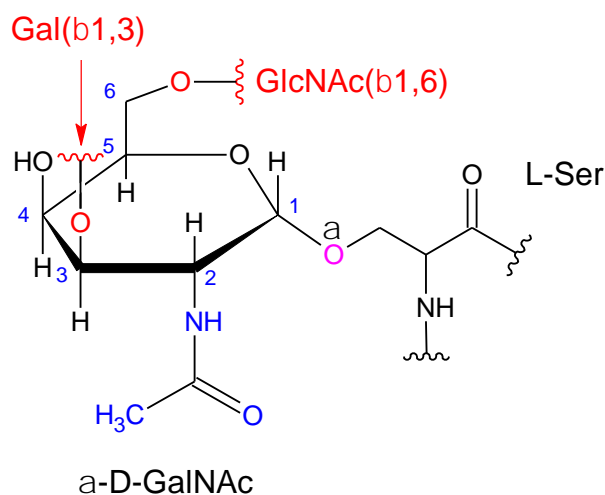
A glikánok sokszínűsége, így értelemszerűen az O-glikánoké is, a fehérjékhez és a nukleinsavakhoz képest hatalmas. Míg a nukleinsavak és a fehérjék lineáris polimerek, amelyek monomerjei egy sablon alapján kapcsolódnak egymáshoz (DNS esetében a sablon maga a DNS, míg fehérjék esetében a DNS-ről átíródó mRNS), addig a glikánok szintéziséhez ilyen sablon nem áll rendelkezésre [Varki, 2015]. Továbbá a fehérjék és a nukleinsavak monomerjeit egyfajta kémiai kötéstípus (fehérjék esetében a peptidkötés, míg a nukleinsavaknál a foszfodiészter kötés) egy adott térállásban kapcsolja össze. A monoszacharidok ezzel ellentétben α vagy β helyzetű glikozidos kötéssel kapcsolódhatnak egymáshoz, és nem csak lineáris, de elágazó struktúrákat is alkothatnak. A komplexitást tovább növeli, hogy egy adott monoszacharidnak több kapcsolódásra alkalmas pozíciója van, például a hexózoknak öt hidroxil-csoportja kapcsolódhat más cukrokhoz, illetve ezeket a hidroxilokat módosíthatják különböző kémiai csoportok (acetyl-csoport, szulfát, foszfát stb.). Szerencsére egy-egy organizmus az eddig megismert monoszacharid építőegységeknek csak egy részhalmazát hasznosítja és a belőlük képződő glikánstruktúrák számossága sem közelíti meg az elméleti maximumot a glikánok szintézisében résztvevő enzimek specifitása miatt [Varki, 2015]. A leggyakrabban előforduló monoszacharid típusok a következők: **pentózok** (öt szénatomos, semleges cukrok, pl.: D-xilóz), **hexózok** (hat szénatomos, semleges cukrok, pl.: D-galaktóz), **hexóزامinok** (hexózok, amelyek 2-es szénatomja egy amino csoporttal módosított, ez az aminocsoport lehet módosíthatlan, de leggyakrabban acetylált, pl.: N-acetyl-D-galaktóزامin), **6-deoxihexózok** (pl.: L-fukóz), **uronsavak** (hexózok, amelyek a 6-os pozícióban karboxiláltak, pl.: D-glükuronsav (GlcA)), **nonuloszon savak** (9 szénatomos, savas cukrok, amelyek közül gerincesekben a leggyakoribb az N-acetyl-neuraminsav (NeuAc)) [Varki, 2015]. A fent említett cukorcsaládok tagjai izomer szerkezetek, és ez a tény még nehezebbé teszi a glikánok analízisét. A cukorszerkezetek optikai

izomériát is mutatnak. Az élőlények csak az ún. D-konfigurációt hasznosítják, kivétel az L-Fuc. Az emberi szervezet a következő alkotórészekből építkezik: Gal, Glc, Man, GalNAc, GlcNAc, Fuc, NeuAc, GlcA és Xyl.



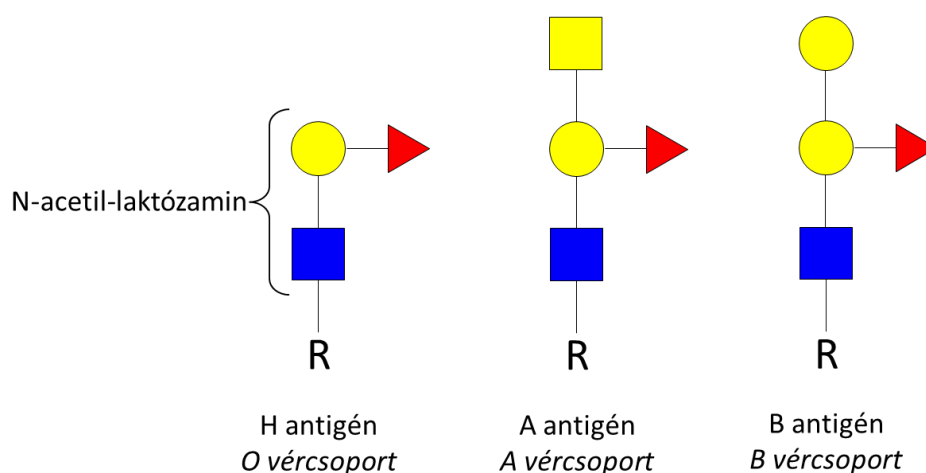
2. ábra A mucin típusú O-glikánok nyolc alapváza.

A mucin típusú O-glikánokban a Gal, GalNAc, GlcNAc, Fuc és NeuAc monoszacharidok fordulhatnak elő. Nyolc alapvázat különböztetünk meg (2. ábra) [Varki, 2015]. A közvetlenül a fehérjéhez α térállásban kötődő GalNAc-hoz a Golgi-készülékben egy Gal β 1-3 pozícióban történő hozzákapcsolásával alakul ki az 1-es alapváz (3. ábra).



3. ábra A mucin típusú O-glikoziláció iniciátor monoszacharidja, a GalNAc, egy α glikozidos kötéssel kapcsolódik a fehérjéhez – itt egy Ser oldalláncához. A monoszacharidhoz további cukoregységek a 3-as és a 6-os pozícióban lévő hidroxilcsoportokon keresztül kapcsolódhatnak.

Az iniciátor GalNAc, és a Gal(β 1,3)GalNAc diszacharid is immunogén (Tn- és T-antigén). Ezen kisebb, “befejezetlen” struktúrák számának megnövekedése a fehérjéken daganatos folyamatokkal hozható összefüggésbe [Varki, 2015; Gill, 2011]. A legegyszerűbb, 1-es alapvázat tartalmazó cukorszerkezet a Gal neuraminsavval történő módosításával keletkezik (NeuAc(α 2,3)Gal(β 1,3)GalNAc). Gyakori az 1-es típusú tetraszacharid is, amikor a GalNAc-hoz is kapcsolódik NeuAc α 2,6 kötésben (A szialil-GalNAc szerkezet önmagában szintén antigén, és előfordulását daganatos betegségekhez kötik [Munkley, 2016]). Az 1-es típusú diszialo szerkezetet Tyr módosításként is azonosították [Halim, 2011]. A 2-es alapváz az 1-es alapvázból keletkezik úgy, hogy a GalNAc-hoz még egy GlcNAc is kapcsolódik β 1,6 pozícióban (2. és 3. ábra). A 2-es alapváz is záródhat NeuAc-cal, illetve hosszabbítható N-acetil-laktózámmal (Gal(β 1,4)GlcNAc) egységekkel és különböző vércsoport-antigénekkal is (4. ábra) [Varki, 2015; Jin, 2017; Rossez, 2012]. Ez a két alapváz fordul elő a leggyakrabban. Például a szérumban O-glikánjainak nagyjából 80%-a NeuAc₁₋₂GalGalNAc és NeuAcGalGlcNAc(NeuAcGal)GalNAc szerkezet, azaz mucin-típusú 1-es és 2-es szerkezetek [Yabu, 2014]. A 3-as és 4-es alapvázak ritkábbak, például a légző és az emésztőrendszer bizonyos szöveteiben fordulnak elő [Varki, 2015]. A fennmaradó alapvázak (alapváz 5-8) ritkák és csak speciális mintákból lettek azonosítva [Varki, 2015]. Kovalens módosítások is előfordulnak a mucin típusú O-glikánokon, mint például a szialsavak O-acetilációja [Varki, 2015; Halim, 2011; Darula, 2019], a Gal és a GlcNAc szulfatálása [Varki, 2015; Solecka, 2016; Pap, 2020].



4. ábra Az ABO vércsoport antigének szerkezete.

A mucin típusú O-glikánok szerepét több biológiai folyamatban is igazolták már: patogénekkal való interakció [Ricciuto, 2008; Guzman-Aranguez, 2010], sejtadhézió [Lasky, 1992; Pan, 2014; Kariya, 2014], illetve proteolitikus folyamatok szabályozása [Pan, 2014; Peng, 2011;

Leuenberger, 2003; Goth, 2015; Goth, 2017]. Az O-glikozilációs mintázat megváltozását, a korábban említett daganatos folyamatokon kívül, több más betegséggel is kapcsolatba hozták. Például az IgA1 molekula hibás O-glikozilációját közölték IgA nefropátiában és Henoch-Schönlein purpura nefritiszben szenvedők esetében is [Allen, 1998].

1.1.2 Makro- és mikroheterogenitás

A különböző PTM-ekre általában bizonyos fokú heterogenitás jellemző. Ez alól az O-glikoziláció sem kivétel. Makroheterogenitáson azt értjük, hogy egy fehérje ismert O-glikozilációs helyei milyen mértékben módosítottak valamilyen O-glikánnal. Azaz előfordul-e, hogy egy adott Thr/Ser a fehérjepopuláció egy részében szabadon marad.

Egy adott glikozilációs hely különféle oligoszacharidokkal lehet módosítva. Azt, hogy egy adott Ser vagy Thr hány különböző glikoformában van jelen mikroheterogenitásként definiáljuk. Az irodalom alapján nincs nagy változatosság ebben a tekintetben [Trinidad, 2013; Medzihradszky, 2015 (A); Medzihradszky, 2015 (B)], habár a gyomor nyálkahártyájában található mucinok O-glikán analízise során [Jin, 2017; Rossez, 2012] ennek pont az ellentéte látszódtott beigazolódni. A kétféle heterogenitás jellemzése összefügg, és komplex feladat, hiszen nemcsak azt kell megállapítani, hogy egy adott glikozilációs hely módosított-e és ha igen, hány különféle szerkezettel, hanem azt is, hogy ezek az individuális glikoformák mennyiségileg hogyan oszlanak meg, és hogyan kombinálódnak egy hosszabb peptiden, és végül a teljes fehérjén. A dolgozatban bemutatott eredmények (4. fejezet) az O-glikozilációs helyek azonosítására, és az azonosított O-glikopeptidek mikroheterogenitásának jellemzésére irányult. A fehérjék makroheterogenitását nem tanulmányoztuk.

1.2 Glikopeptidek analízise: Történeti áttekintés

Már a XX. század elején jelentős volt a szénhidrátok kémiájával, metabolizmusával kapcsolatos kutatás. Biológiai jelentőségüket, azon túl, hogy energiaforrásnak tekinthetők, azonban csak jóval később kezdték felismerni, habár a mucinokat, a nyálkahártyáinkat burkoló/védő erősen O-glikozilált fehérjéket kb. 150 évvel ezelőtt fedezték fel [Varki, 2015] és az N-glikozilációt is leírták az ovalbumin bioanalitikai jellemzése során [Conchie, 1969]. A kémiai és bioanalitikai módszerek fejlődésével kb. a század közepén kezdett tudásunk gyarapodni a különböző glikokonjugátumokról, közöttük a glikoproteinekről is. A kutatásnak ebben a fázisában a szénhidrátokat és a fehérjéket külön tanulmányozták, és ezt a megközelítést ma is alkalmazzák. Az oligoszacharidokat eleinte kémiai módszerekkel szabadították fel, de az N-glikánok eltávolítására, hamarosan találtak egy univerzális enzimet, a peptid-N-glikozidáz F-et

(PNGaseF) [Makino, 1966]. Az enzim az N-glikán legelső GlcNAc egysége és a fehérje aszparagin aminosava között hasítja el az N-glikozidos kötést, így az egész N-glikán eltávolítható. Az O-glikánok intakt módon történő enzimátikus eltávolítására nincsen lehetőség. Több enzim együttes és szekvenciális alkalmazása szükséges ahhoz, hogy elérjük azt, amit az N-glikánok esetében egyetlen enzimmel sikerült. Ugyanakkor a Ser és Thr módosítások érzékenyek a magas pH-ra, így eltávolíthatóak β -eliminációval. Tehát, ha célunk a jelenlévő O-glikánok összességének megismerése, akkor legcélszerűbb a glikopeptideket lúggal (0,1 M NaOH) kezelni, és a szabad glikánokat azonnal, a hasítással egyidőben redukálni (NaBH_4), hogy az egyes alegységek elvesztését (peeling) megelőzzük [Carlson, 1968]. A marha fetuint módosító mucin típusú core 1 és core 2 szerkezetek pontos jellemzéséhez szükség volt a bázissal felszabadított glikánok több lépéses kémiai analízisére és az izolált glikopeptidek szekvenciális enzimátikus hasítására is [Spiro, 1974; Edge, 1987]. Később komplex NMR vizsgálatok vették át a hosszadalmas kémiai és enzimátikus vizsgálatok szerepét [Pollex-Krüger, 1993; Manzi, 2000], és a glikoziláció enzim-rendszerének egyre több komponensét azonosították és jellemezték [Sukeno, 1972; Powell, 1974; Schwyzer, 1977; Beyer, 1979; Kuhns, 1993; Kuhns, 1995; Wang, 1992; Wen, 1992].

Manapság a glikán-populáció vizsgálatára többféle módszer létezik. A redukáló véget megfelelő hidrofób és/vagy kromofór vegyülettel derivatizálva grafit- vagy akár fordított fázison is végzett kromatográfiával frakcionálhatjuk az elegyet, vagy alkalmazhatunk kapilláris elektroforézist. A detektálás tömegspektrometriás adatok alapján, míg az azonosítás retenció idő sztenderdekkel történik. A kvantitatív meghatározás pedig a megfelelő optikai módszerekkel is lehetséges. Az eljárások zömét N-glikoziláció jellemzésére fejlesztették ki. Itt azokat a tanulmányokat említem csak meg, amelyek fontosnak, hasznosnak bizonyultak az O-glikoziláció jellemzésére is. Karlsson és mtsai. humán nyálból izolált MUC5B glikozilációját vizsgálták. A lehasított O-glikánokat grafit-oszlopon kromatográfiásan frakcionálva, negatív ionizációval, on-line LC-MSⁿ (MS² és MS³) analízisnek vetették alá ioncsapdát használva detektornak [Karlsson, 2004]. Hasonló módon jellemezték a gyomornyálka mucinjainak cukorszerkezeteit Jin és mtsai. [Jin, 2017]. Rossez és mtsai. szintén a gyomor nyálkahártya mucinjait vizsgálták. Q-TOF tömegspektrométert használtak detektornak, pozitív és negatív üzemmódban is, és off-line frakcionálást, illetve NMR vizsgálatokat is végeztek [Rossez, 2012]. Yabu és mtsai. a humán szérum O-glikánjait analizálták: az aminopiridinnel derivatizált glikánokat szétválasztották gyenge anioncserélő kromatográfiát alkalmazva, majd az így elválasztott mintákat tovább frakcionálták kétdimenziós kromatográfiával (méret és hidrofobicitás szerint). A tömegspektrometriás analízist egy 3D ioncsapdás készülékkel végezték, illetve a felszabadított

és jelölt cukrokat neuraminidázzal és fukozidázzal is emésztették [Yabu, 2014]. Fontos megemlíteni, hogy a tömegspektrometria önmagában csak limitált szerkezeti információt biztosít (1.5 fejezet), és a felsorolt analízisek során csakis azért lehetséges „azonosítani” a komponenseket, mert korábban gondosan és alaposan jellemzett standard szerkezetek állnak rendelkezésre. Ezeket tanulmányozva sikerült a megfelelő elválasztási módszereket kidolgozni, és rögzíteni a szerkezeti változások hatását pl. az elúciós sorrendre, a retenciós időre.

A glikán-populáció analízise sok mindent elárul a mintában levő fehérjék O-glikozilációjáról. Felfedezhetjük pl., ha egy fiziológiai változás vagy betegség komolyan megváltoztatja a glikoziláció folyamatát. Ugyanakkor egy lényeges információt elveszítünk: nem tudhatjuk meg, mely fehérjék, mely pozíciókban és milyen szerkezetekkel voltak módosítva. Így az is rejtve marad, ha egy glikozilációban bekövetkezett változás, amit egy betegség esetleg előidézett, a glikoproteinek összességét érinti-e, vagy csupán egyetlen vagy néhány fehérjét befolyásol. Ugyanakkor, ha csupán néhány kis mennyiségben jelenlevő fehérje glikozilációja módosul, azt nem feltétlenül vesszük majd észre, ha az összes fehérje cukorszerkezeteit tanulmányozzuk. Emiatt fontos a glikopeptidek vizsgálata.

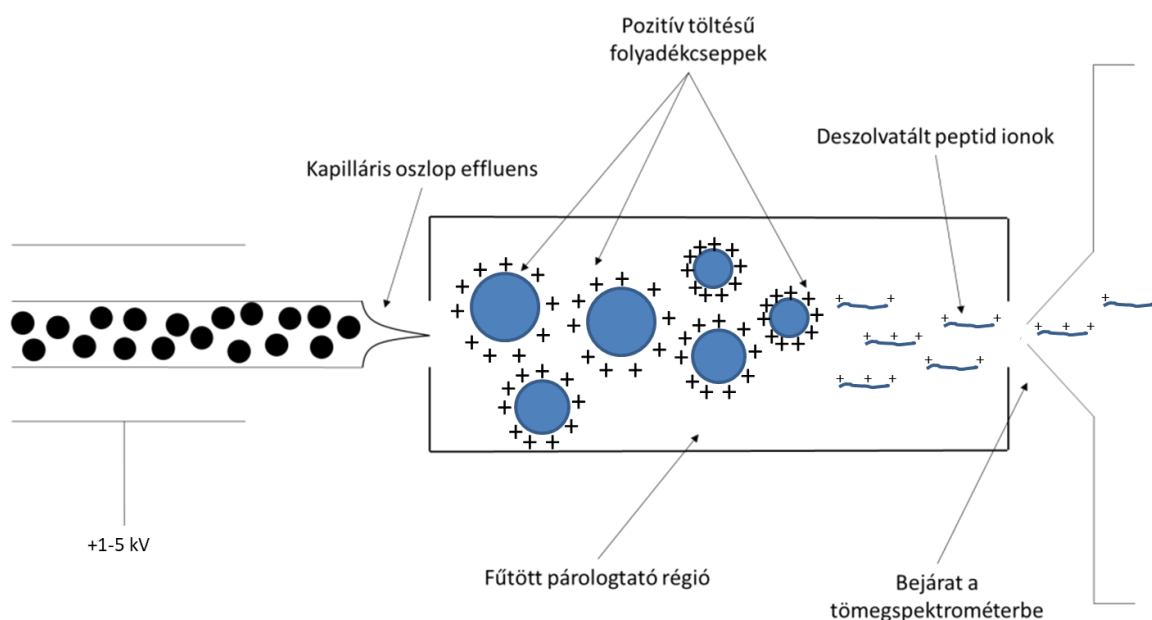
A tömegspektrometriát már az első ún. „lágyszó” ionizációs módszerek (Fast Atom Bombardment, Liquid Secondary Ionization MS) megjelenésétől alkalmazták O-glikoziláció analízisére. Mind a peptid- [Biemann, 1990], mind a glikán-fragmentáció [Domon, 1988] nevezéktana akkor született. Azonban a mennyiségi követelmények és a technikai akadályok miatt csak egyes rekombináns vagy természetes fehérjék glikopeptid-frakcióit analizálták ily módon, eleinte nagy energiájú CID analízissel, négy-szektoros tömegspektrométerekkel [Settineri, 1990; Medzihradszky, 1990; Medzihradszky, 1996], majd ESI-MS/MS-sel: többek között Q-TOF készülékekkel [Chalkley, 2001] és off-line MALDI-TOF technikával is [Huberty, 1993]. Komplex minták nagy volumenű analíziséhez a jelenlegi gyorsabb és pontosabb készülékekre, és újfajta aktiválási módszerekre is szükség volt, amint azt az 1.5 fejezetben ismertetem.

1.3 Tömegspektrometriai alapok, fehérjeazonosítás LC-MS/MS technikával

A tömegspektrometriás analízis alapfeltétele a vizsgálni kívánt anyag ionizálása és gázfázisba juttatása, amelyért a tömegspektrométer ionforrása felel. A proteomikai analízisek során az electrospray ionizációt (ESI) [Fenn, 1989] alkalmazzuk a leggyakrabban. Népszerűségének oka, hogy az ionforrással a poláris molekulák jól ionizálhatók. Ellentétben a korábbi ionizációs technikákkal (EI, CI), az ionizációs folyamat nem vákuumban történik, és nem destruktív, azaz a minta komponensei az ionizáció folyamata során nem vagy csak kismértékben

fragmentálódnak. Az ionizáció során (lásd alább) többszörösen töltött ionok képződnek, azaz elegendő szűkebb tömegtartományt monitorozni. Végül, de nem utolsósorban ez az ionforrás az, ami lehetővé teszi a folyadékkromatográf összekapcsolását a tömegspektrométerrel (LC-MS), amely kapcsolt technika ma a legelterjedtebben alkalmazott, nem csak a proteomika területén.

A HPLC felől érkező kapilláris egy fémből vagy üvegből készült emitterben végződik, amelyre nagyfeszültséget kapcsolnak ($\pm 1-5$ kV). Ez tulajdonképpen maga az ESI ionforrás. Az ionizáció folyamata részleteiben a mai napig nem teljesen ismert, de a legelfogadottabb elmélet [Kearle, 2009] a következőképpen magyarázza az ionképződést (5. ábra), pozitív módban, peptidekre: A HPLC felől érkező savas eluens a túra kapcsolt nagyfeszültség miatt ún. Taylor kúpot alkot az emitter hegyén, amelyről apró folyadékcseppek válnak le. Ezek az apró cseppek egy fűtött kapillárison haladnak keresztül (a Thermo készülékekben), ahol a párolgás miatt a folyadékcseppek mérete egyre csökken. A csökkenő folyadékfelület megnöveli a pozitív töltések sűrűségét a felszínen, amely a zsugorodó csepp szétválását eredményezi (Coulomb-robbanás). Ez a folyamat addig ismétlődik, amíg a folyadék teljesen el nem párolog és nem marad más vissza, mint a gázfázisú, többszörös töltéssel rendelkező fehérje vagy peptid kation.



5. ábra Az ESI ionforrás vázlatos felépítése.

A proteomikában az ESI mellett a MALDI (Matrix-assisted laser desorption/ionization) ionforrás használata is gyakori [Hillenkamp, 1991]. Az ionizálni kívánt anyagot együtt kristályosítják lézerfény által gerjeszthető savas karakterű mátrix-szal (a két leggyakrabban alkalmazott mátrix peptidok elemzésére a dihidroxi-benzoészav és az α -ciano-4-hidroxi-

fahéjsav). A mátrixmolekulák deszorbeálódnak a lézersugár hatására. A deszorpció során a vizsgálandó komponensek is a légtérbe kerülnek, és a mátrixmolekulákkal kölcsönhatásban MH^+ ionokat képeznek, többszörösen töltött ionok ritkán keletkeznek. Az ESI-től eltérően a MALDI nem alkalmas on-line frakcionálással kapcsolt mérésekre, de jól alkalmazható egyszerű elegyek gyors tömegspektrometriás analízisére, illetve off-line gyűjtött frakciók vizsgálatára.

A sikeres ionképzés tehát az első lépés a tömegspektrometriás analízis felé. ESI esetében, a zsugorodó cseppek a potenciál-különbség hatására a leföldelt interface felé haladnak, és egy parányi nyíláson (orifice) lépnek be a további folyamathoz nélkülözhetetlen nagy vákuumba, azaz az analizátor belsejébe. Az analízis központi eleme a tömegmérés, amelyet a tömegspektrométer analizátora tesz lehetővé. Az ionok szétválasztása tömeg/töltés (m/z) alapján történik, amit a detektálás követ. Az analizátorok fontos jellemzői a felbontás, a tömegpontosság, a pásztázási vagy mérési sebesség és a vizsgálható tömegtartomány.

Az analizátor felbontása azt fejezi ki, hogy milyen tömegkülönbséggel tud két iont (egy m és egy $m+\Delta m$ csúcsot) elválasztani. A felbontás egy dimenzió nélküli szám és fordítottan arányos a Δm tömegkülönbséggel, amit felbontóképességnek neveznek. Tehát minél nagyobb egy analizátor felbontása, annál kisebb tömegkülönbséggel rendelkező két iont tud elválasztani és külön csúcsként detektálni. A felbontás és a Δm felbontóképesség közötti matematikai kapcsolatot az A. képlet szemlélteti [Kinter, 2000].

$$R = \frac{m}{\Delta m}$$

A. képlet, ahol R a felbontás, m egy adott ion m/z értéke és Δm a felbontóképesség.

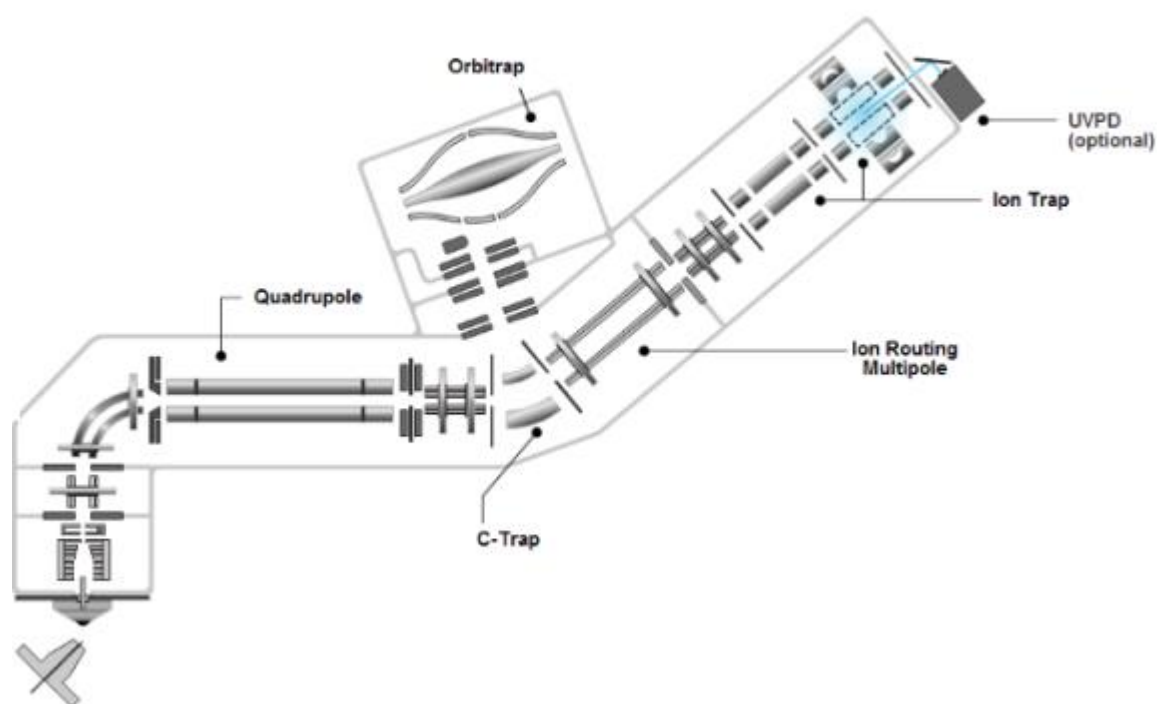
Manapság a felbontóképességet általában a csúcs magasságának felénél megállapított csúcshélességgel (félértékszélesség, az angol nevezéktan szerint full width at half maximum, FWHM) adják meg.

A tömegpontosság azt fejezi ki, hogy az analizátor mekkora eltéréssel képes megmérni egy adott ion tömegét az elméleti tömegértékhez viszonyítva (természetesen megfelelő kalibráció mellett). A mérési hibát megadhatjuk Daltonban (abszolút tömegpontosság), vagy pedig azt határozzuk meg, hogy az eltérés hány milliommód része a mért tömegnek (relatív tömegpontosság) [Kinter, 2000]. Nyilvánvaló, hogy az utóbbi érték sokkal precízebben jelzi a mérések megbízhatóságát,

hiszen ugyanaz a pl. 1 Da különbség kis m/z értéknél óriási hibát jelent, míg egy intakt fehérje molekulatömeg meghatározásánál rendkívüli pontosságot.

A pásztázási sebesség olyan készülékek jellemzője, amelyek a képződött ionokat csak szűkebb tömegtartományban tudják detektálni (pl. szektoros vagy kvadrupól tömegspektrométerek) és arról tájékoztat, hogy egy adott tömegtartomány analízisét mennyi idő alatt lehet végrehajtani. Amikor viszont a képződött ionokat kis- vagy nagyfelbontású (Orbitrap) ioncsapdák segítségével mérjük, akkor az adatgyűjtés sebességét szokás megadni. Az általunk használt készüléknél az MS/MS adatgyűjtés frekvenciája 40 sec^{-1} is lehet a lineáris ioncsapdánál, és 30 sec^{-1} az Orbitrap méréseknél.

A dolgozatban bemutatásra kerülő adatok egy ún. tribrid készülékkel lettek felvéve, amelyben három analizátor található: kvadrupól, kétcellás ioncsapda és egy Orbitrap. A műszerben való elhelyezkedésüket a 6. ábra szemlélteti.



6. ábra Thermo Scientific™ Orbitrap Fusion™ Lumos™ Tribrid™ tömegspektrométer felépítése [Thermo Fisher Scientific, 2017].

A kvadrupól (vagy kvadrupól tömegszűrő) egy olyan analizátor, amely négy fémrúdból áll. A rudak egy képzeletbeli négyzet csúcspontjaiban ülnek és az egymással szemközti rudakra azonos

előjelű feszültséget kapcsolnak. Az ionok az analizátorba lépve spirális mozgást végeznek. A rudakra kapcsolt egyen- és váltófeszültség aránya szabja meg, hogy egy adott m/z -vel rendelkező ion pályája stabil lesz-e vagy sem, azaz sikeresen áthalad-e az analizátoron. A kvadrupólt általában egységnyi felbontáson üzemeltetik, azaz a $\Delta m = 1\text{Da}$ [Kinter, 2000], de képes magasabb felbontás elérésére is, aminek maximuma 3000 körül mozog. Ezzel az analizátorral könnyen elérhető az 1000 Th sec^{-1} pásztázási sebesség (sőt ennél magasabb érték is), így kiválóan alkalmas kromatográfiai rendszerekkel való összekapcsolásra [Kinter, 2000]. A kvadrupól analizátor manapság is népszerű és széles körben, sokféle feladatra használják egyszerű és tandem készülékekben. A mi szempontunkból az volt a fontos, hogy a kvadrupól analizátorok jól használhatók MS/MS kísérletben prekursorionok kiválasztására.

Az ioncsapdák abban a tekintetben hasonlítanak a kvadrupól analizátorokhoz, hogy bennük az ionok oszcilláló mozgást végeznek. Az analizátor neve működésének lényegét is magában hordozza. Az ioncsapdák a megfelelő radiofrekvenciás (RF) feszültség alkalmazásával képesek egy adott m/z tartomány csapdázására. A tömegmérés során a csapdában az RF feszültségérték úgy változik, hogy m/z -függően folyamatosan növeli az ionok oszcillációjának amplitúdóját, amelynek hatására az ionok növekvő m/z érték szerint hagyják el az analizátort és lépnek a detektorba. Ezt a jelenséget tömegszelektív instabilitásnak hívják. Ezt kihasználva lehet az ioncsapdában ionokat izolálni MS/MS kísérlethez is. Az ioncsapdák felbontás tekintetében a kvadrupól analizátorokhoz hasonlóak, tömegpontosságuk tipikusan $0,5\text{-}0,6\text{ Da}$. További jellemzőjük, hogy a legkisebb m/z érték detektálása MS/MS kísérletben a prekursor m/z értékétől függ. Ez a gyakorlatban azt jelenti, hogy az ioncsapda nem képes visszatartani a prekursor m/z értékének egyharmadánál kisebb m/z -vel rendelkező ionokat [Kinter, 2000]. Korábbi munkáink során használtunk ioncsapdát glikopeptidek analízisére, de a dolgozatban ismertetett kutatás során az ioncsapdát csak az ETD aktiválásra (1.5.2.3 fejezet) alkalmaztuk.

Az Orbitrap egy nagy felbontással ($R_{\text{FWHM}} = 500000 @ m/z 200$) és nagy tömegpontossággal ($\leq 5\text{ppm}$) rendelkező analizátor, amely két részből, egy külső hordószerű és egy belső orsó alakú elektródából áll. Az ionok az analizátorba lépve gyűrűt formálnak a belső elektróda körül és ugyanakkor oszcilláló mozgásba kezdenek ugyanezen elektróda mentén. A különböző m/z -vel rendelkező ionok eltérő oszcillációs frekvenciával rendelkeznek, amely a tömegmérés alapjául szolgál. A tömegmérés során a készülék egy látszólagos ionáramot rögzít a külső elektródába épített szenzor segítségével. A látszólagos ionáramot a szenzor előtt elhaladó oszcilláló ionok keltik, amelyet a készülék egy idő-intenzitás koordináta-rendszerben rögzít. Ez egy olyan „hullámkép”, amely elemi hullámok szuperpozíciójából áll. Fourier transzformációval az

összetett hullámkép felbontható különböző frekvenciájú elemi hullámkomponensekre, amelyek tulajdonképpen maguk az oszcilláló ioncsomagok. Miután megtörtént a Fourier transzformáció, a frekvenciakomponenseket a B. képlet segítségével feleltetik meg egy adott m/z értéknek.

$$\omega_z = \sqrt{\frac{k}{m/z}}$$

B. képlet, ahol ω_z az axiális oszcilláció frekvenciája, k a műszerre jellemző állandó és m/z pedig a tömeg/töltés arány.

Az m/z érték és az oszcilláció frekvenciája között fordított arányosság áll fenn. Tehát minél nagyobb egy adott ion oszcillációs frekvenciája, annál kisebb az m/z értéke. Az Orbitrap analizátor felbontóképessége egyenes arányban csökken a növekvő molekulatömeggel. Az aktuálisan elérhető felbontóképessége az eszköznek 1000000 (m/z 200-nál), de az ilyen pontos tömegmérésnek a hosszú analízisidő a hátulütője, azaz a felbontás növelésével egyenes arányban növekszik az analízis ideje. 60000-es felbontási értéket választva az analízis ideje ~100 ms, amely a nano-LC rendszerek csúcs-szélességéhez jól illeszthető. A felbontóképesség megfelelő értéke akkor játszik nagy szerepet, ha a minták sok izobár komponenst tartalmaznak. Ez a jelenség szinte minden komplex minta esetében előfordul [Kinter, 2000; Scigelova, 2006], de a kromatográfias frakcionálás többnyire elválasztja egymástól ezeket a komponenseket. Kutatásunk során a tömegmérés minden esetben az Orbitrap analizátorral történt.

A fentebb leírt típusokon kívül más analizátorok is használatosak a proteomikában. A TOF (time-of-flight), azaz repülési időt mérő analizátor tökéletes a szakaszos ionáramot produkáló MALDI ionforrással párosítva, illetve hibrid készülékekben (Q-TOF) ESI ionizációval is kombinálható. A nagy felbontású, mágneses FT-ICR (Fourier-transform ion cyclotron resonance) készülékek komplex működtetésük és drága fenntartási költségeik miatt mára háttérbe szorultak a proteomikai kutatásban. Azonban speciális, rendkívül nagy felbontású ($R_{FWHM} > 2\,700\,000$ @ m/z 400) FT-ICR készülékeket manapság is építenek [Smith, 2018], és intakt fehérjék, fehérjekomplexek vizsgálatára használják [Deighan, 2020].

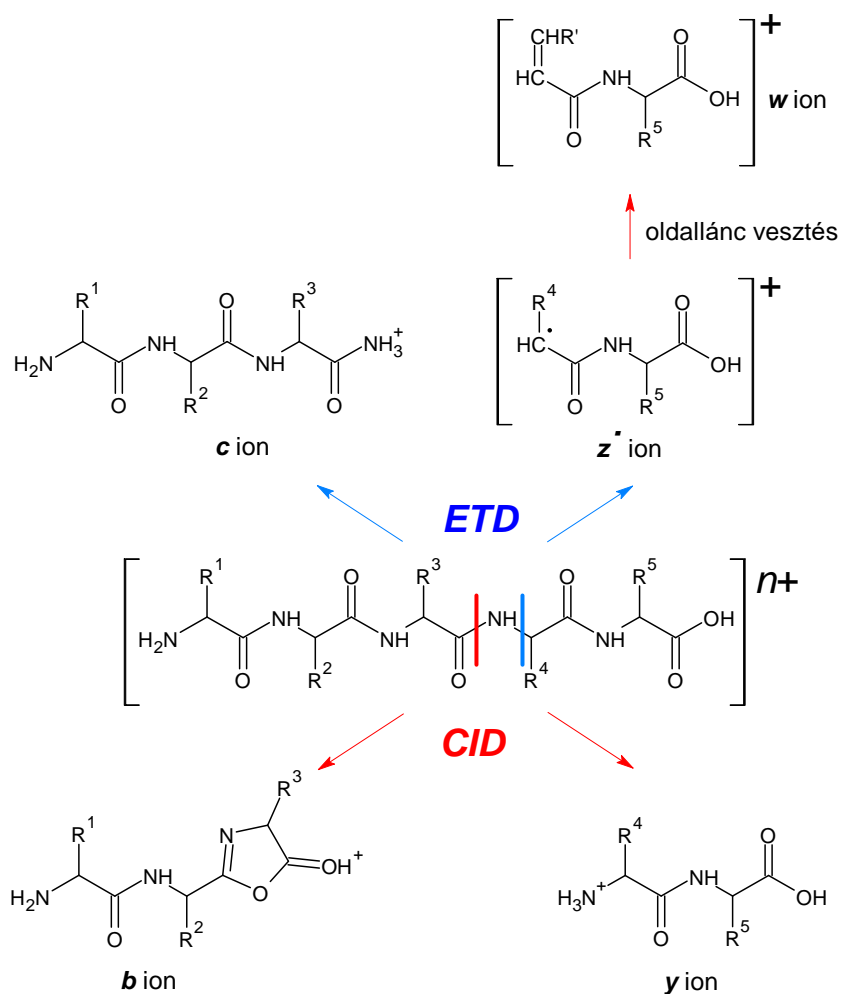
Egy molekula tömegének pontos meghatározása nem elegendő a molekula szerkezetének jellemzésére, ehhez további adatokra van szükség. A klasszikus EI spektrumok gyakran tartalmazták a molekulaiont és a fragmenseket is, egyszerre biztosítva a méret- és szerkezet-meghatározáshoz szükséges adatokat. A „lágy” ionizációs technikák, mint az ESI, első megközelítésben csak a molekulatömeg-meghatározást teszik lehetővé, mivel normál ionizációs

körülmények között fragmentáció nem fordul elő. A szerkezeti információk kinyeréséhez a bennünket érdeklő molekulaiont (prekursor ion) ki kell választani, azaz el kell különíteni a jelenlévő többi iontól, majd aktiválni kell, hogy a megfelelő, információt hordozó darabokra hasadjon, és végül ezeket a fragmenseket detektálni kell. Az így lefolytatott ún. MS/MS analízis akár *de novo* szerkezet-meghatározáshoz is elegendő információt biztosít. A fentebb leírt a feladatsort tandem tömegspektrometriával tudjuk végrehajtani, azaz többnyire két analizátorra van szükségünk. Az első analizátor, esetünkben a kvadrupól, a tömegmérés (MS1) során teljes transzmisszióra van beállítva (a beérkező molekulák tömegét az Orbitrappal határozzuk meg), viszont a fragmentációs lépés (MS2) előtt, amely majd egy ütközési cellában valósul meg, úgy kapcsol be, hogy csak a megfelelő m/z értékű ionok haladjanak át rajta. Az ütközési cellában keletkezett fragmensionok szortírozásáról és detektálásáról a második analizátor, azaz az Orbitrap gondoskodik.

A proteomikában a legelterjedtebben alkalmazott analitikai technika az LC-MS/MS analízis. Ebben az esetben a tandem tömegspektrométerhez egy folyadékkromatográf kapcsolódik, és a kromatográfiás oszlopról az eluens azonnal a tömegspektrométerbe kerül, amely a folyadékkromatográf detektoraként szolgál. A folyadékkromatográfia (LC) során a mintát leggyakrabban hidrofobicitás alapján (fordított fázisú folyadékkromatográfia, RP-HPLC) frakcionálják. Az eluálódó komponensek tandem tömegspektrometriás analízise (MS/MS) két fő lépésben valósul meg. Első lépésben az oszlopról éppen kilépő effluensben található molekulák intakt tömegmérése történik meg (MS spektrum, survey scan, MS1 szint; Orbitrap analizátorral). Ezután a tömegspektrométer egy megfelelően felépített program paramétereit követve az MS spektrumban található adott számú legintenzívebb iont (prekursor ionok) egyenként izolálja, fragmentálja, a keletkezett fragmensionok tömegeit megméri és egy MS/MS spektrumban rögzíti. Ez az úgynevezett adatfüggő mérés (data dependent acquisition, DDA).

Manapság népszerű módszer az adatfüggetlen mérés (data independent acquisition, DIA) is, amelyet kvantitatív célokra használnak. A DIA lényege, hogy a monitorozott MS1 tömegtartományt felosztjuk szélesebb ablakokra (~10 Da) és az ezekben jelenlévő összes prekuzort egyszerre fragmentáljuk. A rögzített MS/MS spektrumok sokkal komplexebbek, mint egy DDA kísérlet esetében. Előnye a DDA módszerhez képest, hogy a monitorozott tömegtartomány összes prekuzoráról van MS/MS szintű információnk, azonban egyben ez a hátránya is a módszernek, mert a bennünket érdeklő információ kinyerése bonyolultabb [Zhang, 2020].

A leggyakrabban alkalmazott fragmentációs technika az ütközéses aktiválás (Collision-induced dissociation, CID), ami kivitelezhető ioncsapdában vagy ütközési cellában (beam-type CID, HCD). Ez utóbbi során a prekursor ion egy ütközési cellában inert gázatomokkal (N₂ (levegő) vagy Ar) ütközik és az ütközés során elnyelt mozgási energia vibrációs energiává alakulva bizonyos kötések hasadásához, azaz a prekursor fragmentációjához vezet. A keletkezett fragmensek további ütközéseknek vannak kitéve, azaz tovább hasadhatnak. Az ütközéses aktiválás a molekulában a leggyengébb kötés(ek) hasadását idézi elő. Peptidek esetében ez leginkább a peptidkötés, azaz a karbonil szénatom és az aminos csoport nitrogén atomja közötti kötés. A keletkező fragmensionok közül a legfontosabbak az ún. szekvencia ionok, amelyek a peptid aminosav-sorrendjének meghatározását teszik lehetővé. A peptid eredeti N-terminálisát tartalmazó fragmenseket b-ionoknak, míg az eredeti C-terminálisát tartalmazókat y-ionoknak nevezzük (7. ábra) [Biemann, 1990]. Ezeket az ionokat a megfelelő terminus felől számozzuk.



7. ábra Peptidek CID aktiválása során a peptidkötés elhasadásával b vagy y szekvenciaionok keletkeznek, amelyeket a peptid N-, illetve C-terminálisa felől számozzuk [Biemann, 1990]. Az ETD aktiváció során c és z[•] ionok [Biemann, 1990] keletkeznek. A z[•] ionból w ion keletkezhet oldalláncvesztéssel [Johnson, 1987].

A szekvencia ionokon kívül – többszörös kötéshasadással – keletkezhetnek még immónium ionok, illetve belső fragmensionok. Az immónium ionok ($^+NH_2=CHR$) az MS/MS spektrum alacsony tömegtartományában detektálhatóak (m/z 30-159), jelenlétük a peptid aminosav összetételéről ad információt. A belső ionok két vagy több aminosav-hosszúságú b-fragmens típusú belső szekvenciárészletek [Kinter, 2000]. A szekvencia-, és belső ionok veszíthetnek vizet vagy ammóniát (ún. szatellit-ionok). Vízvesztésre hajlamosító aminosavak általában a Ser, Thr, Asp, Glu. Ammóniavesztés leggyakrabban Asn, Gln, Lys és Arg aminosavakból lehetséges [Sun, 2008]. A b-ionok CO veszteséssel ún. a-ionokká alakulhatnak, ez igaz a belső ionokra is.

Az ütközéses aktiválás (ioncsapda és „beam-type” CID) mellett használható alternatív aktiválási módszer az elektron transzfer disszociáció (ETD) [Syka, 2004]. Ennél a technikánál egy reagens gyök anion (az általunk használt készülékben fluorantén) és a fragmentálni kívánt, többszörösen töltött prekursorion között elektrontranszfer játszódik le. Az így létrejött gyökion instabil, és az elektronbefogás helyén elhasad. A gyök anion előállítására EI forrást használnak. A fragmentáció során főképpen a peptidgerinc hasad. A peptidkötés nitrogénje és az alfa szén közötti kötés elhasadásával c vagy z^* ionok keletkeznek [Syka, 2004] (7. ábra). A c ionokat a peptid eredeti N-terminálisa, míg a z^* ionokat a peptid eredeti C-terminálisa felől számozzuk. Hidrogénvándorlással $c-1^*$ és $z+1$ fragmensek is képződhetnek, különösen duplán töltött prekursorokból [Bakken, 2004]. Az oldalláncok többnyire érintetlenek maradnak, így a módszer különösen alkalmas könnyen hasadó módosításokat hordozó peptidek (pl. foszfo- vagy glikopeptidek) analízisére [Mikesh, 2006]. A módosítások helye egyértelműen meghatározható a módosított szekvencia ionok alapján.

Az eddig bemutatott aktivációs technikákhoz képest az EThcD egy újfajta, hibrid módszer. Első lépésben az ETD aktiváció történik meg az ioncsapdában, majd az így keletkezett heterogén ionpopulációt az ütközési cellában HCD technikával fragmentálják tovább [Frese, 2012]. Az ETD fragmensek többnyire nem hasadnak tovább (kivételem a w-ion képződés, lásd lejjebb), a másodlagos fragmentáció az aktivált állapotban megrekedt molekulaionokat érinti. A „kombinált” spektrum jelentősebb információ-tartalmával megkönnyíti a vizsgált komponensek azonosítását, és ez különösen igaz „egyoldalú” fragmentációra hajlamosabb komponensekre. A módszert többek között Ser-, Thr-foszforilált peptidek analízisére javasolták, amelyeknél a foszforsav β -eliminációja a preferált fragmentációs lépés. EThcD aktivációval jobb minőségű és nagyobb információ-tartalmú adatokat sikerült generálni, amely javította a foszforiláció helymeghatározását a peptiden belül [Frese, 2013]. Más poszttranszlációs módosításokat hordozó peptidek – köztük a glikopeptidek [Frese, 2012] – analízisére is kezd elterjedté válni

ez az újfajta MS/MS technika. Az EThcD technika során a kiegészítő HCD aktiváció miatt az ETD-ben keletkezett z^+ fragmensek oldallánc vesztéssel w ionokká alakulhatnak (7. ábra) (így pl. megkülönböztethetőek az izomer Leu és Ile aminosavak) [Johnson, 1987; Stults, 1987; Zhokhov, 2017].

A különböző hosszúságú polipeptid-láncok tömegspektrometriás szerkezeti jellemzése mindig a fragmentációs adatok alapján történik. A feladat lehet egy adott fehérje aminosav-sorrendjének igazolása, egy mutáció vagy egy kovalens módosítás detektálása, vagy éppen egy ismeretlen fehérje azonosítása, esetleg *de novo* szekvenálása. Kutatásaim során egyszerre kellett fehérjéket azonosítanom és azok kovalens módosításait jellemeznem. Így most először ismertetem, hogyan történik egy elegendően jelenlévő fehérjék azonosítása a tömegspektrometria segítségével. Általában az analízis első lépése, hogy a jelenlévő fehérjéket specifikus proteázzal (pl.: tripszin) megemésztjük és a keletkezett peptidkeveréket adatfüggő LC-MS/MS technikával vizsgáljuk. Ha bizonyos poszttranszlációs módosítások jellemzése a cél, akkor a keletkezett peptidkeverékből a jellemezni kívánt PTM-re specifikus módszerrel dúsítást kell végezni (1.4 fejezet). Az adatbázis-kereséssel történő peptid- majd fehérje-azonosítás automatizált folyamat [Kinter, 2000], a feladatot speciális szoftverek hajtják végre a felhasználó által betáplált lekeresési paraméterek alapján. Ezek között szerepelnek analitikai/technikai adatok, pl. a használt enzim specificitása, a tömegmérés pontossága, az aktiválás típusa, illetve a mintával kapcsolatos információ, pl. milyen faj(ok)ból izoláltuk a fehérjét, milyen szennyezések, módosítások fordulhatnak elő. A peptidekről készült MS/MS spektrumok azonosítása adatbázis-kereséssel a folyamat központi eleme. Az adatbázis-keresés során egy fehérje-adatbázis (Swissprot, TrEMBL) *in silico* emésztésével “peptideket állítunk elő”. A használt proteáz specificitása és megbízhatósága jelentősen befolyásolja a peptidlista hosszát. Első lépésben a szoftver a mért és az elméleti prekursor ion tömegek összehasonlításával, a tömegmérés pontosságát figyelembe véve, azonosítja a lehetséges jelölteket a generált peptid-adatbázisból. A szűrés után megtartott szekvenciákból a szoftver elméleti MS/MS spektrumokat készít, amit aztán a kísérleti MS/MS adatokhoz hasonlít. Ez a lépés az, ahol egy MS/MS spektrumhoz a szoftver egy adott peptidszekvenciát rendel. Az egymáshoz rendelés megbízhatóságát részben egy szoftver-specifikus pontszám (score), de sokkal inkább egy valószínűség-számításon alapuló mérőszám (pl. E-value) jelzi (ez utóbbit is másképp számítja minden szoftver). Az azonosított MS/MS spektrumokat a szoftver fehérjékké rendezi össze. Sajnos, ez nem egészen egyértelmű, részben, mert ugyanaz a peptid több fehérjéből is származhat, illetve az adatbázisok redundáns szekvenciái miatt. A SwissProt fehérje-adatbázis a legmegbízhatóbb. A fals pozitív arány megbecsülésére az adatbázist szokás ún. decoy szekvenciákkal megduplázni. Ezek az

“álszekvenciák” a valódi fehérjék fordított vagy randomizált aminosav sorrendjével rendelkeznek.

1.4 O-glikopeptidek dúsítása komplex mintákból

A fehérjéket tehát szekvencia alapján peptidekből azonosítjuk. Könnyen belátható, hogy egy komplex fehérje elegy triptikus emésztményében a peptidek zöme nem lesz glikozilálva, még akkor sem, ha a keverék számos glikoproteint tartalmaz, azaz a bennünket érdeklő molekulákat izolálni kell, vagy legalábbis feldúsítani. Miután a szekretált fehérjék jelentős része glikoprotein, a különböző testfolyadékok könnyen hozzáférhető forrásai a glikopeptideknek. Ezért csoportunk először szérumból, majd vizeletből nyert ki glikopeptideket.

A kutatás során intakt glikopeptideket akartunk tanulmányozni, így kémiai átalakítást nem, csak kromatográfiás elválasztást alkalmaztunk. A leggyakrabban alkalmazott módszerrel, hidrofil interakciós folyadékkromatográfia (hydrophilic interaction liquid chromatography, HILIC), jól dúsíthatók olyan glikopeptidek, amelyek magas glikán/peptid aránnyal rendelkeznek. Ez főképp N-glikopeptidekre és a többszörösen módosított O-glikopeptidekre igaz [Hägglund, 2007].

Glikopeptidek dúsíthatók lektin affinitás kromatográfiával is. Az *Artocarpus integrifoliából* származó Jacalin lektin a Gal(β 1,3)GalNAc motívumra specifikus, ahol a GalNAc 6-os szénatomján lévő hidroxil-csoport nem lehet szubsztituált. Számos O-glikozilációval foglalkozó tanulmány alkalmazott Jacalint magzati borjúsérum [Darula, 2009; Darula, 2012], humán plazma [Bai, 2015] és szérum [Darula, 2016] O-glikozilációjának jellemzésére. Ez a lektin jó választás olyan mintákból való dúsításra, amikben a mucin típusú O-glikoziláció core 1-es szerkezete a domináns. Ilyen például a humán plazma és szérum [Yabu, 2014; Yamada, 2010]. Viszont az emberi plazmában előforduló O-glikánok ~20%-a diszialilált mucin típusú core 1-es szerkezet [Yabu, 2014], amelyben a második NeuAc a GalNAc 6-os szénatomján lévő hidroxil-csoportot módosítja, azaz nem dúsítható a fent említett lektinnel. A szialsavak enzimatikusan eltávolíthatóak, de akkor értékes glikán szerkezeti információt veszítünk. Hasonlóképpen kiesnek a mucin core 2-es szerkezetek (lásd 2. ábra). A földimogyoró agglutinin szintén köti a Gal(β 1,3)GalNAc motívumot, de a szialsavak gátolják a kölcsönhatást a lektin és a glikán között [Reisner, 1976]. Ezt a lektint és a VVA lektint (ld. alább) alkalmazták plazma, vérlemezkék és endoteliális sejtek O-glikozilációjának jellemzésére, a szialsavak eltávolítása után [King, 2017]. A *Vicia villosa* (VVA) és a szójabab agglutinin a GalNAc monoszacharidra specifikus, így ezekkel a lektinekkel jól dúsíthatók olyan mintákból származó O-glikoproteinek vagy O-glikopeptidek, ahol a mucin típusú O-glikoziláció elongációja gátolt [Hägglund, 2007]. A mucin

típusú O-glikoziláció felderítésében rendkívül fontos eredményeket értek el egy biológiailag elongáció képtelen sejtvonallal, az ún. SimpleCell módszerrel, ahol a fent említett lektinek igen hasznosnak bizonyultak [Steentoft, 2011; Vakhrushev, 2013]. Létezik olyan széles spektrumú lektin is, amely jól használható általános O-glikopeptid dúsításra, és ez a búzacsíra agglutinin (wheat germ agglutinin, WGA). Ezen lektin az irodalom alapján csak GlcNAc és NeuAc specifikus [Nagata, 1974; Ganguly, 1984], de a tapasztalatok szerint sokkal változatosabb szerkezeteket képes feldúsítani [Chang, 2011, Trinidad 2013]. Ezt a lektint használták O-glikopeptid dúsításra egér szinaptoszómából [Trinidad, 2013], májból [Medzihradzky, 2015 (A); Medzihradzky, 2015 (B)] és humán szérumból is [Darula, 2016; Pap, 2017].

1.5 O-glikopeptidek tandem tömegspektrometriája

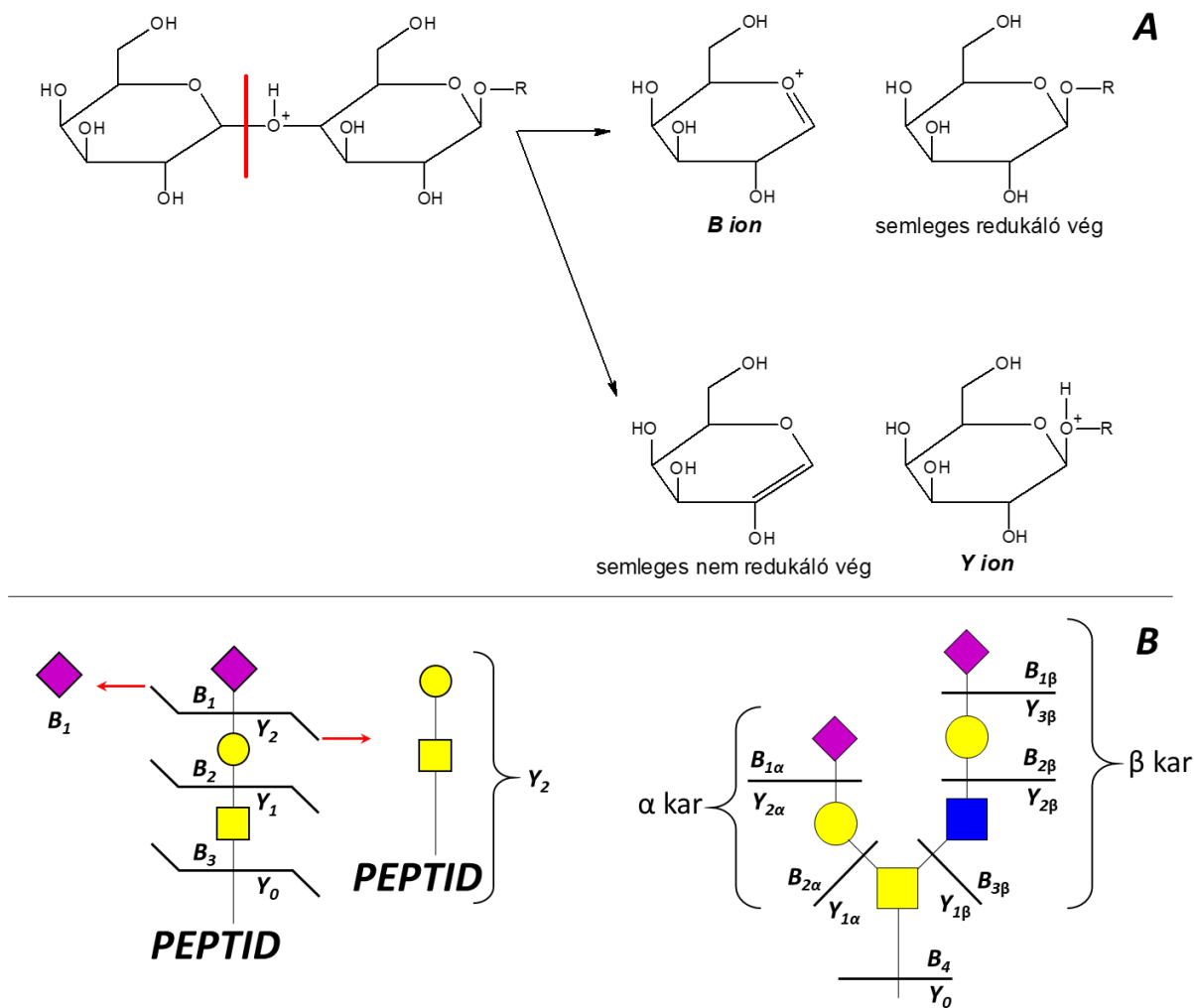
1.5.1 Glikopeptidek fragmentációja ütközéses aktiválás esetén: nevezéktan, fő fragmentation típusok

A glikopeptidek ütközéses aktiválása során kétféle kötés elhasadása lehetséges. Az egyik a peptid kötés, a másik pedig a cukromonomereket összekapcsoló glikozidos kötés. A peptid kötés hasadásakor keletkező fragmenseket a 1.3 fejezetben már tárgyaltam, ebben a fejezetben csak a glikozidos kötéshasadásból származó fragmensek bemutatására fókuszálok. A cukrok pozitív módban történő ütközéses aktivációja során megfigyelhető fragmentációs szabályokat 1988-ban írta le Domon és Costello [Domon, 1988]. A fragmentációs esemény során a glikozidos kötés protonálódik, ami a kötés hasadásához vezet. A folyamat során a glikozidos kötésben részt vevő oxigénatom a cukor redukáló végét tartalmazó fragmenszen marad. A fragmentáció során két fajta fragmentation képződik, amelyet B és Y ionnak nevezünk (8. ábra, A panel). A B ion az ún. cukor oxónium ion. Ez az ion a glikán nem redukáló végét tartalmazza és számozása is innen kezdődik. A komplementer Y ion glikopeptidek esetében a glikán redukáló végéből és a peptidből áll (8. ábra, B panel). Számozása a redukáló vég felől kezdődik.

Glikopeptidek esetében a keletkező Y fragmentationok a prekursor ion töltésállapotától függően több töltésállapotban is megtalálhatóak és a töltéseket maga a peptidrész hordozza, míg a B fragmentationok egyszeresen töltve figyelhetők meg. Elágazó glikánstruktúrák esetében az antennák megkülönböztetésére, balról jobbra haladva, a görög ABC betűit használjuk (8. ábra, B panel).

A B fragmentationok m/z értékét úgy számoljuk ki, hogy az adott fragmensnek megfelelő cukorszerkezet additív tömegéhez egy proton tömegét (1,0073 Da) adjuk. A 8. ábra B paneljén lévő B_1 ion értéke például: $B_1 = 291,0954 + 1,0073 = 292,1027$ lesz. Az Y fragmentationok m/z értékének kiszámításához a peptid MH^+ értékét és az adott Y fragmentationnak megfelelő

cukorszerkezet additív tömegét összeadjuk. Példaként a 8. ábra B paneljén található glikopeptid Y_2 ionja a következőképpen számolandó: $Y_2 = 671,3246$ (peptid MH^+) + $203,0794$ (HexNAc) + $162,0528$ (Hex) = $1036,4568$. A leggyakoribb cukor monomerek monoizotópos additív tömegét az 1. táblázat tartalmazza.



8. ábra **A panel** Cukrok ütközéses aktiválásakor keletkező B és Y fragmensek szerkezete [Domon, 1988]. **B panel** Glikopeptidek esetében az Y ion mindig tartalmazza magát a peptidet. A teljesen deglikozilált Y ion maga a peptid, amelyet Y_0 -val jelölünk. Elágazó cukorszerkezet esetében a görög ABC betűivel jelöljük a különböző karokat és a karok fragmentációjával keletkező B és Y fragmenseket. [Domon, 1988]

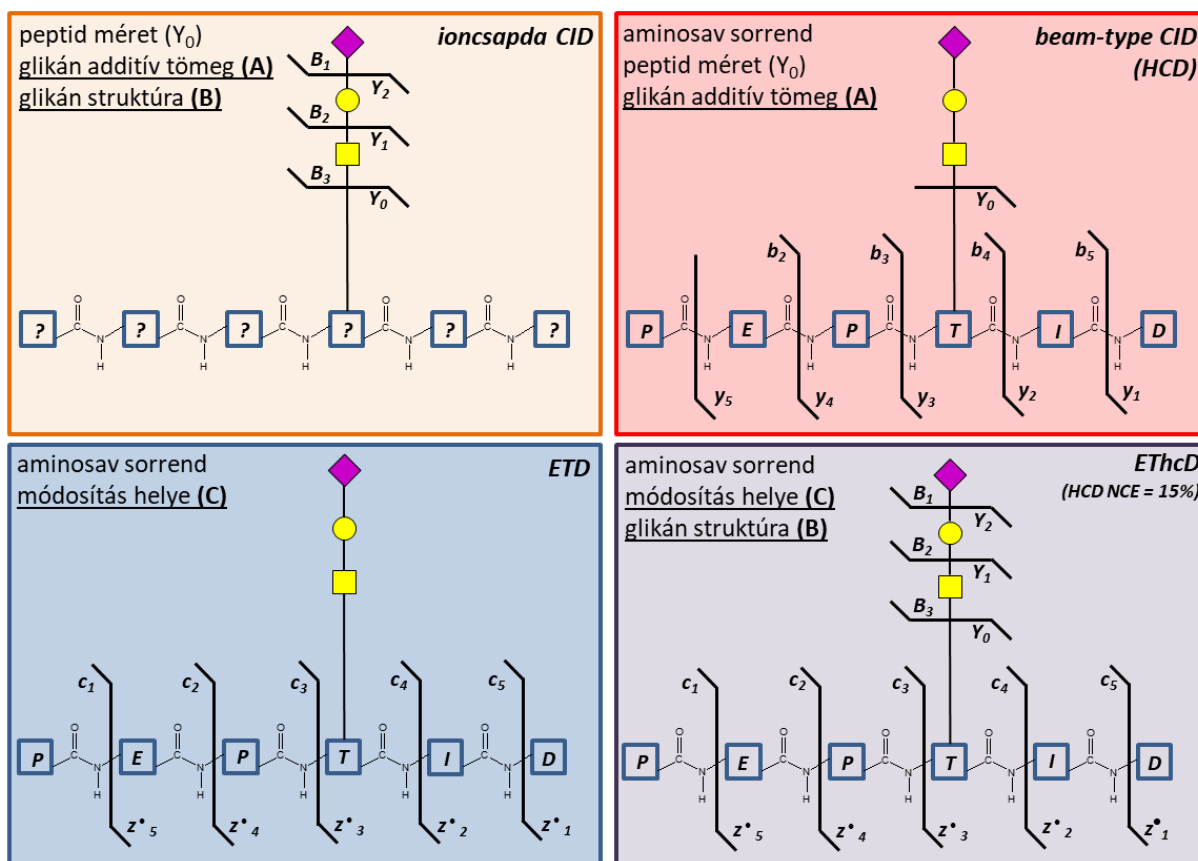
1. táblázat Monoszacharid egységek monoizotópos additív tömege.

Monoszacharid	Monoizotópos additív tömeg [Da]
Hex	162,0528
HexNAc	203,0794
Fuc	146,0579
NeuAc	291,0954

Minden, a dolgozatban bemutatásra kerülő glikopeptid spektrum esetében a glikozidos kötés-hasadásokból származó fragmensionok annotációja a fent bemutatott Domon-Costello nevezéktan alapján történt. A B ionok annotációjára, a Domon-Costello nevezéktan mellett, SNFG szimbólumokat is használtam, amennyiben egyszerűbbé/egyértelműbbé tették a fragmentáció értelmezését.

1.5.2 O-glikopeptidok analízise során alkalmazott fragmentációs technikák

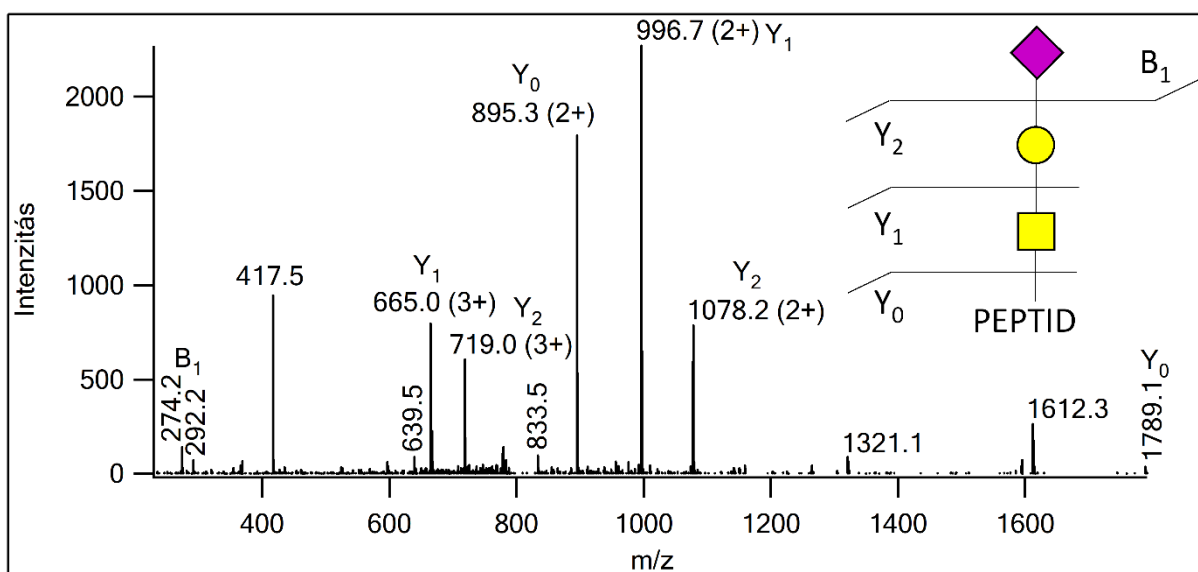
A glikopeptidok két, kémiaiag eltérő és fragmentáció szempontjából is különböző vegyülettípusból épülnek fel. Jellemzésükhöz egyszerre kell a peptidet és a módosító glikán(oka)t azonosítani. A poliszacharidok analízise önmagában is problémás, mert izomer alkotórészeik tömegük alapján nem különböztethetőek meg. Hasonlóképpen nincs egyszerű tömegspektrometriás módszer a glikozidos kötés pontos pozíciójának és térszerkezetének jellemzésére. Tehát szerencsés esetben a következő információk nyerhetők a glikopeptidekről: peptidszekvencia, **(A)** monoszacharid összetétel, **(B)** a glikánt alkotó monoszacharid egységek egymáshoz viszonyított helyzete, **(C)** a módosítás helye. A tömegspektrometriás analízis során több MS/MS kísérlet szükséges különféle fragmentációs technikák alkalmazásával, hogy a két eltérő molekularészletről a lehető legtöbb szerkezeti információt nyerjük (9. ábra).



9. ábra O-glikopeptidok MS/MS analízise során alkalmazott fragmentációs technikák, és információ-tartalmuk.

1.5.2.1 Ütközéses aktiválás/rezonancia aktiválás ioncsapdában (ion trap CID, resonance CID)

Az ütközéses aktiválás a leggyakrabban alkalmazott MS/MS fragmentációs technika. Egy ioncsapdával megoldható a bennünket érdeklő prekursor ion kiválasztása (a nem kívánatos tömegek eltávolításával), és az aktiválás is végrehajtható magában a csapdában. Az ioncsapdában történő ütközéses aktiválás egy ún. radiofrekvenciás gerjesztésen alapuló fragmentációs technika (rezonancia-aktiválás). Mivel csak egy szűk m/z tartományban működik, a képződött fragmensek már nem hasadnak tovább, azaz többnyire egyszeres kötéshasadások történnek. Az ütközéses aktiválás – az ioncsapda CID is – a molekula leggyengébb kötéseit veszi célba. Glikopeptidek esetében ezek a glikozidos kötések, amelyek hasadásából keletkező fragmensek a legintenzívebbek egy ioncsapda CID-val felvett MS/MS spektrumban (10. ábra). A fragmentáció során a töltés leggyakrabban a glikán redukáló végét tartalmazó fragmenszen marad (10. ábra).



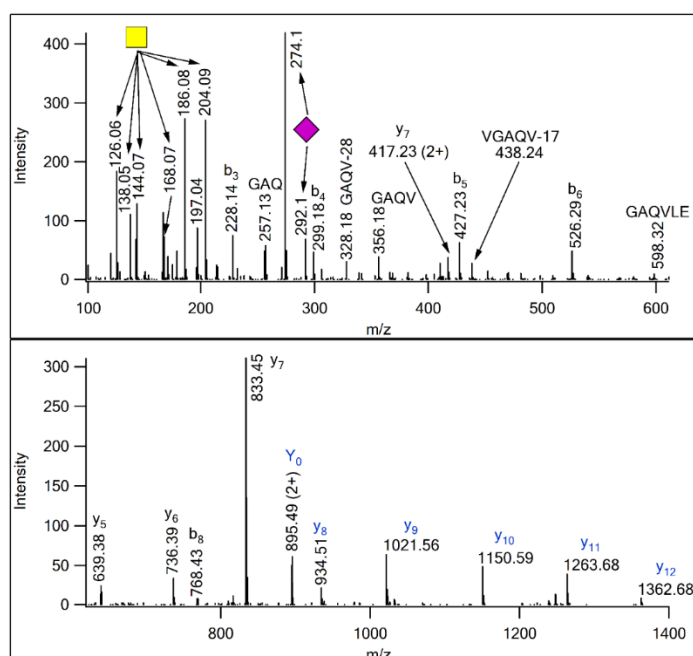
10. ábra m/z 815,731 (3+) prekuzorról felvett ioncsapda CID spektrum. Ez a spektrum egy valós glikopeptid CID spektruma, de az aminosav szekvenciát nem lehet ezekből az adatokból megállapítani.

Az 10. ábrán jól látható, hogy a glikozidos kötéshasadásból származó Y ionok a legintenzívebbek a spektrumban és a prekursor töltésállapotával megegyező vagy kisebb töltésállapottal is előfordulnak. Az Y_0 a módosítatlan peptid, amely egy átrendeződési reakcióval képződik a glikán gázfázisban történő lehasadása során. A reakció után semmi nyoma nem marad annak, melyik aminosav volt módosítva, így a CID spektrumokból nem állapítható meg az O-glikoziláció helye. A 10. ábrán felfedezhető a terminális NeuAc oxónium ionja ($B_1 = m/z$ 292,2) és annak vízvesztéssel keletkezett párja (m/z 274,2) is detektálható. Ez az ion éppen csak belül van a detektálási határon, miután az ioncsapda nem képes visszatartani azokat a CID aktiválás során képződött fragmenseket, amelyek a prekursor ion tömegének kb. 1/3-a alatti

tömeget képviselnek (ez a 10. ábrán látható spektrumban kb. m/z 271). A peptidszekvenciát ezekből az adatokból nem lehet meghatározni, habár az ábrán az m/z 417,5 ion valószínűleg egy szekvencia ion lehet. Összegzőként elmondható, hogy egy O-glikopeptid ioncsapda CID spektrumából megállapítható a módosító O-glikán összetétele (az 10. ábrán bemutatott glikopeptid esetében HexNAcHexNeuAc). Az O-glikán monoszacharid egységeinek kapcsolódási sorrendjét is meghatározhatjuk az egyszeres kötéshasadással keletkezett Y ionok segítségével [Wu, 2014]. Ugyanezen ionok jelzik a szerkezet linearitását/nemlinearitását. Továbbá meghatározható a gázfázisban deglikozilált peptid tömege (Y_0).

1.5.2.2 Ütközéses aktiválás ütközési cellában (beam-type CID, HCD)

Az ütközési cellában végrehajtott ütközéses aktiválás során általában a peptid szekvenciája azonosítható a többszörös ütközésekből adódó (így nem csak glikánfragmentációt okozó) teljesebb peptidgerinc-fragmentáció miatt.



11. ábra m/z 815,731 (3+) prekuzorról felvett beam-type CID (HCD) spektrum. A spektrumból azonosítható a peptid szekvenciája AVGAQVLESTPPPHVMR. A gázfázisban deglikozilált ionok kék színnel jelöltek.

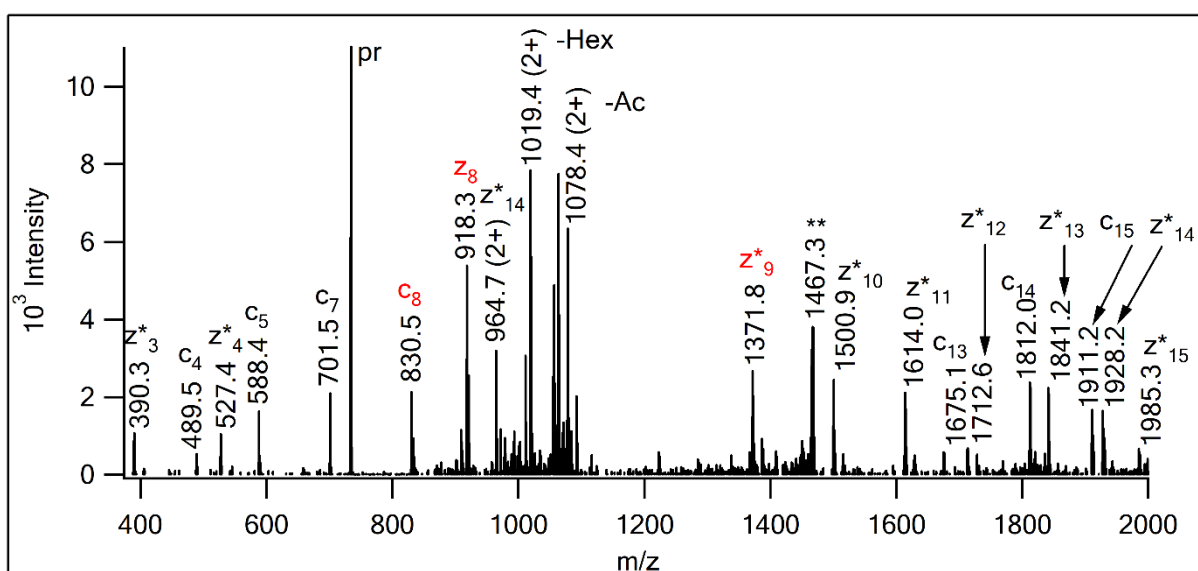
A 11. ábrán bemutatott O-glikopeptid HCD MS/MS spektruma több peptidfragmenst (b és y ionok) is tartalmaz, amelyek segítségével megbízhatóan azonosítható a peptid szekvenciája: AVGAQVLESTPPPHVMR. Természetesen ehhez a meghatározáshoz meg kellett adnunk a lehetséges cukorszerkezetek listáját is. Ütközéses aktiválásnál a glikán lehasadása miatt a glikántömegeket „semleges vesztésként” definiáljuk, azaz a peptidfragmenseket módosítatlanul azonosítjuk. A többszörös ütközések lehetővé tették a peptid szekvenciájának az azonosítását,

viszont a módosító glikánról már kevesebb információ áll rendelkezésre. Az m/z 274 és 292-es ionok szialsav jelenlétére utalnak, míg az m/z 204-es ion HexNAc jelenlétét jelzi. Ez az ion önmagában egyaránt jelenthet GlcNAc-ot vagy GalNAc-ot is. Azonban Halim és mtsai megfigyelték, hogy HCD aktiválás esetén a GalNAc-ból, illetve a GlcNAc-ból keletkező fragmensionok intenzitás mintázata eltérő. GalNAc esetében az m/z 126 és 144 ionok összintenzitása az m/z 138 és 168 összintenzitásához képest magasabb. GlcNAc esetében ezen ionok egymáshoz viszonyított intenzitása fordított. A 11. ábrán az m/z 126 és 144 ionok intenzívebbek az m/z 138 és 168 ionokhoz képest, így GalNAc a helyes válasz [Halim, 2014]. Az adott cukorkompozícióra (HexNAcHexNeuAc) a GalNAcGalNeuAc szerkezet biokémiai ismereteink alapján már valószínűsíthető, de pl. annak eldöntésében is segíthet a fragmensek intenzitásarányának vizsgálata, hogy két core 1 triszacharid vagy egy core 2 hexaszacharid módosít egy vizsgált peptidet. Komplex elegyekben azonban a prekursor ion interferencia miatt a mért arányok nem mindig megbízhatóak. Prekursor ion interferencia minden elegyre jellemző. MS/MS analízisre minden esetben egy szűk tömegtartományt választunk ki, de ez a „prekursorablak” még monoizotópos prekursor választásnál is tartalmazhat több különböző komponens [Falick, 1990]. Ezek a komponensek egyszerre aktiválódnak és fragmentálódhatnak, és az így képződő ionokat detektáljuk az MS/MS spektrumban. Komplex elegyekben a frakcionálás ellenére is előfordul, hogy egyszerre több átfedő ioncsoportot, különböző mennyiségben jelenlévő összetevőt fragmentálunk. Azaz az összes izolált prekursor valamilyen formában hozzájárulhat a kialakult fragmentációs spektrumhoz. Komplex glikopeptid elegyek esetében az egyidejűleg izolált glikopeptid prekursorokból származó oxónium ionok összessége alakítja ki a fragmension intenzitás eloszlási mintázatot, így ebben az esetben a GlcNAc/GalNAc kérdés eldöntése nehezen vagy egyáltalán nem lehetséges. A módosító glikán méretére vonatkozólag közvetett bizonyítékot a deglikozilált peptidion szolgáltat (Y_0 , m/z 895,49 (2+), 11. ábra), amely alapján HexNAcHexNeuAc összetételű O-glikán valószínűsíthető. Az ioncsapda CID és a HCD fragmentációs technikák kombinációjával megbízhatóbb eredmények érhetőek el, mivel a két MS/MS aktivációs módszer eltérő információt szolgáltat ugyanazon molekula két különböző részéről. A 10. és a 11. ábrán bemutatott peptid példájánál maradva ioncsapda CID-val kideríthető a cukor mérete, a szerkezet linearitása és összetétele: HexNAcHexNeuAc, illetve a peptid tömege. HCD-vel a glikán összetétele tovább pontosítható: GalNAcHexNeuAc, továbbá a peptid szekvenciája is meghatározhatóvá válik. A glikán pontos helyzete a peptiden belül sajnos továbbra sem állapítható meg, mert ütközéses aktiválásnál, legyen az CID vagy HCD, a gázfázisú deglikoziláció a preferált fragmentációs esemény. Néha intenzívebb peptid-fragmensionok (tipikusan azok, amelyek a prolin előtti peptidkötés hasadásából keletkeznek) hordozhatják

részben vagy egészben a cukorszerkezetet, lehetővé téve a glikoziláció helyének meghatározását/behatárolását.

1.5.2.3 Gyökös fragmentáció: elektron transzfer disszociáció (ETD)

Az ETD aktivációs technika során a fragmentálni kívánt többszörösen töltött prekursor ion reagens anion (fluorantén) jelenlétében elektrontranszferrel instabil gyökionná alakul, majd a peptidkötés nitrogénje és az alfa szénatom közötti kötés elhasadásával c (N-terminális) vagy z[•] (C-terminális) peptid-fragmensek keletkeznek. Az ETD hatásfokát jelentősen befolyásolja a prekursor ion töltése és töltéssűrűsége: kétszeres töltésű ionok és m/z > ~850 esetén kevésbé működik jól a technika [Good, 2007].



12. ábra m/z 733,692 (3+) prekuzorról lineáris ioncsapdában felvett ETD spektrum. Az adatokból az AQGSQVLES(HexNAcHex)TPPPHVMR peptidszekvencia azonosítható. A módosítás helyét bizonyító fragmensek (z₈, z₉^{*} és c₈) az ábrán piros színnel szerepelnek. A „pr” a prekuzor iont jelöli. A „**” egy 2+ töltésű, koeluálódó, töltésvesztett iont jelöl. A csillaggal jelzett z^{*} fragmensek a hidrogénmigráció termékei, azaz z+1 fragmensek. A spektrum m/z 50-től van felvéve, de helytakarékoság miatt csak az első detektált fragmenstől van ábrázolva, ami m/z 390,3.

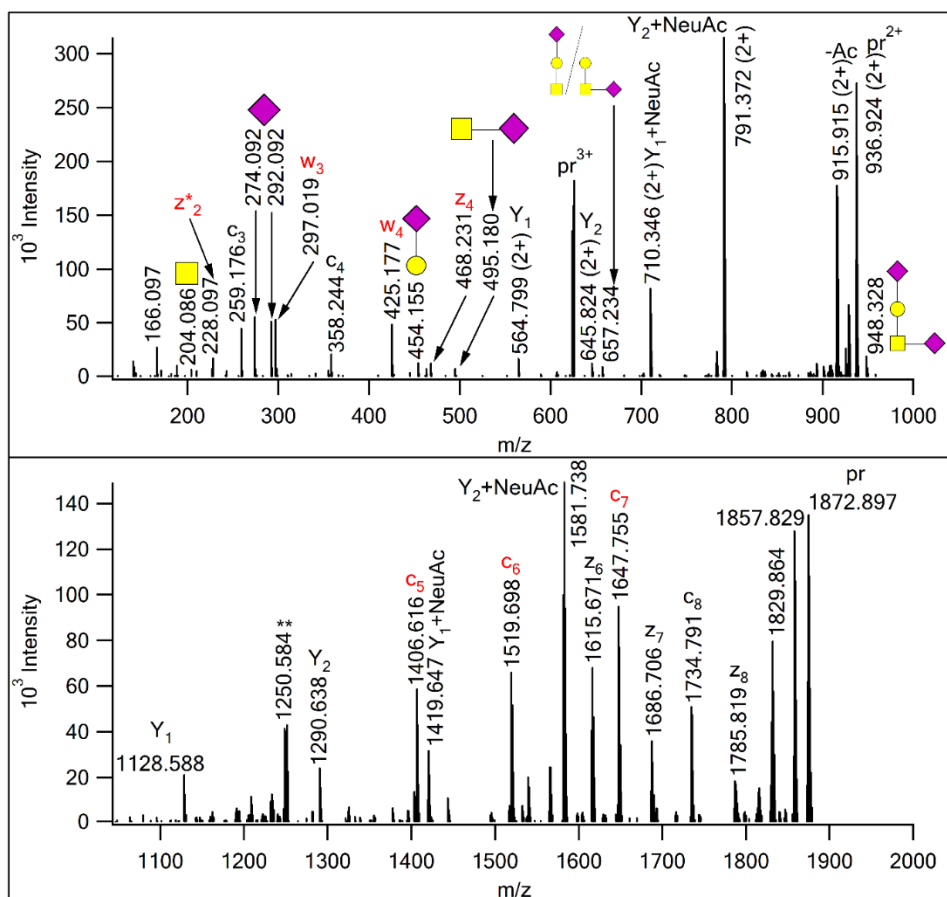
Hidrogén migrációval c-1[•] és z+1 fragmensek is keletkezhetnek [Bakken, 2004]. A technika nagy előnye – amellet, hogy egy újabb módszerrel erősíthetjük meg a peptid szekvenciáját – az, hogy az aminosavak oldalláncai és az itt található módosítások nagyrészt érintetlenek maradnak, így a labilis poszttranszlációs módosítások (amilyen az O-glikoziláció is) helye a peptiden belül meghatározhatóvá válik. Ezért mostanában a legtöbb O-glikozilációval foglalkozó kutatás ETD adatgyűjtést is alkalmaz [Trinidad, 2013; Darula, 2012; King, 2017; Medzihradszky, 2015 (A); Medzihradszky, 2015 (B); Darula, 2016; Pap, 2017; Darula, 2009; Halim, 2012; Halim, 2013; Hoffmann, 2016]. A 12. ábrán egy ETD spektrum látható a már korábban bemutatott peptid

humán homológjáról. Mivel az ETD aktiváció során az aminosav oldalláncok érintetlenek maradnak, a módosító glikánokról valamilyen előzetes információval kell rendelkezünk (azaz a spektrumból *per se* nem azonosítható a módosító glikán), és természetesen a lekeresés paraméterei között definiálnunk kell, milyen módosítások lehetnek jelen. Ebben az esetben is az adatbázis-keresést érdemes csupán a leggyakoribb glikánstruktúrákkal kezdeni, illetve, ha specifikusan csak bizonyos szerkezetekre történt a dúsítás, akkor csak a dúsítási eljárással „kihalászható” szerkezeteket érdemes az adatbázis-keresés első körében megengedni. A 12. ábra ETD spektrumából azonosított peptid humán szérum triptikus emésztményéből származik, amit Jacalin oszlopon dúsítottunk, és neuraminidázzal kezeltünk, emiatt csak két glikán szerkezetet vettünk figyelembe: HexNAc (GalNAc), és HexHexNAc (GalGalNAc) módosításokat. A keresés során a 12. ábrán látható spektrum a humán ITIH2 fehérje AQGSQVLES(HexHexNAc)TPPPHVMR peptidjeként lett megbízhatóan azonosítva. A Ser-4 kizárható a potenciális módosítási helyek közül a c₄-c₈ módosítatlan N-terminális fragmension-sorozat alapján. Mivel a további két glikozilációs hely egymással szomszédos, így egyértelmű helyazonosítást a c₉ és/vagy z₈ fragmensionok detektálása jelentene. Az ETD spektrum alapján a z₈ fragmens módosítatlanul lett detektálva, és a Ser-9 módosítását a glikozilációt hordozó z*₉ fragmens bizonyítja. A C terminális fragmenseket z+1 formában detektáltuk (a 12. ábrán z*-ként jelölve) a kötéshasadás során fellépő hidrogén-migráció miatt [Bakken, 2004]. Továbbá enyhe mértékű glikán-fragmentáció is megfigyelhető (a 12. ábrán az m/z 1019,4 (2+) ion a prekursor ion Hex vesztesét mutatja), amelyet már korábban is leírtak ennél az aktivációs technikánál [Darula, 2009; Darula, 2010].

1.5.2.4 A gyökös és ütközéses aktiválás kombinációja: EThcD

Az EThcD az egyik legújabb MS/MS technika. Első lépésben az ETD aktiváció történik meg az ionsapdában, majd az így keletkezett heterogén ionpopulációt az ütközési cellában (alacsony normalizált ütközési energiájú) HCD technikával fragmentáljuk tovább [Frese, 2012]. A 13. ábrán egy EThcD spektrum látható az m/z 624,950 (3+) prekuzorról, amelyből a humán Protein YIPF3 fehérje AVAVT(HexNAcHexNeuAc₂)LQSH peptidjét azonosítottuk. Az új fragmentációs technika egyesíti az ütközéses és a gyökös aktiváció előnyeit. Az ETD-nek köszönhetően a módosított peptidszekvencia megállapítható a fragmentáció során keletkezett c és z* ionok segítségével. A z₄ és a c₅ fragmensek egyértelműen azonosítják a glikoziláció helyét, azaz a Thr-t. Az ETD lépés során keletkezett C-terminális gyökionok (z*) oldalláncában C-C kötéshasadás történhet (w ionok, nevezéktan: [Biemann, 1990]) a HCD aktiváció hatására. A jelenséget, ami jól használható a Leu/Ile izomer aminosavak megkülönböztetésére, eredetileg nagy energiájú CID kísérletekre írták le [Johnson, 1987], és már korábban megfigyelték EThcD

spektrumokban is [Zhokhov, 2017]. A peptidszekvencia és a módosítás helyének meghatározása mellett a spektrum a glikánra vonatkozó strukturális információt is hordoz B és Y ionok formájában.



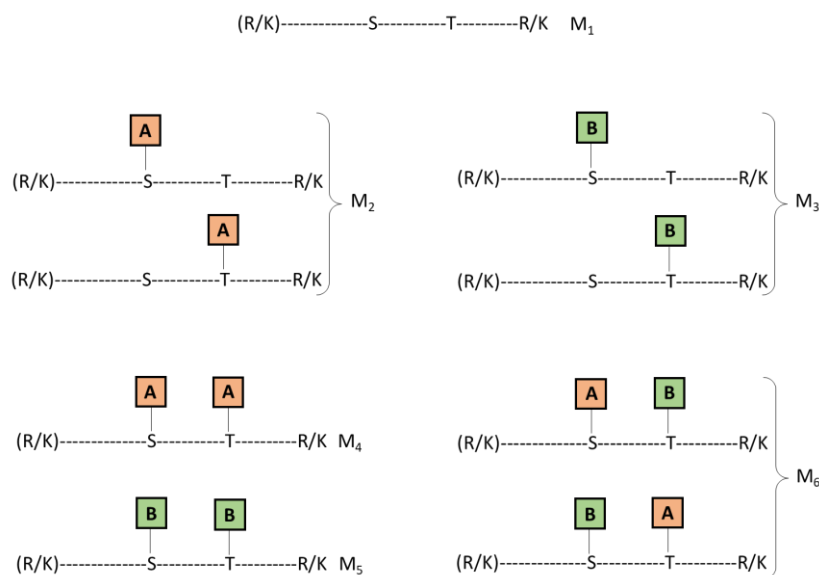
13. ábra m/z 624,950 (3+) prekuzorról készült ETHcD spektrum, amiből a Protein YIPF3 fehérje AVAVT(HexNAcHexNeuAc₂)LQSH peptidje azonosítható. A glikoziláció helye a Thr-5, amelyet az ábrán a pirossal jelölt fragmensionok támasztanak alá. A módosítás egy diszialo mucin 1-es alapvázat tartalmazó tetraszacharid O-glikán.

Fontos megjegyezni, hogy a tandem tömegspektrometriás analízis során a cukromonerek pontos típusa nem állapítható meg a szerkezeti izoméria miatt. Ugyanígy a monomerek kapcsolódásának pontos pozíciójáról és sztereokémiájáról sem szerezhető információ ezekből az adatokból. Így amikor a dolgozatban egy-egy adott cukromonomer pontos típusa vagy kötéspozíciója szerepel, az csakis más módszerekkel kiderített O-glikán szerkezetek és/vagy a bioszintetikus útvonalak ismeretében lehetséges. Vizsgálataink során glikopeptidenként kétféle spektrumot rögzítettünk: HCD és ETHcD spektrumokat. Az ETHcD adatok gyűjtése több időt igényel, mint egy egyszerű ütközéses aktiválás, ami a módosítatlan peptid azonosításához elegendő. Viszont, ahogy ismertettem korábban, a glikopeptidok jellemzéséhez nélkülözhetetlenek a gyökös fragmentációs folyamatok. Ezt a problémát úgy oldottuk meg, hogy valahányszor egy HCD spektrum tartalmazott egy glikopeptidre (pontosabban HexNAc-re)

jellemző fragmensiont, akkor a készülék automatikusan beindította az EThcD adatgyűjtést is (3.4 fejezet) [Zhao, 2011].

1.6 O-glikopeptidok tömegspektrometriás adatainak elemzése

Az O-glikopeptidok tömegspektrometriás mérésrel előállított adatainak kiértékelése a nem glikozilált peptidekhez hasonlóan adatbáziskereső szoftverekkel történik. A prekursoroknak megfelelő glikopeptid-adatbázis a módosítatlan szekvenciákból generált *in silico* listának a sokszorososa. Ennek oka a következő példán szemléltethető. A és B glikán módosíthatja a peptidjeinket, és egy peptid egyszerre maximum két módosítást viselhet. Miután konszenzus szekvencia nem létezik, makro- és mikroheterogenitás viszont igen, egy peptid, amiben csupán 1-1 Ser és Thr fordul elő, 9 különböző szerkezettel, és 6 különböző tömeggel szerepel az *in silico* peptid-listán (14. ábra).



14. ábra Az O-glikopeptidok adatbáziskeresése során a számítógép által generált peptidadatbázis sokszorososa a módosítatlan peptidek keresése során előállítható képest.

A kihagyott vagy nemspecifikus hasítások és egyéb változó módosítások (pl. Met oxidáció) tovább növeli az *in silico* glikopeptid adatbázis méretét. A módosított peptid azonosítása most is a szekvencia ionokból történik. Az adatbázis exponenciális növekedését valamilyen szinten kompenzálni lehet a HCD adatok értelmezésénél úgy, hogy a cukor-módosítást semleges vesztésként értelmezzük, és eleve csupán módosítatlan fragmenseket keresünk. A lekereső programok különböző variációkkal ezt a megközelítést alkalmazzák HCD adatokra. Tehát a módosított peptid személyazonossága és a módosító cukrok összessége meghatározható. Az ET(hc)D adatok értelmezésénél viszont már nincs ilyen könnyítésre lehetőség, a helymeghatározást és valamennyi pozíciós izomert is figyelembe kell venni, hiszen ebben az

esetben egy konkrét izomert kell egy spektrumhoz rendelni. Izomerként azokat a glikoformákat definiáljuk, amelyekben a módosított szekvencia és a cukor-összetétel azonos. Ezek a glikopeptidek lehetnek pozíciós izomerek, amikor is csak a módosítások helye különböző a glikoformákban. De lehetnek olyan esetek is, amikor több glikán kombinált tömege azonos, és ismertetek majd olyan eseteket is, ahol egy adott pozícióban azonos összetételű, de különböző szerkezetű oligoszacharidok módosítják a peptidet. Az ilyen szerkezetek megkülönböztetéséhez a peptid- és cukor-fragmenseket egyaránt figyelembe kellene venni. Miután az ETD fragmentáció gyakran hiányos szekvenciaion sorozatot ad, gyakran a HCD alapú azonosításokhoz hasonlóan, csak a módosított szekvenciát és a cukor additív tömegét tudjuk meghatározni. Más esetekben viszont az ET(hc)D önmagában nem vezet sikeres azonosításhoz, de ha bármilyen egyéb forrásból ismerjük a peptid szekvenciáját, az adatok lehetővé tehetik a módosító glikán(ok) és a glikozilációs helyek azonosítását (4.4 fejezet).

2. CÉLKITŰZÉS

Jelen dolgozat a humán vizeletben található fehérjék mucin típusú O-glikozilációjának tömegspektrometriás analízisével foglalkozó kutatás eredményeit mutatja be. A fehérjék glikozilációjának pozíció-specifikus, mélyreható/átfogó analízise egyre nagyobb jelentőségű az alapvető biológiai folyamatok megértése és gyógyászati célból termeltetett fehérjék minőség-ellenőrzése szempontjából egyaránt. A poszttranszlációs módosítások, többek között a glikoziláció jellemzésére ma a leggyakrabban alkalmazott analitikai technika a tandem tömegspektrometria.

A mucin típusú O-glikopeptidek analízise kihívásokkal teli feladatnak számít. Ezek a kihívások a tömegspektrometriás analízis és az adatkiértékelés esetében is jelen vannak. A dolgozat alapjául szolgáló kutatások során az O-glikopeptidekről gyűjtött tömegspektrometriás adatok kiértékelésére fókuszáltam. Az adatkiértékelésnél a cél az volt, hogy jellemezzem e molekulák fragmentációs sajátosságait egy újfajta MS/MS aktivációs technika, az EThcD esetében és ezt összehasonlítsam a régebbi fragmentációs módszerekkel (ioncsapda CID, HCD, ETD).

A dolgozat alapját képező kutatások kettős célt szolgálnak. Az egyik cél, ahogy azt fentebb már említettem, technikai jellegű az O-glikopeptidek fragmentációs karakterisztikájának tanulmányozásával a középpontban. A másik cél jellegét tekintve biológiai, és a humán vizeletben található fehérjék O-glikozilációjának megismerésére irányul. A két cél nem választható el egymástól. A kutatás analitikai részében megszerzett ismeretek az O-glikopeptidek tandem tömegspektrometriás viselkedéséről alapjaiban meghatározza a felvett adatokból kinyerhető, biológiailag releváns információ megbízhatóságát.

3. ANYAGOK ÉS MÓDSZEREK

3.1 Vizeletminták

A mintavételhez szükséges etikai engedély száma: 1011/16. Minden mintavételt a donorok tájékoztatása és beleegyezése előzött meg. A dolgozatban bemutatásra kerülő eredmények összesen 13 páciensztől származó vizeletminták külön LC-MS/MS analíziséből származnak (2. táblázat). Az Urológiai Klinikáról származó mintákat a mindenkor ügyeletes nővér adta át számomra.

2. táblázat Összefoglaló táblázat a feldolgozott vizeletmintákról.

Sorszám	Nem	Mintavétel helyszíne	Kategória	Analizált WGA frakciók	Eredmények bemutatása
1.	Nő	SZBK – Proteomikai laboratórium	Kontroll (Egészséges)	FTEnd Shoulder GlcNAc	4.1 fejezet
2.	Férfi				4.2 fejezet
3.					4.3 fejezet
4.	Férfi	SZTE – Szent Györgyi Albert Klinikai Központ Kórháza – II. épület Urológiai Klinika	Kontroll (Egészséges)	FTEnd Shoulder GlcNAc	4.4 fejezet
5.			Felszínes hólyagdaganat		
6.					
7.			Előrehaladott hólyagdaganat		
8.					
9.			Hólyaggyulladás		
10.					
11.					
12.					
13.					

3.2 Mintaelőkészítés

3.2.1 Emésztés cellulózmembrán szűrőn (FASP)

A mintaelőkészítéshez és a tripszines emésztéshez kutatócsoportunk által módosított, Wiśniewski és mtsai. által közölt FASP (filter-aided sample preparation) protokollt alkalmaztam [Wiśniewski, 2009; Pap, 2017]. A minták előcentrifugálása (Hettich Universal 16 centrifuga) 5000 g-n, 4 °C-on, 20 percig történt. A felülúszót óvatos pipettázással, az üledék felkavarása nélkül egy 50 ml-es centrifuga csőbe (Corning) helyeztem, az üledéket kidobtam. Az előcentrifugált mintákat ultraszűrővel koncentráltam 5000 g-n, 4 °C-on 20 percig, amelyhez egy 50 ml-es centrifugacsőbe helyezhető cellulózmembrán szűrőegységet (Amicon Ultra 15 ml, MWCO: 10 kDa) használtam. A betöményített mintát (~250 µl) 6 M guanidin-hidroklorid (Merck)/25 mM ABC (Merck) és a becsült fehérje mennyiségéhez (~100 µg/ml, koncentrálatlan vizelet) képest százszoros moláris feleslegben lévő DTT (Merck) jelenlétében denaturáltam és

redukáltam 30 percen át 56 °C-on, 1000 rpm-en egy rázatógépben (Grant-Bio PHMT thermoshaker). Az SH-csoportok alkilezése 30 percig, sötétben, szobahőmérsékleten történt: a reakcióelegyhez a DTT mennyiségéhez képest 1,1 ekvivalens mennyiségű IAM-ot (Merck) adtam. Ezután a felesleges alkilezőszert DTT-vel semlegesítettem. Ezt követően a mintákat három lépésben mostam 4 ml 6 M guanidin-hidroklorid/25 mM ABC oldattal, majd három lépésben 4 ml 25 mM ABC oldattal. A mosóoldattal minden esetben 1 percig rázattam a mintákat majd centrifugálással (5000 g, 4 °C, 10 perc) távolítottam el a mosófolyadékot. Végül a szűrőegységekben lévő mintákhoz a becsült fehérje mennyiségéhez képest 1% (w/w) tripszint (oldalláncvédett, sertés hasnyálmirigyből, Thermo) adtam, majd 37 °C-on, enyhe rázatás mellett inkubáltam a mintákat 12 órán át. Az emésztési idő leteltével az emésztési elegyet átpipettáztam egy 0,5 ml térfogatú cellulózmembrán szűrőegységbe (Amicon Ultra 0,5 ml, MWCO: 10 kDa) és a peptideket centrifugálással (B. Braun Sigma 2k15 centrifuga, 14000 g, 20 perc, szobahőmérséklet) összegyűjtöttem, majd vákuumcentrifugával (Jouan RC1010) beszárítottam.

3.2.2 Lektin affinitás kromatográfia

A lektin affinitás kromatográfiát a kutatócsoportunk által korábban már kidolgozott módszerrel végeztem [Darula, 2016]. A FASP protokollal készített emésztményeket a beszárítás után 100µl 150mM ABC pufferben (pH=7,5) oldottam fel. A dúsítást egy 2mm×250mm (belső átmérő×hossz) rozsdamentes acél oszlopba töltött, aldehid funkcióval ellátott, porózus szilikagyöngyre (POROS 20 AL, Thermo) immobilizált WGA lektinnel (Vector Laboratories) végeztem.

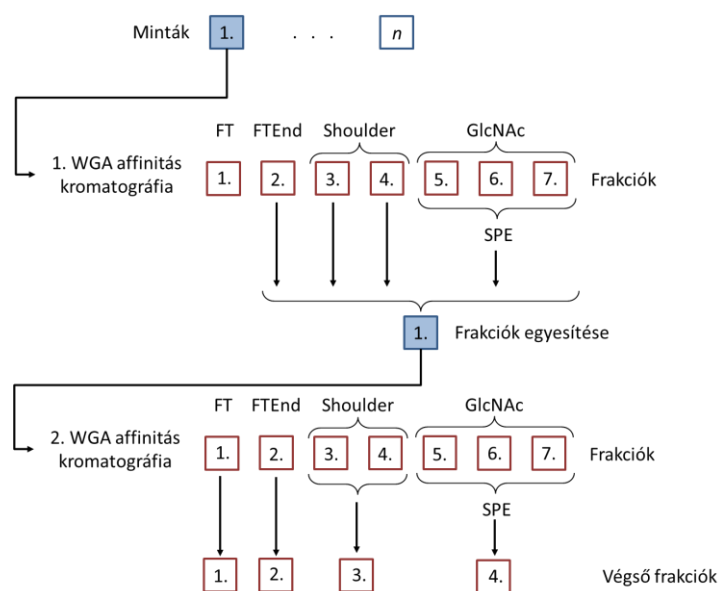
A frakcionáláshoz a teljes mintatérfogatot (100µl) felhasználtam. A mobil fázis 150mM ABC puffer (pH=7,5) volt. A kromatográfiás program hossza 41 perc volt, amelynek során a mozgó fázis térfogatárama a 3. táblázat alapján változott.

3. táblázat WGA affinitáskromatográfia programja.

idő [perc]	térfogatáram [ml/perc]
0	0,125
25,5	0,125
25,6	0,5
40,0	0,5
40,1	0,125

Az erősebben kötődő glikopeptidek leszorítására 200 µl 200 mM GlcNAc (Merck)/150 mM ABC (pH=7,5) oldatot használtam, amelyet minden minta esetében a kromatográfiás program

10,7 percében injektáltam az oszlopra. Az első affinitás kromatográfia során 7 db 3 perces frakciót szedtem mintánként 4,3 – 25,3 perc között (15. ábra). Az 5-7. frakciók GlcNAc tartalmát Waters Sep-Pak C18 Plus SPE oszlopon távolítottam el. Ezután az áteső csúcsot tartalmazó frakció (FT) kivételével a fennmaradó frakciókat egyesítettem, majd vákuumcentrifugával beszárítottam. A beszárítás után a mintákat feloldottam 100 µl 150 mM ABC pufferben (pH=7,5), majd megismételtem a dúsítási lépést az előbbieken leírt protokollt alkalmazva. A kétlépéses lektin affinitás kromatográfia végére mintánként végül 4 frakció keletkezett, melyeket vákuumcentrifugával beszárítottam. Ezek sorrendben a következők: áteső frakció (1, FT), áteső csúcs vége frakció (2, FTEnd), váll frakció (egyesített 3-4, Shoulder) és az GlcNAc-eluált frakció (egyesített, cukor-mentesített 5-7, GlcNAc) (15. ábra).



15. ábra Humán vizeletminták frakcióinak kezelése

3.3 A tömegspektrometriás analízis és az adatkiértékelés során alkalmazott paraméterek

A mintákat nanoLC-MS/MS analízissel vizsgáltam.

3.3.1 nanoLC paraméterek

A WGA frakciók fordított fázisú on-line frakcionálása egy Waters nanoAcquity M-Class UHPLC készülékkel történt. A minták fókuszálása és a sóalanítás egy csapdázó oszlopon (Waters Acquity UPLC M-Class Symmetry C18 180 µm × 20 mm, 5 µm szemcseméret, 100 Å pórusméret, áramlási sebesség: 10 µl/min, oldószerösszetétel: 3 %B) történt, majd egy lineáris gradiens segítségével (B oldószer: 10 %-ról 30 %-ra 60 perc alatt, a teljes programot lásd a 4. táblázatban) egy fordított fázisú nano oszlopon frakcionáltam a mintákat. (Waters Acquity

UPLC M-Class BEH C18 75 μm \times 250 mm, 1,7 μm szemcseméret, 130 Å pórusméret). A frakcionálás során A oldószerként 0,1 % hangyasav/víz, B oldószerként 0,1 % hangyasav/acetonitril (VWR) keverékét használtam. A grádiensprogram teljes hossza 110 perc volt.

4. táblázat nanoLC grádiensprogram.

Idő [perc]	Térfogatáram [nl/perc]	Oldószerarány [%B]
0	300	3
5	300	10
65	300	30
75	300	40
78	300	90
80	300	90
84	300	10
87	300	90
90	300	3

Az on-line kromatográfiás detektor szerepét egy Thermo Scientific™ Orbitrap Fusion™ Lumos™ Tribrid™ tömegspektrométer töltötte be. A mérések során a nano-ESI forrásban alkalmazott spray feszültség értéke +1200 V. A transzferkapilláris hőmérséklete 300 °C volt.

A tömegspektrometriás analízis során minden MS1 mérést (m/z 380-1580, $R = 60000$, Orbitrap, profil mód, pozitív ion mód) maximum 3 másodperces ciklus követett, ahol a készülék a prekursorokat csökkenő intenzitás szerint választotta ki MS/MS kísérletre (minimum intenzitás: 10^6). A $z = 3-5$ és $z = 2$ töltéssel rendelkező prekursorokat külön analizáltuk. A prekursorok kiválasztása kvadrupól analizátorral történt, 2 Da-os ablakkal. HCD spektrum minden kiválasztott prekuzorról készült (AGC célérték: 50000, NCE = 28%), míg EThcD spektrum csak azokról (AGC célérték: 300000, kiegészítő NCE: 15%), amelyek HCD spektrumában a 20 legintenzívebb ion között a HexNAc oxónium ionjának megfelelő m/z 204,0867-es ion detektálható volt (10 ppm tömegpontosságon belül). A készülék minden MS/MS spektrumot az Orbitrap analizátorral vett fel ($R = 15000$, centroid mód). A mérés során a dinamikus kizárás 30 s volt.

3.3.2 Glikopeptidek azonosítása automatizált adatkiértékeléssel (4.1 fejezet)

A nyers mérési fájllokból a Proteome Discoverer szoftverrel (Thermo Scientific, v2.2.0.388) külön HCD és EThcD csúcslistákat (.mgf formátum) állítottam elő a. A szoftver csak azokat a spektrumokat tartotta meg, amelyekben legalább 10 csúcs volt. Az EThcD csúcslistákat tovább

szűrtem a NeuAc oxónium ionjának jelenlétére (m/z 292,1027) a Protein Prospector MS-Filter szoftverének segítségével. A szoftver azokat az EThcD spektrumokat tartotta meg, ahol az ion a 80 legintenzívebb csúcs között volt és 10 ppm tömegpontosságon belül. Az így előszűrt EThcD csúcslistákat a Protein Prospector Batch Tag Web (v5.16.0) és Byonic (v2.13.17) szoftverekkel kerestem le. Az N- és O-glikopeptideket külön keresésben azonosítottam. Az N-glikopeptideket csak a Byonic szoftverrel, míg az O-glikopeptideket mindkét szoftverrel azonosítottam. Mindkét szoftver fehérjeadatbázisként a SwissProt humán adatbázisát használta (2017.9.19. verzió, 20291 szekvencia), amely kiegészült a fehérjék randomizált szekvenciáit tartalmazó decoy adatbázissal. Az adatbázis keresés során a triptikus peptidek mellett szemitriptikus peptideket (N- vagy C-terminálison) is figyelembe vett a szoftver. A kihagyott hasítóhelyek maximális száma: 1. A prekursor ionokra az alkalmazott tömegpontosság 5 ppm, míg a fragmensionokra 10 ppm volt. Állandó módosítás: karbamidometiláció (Cys). Változó módosítások: acetiláció (fehérje N-terminális), ciklizáció (Gln a peptid N-terminálisán), oxidáció (Met), O-glikánok (HexNAcHex, HexNAcHexNeuAc, HexNAcHexNeuAc₂, HexNAc₂Hex₂NeuAc₂ szerkezetek Ser és Thr aminosavakon), N-glikánok (57 humán N-glikán szerkezet (1. melléklet) Asn aminosavon). Egy peptiden maximum 2 változó módosítást engedtem meg. Az N-glikán struktúrák „rare”, az összes többi változó módosítás „common” módosításként lett definiálva. A „common” módosítások szabadon kombinálhatók, míg a „rare” módosításból csak egy szerepelhet a peptid azonosításokban. A Protein Prospectorral végzett lekeresések elfogadási paraméterei a váll és a GlcNAc frakciókra: FDR: 5 (fehérje) és 1% (peptid), illetve 10 és 5% továbbá SLIP score ≥ 6 . Byonic keresések esetében: Maximális fehérje FDR: 1%, PEP2D $< 0,1$.

3.3.3 Oxónium ion intenzitás-mintázat analízise HCD és EThcD spektrumokban (4.2 fejezet)

A HCD és EThcD csúcslisták szűrését 4 cukor oxónium ion jelenlétére az MS-Filter szoftverrel [Medzihradsky, 2015 (A)] végeztem. Ezek rendre m/z 292,1027 (NeuAc⁺), m/z 657,2349 (HexNAcHexNeuAc⁺), m/z 948,3303 (HexNAcHexNeuAc₂⁺) és m/z 1313,4625 (HexNAc₂Hex₂NeuAc₂⁺). A NeuAc (m/z 292) és a HexNAcHexNeuAc (m/z 657) szerkezetek oxónium ionjait a HCD és EThcD spektrumok 20 legintenzívebb ionja között, míg a másik két szerkezetnek megfelelő oxónium ionokat, mind a két spektrumtípus esetében, a 80 legintenzívebb csúcs között vizsgáltam. A szűrés során a beállított tömegpontosság 10 ppm volt.

Az intakt glikánok jelenlétére is megsűrt EThcD csúcslistákat egy olyan adatbáziskereséssel azonosítottam, ahol minden paraméter megegyezett az m/z 292-es EThcD csúcslistákkal végrehajtott adatbáziskeresés paramétereivel.

3.3.4 Új O-glikán struktúrák azonosítása nem definiált módosítás kereséssel EThcD adatokból (4.3 fejezet)

A tömegspektrometriás analízis paraméterei megegyeznek a 3.3.1 fejezetben leírtakkal. A nyers mérési fájljokból külön HCD és EThcD csúcslistákat gyártottam a Proteome Discoverer szoftverrel (v2.2.0.388). A konverzió során csak olyan spektrumok maradtak meg, amelyekben legalább 40 csúcs volt. Az EThcD adatokat a Protein Prospector MS-Filter programjával megszűrtem az m/z 292,0927 ion (NeuAc) jelenlétére, amelynek a 20 legintenzívebb csúcs között kellett lennie, 10 ppm tömegpontosságon belül. Az így megszűrt EThcD csúcslistákkal egy nem definiált módosítás keresést végeztem a Protein Prospector Batch Tag Web szoftver segítségével (v5.16.0). A kereséshez egy szűkített fehérjeadatbázist használtam, amely már a korábban és megbízhatóan azonosított O-glikoproteinek szekvenciáiból állt (418 szekvencia). Az adatbázis keresés során csak triptikus peptideket vettem figyelembe maximum 2 kihagyott hasítóhellyel. A prekursor ionokra 5 ppm, a fragmensionokra 10 ppm tömegpontosságot állítottam be. Állandó módosítás: karbamidometiláció (Cys). Változó módosítások: acetiláció (fehérje N-terminális), ciklizáció (Gln a peptid N-terminálisán), oxidáció (Met), illetve nem definiált módosítás 0-3000 Da tartományban Ser, Thr és Asn aminosavakon. Az eredmények elemzése során nem határoztam meg elfogadási paraméter célértéket, mivel ez egy felfedező jellegű keresés volt.

A nem definiált módosítás keresésekből meghatározott új O-glikán szerkezetekkel újabb adatbázis keresést végeztem a Protein Prospector szoftver segítségével. Ehhez a kereséshez a SwissProt humán fehérjeadatbázisát használtam (2017.9.19. verzió, 20219 szekvencia), amely kiegészült a fehérjék randomizált szekvenciáit tartalmazó decoy adatbázissal. Az adatbáziskeresés során a triptikus peptidek mellett szemitriptikus peptideket is figyelembe vett a szoftver. A kihagyott hasítóhelyek maximális száma: 1. A prekursor és fragmensionokra vonatkozó tömegpontosság: 5, illetve 10 ppm. Állandó módosítás: karbamidometiláció (Cys). Változó módosítások: acetiláció (fehérje N-terminális), ciklizáció (Gln a peptid N-terminálisán), oxidáció (Met), O-glikánok (28 db összetétel (4. melléklet) Ser és Thr aminosavakon). Minden változó módosítást „common”-ként definiáltam. Egy peptiden maximum 2 változó módosítás lehetett. Az adatbázis keresés során a szoftver a spektrumok 80 legintenzívebb csúcsát vette figyelembe (egy spektrum mindkét feléből 40-40 csúcs). A glikopeptid azonosítások elfogadási paraméterei: 5% és 1% FDR fehérje és peptid szinten, illetve SLIP score ≥ 6 .

3.3.5 További új O-glikán struktúrák azonosítása GF-Hunterrel (4.4 fejezet)

A tömegspektrometriás analízis EThcD MS/MS kísérletének AGC célértéke módosult a 3.3.1 fejezetben leírthoz képest. Az új érték 50000. A HCD és EThcD csúcslistákat a 3.3.4 fejezetben ismertetett paraméterekkel generáltam. Az EThcD csúcslisták szűrését az MS-Filter programmal végeztem a 3.3.4 fejezetben leírt paraméterekkel. Az EThcD csúcslisták adatbáziskeresése a 3.3.2 fejezetben ismertetett paraméterekkel történt. A keresési eredmények elfogadási paraméterei: Minimum fehérje és peptid score értékek: 22 és 15, illetve maximum fehérje és peptid E-Value értékek: 0,01 és 0,05.

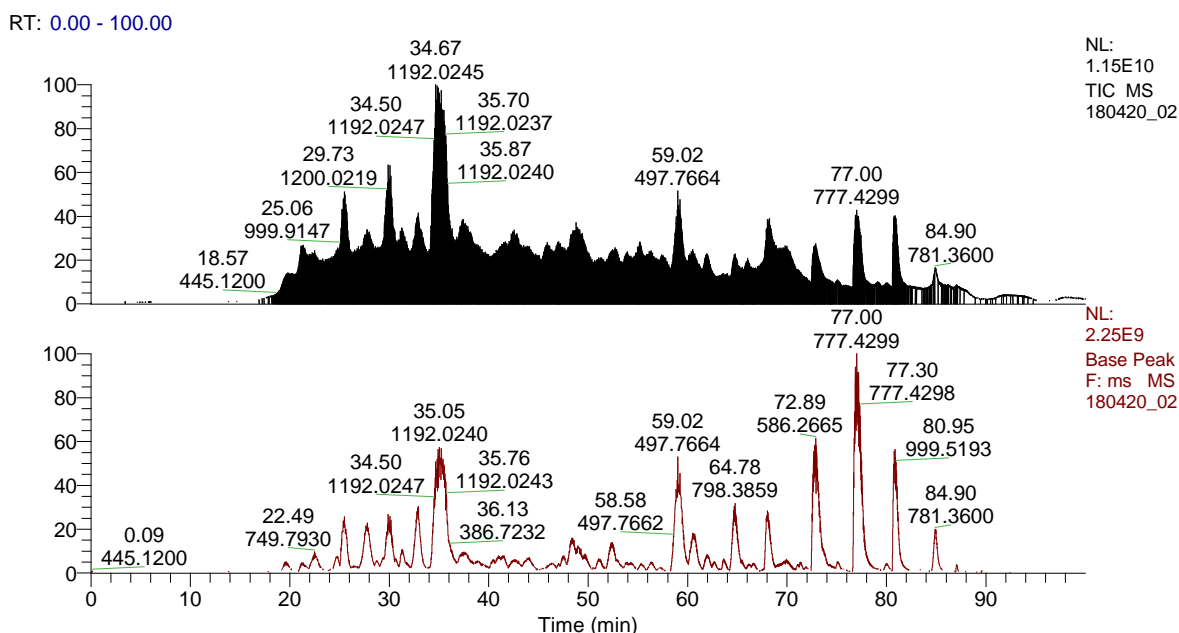
Az elfogadási paramétereknek megfelelő EThcD O-glikopeptid azonosításokból készítettem el a GF-Hunter bemeneti peptid listáját, amely egy .txt formátumú fájl. A listán szereplő peptidszekvenciák nem tartalmazzák az adatbázis keresés során azonosított glikánt, minden más kovalens módosítást azonban igen. Ez a .txt file és a HCD csúcslisták képezik az GF-Hunter bemeneti file-jait. A szoftver a HCD csúcslistákban a bemeneti listán szereplő peptidok Y_0 - Y_1 párjait kereste 10 ppm tömegpontosságon belül, azonos töltésállapotban (1+, 2+, 3+), a 30 legintenzívebb csúcs között. Ha a szoftver által számolt addítv tömeg 10 ppm tömegpontosságon belül volt a beépített glikánlistában szereplő valamely glikán elméleti additív tömegéhez, akkor a szoftver a passzoló glikánösszetételt hozzárendeli megfelelő HCD spektrum – peptid párhoz.

4. EREDMÉNYEK ÉS MEGVITATÁSUK

4.1 Humán vizeletből dúsított glikopeptidek azonosítása automatizált adatkiértékeléssel

A humán vizeletben található fehérjék O-glikozilációjának megismerését három egészséges donor vizeletének vizsgálatával kezdtem. Ehhez a vizeletben található fehérjéket tripszinnel emésztettem, majd két körben WGA affinitás kromatográfiával glikopeptideket dúsítottam. A kétszeres lektinkromatográfia célja a minta nem specifikus háttérének csökkentése volt. A kromatográfia során több frakciót is gyűjtöttem (3.2.2 fejezet). Az áteső frakcióban a WGA lektinnel nem kölcsönható peptidek találhatóak. LC-MS/MS analízist az FTEnd (2), Shoulder (3-4) és a GlcNAc (5-7) frakciókkal végeztem. Az áteső csúcs végi frakcióban jelennek meg a lektinnel gyengén kölcsönható egyszeresen glikozilált O-glikopeptidek, ezek maximuma a shoulder frakcióban található, míg a GlcNAc-nal eluált frakciókban a lektinhez erősebben kötődő, főleg többszörösen módosított O-glikopeptidek találhatóak.

A futások összionáram és báziscsúcs kromatogramját a 16. ábra szemlélteti egy kontroll mintán bemutatva.



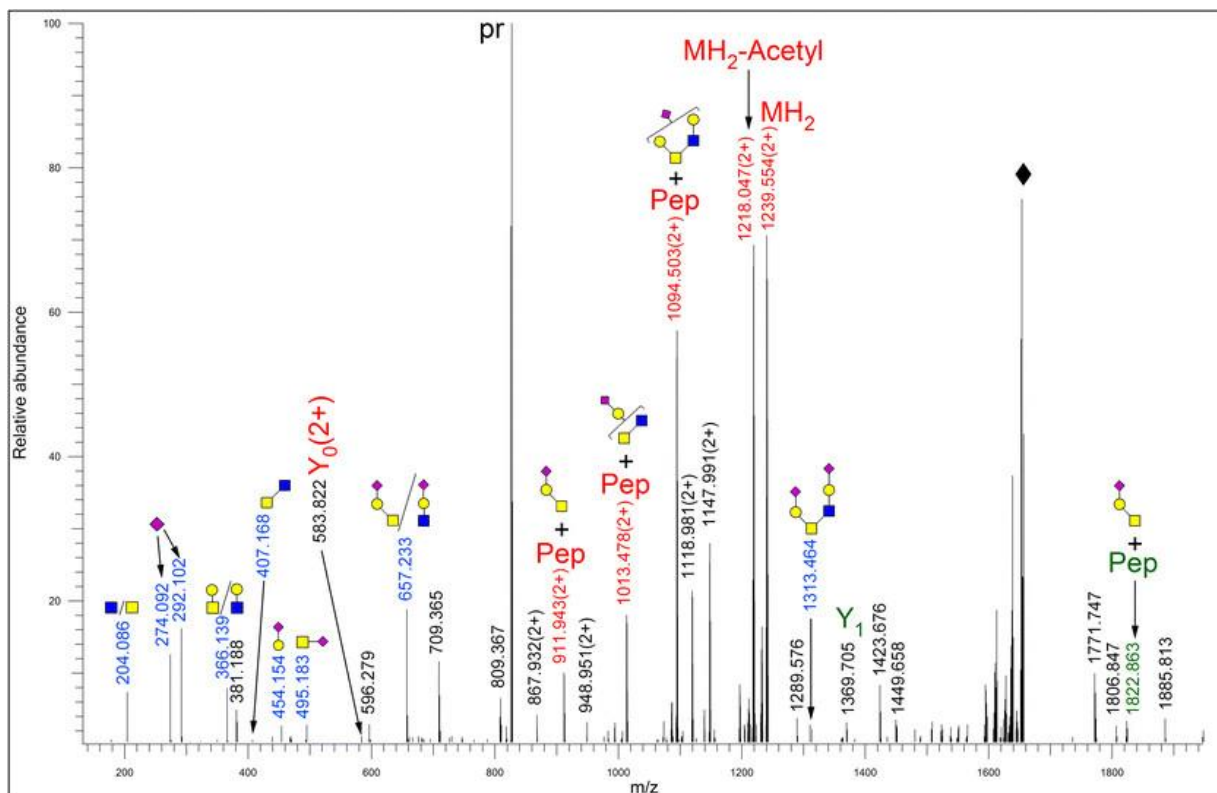
16. ábra Kontroll minta összionáram (felső panel) és báziscsúcs (alsó panel) kromatogramja.

A glikopeptidek dúsítására használt WGA egy nonspecifikus lektin, így az izolált peptideken jelen lévő glikánokat nem lehet megjósolni. Az MS/MS adatok előzetes szűrése során megállapítottam, hogy az EThcD spektrumok döntő többsége tartalmazza az m/z 292.103 iont, ami szialilált glikán struktúrák jelenlétére utal, így az adatkiértékelés során ezekre a struktúrákra

fókuszáltam. Az O-glikopeptidek azonosítására két szoftvert (Byonic [Bern, 2012], Protein Prospector [Baker, 2011]) használtam. Potenciális módosításként di-, mono- és aszialo core 1-es és diszialo core 2-es alapszerkezeteket vettem figyelembe, feltételezve, hogy a vizeletben jelenlévő fehérjék hasonló O-glikán mintázattal rendelkeznek, mint a plazma fehérjék [Yabu, 2014]. N-glikopeptidek azonosítására csak a Byonic szoftvert használtam. Az N-glikán adatbázis a Byonic belső N-glikán adatbázisa, amely a human plazmában előforduló 56 leggyakoribb szerkezetet tartalmazza. Az elfogadási paramétereknek megfelelő O- és N-glikopeptid azonosítások a 2. és 3. mellékletben találhatóak. Összességében több O-glikopeptidet sikerült azonosítani, mint N-glikopeptidet. Az O-glikopeptid azonosítások a core struktúrák szintjén 67%-ban átfednek a Halim és mtsai. által közölt adatokkal, akik szintén humán vizeletből analizálták az O-glikopeptideket. Az N-glikopeptid azonosítások átfedése alacsonyabb (20%) [Halim, 2012]. A kutatócsoport hidrazid funkciót tartalmazó gyöngyökön dúsított glikoproteineket a szialsavak perjódsavas oxidációját követően. Az oszlophoz kötött fehérjéket tripszinnel emésztették, a módosítatlan peptideket mosással eltávolították, majd a glikopeptideket hangyasavas hidrolízissel (lehasítva a szialsavat) izolálták. A mintákat LC-MS/MS technikával analizálták. CID-MS² segítségével a peptideket módosító glikán összetételét határozták meg, majd CID-MS³ kísérlettel a CID-MS²-ben keletkezett deglikozilált peptid iont (Y₀) fragmentálták a peptidszekvencia azonosítása érdekében. ECD-MS² kísérlettel az O-glikán peptiden belüli pozícióját kívánták meghatározni. A fent vázolt eljárás következtében az eredeti glikán struktúra nem határozható meg a szialsavak eltávolítása miatt (pl. egy mono- és diszialilált core 1 struktúrát hordozó glikopeptid egyaránt a GalNAcGal glikánt hordozza a mintaelőkészítés után). Az eredmények összehasonlítása során azokat a glikopeptideket tekintettük azonosnak, ahol az általunk meghatározott glikán aszialo része egyezett Halim és mtsai. adataival. Többnyire a módosítás helye is megegyezett, bár a különböző szoftverek segítségével végzett PTM lokalizáció megbízhatósága kérdéses.

Az EThcD spektrumok azonosítási rátája alacsony. A két vizsgált frakció közül az áteső csúcs váll frakciójából származott több azonosítás. O-glikopeptidek tekintetében a két adatbáziskereső szoftver hasonlóan teljesített. A Byonic összesen 556 spektrumot azonosított, míg a Protein Prospector 552-t. Mindkét szoftver által azonosított spektrumok száma 343. Ezeket a számokat kombinálva az N-glikopeptid azonosításokkal (328 PSM) az azonosítási ráta, az összes betáplált EThcD spektrumra vonatkozóan, kb. 5%. A GlcNAc-nal eluált frakció esetében a Byonic 326, míg a Protein Prospector 304 spektrumot azonosított. Az N-glikopeptid azonosítások száma 109. Így a GlcNAc frakció azonosítási rátája kb. 2,5%. Az alacsony azonosítási ráta egyik oka az EThcD aktiválásra kiválasztott prekursorok alacsony töltéssűrűsége [Good, 2007]. A prekursor

ionok több, mint 90%-a m/z 900-nál nagyobb. Az m/z 900 alatti prekursorok esetében a spektrumok 13%-át sikerült azonosítani, míg az m/z 900 feletti prekursorok esetében ez az érték csak 2%. A többszörös glikoziláció vagy egy nagy méretű glikán is akadályozhatja az azonosítást, mert: 1) több könnyen hasadó glikozidos kötés van jelen és 2) kevesebb protonálható csoport.



17. ábra m/z 827,037 (3+) azonosítatlan glikopeptidről készült EThcD spektrum. A peptidhez egy core 2-es hexaszacharid kapcsolódik. A peptid szekvenciájának azonosításához szükséges fragmensek hiányoznak. Az oxónium ionok kék színnel jelöltek, míg a töltésvesztett prekursorokból származó cukorvesztések pirossal (2+) és zölddel (1+). Y_0 a deglikozilált peptidet, míg Y_1 a GalNAc-nal módosított peptidet jelöli. A rombuszsal jelölt csúcs egy kétszeres töltésű koelválódó prekursor töltésvesztett formája.


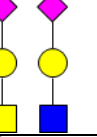

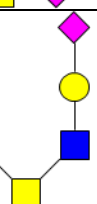
Ebből jól látható, hogy az ETD alacsony hatásfoka limitáló tényező az EThcD spektrumok azonosításában is, továbbá a kiegészítő aktiválásként alkalmazott HCD sem segíti elő a peptidszekvencia meghatározásához szükséges peptidfragmensek keletkezését. Sok EThcD spektrum esetében a gyökös fragmentációból származó szekvencia ionokból egyet sem, vagy csak keveset detektáltunk. A spektrumokban a domináns fragmensionok a glikán fragmentációjából származnak, amelyek a kiegészítő HCD aktiválás során keletkeznek. A glikánfragmentáció dominánssá válása kifejezetten jellemző nagyobb és/vagy több O-glikánnal módosított peptidek esetében (17. ábra).

A 17. ábrán egy azonosítatlan glikopeptid EThcD spektruma látható. A spektrumban megfigyelhető m/z 1313,464 ion egy $\text{HexNAc}_2\text{Hex}_2\text{NeuAc}_2$ összetételű glikánra utal, amelyből arra lehet következtetni, hogy a peptid egy diszialo core 2-es szerkezettel módosított. A glikán additív tömegét igazolja a deglikozilált peptid-ion (Y_0) m/z 583,822 (2+) jelenléte is. A spektrumban nem található elegendő szekvencia ion a peptid aminosav-sorrendjének meghatározásához. A sikeresebb spektrum azonosítás ellen dolgozik a komplex elegyekben megfigyelhető prekursor ion interferencia is. A legtöbb EThcD spektrumban megfigyelhetőek olyan ion klaszterek, amelyek más töltésállapotban lévő prekursorok töltésvesztett formáinak felelnek meg (17. ábra, rombuszsal jelölt csúcs). Az automatizált adatkiértékelés során is tetten érhető ez a jelenség. Az EThcD adatokból gyakran sikerült nem glikozilált peptideket is azonosítani annak ellenére, hogy az EThcD spektrumok felvétele a HCD-ben megjelenő diagnosztikus cukor oxónium ionhoz (m/z 204,087) volt kötve.

4.2 Oxónium ionok intenzitás-mintázata alacsony ütközési energiával (HCD NCE = 15%) felvett EThcD spektrumokban

Az EThcD spektrumok manuális ellenőrzése során felfigyeltünk arra, hogy nagyobb oxónium ionok is megjelennek a HCD spektrumban megfigyeltékhez képest. Némely esetben a peptidet módosító teljes cukorszerkezetnek megfelelő oxónium ion is detektálható (17. ábra).

5. táblázat A HCD és EThcD csúcslisták összehasonlítására használt oxónium ionok szerkezete és monoizotópos tömege.

Oxónium ion	Szerkezet	Monoizotópos tömeg	Glikán típusa
$(\text{NeuAc})^+$		292,1027	O, N
$(\text{HexNAcHexNeuAc})^+$		657,2349	O, N
$(\text{HexNAcHexNeuAc}_2)^+$		948,3303	O
$(\text{HexNAc}_2\text{Hex}_2\text{NeuAc}_2)^+$		1313,4625	O

A Protein Prospector MS-Filter programjával vizsgáltuk, hogy ez mennyire általános jelenség: a HCD és az EThcD adatokat „megtisztítottuk” az 5. táblázatban található cukor oxónium ionok

jelenlétére. Relatív intenzitásukat is igyekeztünk megbecsülni úgy, hogy limitáltuk a legintenzívebb ionok számát, amelyek között a keresett oxónium ionnak szerepelnie kellett (3.3.3 fejezet, 6. táblázat). A HCD és EThcD spektrumok 20 legintenzívebb fragmensét vizsgálva, a NeuAc oxónium ionjának előfordulási gyakorisága megegyező (6. táblázat). A monoszialo core-1 oligoszacharid oxónium ionja (m/z 657) azonban 50%-al gyakrabban fordul elő az EThcD spektrumokban, mint a HCD adatokban (6. táblázat).

6. táblázat A 5. táblázatban feltüntetett oxónium ionok előfordulási gyakorisága (%) a HCD és EThcD csúcslistákban 10 ppm tömegpontossággal, az adott számú legintenzívebb ion között (**Top X**). V és G a WGA kromatográfia során gyűjtött váll és GlcNAc frakciókat, míg az 1, 2 és 3 a különböző mintákat jelöli.

Minta	292		657		948		1313	
	Top 20		Top 20		Top 80		Top 80	
	HCD	EThcD	HCD	EThcD	HCD	EThcD	HCD	EThcD
V1	85	88	44	87	0,41	23	0	0,16
V2	83	86	45	89	0,49	17	0	0,21
V3	95	95	53	91	2	36	0	0,59
G1	96	93	47	92	1,4	33	0	0,59
G2	95	94	50	95	1,4	30	0	0,3
G3	96	97	66	96	14	57	0	0,12

A nagyobb alapszerkezetek oxónium ionjainak az előfordulási gyakorisága a triszacharidéhoz képest alacsonyabb már EThcD-ben is, de a HCD adatokhoz képest mindenképpen magasabb (6. táblázat). Az m/z 948 ion (diszialo core 1 tetraszacharid) előfordulási gyakoriságának növekedése (mindkét aktiválási módszer esetében) akkor állt meg, ha a 80 legintenzívebb fragmenst vettük figyelembe. Az m/z 1313 ion (diszialo core 2 hexaszacharid) gyakorisága volt a legkisebb (< 1%), ami nem csak a struktúra „törékenységét” jelzi, de a humán szérumban végzett szabad glikánok vizsgálatából az is sejthető, hogy kisebb az előfordulási gyakorisága [Yabu, 2014]. A diszialisált core 2-es hexaszacharid (m/z 1313) további fragmensei lehetnek még: HexNAc₂Hex (m/z 569), HexNAc₂Hex₂ (m/z 731), HexNAc₂HexNeuAc (m/z 860), HexNAc₂Hex₂NeuAc (m/z 1022), ám ezek közül egyik fragmens sem volt számottevő mértékben detektálható. Ezek az új megfigyelések segíthetik az O-glikopeptidek azonosítását, mivel annak eldöntése, hogy egy szekvencia egy nagyobb oligoszachariddal vagy két kisebbel van módosítva sok esetben nehezen dönthető el.

Tapasztalataink alapján új megközelítést alkalmaztunk annak érdekében, hogy több tetra-, illetve hexaszachariddal módosított peptidet azonosíthassunk. A csúcslistákban csak olyan

spektrumokat hagytunk, amelyek a nagyobb glikánokra jellemző fragmenseket tartalmaztak. A tetraszacharidra keresve az m/z 292-es ion mellett további feltétel volt, hogy a spektrum tartalmazza az m/z 948-as iont is, amely az intakt tetraszacharidnak felel meg. A hexaszacharidra keresve az m/z 292-es ion mellett az EThcD spektrumban szerepelnie kellett az m/z 657-es és ezzel egyidejűleg az m/z 1313 (a teljes szerkezetnek megfelelő oxónium ion) ionnak is. Ezzel a lekeresési stratégiával az azonosított O-glikopeptidek 90%-a a csúcslisták szűrésénél használt intakt oxónium ionnak megfelelő O-glikán szerkezettel illett az adatokhoz. Ennek ellenére az azonosítási ráta nem javult jelentősen azokhoz az eredményekhez képest, amelyek a csak m/z 292 ion jelenlétére szűrt EThcD csúcslistákkal végzett adatbáziskeresésekből származtak (2. melléklet, Összefoglaló fül). A Protein Prospector az új, tetraszacharid jelenlétére is szűrt csúcslistákkal összesen 270 (váll frakció) és 100 (GlcNAc frakció) spektrumot, míg a Byonic összesen 217 (váll frakció) és 77 (GlcNAc frakció) spektrumot azonosított (7. táblázat). Mindkét szoftver azonosított olyan spektrumokat is, amelyek a csak m/z 292 ion jelenlétére szűrt EThcD csúcslistákkal végzett lekeresésekkel nem jöttek ki. A Protein Prospector 53 (váll frakció) és 26 (GlcNAc frakció) EThcD spektrumot azonosított újonnan, tetraszacharid szerkezettel módosított peptidekként. A Byonic csupán 2 és 10 új azonosítást produkált a váll és GlcNAc frakciókból összesen (8 táblázat).

7. táblázat A nagyobb oxónium ionok (m/z 948 és 1313) jelenlétére szűrt csúcslistákkal végzett adatbáziskeresések eredményeinek összehasonlítása. PP: Protein Prospector.

Keresések	Váll frakció			GlcNAc eluált frakció		
	Σ EThcD spektrum	Azonosítva	Egyedi azonosítás	Σ EThcD spektrum	Azonosítva	Egyedi azonosítás
<i>Byonic (292)</i>	20965	556	104	20452	326	93
<i>Byonic (292+948)</i>	5530	217	2	8266	77	10
<i>Byonic (1313)</i>	95	17	5	97	0	0
<i>PP (292)</i>	20965	552	53	20452	304	84
<i>PP (292+948)</i>	5530	270	155	8266	100	26
<i>PP (1313)</i>	95	22	7	97	7	1

Az m/z 292 + 1313 ionokra szűrt EThcD csúcslistából (95 spektrum, váll frakció) a Protein Prospector 22 spektrumot azonosított 16 egyedi peptidszekvenciával. A Byonic ugyanezt a csúcslistát használva 17 spektrumot azonosított 12 egyedi peptidszekvenciával (7. táblázat). A Protein Prospectorral való átfedés 12 azonosítás, amely 7 egyedi O-glikopeptidnek felel meg (7. táblázat). A legjobb azonosításhoz tartozó spektrumot a 18. ábra mutatja be. A csak az m/z 292-es ion jelenlétére szűrt csúcslistákkal végzett lekeresésekkel (Protein Prospector és Byonic) összesen 110 EThcD spektrumot azonosítottunk hexaszachariddal módosított glikopeptidként a váll frakcióból. Ezeknek a spektrumoknak egy jelentős része a módosító glikán tekintetében nem

4.3 Új O-glikán struktúrák azonosítása

Ahhoz, hogy növeljük az azonosított spektrumok számát és megtudjuk milyen más O-glikán szerkezettel módosított peptidek rejtőznek az adatokban, egy ún. nem definiált módosítás keresést végeztünk, amelynek során a módosítás tömege legfeljebb 3000 Da lehetett Ser, Thr és Asn aminosavakon. A keresést a már korábban azonosított fehérjék részhalmazán végeztük. (3.3.4 fejezet). Az ilyen lekeresések eredményeinek értékelése során célszerű a korábban már megbízhatóan, nagy pontszámmal, több mintában is előforduló glikopeptidek új módosításait megvizsgálni, különös tekintettel az egyetlen helyen módosítható szekvenciákra. Ezekben az esetekben a legegyszerűbb ugyanis az új cukor-összetételeket felismerni a mikroheterogenitás alapján. Az így detektált tömegkülönbségek azt sugallták, hogy az eddig megengedett szialoszerkezetekre acetilcsoport(ok) és/vagy extra NeuAc került, és ezt az MS/MS spektrumok ellenőrzése igazolta is.

8. táblázat Neu5,9Ac₂-t és/vagy disziálsavat tartalmazó O-glikánt hordozó glikopeptidek.¹

Uniprot Acc #	Fehérje neve	Peptid ²	(4) ³		(5) ³			(6) ³	
			1	2	0	1	2	3*	1
Q9GZM5	Protein YIPF3	³⁴² AVAV TL QSH ³⁵⁰	+	+	+	+	+	+	+
P01344	Insulin-like growth factor II	⁹³ DVSTP P TVLPDNFPR ¹⁰⁷			+	+			
P78423	Fractalkine	¹⁷⁵ AQDGGPVG T ELF ¹⁸⁶				+			
		¹⁷⁵ AQDGGPVG T ELFR ¹⁸⁷				+			
		²⁴⁷ VWGQGG Q SPRPE ²⁵⁷					+		
		²⁴⁷ VWGQGG Q SPRPE N SLER ²⁶²			+	+			
		²⁶³ EEMGPVPA H TDAF ²⁷⁵			+	+			
Q9BZL3	Small integral membrane protein 3	¹ MDAV S QVPMEVVLPK ¹⁵	+			+			
P98160	Basement membrane-specific heparan sulfate proteoglycan core protein	³³ SLPEDIETV T ASQMR ⁴⁷			+				
		⁴⁸ W T HSYL ⁵³				+			
Q13508	Ecto-ADP-ribosyltransferase 3	³¹⁶ ILE P TQIPGMK ³²⁶	+						
Q9H2B2	Synaptotagmin-4	⁹ EEFDEI P TVVGF ²¹				+			
Q8TBP5	Membrane protein FAM174A	⁵⁰ P R TLPLPPG P TPAQQPGR ⁶⁸	+						
P09603	Macrophage colony-stimulating factor 1	³⁷⁰ VGPVR P TGQDWNHTPQK ³⁸⁶			+				
P14209	CD99 antigen	²⁹ SDALPDNENKK P TAIPK ⁴⁵			+				
P60022	Beta-defensin 1	²² GNFL T GLGHR ³¹			+				

¹Protein Prospectorral azonosítva és manuálisan validálva. ²A pirossal jelölt Ser/Thr aminosavak a glikoziláció helyét jelölik. ³A glikánban található monoszacharid egységek számát jelöli, amelyek a szialosavakat is tartalmazzák. A következő sorban szereplő számok a glikánban található Neu5,9Ac₂-k számát jelölik. A feltételezett struktúrákat

az 4. melléklet tartalmazza. *valójában egy NeuAcNeu5,9Ac₂Neu4,5,9Ac₃ kombinációt jelöl (ld. lentebb és a 22. ábrát).

Ismert, hogy polisziálsav láncok és disziálo-csoportok léteznek [Varki, 2015], de tömegspektrometriás detektálásukat intakt glikopeptidek esetében (részben, mert a szíálsavat többnyire eltávolítják) [Halim, 2012; Steentoft, 2011] eddig nem közölték. Az acetyl-csoport egy közismert fehérje poszttranszlációs módosítás, de az adatok elemzése során bizonyítást nyert, hogy a cukorszerkezet acetylált, nem pedig a peptid: a HCD és EThcD spektrumokban az m/z 334,113 és a vízvesztéssel keletkezett m/z 316,103 ionok Neu5,9Ac₂ jelenlétét jelzik. Ezekből az adatokból természetesen nem lehet meghatározni, hogy a NeuAc mely hidroxil-csoportja acetylzódott, de a fent említett struktúra a legvalószínűbb [Varki, 2015]. A Protein Prospector MS-Filter programjával, a két fragmens (m/z 316, 334) együttes jelenlétére szűrve, az EThcD spektrumok nagyjából 5%-a eredhet Neu5,9Ac₂-tartalmú glikoformából.

Új adatbázis keresést végeztünk egy olyan glikán adatbázissal, amelyben már disziálsavat és Neu5,9Ac₂-t tartalmazó O-glikán struktúrákat is felsoroltunk (3.3.4 fejezet, 4. melléklet). Az azonosított spektrumokat manuálisan ellenőriztük. A disziálsavat, illetve a Neu5,9Ac₂-t tartalmazó O-glikánnal módosított peptidek listáját a 8. táblázat tartalmazza.

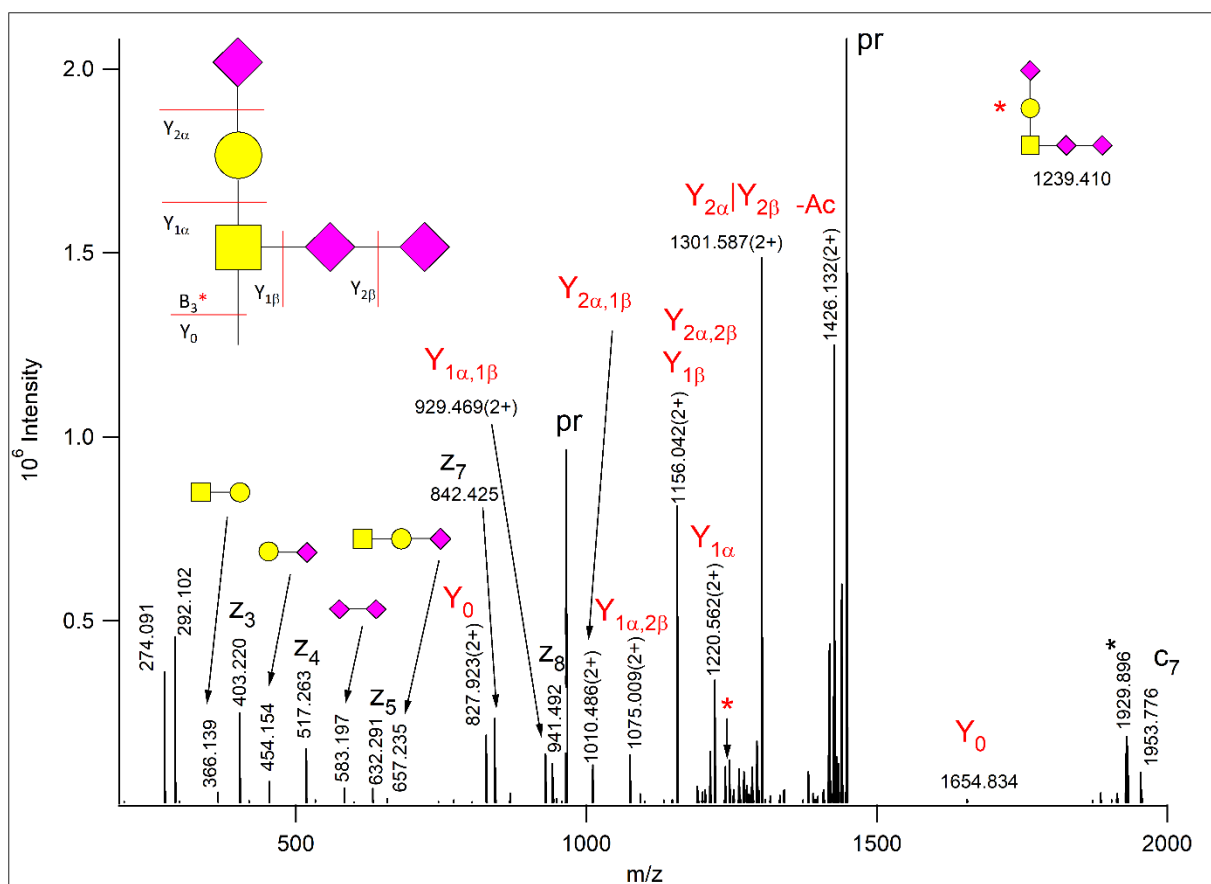
4.3.1 Glikán fragmentációs szabályok alacsony ütközési energiával (NCE = 15%) felvett EThcD spektrumokban

Amint azt a 4.2 fejezetben már részletesen bemutattam, a HCD és EThcD spektrumok másfajta oxónium ion eloszlást mutatnak. Ez utóbbiakban az aktiválást nagyobb tömegű oxónium ionok is túlélnek. Sokszor a teljes cukorszerkezetnek megfelelő oxónium ion is detektálható, még a hexaszacharid (HexNAc₂Hex₂NeuAc₂) esetében is. Az újonnan felfedezett glikoformák adatainak gondos vizsgálatával arra a következtetésre jutottunk, hogy az általunk használt EThcD körülmények között, akárcsak az ioncsapda CID-ben, az egyszeres kötés-hasadás preferált. Így az oxónium ionok az Y ionokkal együtt lehetővé teszik a glikán topológiájának meghatározását. Ez a megfigyelés jól használható volt arra, hogy meghatározzuk a disziálsav és a Neu5,9Ac₂ kapcsolódási helyét a cukorszerkezeten belül.

4.3.1.1 Disziálsavat tartalmazó glikánok

A Neu5,9Ac₂-t tartalmazó O-glikánnal módosított glikopeptidek és a disziálsavat tartalmazó core 1-es O-glikánnal módosított glikopeptidek között átfedés észlelhető (8. táblázat). A disziálsav terminális egysége lehet "normál" NeuAc, Neu5,9Ac₂ vagy Neu4,5,9Ac₃.

Disziálsavat tartalmazó O-glikánra utalnak az EThcD spektrumban megfigyelhető m/z 583,197, illetve m/z 625,210 ionok, amelyek a $(\text{NeuAc})_2$, illetve NeuAcNeu5,9Ac_2 struktúráknak felelnek meg. Egy pentaszachariddal módosított peptid EThcD spektruma látható a 19. ábrán. A peptidet módosító glikán három “normál” NeuAc -t hordoz. Ezt a $(\text{NeuAc})_2$ és a HexNeuAc lehasadását jelző Y ionok jelzik (m/z 1156,04 (2+), 1220,56 (2+)). Ezek az ionok arról is tájékoztatnak, hogy a glikán elágazó és a disziálsav egység a GalNAc -hoz kötődik. A két szílsav egymáshoz kapcsolódását a disziálsav oxónium ionja támasztja alá (m/z 583,197). A teljes cukorszerkezetet reprezentáló oxónium ion m/z 1239,41-nél megerősíti, hogy a peptidet egyetlen O-glikán módosítja.

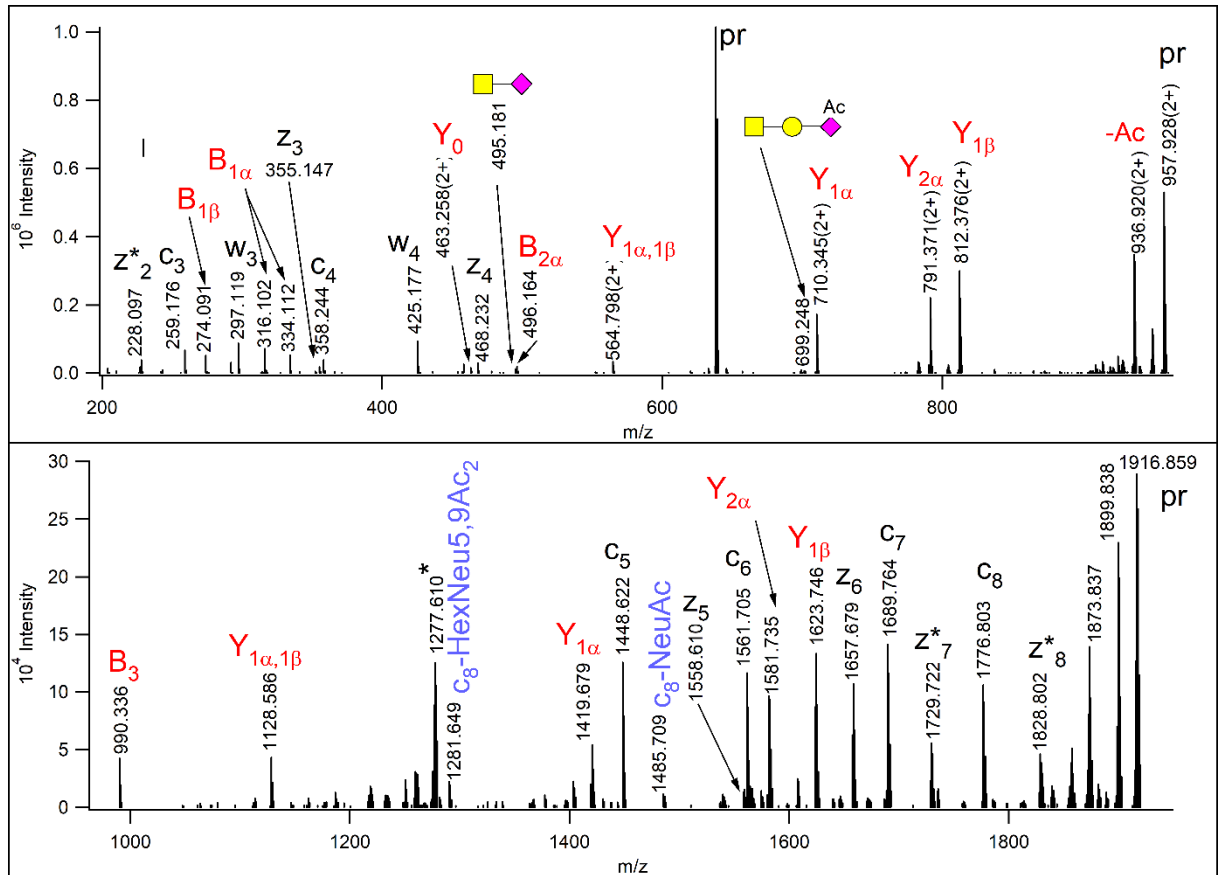


19. ábra m/z 965,092 (3+) perkurzorról készült EThcD spektrum, amely DVSTPPT(HexNAcHexNeuAc₃)VLPDNFPR glikopeptidként lett azonosítva. A prekursor iont és töltésvesztett formáját „pr” jelöli. Az ábra bal felső sarkában bemutatott cukorszerkezeten illusztrálom a fragmentációt, az intakt glikán oxónium ionját az ábrában piros csillag jelzi. A fekete csillaggal jelölt ion egy kétszeres töltésű koeluláló molekula töltésvesztett formája.

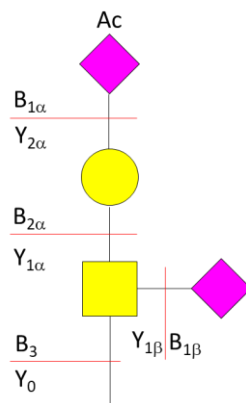
4.3.1.2 O-acetil-neuraminsavat tartalmazó glikánok

Az EThcD adatok alapján a Neu5,9Ac_2 mindig terminális pozícióban található, amelyet alátámasztanak a töltésvesztett prekursorokból Neu5,9Ac_2 veszttel keletkező Y ionok. Ez tulajdonképpen várható is volt, mert a módosítás az amúgy lánczáró NeuAc maszkírozására szolgál [Varki, 2015]. Az egy Neu5,9Ac_2 -t tartalmazó core 1-es tetraszacharidok esetében a diacetilált neuraminsav a galaktózhoz kapcsolódott, amelyet bizonyított az EThcD spektrumban

található, HexNeu5,9Ac₂ jelenlétét jelző B₂ ion (m/z 496,166) és az ezzel komplementer Y_{1α} ion (m/z 710,345 (2+)) (20. ábra, 1. séma). Az intakt tetraszacharidot is detektáltuk (m/z 990,336).

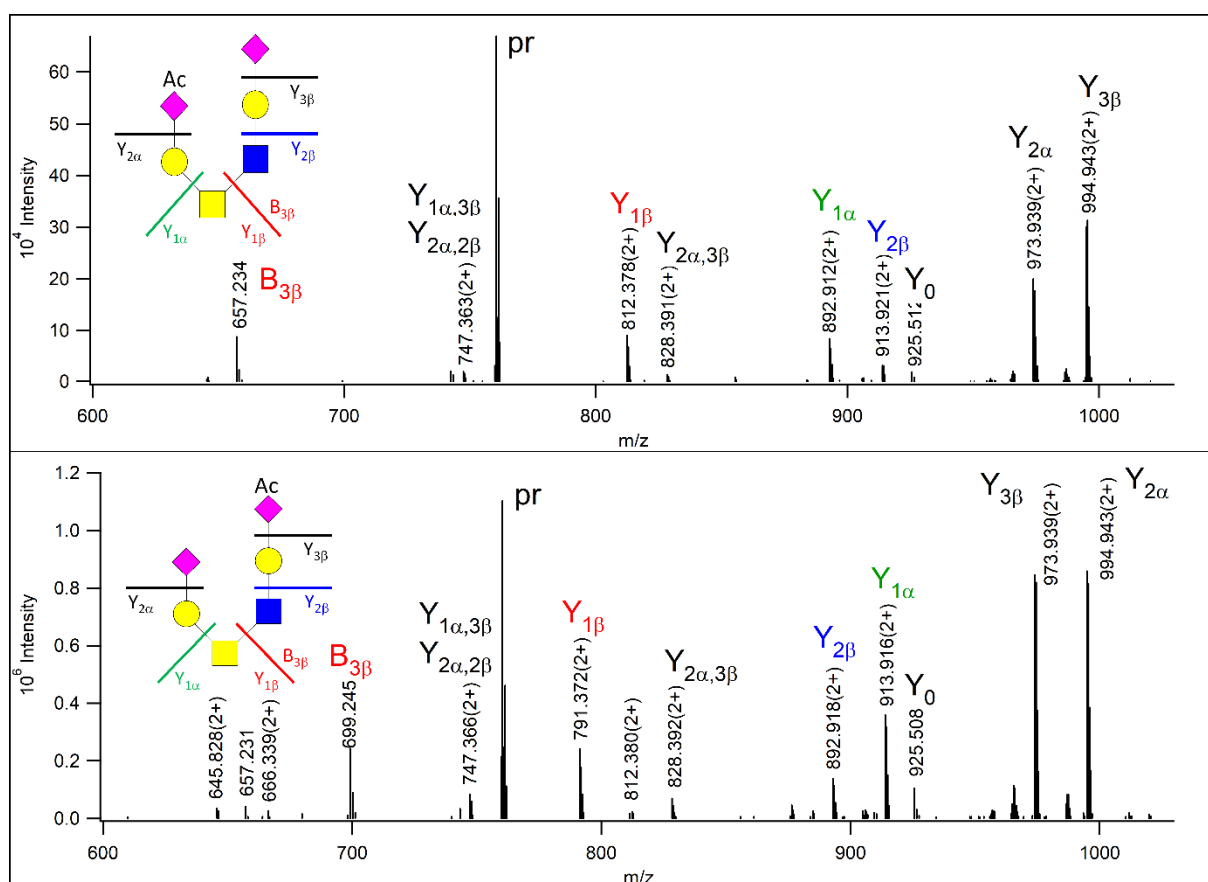


20. ábra m/z 638,954 (3+) prekuzorról készült EThcD spektrum, amely AVAVT(HexNAcHexNeuAcNeuAcAc)LQSH glikopeptidként került azonosításra. A Neu5,9Ac₂ a galaktózt módosítja (1. séma). A hidrogénmigrációval keletkezett z+1 peptidfragmensek az ábrán csillaggal jelöltek.



1. séma A 20. ábrán bemutatott, EThcD spektrumból azonosított AVAVTLQSH peptidet módosító glikán fragmentációját illusztráló séma.

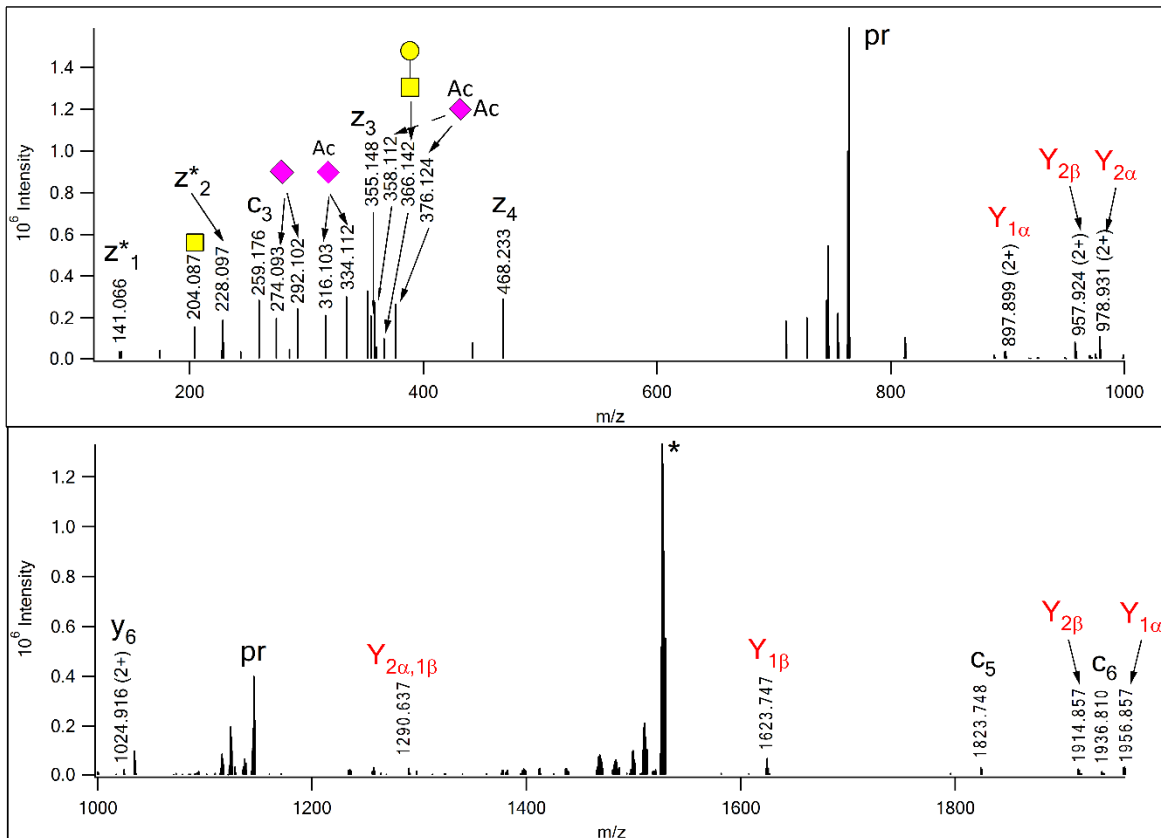
Az egy Neu5,9Ac₂-t tartalmazó diszialo core 2-es hexaszacharid esetében a Neu5,9Ac₂ a glikán mindkét karját lezárhatja. Ezt az esetet szemlélteti a 21. ábra. Az adatbáziskeresés során a Protein Prospector megegyező peptidszekvenciával és cukorösszetétellel azonosította az m/z 760,664 (3+) prekuzorról két különböző időpontban felvett EThcD spektrumot: a Protein YIPF3 (Q9GZM5) fehérje AVAVT(HexNAc₂Hex₂NeuAcNeuAcAc)LQSH peptidjeként. Mindkét spektrumban megtalálható a hexaszacharid oxónium ionja (m/z 1355,46, 8. melléklet). A Neu5,9Ac₂ mindkét esetben galaktózhoz kötött, amelyet az m/z 496,166 ion bizonyít (8. melléklet). A két, eltérő időpontban felvett EThcD spektrumot jobban szemügyre véve kiderül, hogy az azonosság csak a peptid szekvenciájára, a módosítás helyére és a glikán összetételére érvényes, viszont az oligoszacharidok szerkezete különböző. A korábban eluálódó, intenzívebb glikoforma esetében a Neu5,9Ac₂ a GlcNAcGal karon található (a 21. ábra alsó paneljén β karként jelölve).



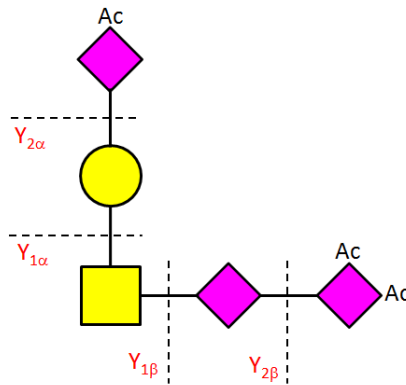
21. ábra m/z 760,664 (3+) prekuzorról készült EThcD spektrum részlet 14,95 percnél (alsó panel), illetve 18,29 percnél (felső panel). Mindkét spektrum a Protein YIPF3 fehérje AVAVT(HexNAc₂Hex₂NeuAcNeuAcAc)LQSH peptidjeként lett azonosítva. A spektrum manuális ellenőrzése izomer O-glikán szerkezetek jelenlétét tárta fel. Az izomer szerkezetek megkülönböztetésében diagnosztikus értékkel bíró fragmensek és a mögöttük meghúzódó kötéshasadásokat azonos színnel jelöltem.

Ez a kar egyszeres kötés-hasadással eltávolítható a hexaszacharidról, amit a spektrumban az m/z 699,245-nél ($B_{3\beta}$) detektálható HexNAcHexNeuAcAc oxónium ion jelez. Ennek az oxónium ionnak a komplementer párja ($Y_{1\beta}$) is látható a spektrumban (m/z 791,372 (2+)). Ezek az ionok a később eluálódó glikoformáról felvett EThcD spektrumban (21. ábra, felső panel) hiányoznak. Ebben az esetben az O-acetil-sziálsav az α kart zárja le. Így amikor a GalNAc és a GlcNAc közötti kötés elhasad, egy intenzív $B_{3\beta}$ ion keletkezik m/z 657,234-nél. Az ennek a B ionnak megfelelő komplementer $Y_{1\beta}$ -t is detektáltuk (m/z 812,378 (2+)). A két spektrumban található többi Y fragmens glikánkompozíciója megegyező. Az $Y_{1\alpha}$ és az $Y_{2\beta}$ fragmensek relatív intenzitása a két glikoformáról készült EThcD spektrumban eltérő. Az adatok alapján feltételezhető, hogy a GalNAc és a Gal közötti kötés hasadása preferáltabb, mint a GlcNAc és a Gal közötti kötésé.

Egy EThcD spektrum adatainak értelmezésekor a Protein Prospector a peptiden olyan pentaszacharidot azonosított, amelyben három O-acetil-sziálsav található. Habár a peptidet módosító glikán additív tömege megfelel a fenti összetételnek, ez a meghatározás mégis egészen valószínűtlennek tűnik, mert ez azt jelentené, hogy a peptidet módosító O-glikánban nemcsak a terminális sziálsavakat módosította az acetil-transzferáz enzim. Az EThcD spektrum alapos ellenőrzése három különböző sziálsav jelenlétét támasztja alá (22. ábra, 2. séma). A galaktózt egy Neu5,9Ac₂ zárja, míg a GalNAc-hoz először egy "normál" NeuAc kapcsolódik, amelyet egy két „extra” acetil-csoporttal rendelkező NeuAc zár. Ez a szerkezet nagy valószínűséggel egy O-4,9-diacetil-N-5-acetil-neuraminsav (Neu4,5,9Ac₃) [Varki, 2015]. Jelenlétét a belőle képződő oxónium ion jelzi m/z 376,124-nél illetve ennek vízvesztett párja (m/z 358,112) is megfigyelhető. A Neu4,5,9Ac₃ terminális pozícióját és azt, hogy normál NeuAc-hoz kötődik, az egymás utáni glikozidos kötés-hasadásból származó $Y_{2\beta}$ és $Y_{1\beta}$ fragmensek igazolják. A HexNeu5,9Ac₂ terminális struktúra jelenlétére a $Y_{1\alpha}$ fragmens szolgál bizonyítékként (m/z 897,899 (2+)). Ez egyben egyértelművé teszi, hogy a diszialil egység a GalNAc-on foglal helyet.



22. ábra m/z 763,992 (3+) prekuzorról készült EThcD spektrum, amelyet manuális adatkiértékeléssel AVAVT(HexNAcHexNeuAcNeuAcAcNeuAcAcAc)LQSH glikopeptidként azonosítottunk. A peptid a Thr-346-os pozícióban core 1-es típusú O-glikánnal módosított, amelyben a Gal Neu5,9Ac₂, míg a GalNAc NeuAcNeu4,5,9Ac₃ kötött. A GalNAc módosításai a glikán β karjának felelnek meg. A csillaggal jelzett ion egy kétszeres töltésű koeluálódó molekula töltésvesztett formája.



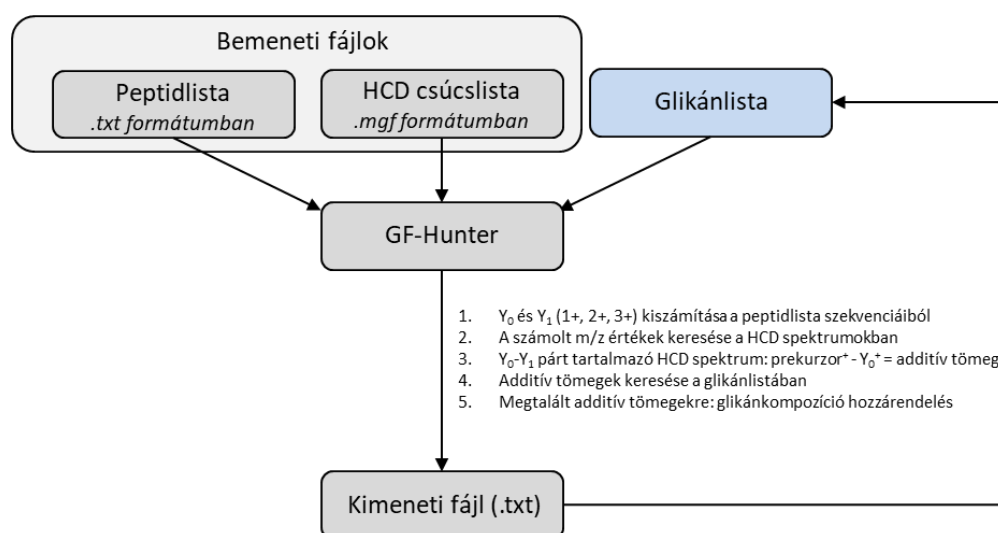
2. séma A 22. ábrán bemutatott EThcD spektrumból azonosított AVAVTLQSH peptidet módosító glikán fragmentációját illusztráló séma, amely három különböző szialavat tartalmaz.

4.4 További új O-glikán struktúrák azonosítása a HCD csúcslisták szűrésével

A HCD adatoktól függő EThcD adatgyűjtési stratégia az eddigi legteljeskörűbb módszer a glikopeptid tandem tömegspektrometriás jellemzésére. Az ezidáig bemutatott eredmények az EThcD spektrumban foglalt strukturális információk bemutatására fókuszáltak, amelyekből

optimális esetben meghatározható a peptid szekvenciája és a módosítás pontos helye, köszönhetően az ETD aktivációnak. Az alacsony NCE értékkel (15 %) végzett kiegészítő HCD aktiválás főként olyan B és Y fragmensek keletkezését eredményezi, amelyek egyszeres kötésű hasadási termékek, így a glikánról strukturális információt szolgáltatnak. Csak az automatizált adatkiértékelésre hagyatkozva azonban csak olyan O-glikopeptideket azonosíthatunk, amelyek a keresés során egy előre meghatározott glikán adatbázisban szereplő struktúrákkal módosítottak. Tehát a cél a glikánadatbázisnak a mintához történő „finomhangolása”. Ezt pedig úgy tehetjük meg, hogy a HCD adatokat is be vesszük az adatkiértékelés folyamatába. A magasabb NCE értékkel (28 %) felvett HCD spektrumokban nagyobb fokú glikán- és emellett peptid fragmentáció is megfigyelhető. Egy ilyen HCD spektrumban csaknem mindig megjelenik a gázfázisban deglikozilált peptid ion (Y_0), amely értékes információforrás. Ennek és a prekursor ion m/z értékének és töltésének ismeretében kiszámítható a „csupasz” peptidet módosító csoport(ok) additív tömege. Az így meghatározott tömegekből új cukorösszetételeket azonosíthatunk, amelyek új glikán struktúrákat vagy több glikán kombinációját képviselik. Ahhoz, hogy az additív tömegeket meg tudjuk határozni, ismernünk kell és meg kell találnunk a HCD spektrumokban az Y_0 iont. Az új módosítások felderítéséhez az alapot a leggyakoribb glikánokkal megbízhatóan azonosított glikopeptidek szolgáltatják. Ezen peptidok listájához a leggyakoribb glikán-szerkezetekkel végzett adatbázis kereséssel juthatunk (3.3.5 fejezet). A további feladatok elvégzéséhez pedig szoftvert fejlesztettünk (GF-Hunter, 23. ábra), amely a HCD adatokban Y_0 - Y_1 párokat keres azonos töltésállapotban, adott tömegpontosságán belül és egy megadott intenzitásküszöb felett. A két fragmens együttes jelenlétét előzetes tesztelési eredményeink emelték szükségszerű követelménnyé: ha csupán az Y_0 jelenlétéhez ragaszkodtunk, a módszer nem volt eléggé szelektív. Hasonlóképpen optimalizálni kellett a többi paramétert is, amelyet egy LC-MS kísérlet adataival végeztünk, gondosan mérlegelve mit nyerünk, vagy veszítünk a változtatások következményeként. Az optimalizáció eredményeként a HCD adatokban az Y_0 - Y_1 párnak 10 ppm tömegpontossággal a 30 legintenzívebb csúcs között kellett szerepelnie. A feltételeket teljesítő spektrumokból a szoftver meghatározza a peptidet módosító csoport(ok) additív tömegét a prekursor ion és a detektált Y_0 ion különbségként. Ha ez az érték 10 ppm tömegpontosságán belül egy adott glikán összetétel elméleti tömegével megegyezik, akkor ezt a szoftver jelzi. Emellett listázza az összes olyan HCD spektrumot is, amelyeknél nem sikerült valamely előre definiált glikán kompozícióhoz rendelni a meghatározott additív tömeget. Ezek az ismeretlen tömegértékek új glikánstruktúrákat jelenthetnek. Az első glikánlistát korábbi tapasztalataink alapján a leggyakoribb mucin-típusú szerkezetekből állítottuk össze, és folyamatosan bővítettük újabb szerkezetek azonosítása után. Sajnos az additív tömegek

automatikus cukor-szerkezetté fordítása nem oldható meg egyszerűen, így pillanatnyilag csak iteratív módon és manuálisan bővíthetjük a figyelembe veendő O-glikánok és azok kombinációinak listáját.

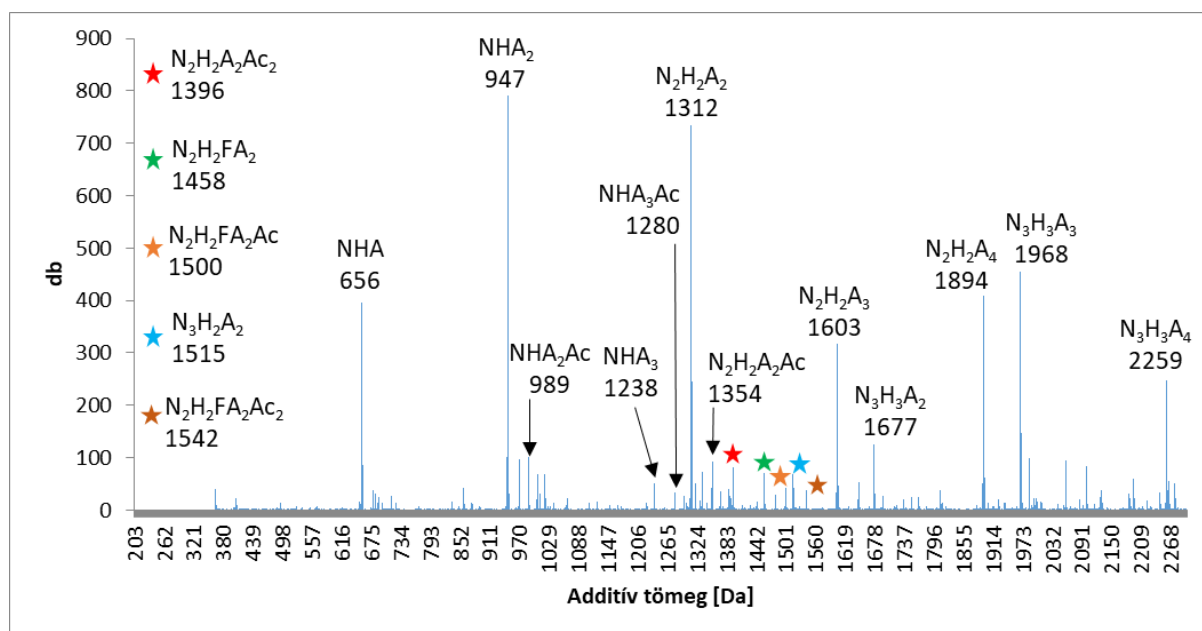


23. ábra A GF-Hunter egy peptidlista segítségével a HCD csúcslistákban Y_0 - Y_1 párokat keres. Az így elkülönített spektrumokra kiszámolja a prekursor ion és a megtalált Y_0 ion tömegének különbségét majd ezt a beépített glikánlistában szereplő elméleti tömegértékekhez hasonlítja. A glikánlista bővíthető az eredményekből meghatározott új glikán-összetételekkel.

A szoftver teszteléséhez használt adatok 10 férfi donor vizeletéből dúsított glikopeptidegyek tömegspektrometriás analiziséből származtak. Az LC-MS/MS analizist a páciensenként gyűjtött WGA affinitás kromatográfia áteső csúcs vége (FTEnd), váll (Shoulder), illetve a GlcNAc eluált frakciókkal végeztem, összesen 60 LC-MS/MS adat sort analizáltam. Az elsődleges adatbázis keresésben a humán szérumban leggyakrabban előforduló 4 O-glikánt vettük csak figyelembe (HexNAcHex, HexNAcHexNeuAc₁₋₂ és HexNAc₂Hex₂NeuAc₂). A megbízhatónak tekinthető eredményekből elkészítettük a GF-Hunter input peptid listáját, amely tartalmazza az azonosított peptideket cukorszerkezet(ek) nélkül, de minden más kovalens módosítással (Met oxidáció, C karbamidometilzés, Peptid N-terminális acetiláció, Peptid N-terminális Gln piroglutaminsavvá alakulása). A szoftver tesztelésének első fázisában a glikán adatbázis a legismertebb alapszerkezeteket, azok kettős-hármas kombinációit és a 4.3 fejezetben ismertetett diszialo- és Neu5,9Ac₂-tartalmú glikánokat tartalmazta. Ez a lista azonban jelentősen kibővült a tesztelés eredményeinek értelmezése után.

A GF-Hunter minden mintát figyelembe véve összesen 13764 HCD spektrumban talált a bemeneti listán lévő peptidhez tartozó Y_0 - Y_1 párt (5. melléklet). Ebből összesen 4387 spektrumhoz sikerült glikánösszetételt is rendelni (Na addukt képződést is figyelembe véve) a számolt additív tömeghez. Korábbi eredményeinkkel összhangban [Pap, 2018; Darula, 2019] a

leggyakoribb additív tömegek a már jól ismert mono- és diszialo core 1-es és a diszialo core 2-es szerkezeteknek feleltethetők meg (24. ábra). Ezen szerkezetek kombinációinak megfelelő additív tömegek (1603, 1894, 1968, 2259) is gyakoriak, amelyek főként hosszú, több potenciális módosítási helyet tartalmazó peptideken fordulnak elő. A 24. ábrán a leggyakoribb, közönséges O-glikánoknak megfelelő additív tömegek mellett megjelennek O-acetilált szialosavat (989, 1354, 1396, 1500, 1542), diszialsavat (1238) és a kettő kombinációját (1280) jelző tömegek is. Emellett az eredmények kiértékelése során azonosítottunk Fuc-tartalmú O-glikánokat, ezek között leginkább vércsoport antigént tartalmazó struktúrákat is.



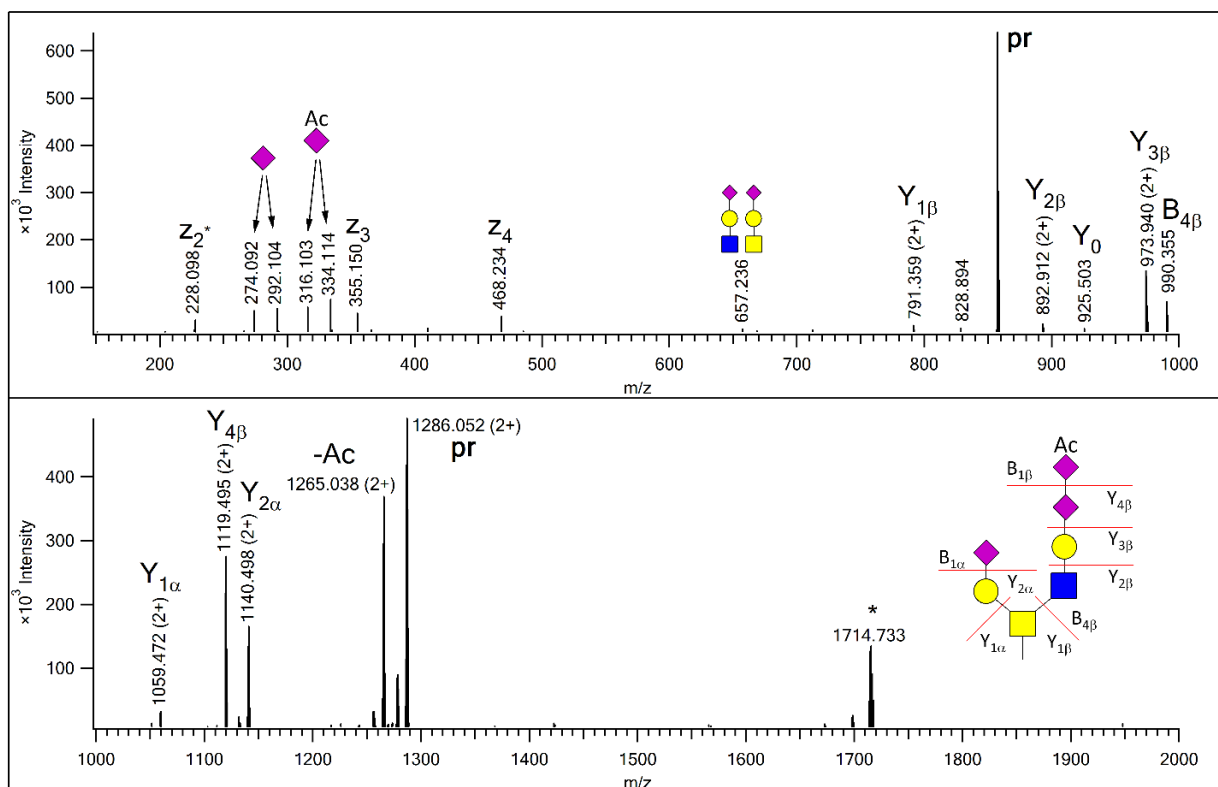
24. ábra Összesítő ábra a GF-Hunterrel detektált módosítások additív tömegének eloszlásáról. Az ábrán a csúcsfeliratok az additív tömeg felett az ahhoz rendelt glikán összetételt jelenítik meg. N: HexNAc, H: Hex, F: Fuc, A: NeuAc, Ac: acetyl-csoport

A GF-Hunter eredmények rávilágítanak arra is, hogy a WGA lektin affinitás kromatográfia során az áteső csúcs végi (FTend) frakcióban O-glikopeptidek kevesebb számban találhatóak, habár a leggyakoribb O-glikán struktúrákkal (HexNAc₁Hex₁NeuAc₁ és HexNAc₁Hex₁NeuAc₂) egyszeresen módosított O-glikopeptidek már itt is megjelennek (~25%). Azonban az ezeknek a struktúráknak megfelelő additív tömeggel „azonosított” HCD spektrumok ~50%-a a Shoulder, ~25%-a pedig a GlcNAc frakcióban található (9. táblázat). A HexNAc₂Hex₂NeuAc₂ O-glikánnal módosított peptidek a Shoulder (56%) és GlcNAc (36%) frakciók között oszlanak meg, az FTend frakcióban előfordulásuk elenyésző (9. táblázat). A kombinációknak megfelelő struktúrák szinte kizárólag a GlcNAc frakcióban találhatóak meg.

9. táblázat A leggyakoribb O-glikánoknak és azok kombinációinak megfelelő összetételek előfordulása a WGA lektin affinitás kromatográfia során gyűjtött frakciókban.

	GlcNAc	Shoulder	FTEnd	Összesen
GF-Hunter által „azonosított” HCD spektrumok összesen	6812	4838	2114	13764
<i>Leggyakoribb O-glikánok</i>				
HexNAc(1)Hex(1)NeuAc(1)	73	139	74	286
HexNAc(1)Hex(1)NeuAc(2)	178	324	158	660
HexNAc(2)Hex(2)NeuAc(2)	236	373	54	663
<i>Kombinációik</i>				
HexNAc(2)Hex(2)NeuAc(3)	135	98	3	236
HexNAc(2)Hex(2)NeuAc(4)	263	84	1	348
HexNAc(3)Hex(3)NeuAc(3)	334	19	0	353
HexNAc(3)Hex(3)NeuAc(4)	187	4	0	191
HexNAc(3)Hex(3)NeuAc(5)	31	0	0	31
HexNAc(3)Hex(3)NeuAc(6)	10	0	0	10
HexNAc(4)Hex(4)NeuAc(1)	12	1	0	13
HexNAc(4)Hex(4)NeuAc(2)	9	2	0	11
HexNAc(4)Hex(4)NeuAc(3)	9	3	0	12
HexNAc(4)Hex(4)NeuAc(4)	36	0	0	36

A szoftver segítségével felfedezett új O-glikán szerkezeteket a Protein YIPF3 fehérje AVAVTLQSH peptidjén mutatom be. Az adatok alapján ez a peptid minden minta jelentős komponense és csak egyetlen O-glikánnal módosított (bár a peptidben két potenciális O-glikozilációs hely is található), amely mindig a 346-os Thr-hez kötődik. A tíz donor HCD adataiból keletkezett GF-Hunter eredményeket összesítve összesen 1878 spektrumban találtuk meg az AVAVTLQSH peptid Y₀-Y₁ párját. Ebből 876 spektrumhoz rendelhető egyértelmű glikán kompozíció is a megállapított additív tömeg alapján (5. melléklet). A legtöbb esetben ezek az additív tömegek egyedi oligoszacharid struktúrákat képviselnek, de izomer struktúrákat is azonosítottunk a glikán-fragmentáció alapján (ld. 4.3.1.2 fejezet 21. ábra). Előzőleg (4.3 fejezet) beszámoltam olyan O-glikán szerkezetekről, amelyek vagy disziálsavat, vagy O-acetilált N-acetil-neuraminsavat tartalmaztak. Az új, HCD adatokat is kiaknázó módszerrel sikerült az AVAVTLQSH peptid olyan glikoformáját azonosítani, amely mindkét végcsoportot tartalmazza egy core 2-es alapszerkezeten (25. ábra).

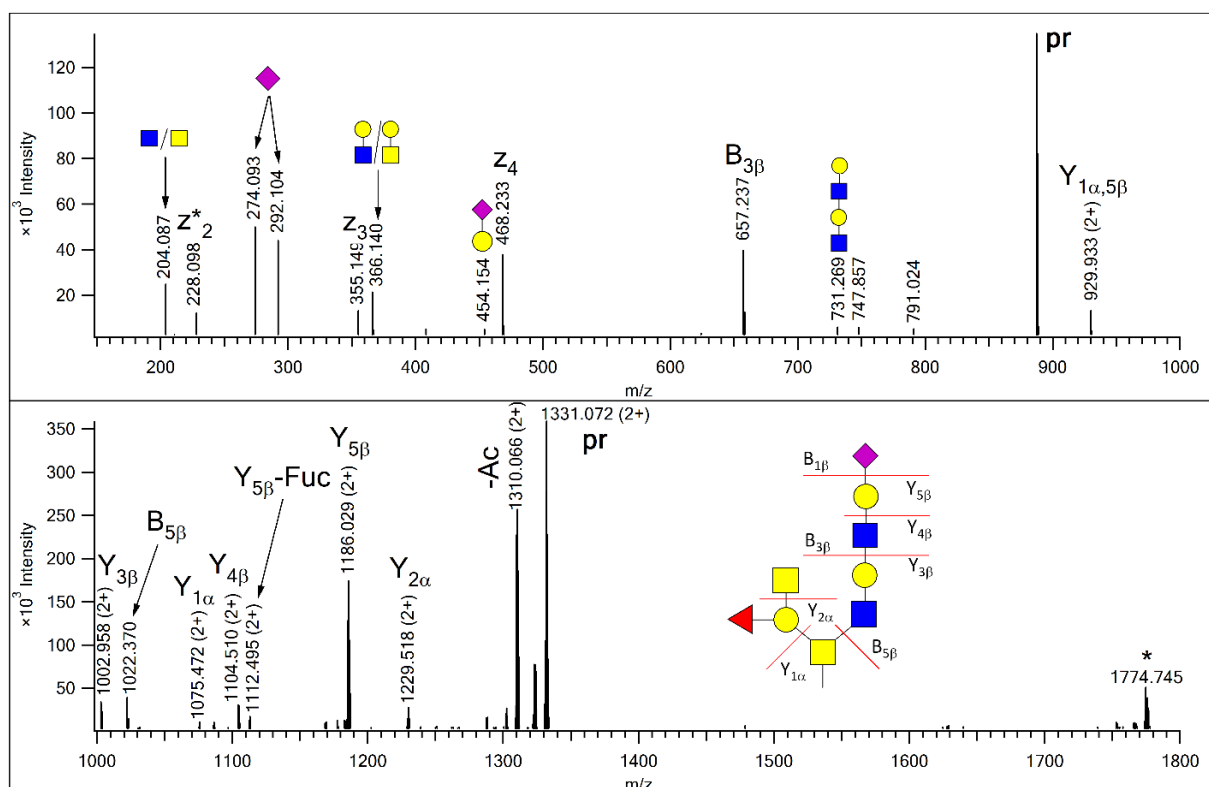


25. ábra m/z 857,695 (3+) prekurzorról készült EThcD spektrum. HCD adatokból a GF-Hunter az AVAVTLQSH peptid szekvenciát és a HexNAc₂Hex₂NeuAc₂NeuAcAc cukor-összetételt azonosította. A csillaggal jelölt ion egy koelualódó molekula töltésvesztett formája.

A peptidet módosító O-glikán GlcNAc karján található a Neu5,9Ac₂-t tartalmazó disziálsav, amelyet a B_{4β} fragmens (m/z 990,355, 25. ábra) bizonyít. Az O-acetilált disziálsav terminális helyzetét a GlcNAc karon belül az Y_{3β} fragmens (m/z 973,940 (2+), 25. ábra) bizonyítja, míg a Neu5,9Ac₂ terminális pozícióját a disziálsav struktúrában belül a Y_{4β} fragmens támasztja alá (m/z 1119,495 (2+), 25. ábra). A Neu5,9Ac₂ jelenlétét továbbá igazolja az m/z 334,114-es oxónium ion és vízvesztéssel keletkezett párja is (m/z 316,103). A módosítatlan z₂-z₄ ionok egyértelműen jelzik, hogy csak a Thr-346 glikozilált; az elágazó cukor szerkezetet az Y_{1α} és Y_{1β} ionok bizonyítják. Az O-acetilált disziálsav jelenlétére utaló m/z 625,209 iont (ld. 4.3.1.1 fejezet) és az intakt glikánt képviselő B₆ fragmenst (m/z 1646,568) nem detektáltuk.

A vizeletmintákból azonosított Fuc-tartalmú oligoszacharidok zöme AB0 vércsoport antigénnel lezárt struktúrákat képviselt. A vércsoport antigének nem csak a vörösvértestek felszínén találhatóak, hanem a szekretált fehérjéken, illetve szabad glikánokként is előfordulnak, amennyiben az egyén pozitív szekretor státusszal rendelkezik [Schneider, 2017]. A vércsoport epitópok szintézise során a kiindulási laktóz-amin egységre egy fukóz kerül (H-antigén), amelyet egy fukozil-transzferáz enzim végez. Ennek az enzimnek két alléja létezik. A FUT1 allél a vörösvértestekben aktív, míg a FUT2 allél az epitél sejtekben lehet aktív. Ha az egyén aktív

FUT2 alléllal rendelkezik, akkor a szekretált fehérjékre is kerülhet vércsoport antigén (A, B, H), illetve szabad glikánokként is előfordulhatnak ezen antigének a testfolyadékokban. Ilyen oligoszacharid szerkezetek jelenlétét kutatócsoportunkon kívül eddig még senki sem demonstrálta tömegspektrometriás adatokból intakt glikopeptideken. Core 2 alapszerkezeteken az összes lehetséges AB0 vércsoport antigént azonosítottuk: HexNAc₂Hex₂FucNeuAc (H antigén), HexNAc₃Hex₂FucNeuAc (A antigén, két izomer-szerkezetet jellemeztünk, a vércsoport epitóp kötődhet a GalNAchoz is, és a GlcNAchoz is), HexNAc₂Hex₃FucNeuAc (B antigén) (10. táblázat). Illusztrációként viszont egy A-antigént hordozó, meghosszabbított core 2 alapszerkezettel módosított glikopeptidet mutatok be (26. ábra).



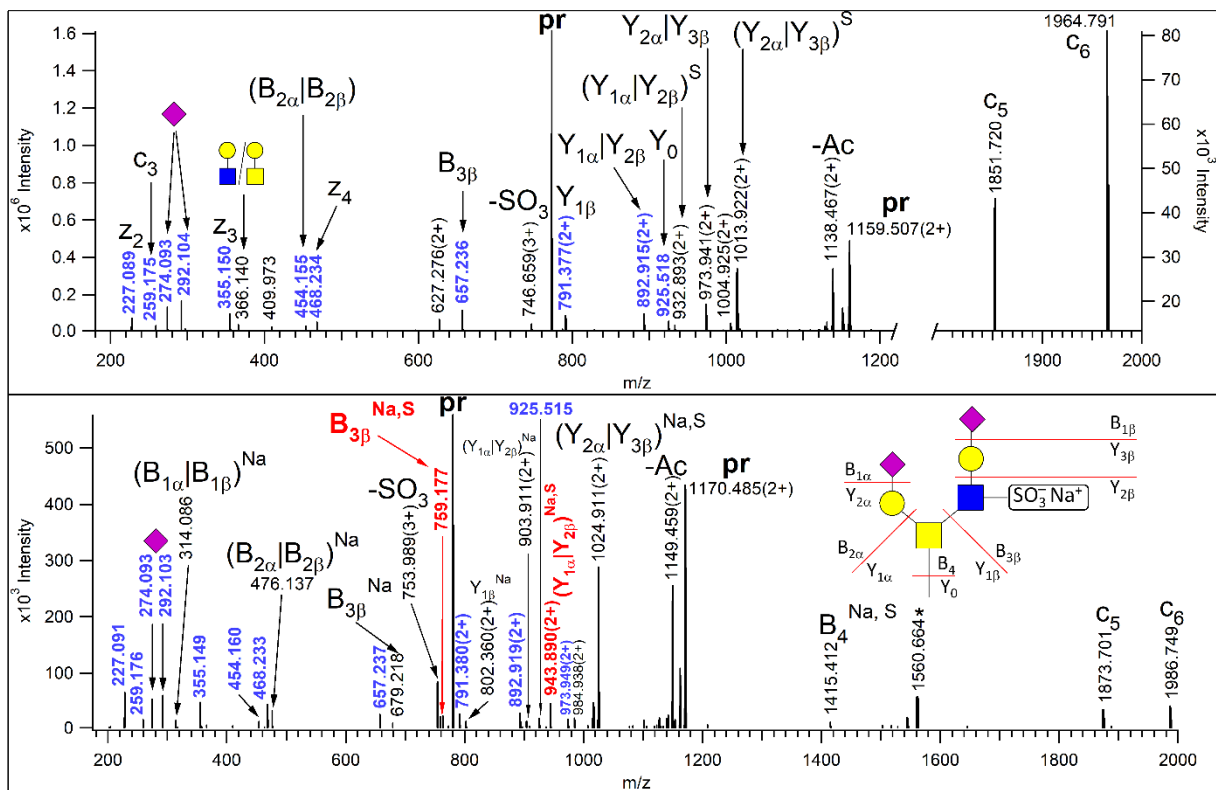
26. ábra m/z 887,718 (3+) prekuzorról készült EThcD spektrum. HCD adatokból a GF-Hunter az AVAVTLQSH peptidet és a HexNAc₄Hex₃FucNeuAc glikánt azonosította. A csillaggal jelölt ion egy koelválódó molekula töltésvesztett formája.

Az A antigén a GalNAc-hoz kapcsolódik, a másik (β) kar szerkezetét egyértelműen jelzik a B_{5β}, B_{3β} és B_{1β} ionok. Az antigén jelenlétét tovább erősíti Y_{2α} ion (m/z 1229,518 (2+)), amely a kétszeres töltésű töltésvesztett prekuzorból egy HexNAc veszteségre utal, azaz a HexNAc terminális pozícióját bizonyítja. A második legnagyobb felfedezett glikánstruktúra (glikán névleges tömeg: 2391 Da) egyik antennája ugyancsak A-antigént hordoz (6. melléklet, 10. táblázat).

10. táblázat A GF-Hunter segítségével azonosított, vércsoport epitópotkat tartalmazó O-glikánok. N: HexNAc, H: Hex, F: Fuc, A: NeuAc

Vércsoport	Monoizotópos additív tömeg	Monoszacharid összetétel	Szerkezet
O	1167,4177	N ₂ H ₂ F ₁ A ₁	
A	1370,4971	N ₃ H ₂ F ₁ A ₁	
	1735,6293	N ₄ H ₃ F ₁ A ₁	
	2391,8569	N ₅ H ₄ F ₁ A ₂	
B	1329,4705	N ₂ H ₃ F ₁ A ₁	

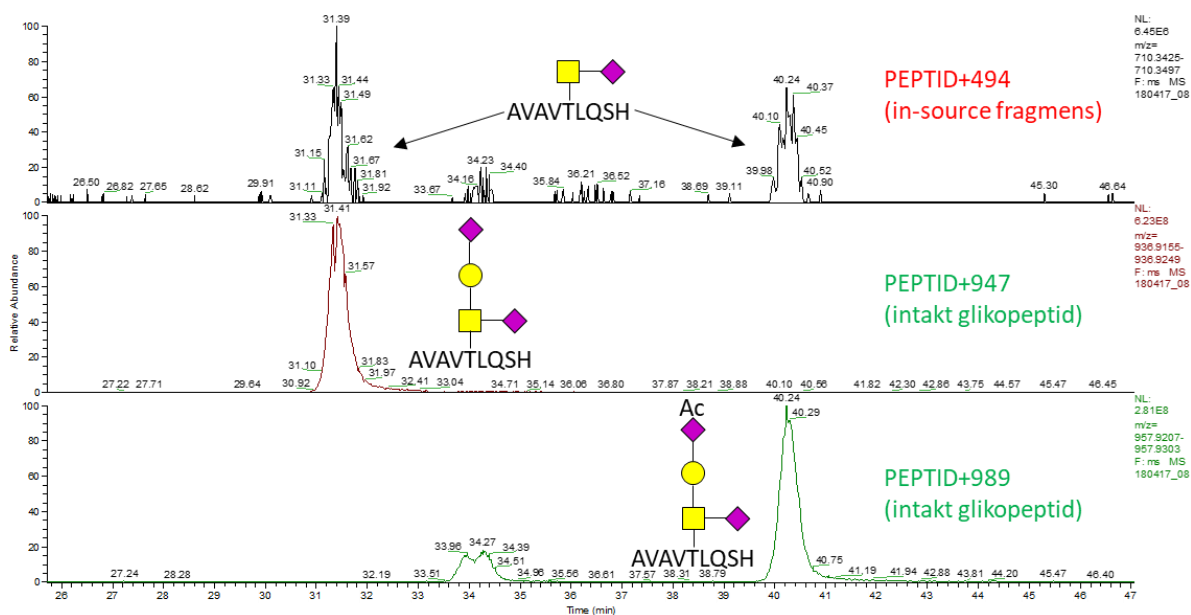
Végül, de nem utolsósorban szulfatált O-glikopeptidet is sikerült ezzel az új eljárással azonosítani. A peptid egy szulfatált, diszialo core 2-es O-glikánnal módosított (27. ábra, felső panel). A szulfo-csoport pozícióját nem sikerült meghatározni ebből az EThcD spektrumból, az aktiválás során SO₃ elimináció történt, amely nem hagy nyomot az előzőleg módosított cukron, akárcsak az O-glikán elimináció ütközéses aktiválás hatására. Ugyanennek a szulfatált O-glikopeptidnek azonban a Na-adduktjáról is készült EThcD spektrum (27. ábra, alsó panel). A 27. ábra felső paneljén az Y_{2β} és B_{3β} fragmenseket csak módosítás nélkül detektáltuk, míg az alsó panelen ugyanezek az ionok a Na-addukt képződés miatt részben megtartották a szulfo-csoportot. Így a szulfatálás helye egyértelműen a glikán β karjában lévő GlcNAc.



27. ábra m/z 773,315 (3+) (felső panel) és m/z 780.640 (3+) (alsó panel) prekursorokról készült EThcD spektrumok. HCD adatok alapján a glikopeptidek az AVAVTLQSH peptid különböző glikoformái, a felső panelen a glikán összetétele HexNAc₂Hex₂NeuAc₂Sulfo, az alsó panelen ugyanezen glikán nátrium adduktja. Az alsó panelen a csillaggal jelölt ion egy koelválódó molekula töltésveszített formája. A kékekkel jelölt ionokat a felső panelen annotáltam. A pirossal jelölt ionok bizonyítják a szulfát glikánon belüli pozícióját. Ha egy ion többféleképpen is annotálható, akkor a lehetséges annotációkat függőleges vonal választja el.

Az adatok elemzése során olyan glikoformákat is azonosítottunk, amelyek valószínűleg in-source fragmentáció termékei. A különböző savasságú komponensek fordított fázisú kromatográfiás elválasztás során nem elválódnak együtt, amennyiben az ionpár-képző hangyasav [Trinidad, 2013; Medzihradszky, 2015 (A)], tehát ha különböző számú szialsavat hordozó glikoformákat egyidejűleg detektálunk, akkor ez in-source fragmentációra utal. A 494, 859, 901, 1021, 1063 Da nominál tömegek rendre HexNAcNeuAc, HexNAc₂HexNeuAc, HexNAc₂HexNeuAcAc, HexNAc₂Hex₂NeuAc, HexNAc₂Hex₂NeuAcAc összetételt jeleznek és ezek a glikoformák a HexNAcHexNeuAc₂ (947), HexNAcHexNeuAcNeuAcAc (989), HexNAc₂Hex₂NeuAc₂ (1312), HexNAc₂Hex₂FucNeuAc₂ (1458), HexNAc₂Hex₂NeuAcNeuAcAc (1354) szerkezeteket hordozó glikopeptidek mellett jelennek meg, tehát feltehetően ezen nagyobb molekulákból képződtek az ionforrásban lezajló ütközések hatására. Ezt az esetet szemlélteti a 28. ábra. A HexNAcNeuAc glikán szerkezetet hordozó AVAVTLQSH peptidet két retenciós idő közelében (31,4 és 40,2 perc) is detektáltuk (28. ábra felső panel), melyek megegyeznek ugyanezen peptid különböző tetraszacharidokkal (HexNAcHexNeuAc₂, 28. ábra középső panel és HexNAcHexNeuAcNeuAcAc, 28. ábra alsó

panel) módosított glikoformáinak. A HexNAcNeuAc szerkezet mindkét tetraszacharidból képződhet, a megfelelő glikoforma párok koelúciója pedig egyértelműen in-source fragmentációra utal.

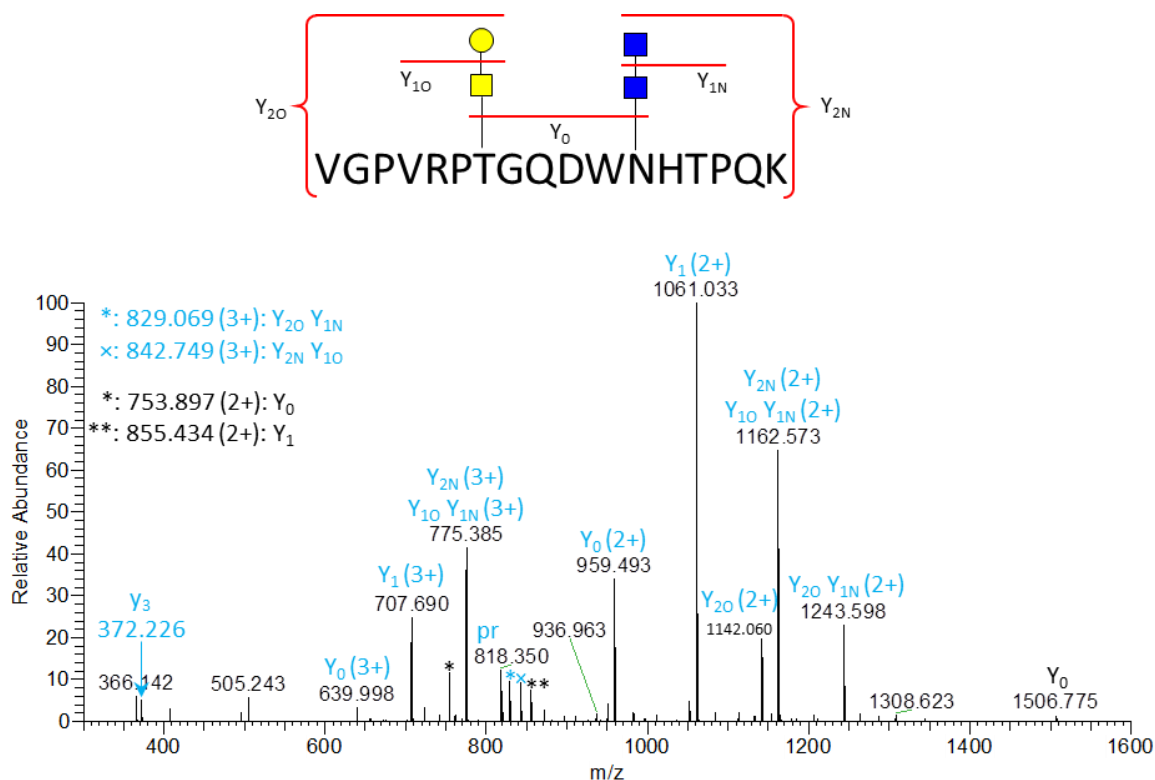


28. ábra Az in-source fragmentációval keletkezett glikopeptidek együtt eluálódnak a fragmentálódó glikoformával.

Sikereink ellenére a GF-Hunter bemeneti listáján található peptidekhez rendelt spektrumok nagyjából kétharmadánál (9377 spektrum) a számolt additív tömeget nem sikerült automatikusan glikán összetételhez rendelni. A szoftver nagy számban azonosított olyan additív tömegeket, amelyek 2 Da-os ablakban helyezkednek el a megbízhatóan azonosított O-glikán struktúrák additív tömegei körül. Miután a glikopeptidek rendszerint nagy molekulatömegű vegyületek, pontos tömegük meghatározása nem egyszerű feladat, hiszen a rendszerint nem túl intenzív, többszörösen töltött prekursor ion monoizotópos csúcsa gyakran beleolvad a háttérbe. A csúcslista-készítő szoftverek emiatt nem ritkán hibáznak a monoizotópos csúcs kiválasztásánál. A hibás monoizotópos tömeg meghatározás a pontos glikoformák meghatározásakor is problémát okozhat: Például a 1313,4711 Da tömeg a HexNAc₂Hex₂Fuc₂NeuAc glikán összetételnek felel meg, ez ~1 Da-nal nagyobb a 1312,4552 Da tömegnél, amely a HexNAc₂Hex₂NeuAc₂ összetételnek feleltethető meg. Az adatok ellenőrzése során kiderült, hogy hibás monoizotópos csúcs kiválasztás eredménye a 1313-as additív tömeg, és minden ilyen esetben a core-2 hexaszacharid a helyes szerkezet.

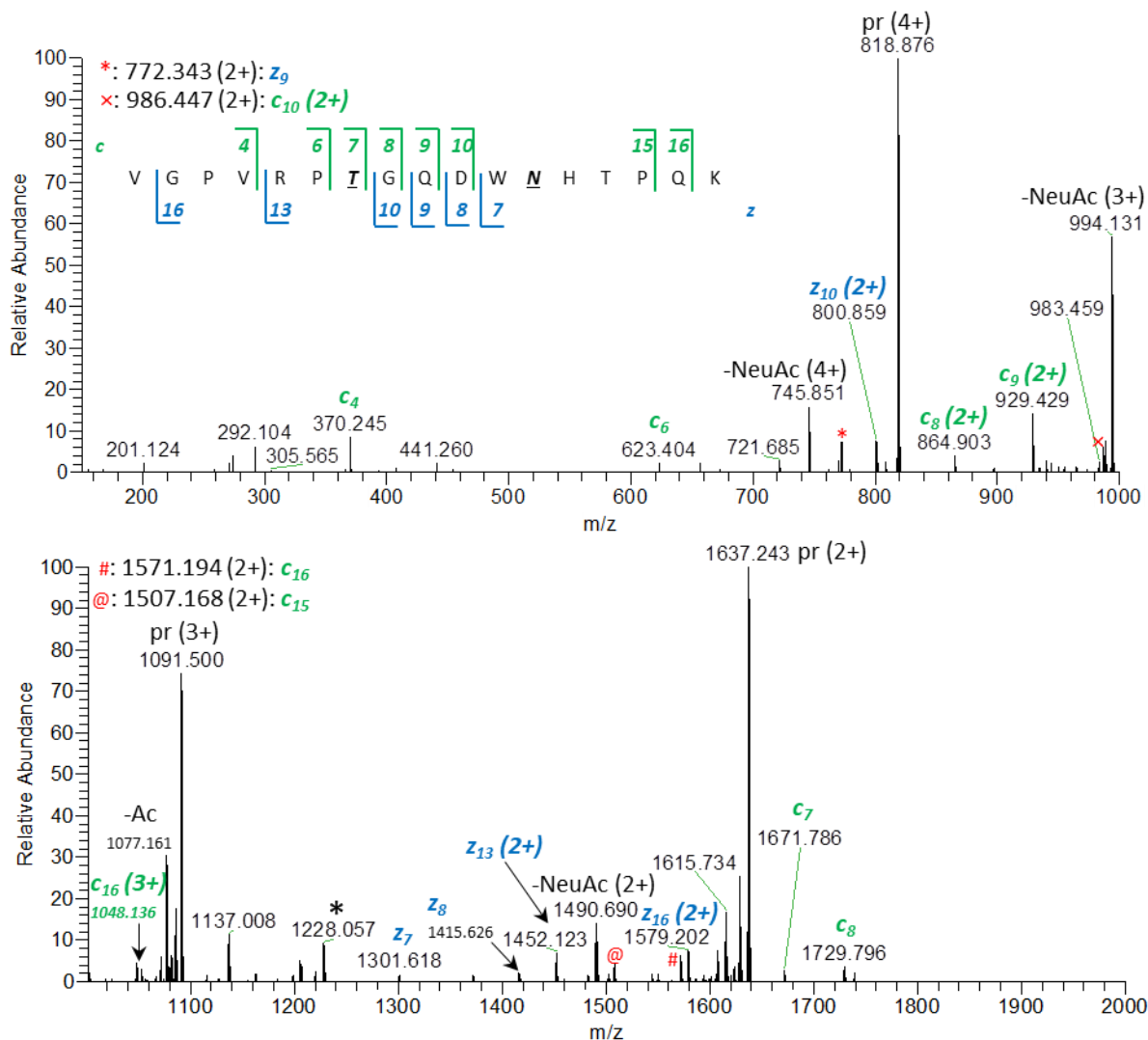
A GF-Hunter bemeneti peptidlistájában szereplő peptidek közül 8 szekvenciájában megtalálható az N-glikozilációs motívum (NXT|S). Ezek a peptidek az adatbáziskeresés során megbízhatóan lettek azonosítva valamilyen egyszerűbb mucin típusú O-glikán struktúrával. A GF-Hunter

eredmények azonban rávilágítanak arra, hogy ezek a peptidok nem csak O-glikoziláltak lehetnek, hanem egyszerre hordozhatnak O- és N-glikánt is. A két fajta glikán együttes jelenléte egy peptiden újabb példa az automatikusan nem megmagyarázható additív tömegre. Például 21 HCD spektrumhoz a szoftver a VGPVRPTGQDWNHTPQK peptidet rendelte, a számolt additív tömeg $\sim 1353,5$ Da. Ez a tömeg a HexNAc₃HexNeuAc₂ glikánösszetételnek feleltethető meg, amely felbontható több glikánra is: egy mucin típusú diszialo core 1 szerkezetre (HexNAcHexNeuAc₂) és két HexNAc-re, amelyek egyenként lehetnek úgy O-hez mint N-hez kötött szerkezetek, illetve összekapcsolódva inkább az N-glikánok legelső két GlcNAc monoszacharidját képviselik. A HCD spektrumban (29. ábra) a glikozidos kötések hasadásából keletkezett Y fragmensek dominálnak, de így is sejthető a kétféle glikántípus együttes jelenléte. Míg O-glikopeptidok esetében (ütközéses aktiválás esetén) az Y₀ ion az intenzívebb, addig N-glikopeptidok esetében ez az Y₁ ion. Így a HCD spektrumban (29. ábra) megfigyelt intenzív Y₁ ion (m/z 1061,033 (2+)) N-glikán jelenlétét valószínűsíti. Viszont a GF-Hunter által megállapított glikánösszetételből nem alkotható egyetlen „értelmes” N-glikán szerkezet, tehát az alternatíva valóban az O-kapcsolt tetraszacharid a két potenciális O-glikozilációs hely (Thr-7 és Thr-14) egyikén.



29. ábra m/z 818,371 (4+) prekuzorról készült HCD spektrum, amelyhez a GF-Hunter a VGPVRPTGQDWNHTPQK peptidszekvenciát és a HexNAc₃HexNeuAc₂ glikán összetételt rendelte. A sémán az O-glikánon lévő két NeuAc-t elhagytam a könnyebb áttekinthetőség érdekében. A feketén szedett ionok egy koelválódó másik glikopeptidból keletkeztek. Az alsó indexben található N és O jelölések az N-, illetve az O-glikánban bekövetkezett kötéshasadásokat jelentik.

Az O-glikán tényleges jelenlétére az EThcD spektrum szolgált (30. ábra) bizonyítékot. A módosított c_7 ion egyértelműen jelzi az O-glikán helyét, ami a Thr-7. A csonka N-glikán jelenlétére (Asn-12) az EThcD spektrumban közvetlen bizonyítékunk nincsen. A z_7 -es fragsens jelzi, hogy az Asn-12 vagy a Thr-14 módosítva lehet egy HexNAc₂ szerkezettel vagy esetleg mindkét aminosav egy-egy HexNAc monoszachariddal. Azonban a peptidben jelenlévő N-glikozilációs motívum és a HCD spektrumban megfigyelt intenzív Y_1 ion erősen valószínűsíti az N-glikán jelenlétét.



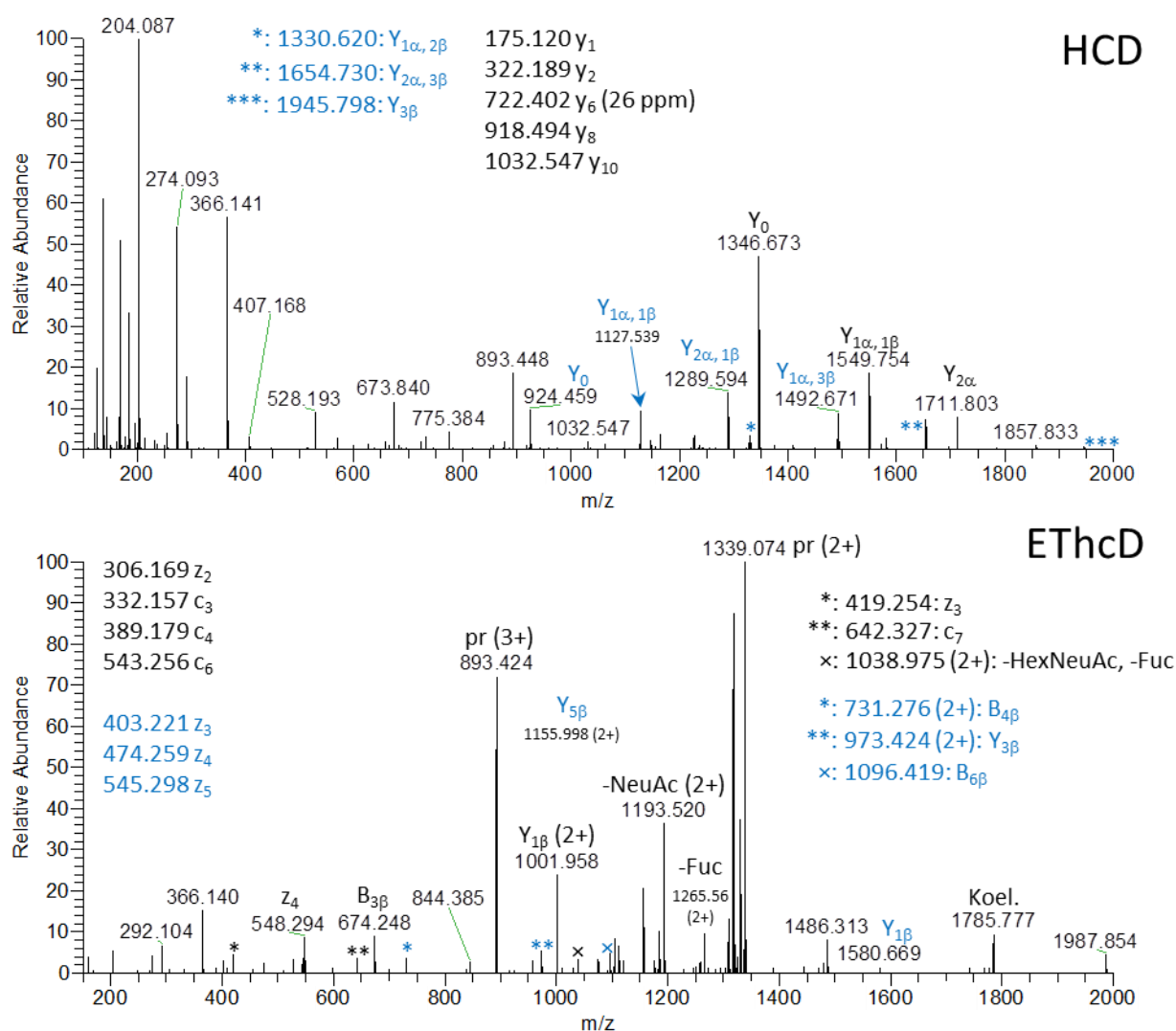
30. ábra m/z 818,371 (4+) prekuzorról készült EThcD spektrum, amelyet VGPVRPT(HexNAcHexNeuAc₂)GQDWN(HexNAc₂)HTPQK glikopeptidként azonosítottunk. Az ábrán a peptidszekvenciában vastagon, dőlttel szedett és aláhúzott aminosavak a két glikoziláció helyét jelölik. A töltésvesztett prekuzorokat „pr” jelöli, amelyekből az acetyl veszteséget -Ac, míg a -NeuAc az N-acetyl-neuraminsav veszteséget jelöli. Fekete csillaggal egy koelualódó prekuzor töltésvesztett formáját jelöltem.

A GF-Hunter által generált adatok rávilágítanak a komplex minták analizéséből fakadó adatértelmezési nehézségekre is. A szoftver alaptulajdonsága, hogy a bemeneti listán szereplő

peptideket az azoknak megfelelő Y_0 - Y_1 pár alapján azonosítja, ami nem más, mint molekulatömeg alapján történő azonosítás. A minta komplexitása miatt a GF-Hunter bemeneti listáján szereplő peptidek között több olyan peptid pár is van, melyek Y_0 és Y_1 ionjainak elméleti monoizotópos tömegei illenek 10 ppm tömegpontosságon belül egy HCD spektrumban detektált Y_0 - Y_1 párra. Összesen ~750 olyan HCD spektrum szerepel az eredmények között, amelyekben egy megtalált Y_0 - Y_1 párhoz több, az adatbáziskeresés során azonosított peptidszekvenciát rendelt a szoftver. Ezen spektrumok több mint fele (~460 db) két olyan peptidpárhoz (GSQVLESTPPPHVMR – PASAPPATSNPDPAVSR, AILPASAPPATSNPDPAVSR – MGQGTRPSPTVTPRPPR) tartozik, ahol a nagy tömegpontosság sem segített a pontos szekvencia eldöntéséhez. A két peptidpár MH^+ értéke között a tömegkülönbség 11 és 2 ppm. Megfelelő peptidfragmentáció esetén természetesen eldönthető, hogy a párok közül mely peptidszekvencia a helyes. Néhány HCD spektrum vizsgálatokor pedig kiderült, hogy habár egy input peptid Y_0 - Y_1 párjának megfelelő tömegű ionok valóban voltak a spektrumban, ezek bizonyos ionklaszterek második izotópcúcsainak feleltek meg.

A fenti példákban jól látszik, hogy a glikopeptid azonosításhoz nem elegendő önmagában a tömegmérés pontosságára hagyatkozni, mindenképpen szükséges a fragmentációs adatok figyelembevétele is. Az eredményeken úgy lehetséges javítani, hogy a GF-Hunterrel valószínűsített új glikoformák HCD és EThcD spektrumait a peptid fragmentumok alapján is értékelni kell. Bizonyos HCD spektrumokban több különböző molekulatömegű peptidhez tartozó Y_0 - Y_1 párok is megtalálhatóak voltak, ami prekursor ion interferenciát jelez. Komplex elegyek esetében nem ritka az az eset, amikor egy MS/MS spektrum kettő, három vagy akár több prekuzorról készül. Az ilyen esetek azonosítására vannak megfelelő szoftverek [Chen, 2010; Wang, 2011] és glikopeptidek esetében is végezhető ilyen adatkiértékelés, habár ehhez még manuális munka szükséges. A 31. ábra egy olyan HCD és EThcD spektrumpárt mutat be, ami két glikopeptidről készült. A HCD spektrumban az intenzívebb glikopeptidhez a GF-Hunter az AQDGGPVGTELFR szekvenciát és a HexNAc₂Hex₃FucNeuAc glikán összetételt, míg a másik glikopeptidhez a DFTAAFPR szekvenciát rendelte. Ebben az esetben a GF-Hunter glikán összetételt nem tudott hozzárendelni a számított additív tömeghez (1751,6734), mivel az nem volt elég pontos (+28 ppm) a legközelebbi összetételhez, ami a HexNAc₄Hex₄NeuAc (1751,6242). Az adatok manuális kiértékelése azonban alátámasztotta, hogy a DFTAAFPR szekvencia valóban egy HexNAc₄Hex₄NeuAc O-glikánnal módosított. A HCD spektrumban az intenzívebb glikopeptid (AQDGGPVGTELFR) szekvencia ionjai vannak jelen, a DFTAAFPR szekvenciát csak az Y_0 ion jelenléte támasztja alá (31. ábra, HCD spektrum). Az EThcD spektrum lehetővé teszi mindkét glikopeptid strukturális jellemzését. Az AQDGGPVGTELFR

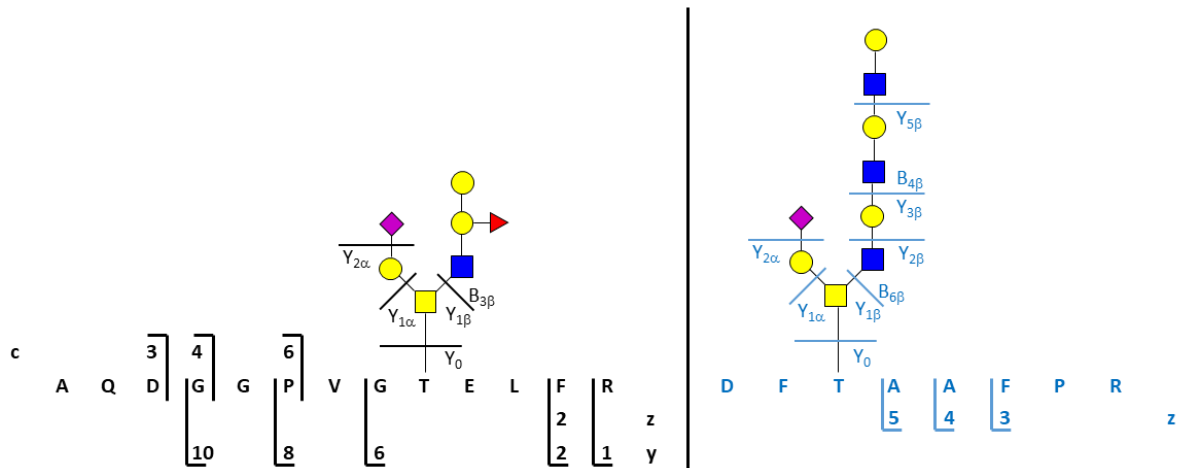
peptidet egy B antigént hordozó core 2 alapszerkezetű O-glikán módosítja. A B antigén a glikán β karján található, amelyet alátámaszt az $Y_{1\beta}$ - $B_{3\beta}$ komplementer ionpár (31. ábra, EThcD spektrum, 3. séma).



31. ábra m/z 892,718 (3+) prekuzorról felvett HCD (felső panel) és EThcD (alsó panel) spektrumok. A GF-Hunter a HCD spektrumhoz az AQDGGPVGTELFPR szekvenciát és a HexNAc₂Hex₃FucNeuAc összetételt, illetve a DFTAAFPR szekvenciát és a 1751,6734 additív tömeget rendelte. Az EThcD spektrumban a „Koel.” felirat egy koelualódó prekuzor töltésvesztett formáját jelöli. A kis intenzitású peptid szekvencia ionokat és glikán fragmensionokat (B, Y) mindkét spektrumon szövegdozokban tüntettem fel. A kékkel jelölt fragmensionok (peptid és glikán) a DFT(HexNAc₄Hex₄NeuAc)AAFPR, míg a feketével jelöltek az AQDGGPVGTELFPR(HexNAc₂Hex₃FucNeuAc)ELFR glikopeptidhez tartoznak.

A fukozilált glikán struktúra jelenlétét tovább erősíti a 2+ töltésvesztett prekuzorból detektált fukózvesztés m/z 1265,56-nál (31. ábra, EThcD spektrum, 3. séma). A módosítás lokalizálása indirekt módon, az egyetlen lehetséges O-glikozilációs hely miatt lehetséges, mert az EThcD spektrum nem tartalmaz diagnosztikus peptid szekvencia ionokat. A DFTAAFPR peptid egy meghosszabbított core 2-es alapszerkezetű O-glikánt hordoz. A glikán β karja három laktóz-

amin egységből áll, amelyet az $Y_{1\beta}$ - $B_{6\beta}$ komplementer ionpár bizonyít (31. ábra, EThcD spektrum, 3. séma). A peptidben csak egy potenciális módosítási hely található (Thr-3).



3. séma A 31. ábra spektrumaiból azonosított glikopeptidek fragmentációját szemléltető séma.

Az GF-Hunterrel végül 57 db O-glikán szerkezetet sikerült azonosítani (7. melléklet). A 4.3.1.1 és 4.3.1.2 fejezetekben tárgyalt O-acetilált N-acetil-neuraminsavat és disziálsavat tartalmazó szerkezeteket is sikerült a szoftverrel megtalálni. Az új szerkezetek azonosítása azonban így is hosszadalmas munka eredménye. A szoftver jól használható arra, hogy a HCD adatok alapján kijelölje azokat a HCD-EThcD spektrum párokat, amelyek az adatbázis kereséssel azonosított glikopeptidek potenciális, új glikoformái lehetnek. A szoftver nem helyettesíti az adatbázis kereső szoftvereket, csupán kiegészíti azokat. Használatával olyan spektrumokat is figyelembe tudunk venni, amelyek adatbázis kereséssel nem azonosíthatók a limitált, egyes esetekben hiányzó peptid gerinc fragmentáció miatt. A megtalált O-glikán szerkezetek nagy száma jól mutatja, hogy mennyire fontos a HCD adatok figyelembevétele az O-glikopeptidek adatkiértékelésének folyamatában. A legideálisabb adatkiértékelés az, amikor a HCD és az EThcD spektrumokban található információkat együttesen vesszük figyelembe. Olyan esetek is előfordulhatnak (in-source fragmentáció, szerkezeti izomerek), ahol nem elég csak az MS/MS adatokat használni, hanem szükséges az MS1 jelet és a retenciós időt is figyelembe venni a minél pontosabb adatinterpretáció érdekében.

5. KONKLÚZIÓ ÉS KITEKINTÉS

A humán vizeletben található fehérjék O-glikozilációjának megismerését a gyűjtött tömegspektrometriás adatok legegyszerűbb elemzésével kezdtük. Az általános, rutinszerűen végrehajtott adatbáziskeresés során az azonosított EThcD spektrumok aránya egészen alacsonynak adódott (~5%). Ennek egyik oka az EThcD alacsony határfoka, amely glikopeptidek esetében még inkább előtérbe kerül a módosítatlan peptidekhez képest nagyobb méretük, és emiatt alacsonyabb töltéssűrűségük miatt. Az azonosítási ráta alacsony értékéhez természetesen nem csak az EThcD alacsony határfoka járulhat hozzá, hanem az is, hogy az adatok értelmezésekor nem minden, a mintában jelenlévő O-glikán szerkezetet definiáltunk az adatbáziskereséshez. Ennek a problémának az orvoslása során jutottunk el oda, hogy felismertük mindazokat a fragmentációs szabályszerűségeket, amelyeket az eredményekben bemutattam.

A rutinszerűen elvégzett adatbáziskeresésből származó eredmények kiértékelésekor figyeltünk fel először arra, hogy az alacsonyabb kiegészítő NCE-t alkalmazó EThcD aktivációt a nagyobb cukor oxónium ionok is túlélnek. A nagyobb cukor oxónium ionok figyelembevételével eldönthetővé válik az a kérdés (a spektrumok egy kisebb halmazán), hogy egy peptid több kisebb, vagy esetleg egy nagyobb glikánstruktúrával módosított-e. Sajnos az adatbázis kereső programok az EThcD spektrumokat ilyen szemszögből nem vizsgálják. A lekeresés során a peptid szekvencia azonosítása a cél és ennek megfelelően az adatbázis kereső szoftverek az EThcD spektrumokban peptid-fragmensek után kutatnak. A peptidet módosító glikánt/glikánokat csak additív tömeg "szintjén" kezelik, a belőlük képződött B és Y fragmensek információ-tartalmát nem hasznosítják kellőképpen.

Hogy mennyire fontos a peptid fragmentáció mellett a glikán fragmentáció figyelembevétele is, azt mi sem bizonyítja jobban, mint a 4.3 és 4.4 fejezetben bemutatott O-glikán struktúrák változatossága. Az azonosított O-glikánok strukturális jellemzését az EThcD spektrumokban található egyszeres kötéshasadásból származó Y és B ionok tették lehetővé. Figyelembevételükkel több esetben sikerült azonos cukorösszetételű és azonos peptidszekvenciával rendelkező glikopeptidről kideríteni, hogy valójában szerkezeti izomerek, amelyet tovább erősített a glikoformák eltérő retenciós ideje is. Ugyancsak az egyszeres kötéshasadások miatt sikerült az O-acetilált szialinsavak és vércsoport antigének glikánon belüli pozícióit meghatározni.

Az EThcD adatok mellett a HCD adatok figyelembevételével még több, újabb O-glikán struktúrát sikerült felfedezni, jellemezni. Az erre a célra fejlesztett GF-Hunter szoftver jól

használható arra, hogy a korábban már adatbázis kereséssel azonosított O-glikopeptidek újabb glikoformáit megelhessük. A szoftver segítségével egy szulfatált és további több vércsoport antigénnel lezárt O-glikánt is sikerült azonosítani. Továbbá olyan EThcD spektrumokat is figyelembe tudunk venni, amelyek egy normál adatbázis keresés során nem lennének azonosíthatóak, a hiányos peptidfragmentáció miatt. Ezek a spektrumok azonban nem használhatatlanok, hiszen egy HCD spektrummal kombinálva azonosíthatóvá válhat a peptidszekvencia, míg az EThcD spektrumból a peptidet módosító glikán szerkezetéről szerezhetünk információt. Például az AVAVTLQSH peptid új glikoformáinak vizsgálata során gyakran kellett hiányos HCD és EThcD adatok együttes információ-tartalmára támaszkodnunk, amelyek együttesen megfelelő bizonyítékot nyújtottak a peptid szekvenciájára és az azt módosító O-glikánra is.

A dolgozatban bemutatott eredmények alapján elmondható, hogy a sikeresebb, automatizált O-glikopeptid azonosításhoz az EThcD spektrumokban a peptid fragmentáció mellett a glikán fragmentáció értelmezése is szükséges. Jelenleg az adatbázis kereső szoftverek az EThcD-ben található glikánfragmensek információ-tartalmát nem használják ki a spektrum azonosítása során. Továbbá szükséges az EThcD és HCD spektrumok információ-tartalmának együttes figyelembevétele a sikeresebb és megbízhatóbb azonosítás érdekében. Jelenleg ezt csak egyetlen szoftverbe építették be [Lu, 2020]. A szoftver a HCD spektrumból azonosítja a peptidet majd a prekursor és a módosítatlan peptidszekvencia alapján megállapítja a peptiden található additív tömeget. Az additív tömeghez a szoftver egy glikán adatbázisból rendel lehetséges cukor összetételeket, majd az EThcD spektrum alapján igyekszik ezt az összetételt lokalizálni a peptiden belül, akár úgy is, hogy többféle glikán kombinációt vesz figyelembe.

Jól látható tehát, hogy az O-glikopeptidek tömegspektrometriás elemzése nem egyszerű feladat. A módosítatlan peptidekhez képest sokszor nem elég egyetlen fragmentációs technikával felvett MS/MS spektrum elemzése a sikeres és megbízható azonosításhoz. Ehelyett több fragmentációs technika alkalmazása szükséges a glikán és a peptidrész minél pontosabb jellemzéséhez. A rendelkezésre álló adatbázis kereső szoftverek még nem optimalizáltak teljes mértékben az ilyen feladatokhoz, ami miatt a manuális spektrum kiértékelés elkerülhetetlen. A dolgozatban bemutatott eredmények azonban vázolják mindazokat a faktorokat, amelyeket a szoftverfejlesztőknek figyelembe kellene venni a fejlesztés során. Terveink között szerepel adataink teljes összehasonlító analízise a rendelkezésünkre álló szoftverekkel, és az eredmények részletes és kritikus megosztása a glikoproteomikával foglalkozó kutatóközösséggel.

6. ÖSSZEFOGLALÓ

A fehérjék glikozilációja a leggyakoribb és legváltozatosabb poszttranszlációs módosítás. A fehérjékhez kapcsolódó oligoszacharidok szerepe az élő szervezetben igen sokrétű. Részt vesznek különböző immunfolyamatokban, befolyásolják a fehérjék térszerkezetét, féléletidejét. Szerepük kulcsfontosságú a sejt-sejt kapcsolatokban is. A glikoziláció két leggyakrabban előforduló típusa közül az N-glikozilációról áll rendelkezésre nagyobb ismeretanyag, míg az O-glikoziláció a kevésbé jellemzett PTM-ek közé tartozik. Az analitikai, illetve az informatikai eszközök és módszerek fejlődésével egyre nagyobb mennyiségű adatot vagyunk képesek előállítani és feldolgozni. Ez kifejezetten igaz az N- és O-glikoziláció tandem tömegspektrometriás analízisére is, ahol az adatfelhalmozás korszakát éljük kvalitatív és kvantitatív szempontból egyaránt. Az O-glikoziláció tömegspektrometriás analízise több szinten is kihívásokkal teli feladatnak számít, amelyek közül a dolgozat a tömegspektrometriás adatkiértékelés kérdéskörét érinti.

A poszttranszlációs módosítások jellemzése fragmentációs adatok alapján, azaz MS/MS spektrumok felvételével és kiértékelésével történik. Ez az O-glikoziláció analízisének sincs másképpen, azonban az O-glikopeptidek fragmentációs tulajdonságai miatt többféle fragmentációs technika alkalmazása szükséges a molekula lehető legteljesebb jellemzése érdekében. Az O-glikopeptidek két, kémiaiailag eltérő és fragmentáció szempontjából is különböző vegyülettípusból épülnek fel. Jellemzésükhöz egyszerre kell a peptidet és a módosító glikán(oka)t azonosítani. A poliszacharidok analízise eleve problémás, mert izomer alkotórészeik tömegük alapján nem megkülönböztethetőek. Hasonlóképpen nincs egyszerű tömegspektrometriás módszer a glikozidos kötés pontos pozíciójának és térszerkezetének jellemzésére. Tehát szerencsés esetben a következő információk nyerhetők a glikopeptidekről: (1) peptidszekvencia, (2) a glikán monoszacharid összetétele, (3) a glikánt alkotó monoszacharid egységek egymáshoz viszonyított helyzete és végül (4) a módosítás helye.

Ütközéses aktiváláskor, glikopeptidek esetében, a preferált fragmentációs lépés a glikozidos kötéshasadás. Ez ioncsapda CID során uralkodó fragmentációs eseményként a glikán szerkezeti jellemzését teszi lehetővé, azonban a peptid szekvenciája ismeretlen marad. A peptid tömegéről szerezhetünk információt a glikán gázfázisú eliminációja során keletkezett deglikozilált peptidionnak (Y_0) köszönhetően, illetve a peptidet módosító glikánok additív tömege is meghatározható ezen ion és a prekursor ion segítségével, amiből a monoszacharid összetétel is megállapítható. HCD aktivációkor a többszörös ütközéseknek köszönhetően nem csak a glikozidos kötések, de a peptidkötés is elhasadhat. Az így keletkezett *b* és *y* szekvenciaionok

segítésével a peptid aminosav sorrendje kideríthető, azonban a többszörös ütközések miatti erősebb glikánfragmentáció nem teszi lehetővé annak szerkezeti jellemzését. Az Y_0 ion, ahogyan az ioncsapda CID-ban is, itt is tájékoztat minket a módosító glikán(ok) additív tömegéről, amelyből a monoszacharid összetétel meghatározható. Azonban ioncsapda CID és HCD adatokból sem állapítható meg a módosítás helye, mivel mindkét aktivációs technika esetében a preferált fragmentációs esemény a gázfázisú deglikoziláció.

A labilis poszttranszlációs módosítások helye a peptiden belül ütközéses aktiválással tehát nem határozható meg. Ennek a problémának az áthidalására az ETD aktiváció jelent megoldást, amely egy elektron-transzferrel képződött gyökion bomlásán alapuló fragmentációs technika. A fragmentáció során a peptid kötés nitrogénje és az alfa szén közötti kötés elhasadásával c , illetve $z\bullet$ szekvenciaionok keletkeznek. A technika előnye, hogy a fragmentáció során az aminosav oldalláncok épen maradnak, ezért a megfelelő fragmensek hordozzák a módosító glikánt is, így a módosítás helye meghatározható. Azonban a glikán szerkezetére vonatkozólag az ETD spektrum nem szolgál információval. Továbbá a fragmentáció határfoka is rosszabb az ütközéses aktiválásnál tapasztalható képest, mivel az ETD aktiválás erősen töltéssűrűség függő. Minél nagyobb a prekursorion töltéssűrűsége, annál hatékonyabb a fragmentáció. A glikopeptidek jellemzően alacsony töltéssűrűségű molekulák a módosítatlan peptidekhez képest, mivel a módosító glikán úgy növeli a molekula tömegét, hogy közben több töltést nem ad hozzá a molekulához.

Az EThcD aktiváció egy kombinált fragmentációs technika, amely során először az ETD, majd a teljes ionhalmazon HCD aktiváció történik. A kiegészítő HCD lépés elsősorban a már aktivált, de még el nem hasadt prekursor ionokat érinti. Az ETD lépésnek köszönhetően lehetőség nyílik a módosított peptidszekvencia azonosítására, míg a HCD aktiváció szerkezeti információval szolgál a peptidet módosító glikánról. Előnye, hogy egy spektrumból szerezhethetünk információt a peptid szekvenciájáról, a módosítás helyéről és a glikán szerkezetéről.

Dolgozatom a humán vizeletből dúsított O-glikopeptidek EThcD aktivációval felvett tandem tömegspektrometriás adatainak kiértékelésével foglalkozik. A fragmentációs adatok értelmezésének kettős célja van. Az első technikai jellegű, és az O-glikopeptidek EThcD-ben mutatott fragmentációs sajátosságainak megismerése áll a középpontban. Azonban emellett a vizeletben található fehérjék O-glikozilációjának megismerése is cél. Ezek a célok nem választhatók el egymástól. A kutatás analitikai részében megszerzett ismeretek az O-glikopeptidek tandem tömegspektrometriás viselkedéséről alapjaiban meghatározzák a felvett

adatokból kinyerhető, biológiailag releváns információk megbízhatóságát. A kutatások során elért eredményeket pontokba szedve alább ismertetem.

1.) EThcD aktivációval nagyobb szacharid oxónium ionok is képződhetnek

Az adatértelmezést három egészséges donor vizeletmintáiból dúsított glikopeptid-elegyek EThcD adataival kezdtem. A vizeletben található glikoproteinek O-glikozilációja kevésbé ismert, így a rutin adatbázis keresés során változó módosításként a humán szérumban előforduló leggyakoribb O-glikánokat definiáltam. Az eredmények értékelése közben figyeltem fel arra, hogy az azonosított EThcD spektrumokban nagyobb szacharid oxónium ionok is képződtek az alacsony energiájú kiegészítő HCD aktiválás hatására. Ezek az ionok értékes információt hordoznak arról, hogy a peptid több kisebb vagy egy nagyobb O-glikánnal módosított-e. Megvizsgáltam, hogy a leggyakoribb O-glikánoknak megfelelő intakt oxónium ionok milyen gyakran detektálhatók az EThcD spektrumokban. Azt találtam, hogy a nagyobb oxónium ionok, mint pl. a diszialilált core 1 szerkezetre jellemző m/z 948, vagy a core 2 hexaszacharid jelenlétére utaló m/z 1313 ionok az EThcD spektrumok többségében csak gyenge jelet adnak, így detektálásuk erősen függ a prekursor ion intenzitásától, viszont jól használhatók a glikán azonosítások megerősítésére vagy cáfolatára.

2.) A vizeletben található fehérjék O-glikozilációjának felfedezése nem definiált módosítás kereséssel

Az általános adatbázis keresés azonosítási rátája a teljes betáplált EThcD spektrumszámhoz képest alacsony volt (~5%), amelynek két fő oka lehetséges. Az első az EThcD aktiváció alacsony hatásfoka az alacsony töltéssűrűségű prekursorok miatt. A második pedig az, hogy nagy valószínűséggel az adatbázis kereséskor nem minden, a vizeletben jelenlévő O-glikánt definiáltam. Annak érdekében, hogy feltárjam, milyen más O-glikánok módosítják a vizeletben előforduló glikopeptideket a szérumban leírt leggyakoribb szerkezeteken túl, ún. nem definiált módosítás adatbázis keresést végeztem az EThcD spektrumokon, ahol egy bizonyos tömegtartományon (0-3000 Da) belüli módosítást engedélyeztem csakis glikozilálható aminosavakon. Az így jelentett additív tömegekből hisztogram készíthető, és a gyakoribb tömegek valóban különböző szacharid-összetételeknek felelnek meg. Az így talált módosítások pontosabb jellemzése az EThcD spektrumok manuális kiértékelésével érhető el.

3.) Az EThcD technika segítségével meghatározható a peptidet módosító glikán monoszacharid egységeinek relatív pozíciója

A manuális adatkiértékelés során a nem definiált módosítás keresésből azonosított additív tömegek egy része mono- és di-O-acetil-sziálsavat, illetve disziálsavat tartalmazó O-glikánoknak felelt meg. A nagyobb szacharid oxónium ionok detektálása már jelezte azt, hogy az alacsony energiájú kiegészítő HCD aktiválás glikopeptidek esetében az egyszeres glikozidos kötés hasadásoknak kedvez. Az O-acetilált és disziálsavak O-glikánon belüli pozíciójának meghatározását az egyszeres glikozidos kötés hasadások tették lehetővé. Tehát kijelenthető, hogy az alacsony energiájú kiegészítő HCD aktiválással felvett EThcD technika alkalmas a módosított peptid szekvencia mellett a peptidet módosító O-glikán szerkezeti jellemzésére. Ennek további pozitív hozadéka az, hogy ugyanazon glikopeptid szerkezeti izomerei is megkülönböztethetőek, ha az on-line fordított fázisú kromatográfia során elválnak egymástól. Ezt a dolgotatban be is mutatom egy izomer páron, ahol az egyetlen különbség a glikánon található acetyl-csoport pozíciójában van. A nem definiált módosítás kereséssel és az EThcD spektrumok manuális elemzésével 12 új O-glikán struktúrát sikerült azonosítani a vizeletben található glikoproteineken. Ezek között, a már korábban említett szerkezeteken kívül, azonosítottam szialil-Lewis^{A/X} epitópot tartalmazó O-glikánokat is.

4.) Az EThcD és HCD adatok együttes elemzése szükséges

Az EThcD adatok mellett a HCD adatok bevonásával sikerült újabb O-glikán szerkezetekkel módosított glikopeptidek azonosítása. HCD aktiváció során a magasabb normalizált ütközési energia miatt a peptidet módosító glikán(ok) gázfázisú eliminációja a preferált fragmentációs esemény. Ez a HCD spektrumban a deglikozilált peptidion megjelenését (Y_0) eredményezi. Segítségével nem csak a módosítatlan peptid molekulatömegét, de az azt módosító glikán(ok) additív tömegét is meghatározhatjuk. Az additív tömeg ismeretében pedig a monoszacharid összetétel meghatározható. A HCD adatokból tehát előrejelezhetőek EThcD adatokból már azonosított glikopeptidek új glikoformái. Ezt a hipotézist a mintákból leggyakrabban azonosított peptiddel (³⁴²AVAVTLQSH³⁵⁰, Protein YIPF3) és néhány minta adataival manuálisan teszteltem. Az eredmények meggyőzőek voltak, az adathalmaz nagysága viszont nyilvánvalóvá tette, hogy automatizálásra van szükség.

5.) Szoftverfejlesztés új glikoformák azonosítására HCD adatokból

A HCD spektrumok információ-tartalmának kiaknázására és a potenciális glikoformák előrejelzésének gyorsítására szoftvert fejlesztettünk (GF-Hunter), amely már a korábban megbízhatóan (pl. EThcD spektrumokból) azonosított O-glikopeptidek Y_0 - Y_1 ionpárjait keresi

a HCD spektrumokban. A minta komplexitása miatt a szűrés során egyetlen ion jelenlétére (Y_0) támaszkodni nem elegendő. Mindenképpen szükséges az Y_0 ion mellett annak Y_1 (peptid+HexNAc) párját is keresni a HCD spektrumokban. Az ionpár jelenlétének megkövetelése azonos töltésállapotban csökkenti a fals pozitív találatok arányát. A szoftver, az elfogadási paramétereknek megfelelő HCD spektrumhoz hozzárendeli a megtalált Y_0 - Y_1 ionpárhoz tartozó peptidszekvenciát, valamint a prekursor ion és az Y_0 ion tömegének felhasználásával kiszámítja a módosítás additív tömegét. Ha a módosítás tömege illik az előre megadott glikán adatbázis valamely tagjára, akkor az additív tömeget glikánként azonosítja a szoftver, melyet szükséges az adott HCD spektrum EThcD párjának manuális ellenőrzésével validálni. A szoftverbe épített glikán adatbázis iteratív módon bővíthető, így a korábban felfedezett O-glikánokat a szoftver a következő szűrés alkalmával már képes azonosítani.

6.) Új glikán szerkezetek felfedezése a GF-Hunter segítségével

A szoftverrel a tesztelést követően 10 donor vizeletéből dúsított glikopeptid-elegyek tömegspektrometriás adatait elemeztem. A HCD adatok automatizált szűrésével új oligoszacharid-kompozícióknak megfelelő additív tömegeket azonosítottam. A szoftver segítségével további 17 új O-glikán struktúra jellemzésére került sor a HCD és EThcD adatok együttes felhasználásával. Ezek között olyan új O-glikánok szerepelnek a már korábban azonosított O-acetil szialsavat és diszialsavat tartalmazó szerkezetek mellett, amelyek AB0 vércsoport antigént, poli-N-acetil-laktózamin egységet és szulfatálást tartalmaznak.

A dolgozatban bemutatott eredmények egy hatalmas és reprezentatív tömegspektrometriás adathalmaz szisztematikus elemzésével keletkeztek. Ez az adathalmaz olyan intakt humán mucin-típusú O-glikopeptidek analízisével készült, amelyeket nem laboratóriumi sejtvonalakból, hanem 10 humán donor vizeletéből izoláltam. Az adatok kiértékelése azt mutatja, hogy a vizeletben található fehérjék igen változatos glikánkészlettel módosítottak. A dolgozatban bemutatott adatkiértékelési módszerek még mindig erősen támaszkodnak a manuális adatelemzésre és az adatbázis kereső szoftverek O-glikopeptid azonosításainak felülvizsgálatára. Mindezek ellenére a vizeletben található fehérjéken összesen 57 különböző O-glikánt sikerült azonosítani O-glikopeptidek tandem tömegspektrometriás analízisével. Kutatócsoportunk volt az első, amely ezeket a szerkezeteket konkrét fehérjéken, helyspecifikusan azonosította, pl. vércsoport antigének jelentését bizonyos szekretált fehérjéken.

6. SUMMARY

Glycosylation is among the most common post-translational modifications of proteins. Its roles in living organisms are very diverse. Glycans linked to proteins participate in various immune processes, influence the 3D structure and half-life of proteins. They play a key role in cell-cell adhesion. Of the two most common types of glycosylation, more is known about N-glycosylation, while O-glycosylation is the less characterized PTM. With the development of analytical methods and bioinformatic tools we are able to acquire and process huge datasets, and significant amount of qualitative and quantitative information has been compiled on both types of glycosylation. At the same time O-glycosylation remains a rather challenging task. My dissertation is addressing mass spectrometric data interpretation, i.e., one of the aspects that makes this PTM difficult to tackle.

Post-translational modifications are characterized by fragmentation data, i.e., by recording and evaluating MS/MS spectra. For the analysis of O-glycosylation, however, due to the fragmentation properties of O-glycopeptides, data acquired using multiple activation techniques are required. O-glycopeptides are composed of two types of compounds that are chemically different and exhibit different fragmentation characteristics. During data analysis not just the peptide but also the modifying glycan(s) have to be identified. The analysis of polysaccharides is inherently problematic because their isomeric constituents cannot be distinguished by their mass. Similarly, there is no simple mass spectrometric method to characterize the exact position and stereochemistry of a glycosidic bond. Thus, under optimal conditions the following information can be obtained about glycopeptides: (1) peptide sequence, (2) monosaccharide composition of the glycan(s), (3) the relative position of the glycan-forming monosaccharide units and (4) the site of modification.

The preferred fragmentation step during the collisional activation of glycopeptides is the cleavage of the glycosidic bond. In ion trap CID this allows the structural characterization of the glycan. However, the peptid sequence remains unknown. The mass of the nonglycosylated peptide can be determined from the detected deglycosylated peptide ion (Y_0) formed during the gas phase elimination of the glycan. Furthermore, the additive mass of the glycan(s) can be calculated as the difference between the precursor ion and the Y_0 ion. From the additive mass the monosaccharide composition can also be determined. Upon HCD activation, not only glycosidic bonds but also peptide bonds can be cleaved due to multiple collisions. From the resulting *b* and *y* sequence ions, the amino acid sequence of the peptide can be elucidated. However, the stronger glycan fragmentation due to the multiple collisions does not permit the

structural characterization of the glycan(s). The Y_0 ion, like in the ion trap CID, informs us again about the additive mass of the modifying glycan(s) from which the monosaccharide composition can be determined. However, modification site assignment is not possible from these data as the preferred fragmentation event for both activation techniques is the gas phase deglycosylation.

The site of labile post-translational modifications within the peptide cannot be determined by collisional activation. To overcome this problem, ETD activation offers a solution which is a fragmentation technique based on the decomposition of a radical ion formed via electron transfer. During fragmentation, the cleavage between the nitrogen atom of the peptide bond and the alpha carbon results in c and z^{\bullet} sequence ions. Since the amino acid side chains remain intact during fragmentation, the corresponding sequence ions carry the modifying glycan, hence the site of modification can be determined. However, the ETD spectrum does not provide information on the structure of the glycan. Furthermore, the efficiency of fragmentation is worse than what is observed for collisional activation. This is due to the fact that ETD activation is strongly dependent on charge density. The higher is the charge density of the precursor ion, the more efficient is the fragmentation. Glycopeptides typically have low charge density compared to unmodified peptides because the modifying glycan increases the weight of the molecule without adding more charges to it.

ETHcD activation is a combined fragmentation technique in which ETD is performed first and then HCD on the entire ion set. The additional HCD step primarily affects the already activated but still intact precursor ions. The ETD step enables the identification of the modified peptide sequence and modification site assignment, while HCD activation provides structural information about the modifying glycan. The advantage of this combined activation technique is that all the above information can be obtained from a single spectrum.

During my thesis work I studied ETHcD data acquired from glycopeptides enriched from human urine. My research presented in this dissertation had two specific aims. The first was technical and focused on the fragmentation characteristics of O-glycopeptides in ETHcD. The goal was to characterize the O-glycosylation of urinary proteins. These goals are intertwined. The knowledge gained in the analytical part of the research fundamentally determines the reliability of biologically relevant information extracted from the recorded data. My achievements are presented below.

1.) During EThcD activation, larger saccharide oxonium ions may also survive

Firstly, the two-round WGA lectin affinity chromatography was used to enrich glycopeptides from the urine of three healthy donors and then the mass spectrometric analysis of the enriched glycopeptide mixtures was carried out. As a first step in data evaluation, a general database search was performed with the EThcD data. The O-glycosylation of urinary proteins is less well known, so the three most common O-glycan structures found in human serum were defined. During the evaluation of the search results the presence of larger saccharide oxonium ions in the EThcD spectra were noticed. This phenomenon is due to the low-energy supplemental HCD activation. These ions provide valuable information on whether the peptide is modified with multiple smaller or one larger O-glycan. A statistical analysis was conducted to answer the question: How often intact oxonium ions corresponding to the most common O-glycans can be detected in the EThcD spectra? It was found that larger oxonium ions such as m/z 948 corresponding to the disialylated core 1 structure or m/z 1313 indicating the presence of the core 2 hexasaccharide gave only a weak signal in most EThcD spectra. Thus, their detection strongly depends on the precursor ion's intensity, but they can be used to confirm or refute glycan assignments.

2.) Expansion of the glycan database using undefined modification database search

The identification rate of the general database search was low relative to the total number of EThcD spectra submitted (~5%). This low identification rate has two possible reasons. The first is the low efficiency of EThcD activation due to low charge density precursors. The second is that not all O-glycans present in the urine were defined in the search. In order to explore what other O-glycans can modify the urinary proteins in addition to the most common structures described in serum, a so-called unspecified modification database search was performed with the EThcD spectra, where modifications within a certain mass range (0-3000 Da) was allowed only on Ser and Thr residues. A histogram was constructed from the additive masses reported by this search, and the most common additive masses indeed corresponded to different oligosaccharide compositions. A more accurate characterization of the modifications can be achieved by the manual evaluation of the EThcD spectra.

3.) The EThcD technique can be used to determine the relative position of the monosaccharide units of the modifying glycan

During the manual data evaluation, some of the additive masses identified from the undefined modification search corresponded to O-glycans containing mono- and di-O-acetylsialic acid and disialic acid. The presence of larger saccharide oxonium ions has already indicated that low-

energy supplemental HCD activation favors single glycosidic bond cleavages. This observation i.e., the preferred single bond cleavages, permitted the determination of the position of O-acetylated sialic acids and disialic acids within the O-glycan structure. Thus, it can be concluded that the EThcD technique with low-energy supplemental HCD activation facilitates the structural characterization of the modifying O-glycan. A further benefit of this is that the structural isomers of the same glycopeptide can also be distinguished if these are separated by on-line reversed-phase chromatography, as it was demonstrated in the dissertation with the assignment of an isomeric pair. A total of 12 new O-glycan structures were identified on urinary glycoproteins applying the unspecified modification search and manual data interpretation. Among these, O-glycans containing sialyl-Lewis^{AX} epitope were also identified, in addition to the previously characterized structures.

4.) The combined analysis of EThcD and HCD data is required

In addition to EThcD data, additional O-glycan structures were identified by considering HCD data. During HCD activation, gas phase elimination of the modifying glycan(s) is the preferred fragmentation event resulting in the detection of the deglycosylated peptide ion (Y_0). This can be used to determine not only the molecular weight of the unmodified peptide, but also the additive mass of the glycan(s). Knowing the additive mass, the monosaccharide composition can also be calculated. Thus, new glycoforms of glycopeptides previously assigned from EThcD data can be predicted from the HCD spectra. This hypothesis was tested and confirmed with the most commonly identified peptide (³⁴²AVAVTLQSH³⁵⁰, Protein YIPF3) on a few datafiles, but it also became evident that automatization of the process is required because of the sheer amount of data.

5.) Software development for the identification of new glycoforms from HCD data

In order to exploit the information content of the HCD spectra and to accelerate the prediction of potentially new glycoforms, we have developed a software (GF-Hunter) in cooperation that searches for Y_0 - Y_1 ion pairs of previously and reliably identified (e.g., from EThcD spectra) O-glycopeptides in HCD spectra. Due to the complexity of the samples, it is not sufficient to rely on the presence of a single ion (Y_0) during the filtering process. In addition to the Y_0 ion, it is absolutely necessary to look for its Y_1 (peptide + HexNAc) pair in the HCD spectra. Requiring the presence of the ion pair in the same charge state reduces the rate of false positive hits. When the software identifies a spectrum fulfilling the acceptance criteria, it assigns the peptide sequence of the detected ion pair to that HCD spectrum and calculates the additive mass of the modification using the mass of the precursor ion and the Y_0 ion. If the calculated additive mass

matches to an entry in the predefined glycan database, the additive mass is assigned as a glycan, which needs to be validated by manually checking the EThcD pair of the corresponding HCD spectrum. The glycan database built into the software can be iteratively expanded so that previously discovered O-glycans can be identified by the software at subsequent screening.

6.) Finding new structures with GF-Hunter

After testing, the mass spectrometric data of glycopeptide mixtures enriched from the urine of 10 patients was analyzed. Several missed glycoforms of previously identified glycopeptides were indicated by the automated screening of the patients' HCD data. The presence of 17 novel glycan structures was confirmed by manual inspection of the corresponding EThcD data. These novel O-glycans feature blood group antigens, poly-N-acetyl-lactosamine units, and sulfate.

A large and representative mass spectrometric dataset was generated by the analysis of intact human mucin-type O-glycopeptides enriched, not from laboratory cell lines, but from the urine of 10 individuals. Deciphering the data revealed that the glycan repertoire of the urinary glycoproteins is highly diverse. The data interpretation methods presented in the dissertation still rely heavily on manual data evaluation and the inspection of O-glycopeptide identifications delivered by the database search engines. Nevertheless, a total of 57 different O-glycans on urinary proteins were identified from the tandem mass spectrometric data of O-glycopeptides. Our research group was the first who identified these structures on specific proteins and positions, for example, blood group antigens site-specifically on specific secreted proteins.

7. KÖSZÖNETNYILVÁNÍTÁS

Elsősorban szeretnék köszönetet mondani **Dr. Darula Zsuzsannának** és **Dr. Medzihradszky F. Katalinnak**, akik az elmúlt 5 évben egyengették utamat. Szakmai útmutatásuk és az 5 év alatt nyújtott témérdek mennyiségű segítségük nagyban hozzájárult ahhoz, hogy ez a dolgozat elkészülhessen.

Köszönettel tartozom a Proteomikai Kutatócsoport többi munkatársának is: **Dr. Hunyadi-Gulyás Évának**, **Dr. Klement Évának**, **Dr. Pettkó-Szandtner Aladárnak** és **Bálóné Árva Ágnesnek**, akikhez bármilyen problémával fordulhattam, mindig elláttak hasznos tanácsokkal.

Köszönöm a Biokémiai Intézet igazgatójának, **Dr. Horváth Péternek**, hogy lehetőséget biztosított az intézetben a doktori tanulmányaim elvégzéséhez.

Nagy köszönettel tartozom **Családomnak**, akik támogattak egyetemi és doktori tanulmányaim alatt. Áldozatos munkájukkal Ők is hozzájárultak a dolgozat elkészüléséhez. Még egyszer köszönöm Nekik.

Végül, de nem utolsósorban köszönet illeti minden **barátomat** és **ismerősömet**, akik biztattak és támogattak a dolgozat írása közben.

A kutatómunkát az **SZTE Biológia Doktori Iskola** (Ph.D. ösztöndíj), két **GINOP pályázat** (GINOP-2.3.2-15-2016-00001, GINOP-2.3.2-15-2016-00020), egy **GOP program** (GOP-1.1.1-11-2012-0452) és a **Hungarian Centre of Excellence for Molecular Medicine (HCEMM) szervezet** (EU Horizon 2020, Grant No.: 739593) támogatta. Köszönettel tartozom a Protein Prospector szoftver fejlesztőinek, **Peter Bakernek** és **Robert Chalkleynak**. Köszönöm a Protein Prospector szoftver futtatásához szükséges informatikai háttérrel az **ELKH Cloudnak**.

8. IRODALOMJEGYZÉK

Allen, A. C., Willis, F. R., Beattie, T. J., & Feehally, J. (1998). Abnormal IgA glycosylation in Henoch-Schönlein purpura restricted to patients with clinical nephritis. *Nephrology, dialysis, transplantation: official publication of the European Dialysis and Transplant Association - European Renal Association*, 13(4), 930–934. <https://doi.org/10.1093/ndt/13.4.930>

Anderson, N. L., and Anderson, N. G. (2002) The human plasma proteome: history, character, and diagnostic prospects. *Mol. Cell. Proteomics* 1, 845–867

Bai, X., Li, D., Zhu, J., Guan, Y., Zhang, Q., and Chi, L. (2015) From individual proteins to proteomic samples: characterization of O-glycosylation sites in human chorionic gonadotropin and human-plasma proteins. *Anal. Bioanal. Chem.* 407, 1857–1869

Baker, P. R., Trinidad, J. C., and Chalkley, R. J. (2011) Modification site localization scoring integrated into a search engine. *Mol. Cell. Proteomics* doi: 10.1074/mcp.M111.008078

Bakken, V., Helgaker, T., and Uggerud, E. (2004) Models of fragmentations induced by electron attachment to protonated peptides. *Eur. J. Mass Spectrom.* 10, 625–638

Bern, M., Kil, Y. J., and Becker, C. (2012) Byonic: advanced peptide and protein identification software. *Curr. Protoc. Bioinformatics* Chapter 13:Unit13.20. doi: 10.1002/0471250953.bi1320s40

Beyer, T. A., Rearick, J. I., Paulson, J. C., Prieels, J. P., Sadler, J. E., & Hill, R. L. (1979). Biosynthesis of mammalian glycoproteins. Glycosylation pathways in the synthesis of the nonreducing terminal sequences. *The Journal of Biological Chemistry*, 254(24), 12531–12534.

Biemann, K. (1990) Appendix 5. Nomenclature for peptide fragment ions (positive ions). *Methods Enzymol.* 193, 886–887

Carlson D. M. (1968). Structures and immunochemical properties of oligosaccharides isolated from pig submaxillary mucins. *The Journal of Biological Chemistry*, 243(3), 616–626.

Chalkley, R. J., & Burlingame, A. L. (2001). Identification of GlcNAcylation sites of peptides and alpha-crystallin using Q-TOF mass spectrometry. *Journal of the American Society for Mass Spectrometry*, 12(10), 1106–1113. [https://doi.org/10.1016/s1044-0305\(01\)00295-1](https://doi.org/10.1016/s1044-0305(01)00295-1)

Chang, C. F., Pan, J. F., Lin, C. N., Wu, I. L., Wong, C. H., and Lin, C. H. (2011) Rapid characterization of sugar-binding specificity by in-solution proximity binding with photosensitizers. *Glycobiology* 21, 895–902

Chen, X., Drogaris, P., & Bern, M. (2010). Identification of tandem mass spectra of mixtures of isomeric peptides. *Journal of Proteome Research*, 9(6), 3270–3279. <https://doi.org/10.1021/pr100205k>

Christlet, T. H., and Veluraja, K. (2001) Database analysis of O-glycosylation sites in proteins. *Biophys. J.* 80, 952–960

Clauser, K. R., Baker, P., and Burlingame, A. L. (1999) Role of accurate mass measurement (+/- 10 ppm) in protein identification strategies employing MS or MS/MS and database searching. *Anal. Chem.* 71, 2871–2882

Conchie, J., Hay, A. J., Strachan, I., & Levvy, G. A. (1969). The enzymic degradation of ovalbumin and its glycopeptides. *The Biochemical Journal*, 115(4), 717–723. <https://doi.org/10.1042/bj1150717>

Darula, Z., and Medzihradszky, K. F. (2009) Affinity enrichment and characterization of mucin core-1 type glycopeptides from bovine serum. *Mol. Cell. Proteomics* 8, 2515–2526

Darula, Z., Chalkley, R. J., Baker, P., Burlingame, A. L., and Medzihradszky, K. F. (2010) Mass spectrometric analysis, automated identification and complete annotation of O-linked glycopeptides. *Eur. J. Mass Spectrom. (Chichester)* 16, 421–428

Darula, Z., Sherman, J., & Medzihradszky, K. F. (2012). How to dig deeper? Improved enrichment methods for mucin core-1 type glycopeptides. *Molecular & cellular proteomics: MCP*, 11(7), O111.016774. <https://doi.org/10.1074/mcp.O111.016774>

Darula, Z., Sarnyai, F., & Medzihradzsky, K. F. (2016). O-glycosylation sites identified from mucin core-1 type glycopeptides from human serum. *Glycoconjugate Journal*, 33(3), 435–445. <https://doi.org/10.1007/s10719-015-9630-6>

Darula, Z., Pap, Á., & Medzihradzsky, K. F. (2019). Extended Sialylated O-Glycan Repertoire of Human Urinary Glycoproteins Discovered and Characterized Using Electron-Transfer/Higher-Energy Collision Dissociation. *Journal of Proteome Research*, 18(1), 280–291. <https://doi.org/10.1021/acs.jproteome.8b00587>

Deighan, W. I., Winton, V. J., Melani, R. D., Anderson, L. C., McGee, J. P., Schachner, L. F., Barnidge, D., Murray, D., Alexander, H. D., Gibson, D. S., Deery, M. J., McNicholl, F. P., McLaughlin, J., Kelleher, N. L., & Thomas, P. M. (2020). Development of novel methods for non-canonical myeloma protein analysis with an innovative adaptation of immunofixation electrophoresis, native top-down mass spectrometry, and middle-down de novo sequencing. *Clinical Chemistry and Laboratory Medicine*, cclm-2020-1072. Advance online publication. <https://doi.org/10.1515/cclm-2020-1072>

Domon, B., Costello, C.E. A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra of glycoconjugates. *Glycoconjugate J.* 5, 397–409 (1988). <https://doi.org/10.1007/BF01049915>

Edge, A. S., & Spiro, R. G. (1987). Presence of an O-glycosidically linked hexasaccharide in fetuin. *The Journal of Biological Chemistry*, 262(33), 16135–16141.

Falick, A. M., Medzihradzsky, K. F., & Walls, F. C. (1990). Artifacts in four-sector tandem mass spectrometry. *Rapid Communications in Mass Spectrometry: RCM*, 4(9), 318–322. <https://doi.org/10.1002/rcm.1290040905>

Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., & Whitehouse, C. M. (1989). Electrospray ionization for mass spectrometry of large biomolecules. *Science (New York, N.Y.)*, 246(4926), 64–71. <https://doi.org/10.1126/science.2675315>

Frese, C. K., Altelaar, A. F., van den Toorn, H., Nolting, D., Griep-Raming, J., Heck, A. J., and Mohammed, S. (2012) Toward full peptide sequence coverage by dual fragmentation combining

electron-transfer and higher-energy collision dissociation tandem mass spectrometry. *Anal. Chem.* 84, 9668–9673

Frese, C. K., Zhou, H., Taus, T., Altelaar, A. F., Mechtler, K., Heck, A. J., & Mohammed, S. (2013). Unambiguous phosphosite localization using electron-transfer/higher-energy collision dissociation (EThcD). *Journal of Proteome Research*, 12(3), 1520–1525. <https://doi.org/10.1021/pr301130k>

Ganguly, P., & Fossett, N. G. (1984). The role of sialic acid in the activation of platelets by wheat germ agglutinin. *Blood*, 63(1), 181–187.

Gill, D. J., Clausen, H., & Bard, F. (2011). Location, location, location: new insights into O-GalNAc protein glycosylation. *Trends in Cell Biology*, 21(3), 149–158. <https://doi.org/10.1016/j.tcb.2010.11.004>

Good, D. M., Wirtala, M., McAlister, G. C., & Coon, J. J. (2007). Performance characteristics of electron transfer dissociation mass spectrometry. *Molecular & cellular proteomics: MCP*, 6(11), 1942–1951. <https://doi.org/10.1074/mcp.M700073-MCP200>

Goth, C. K., Halim, A., Khetarpal, S. A., Rader, D. J., Clausen, H., & Schjoldager, K. T. (2015). A systematic study of modulation of ADAM-mediated ectodomain shedding by site-specific O-glycosylation. *Proceedings of the National Academy of Sciences of the United States of America*, 112(47), 14623–14628. <https://doi.org/10.1073/pnas.1511175112>

Goth, C. K., Tuhkanen, H. E., Khan, H., Lackman, J. J., Wang, S., Narimatsu, Y., Hansen, L. H., Overall, C. M., Clausen, H., Schjoldager, K. T., & Petäjä-Repo, U. E. (2017). Site-specific O-Glycosylation by Polypeptide N-Acetylgalactosaminyltransferase 2 (GalNAc-transferase T2) Co-regulates β 1-Adrenergic Receptor N-terminal Cleavage. *The Journal of Biological Chemistry*, 292(11), 4714–4726. <https://doi.org/10.1074/jbc.M116.730614>

Guzman-Aranguez, A., & Argüeso, P. (2010). Structure and biological roles of mucin-type O-glycans at the ocular surface. *The Ocular Surface*, 8(1), 8–17. [https://doi.org/10.1016/s1542-0124\(12\)70213-6](https://doi.org/10.1016/s1542-0124(12)70213-6)

Hägglund, P., Matthiesen, R., Elortza, F., Højrup, P., Roepstorff, P., Jensen, O. N., and Bunkenborg, J. (2007) An enzymatic deglycosylation scheme enabling identification of core fucosylated N-glycans and O-glycosylation site mapping of human plasma proteins. *J. Proteome Res.* 6, 3021–3031

Halim, A., Brinkmalm, G., Rüetschi, U., Westman-Brinkmalm, A., Portelius, E., Zetterberg, H., Blennow, K., Larson, G., and Nilsson, J. (2011) Site-specific characterization of threonine, serine, and tyrosine glycosylations of amyloid precursor protein/amyloid beta-peptides in human cerebrospinal fluid. *Proc. Natl. Acad. Sci. U.S.A.* 108, 11848–11853

Halim, A., Nilsson, J., Rüetschi, U., Hesse, C., & Larson, G. (2012). Human urinary glycoproteomics; attachment site specific analysis of N- and O-linked glycosylations by CID and ECD. *Molecular & cellular proteomics: MCP*, 11(4), M111.013649. <https://doi.org/10.1074/mcp.M111.013649>

Halim, A., Rüetschi, U., Larson, G., and Nilsson, J. (2013) LC-MS/MS characterization of O-glycosylation sites and glycan structures of human cerebrospinal fluid glycoproteins. *J. Proteome Res.* 12, 573–584

Halim, A., Westerlind, U., Pett, C., Schorlemer, M., Rüetschi, U., Brinkmalm, G., Sihlbom, C., Lengqvist, J., Larson, G., and Nilsson, J. (2014) Assignment of saccharide identities through analysis of oxonium ion fragmentation profiles in LC-MS/MS of glycopeptides. *J. Proteome Res.* 13, 6024–6032

Hillenkamp, F., Karas, M., Beavis, R. C., & Chait, B. T. (1991). Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers. *Analytical Chemistry*, 63(24), 1193A–1203A. <https://doi.org/10.1021/ac00024a002>

Hoffmann, M., Marx, K., Reichl, U., Wührer, M., and Rapp, E. (2016) Site-specific O-glycosylation analysis of human blood plasma proteins. *Mol. Cell. Proteomics* 15, 624–641

Huberty, M. C., Vath, J. E., Yu, W., & Martin, S. A. (1993). Site-specific carbohydrate identification in recombinant proteins using MALD-TOF MS. *Analytical Chemistry*, 65(20), 2791–2800. <https://doi.org/10.1021/ac00068a015>

Jin, C., Kenny, D. T., Skoog, E. C., Padra, M., Adamczyk, B., Vitizeva, V., Thorell, A., Venkatakrisnan, V., Lindén, S. K., & Karlsson, N. G. (2017). Structural Diversity of Human Gastric Mucin Glycans. *Molecular & cellular proteomics: MCP*, 16(5), 743–758. <https://doi.org/10.1074/mcp.M116.067983>

Johnson, R. S., Martin, S. A., Biemann, K., Stults, J. T., & Watson, J. T. (1987). Novel fragmentation process of peptides by collision-induced decomposition in a tandem mass spectrometer: differentiation of leucine and isoleucine. *Analytical Chemistry*, 59(21), 2621–2625. <https://doi.org/10.1021/ac00148a019>

Kannagi, R., Yin, J., Miyazaki, K., & Izawa, M. (2008). Current relevance of incomplete synthesis and neo-synthesis for cancer-associated alteration of carbohydrate determinants--Hakomori's concepts revisited. *Biochimica et Biophysica Acta*, 1780(3), 525–531. <https://doi.org/10.1016/j.bbagen.2007.10.007>

Kannagi, R., Sakuma, K., Miyazaki, K., Lim, K. T., Yusa, A., Yin, J., & Izawa, M. (2010). Altered expression of glycan genes in cancers induced by epigenetic silencing and tumor hypoxia: clues in the ongoing search for new tumor markers. *Cancer Science*, 101(3), 586–593. <https://doi.org/10.1111/j.1349-7006.2009.01455.x>

Kariya, Y., Kanno, M., Matsumoto-Morita, K., Konno, M., Yamaguchi, Y., & Hashimoto, Y. (2014). Osteopontin O-glycosylation contributes to its phosphorylation and cell-adhesion properties. *The Biochemical Journal*, 463(1), 93–102. <https://doi.org/10.1042/BJ20140060>

Karlsson, N. G., Schulz, B. L., & Packer, N. H. (2004). Structural determination of neutral O-linked oligosaccharide alditols by negative ion LC-electrospray-MSn. *Journal of the American Society for Mass Spectrometry*, 15(5), 659–672. <https://doi.org/10.1016/j.jasms.2004.01.002>

Kebarle, P., & Verkerk, U. H. (2009). Electrospray: from ions in solution to ions in the gas phase, what we know now. *Mass Spectrometry Reviews*, 28(6), 898–917. <https://doi.org/10.1002/mas.20247>

King, S. L., Joshi, H. J., Schjoldager, K. T., Halim, A., Madsen, T. D., Dziegiel, M. H., Woetmann, A., Vakhrushev, S. Y., & Wandall, H. H. (2017). Characterizing the O-glycosylation

landscape of human plasma, platelets, and endothelial cells. *Blood Advances*, 1(7), 429–442.

<https://doi.org/10.1182/bloodadvances.2016002121>

Michael Kinter, Nicholas E. Sherman, Protein Sequencing and Identification Using Tandem Mass Spectrometry, 2000 *John Wiley & Sons, Inc.*, ISBN: 9780471322498

Kuhns, W., Rutz, V., Paulsen, H., Matta, K. L., Baker, M. A., Barner, M., Granovsky, M., & Brockhausen, I. (1993). Processing O-glycan core 1, Gal beta 1-3GalNAc alpha-R. Specificities of core 2, UDP-GlcNAc: Gal beta 1-3 GalNAc-R(GlcNAc to GalNAc) beta 6-N-acetylglucosaminyltransferase and CMP-sialic acid: Gal beta 1-3GalNAc-R alpha 3-sialyltransferase. *Glycoconjugate Journal*, 10(5), 381–394. <https://doi.org/10.1007/BF00731043>

Kuhns, W., Jain, R. K., Matta, K. L., Paulsen, H., Baker, M. A., Geyer, R., & Brockhausen, I. (1995). Characterization of a novel mucin sulphotransferase activity synthesizing sulphated O-glycan core 1,3-sulphate-Gal beta 1-3GalNAc alpha-R. *Glycobiology*, 5(7), 689–697. <https://doi.org/10.1093/glycob/5.7.689>

Lasky, L. A., Singer, M. S., Dowbenko, D., Imai, Y., Henzel, W. J., Grimley, C., Fennie, C., Gillett, N., Watson, S. R., & Rosen, S. D. (1992). An endothelial ligand for L-selectin is a novel mucin-like molecule. *Cell*, 69(6), 927–938. [https://doi.org/10.1016/0092-8674\(92\)90612-g](https://doi.org/10.1016/0092-8674(92)90612-g)

Leuenberger, B., Hahn, D., Pischitzis, A., Hansen, M. K., & Sterchi, E. E. (2003). Human meprin beta: O-linked glycans in the intervening region of the type I membrane protein protect the C-terminal region from proteolytic cleavage and diminish its secretion. *The Biochemical Journal*, 369(Pt 3), 659–665. <https://doi.org/10.1042/BJ20021398>

Lu, L., Riley, N. M., Shortreed, M. R., Bertozzi, C. R., & Smith, L. M. (2020). O-Pair Search with MetaMorpheus for O-glycopeptide characterization. *Nature Methods*, 17(11), 1133–1138. <https://doi.org/10.1038/s41592-020-00985-5>

Makino, M., Kojima, T., & Yamashina, I. (1966). Enzymatic cleavage of glycopeptides. *Biochemical and Biophysical Research Communications*, 24(6), 961–966. [https://doi.org/10.1016/0006-291x\(66\)90344-5](https://doi.org/10.1016/0006-291x(66)90344-5)

Manzi, A. E., Norgard-Sumnicht, K., Argade, S., Marth, J. D., van Halbeek, H., & Varki, A. (2000). Exploring the glycan repertoire of genetically modified mice by isolation and profiling of the major glycan classes and nano-NMR analysis of glycan mixtures. *Glycobiology*, 10(7), 669–689. <https://doi.org/10.1093/glycob/10.7.669>

Maynard, J. C., Burlingame, A. L., & Medzihradszky, K. F. (2016). Cysteine S-linked N-acetylglucosamine (S-GlcNAcylation), A New Post-translational Modification in Mammals. *Molecular & Cellular Proteomics: MCP*, 15(11), 3405–3411. <https://doi.org/10.1074/mcp.M116.061549>

Medzihradszky, K. F., Gillece-Castro, B. L., Settineri, C. A., Townsend, R. R., Masiarz, F. R., & Burlingame, A. L. (1990). Structure determination of O-linked glycopeptides by tandem mass spectrometry. *Biomedical & Environmental Mass Spectrometry*, 19(12), 777–781. <https://doi.org/10.1002/bms.1200191205>

Medzihradszky, K. F., Gillece-Castro, B. L., Townsend, R. R., Burlingame, A. L., & Hardy, M. R. (1996). Structural elucidation of O-linked glycopeptides by high energy collision-induced dissociation. *Journal of the American Society for Mass Spectrometry*, 7(4), 319–328. [https://doi.org/10.1016/1044-0305\(95\)00682-6](https://doi.org/10.1016/1044-0305(95)00682-6)

Medzihradszky, K. F., Kaasik, K., & Chalkley, R. J. (2015 (A)). Characterizing sialic acid variants at the glycopeptide level. *Analytical Chemistry*, 87(5), 3064–3071. <https://doi.org/10.1021/ac504725r>

Medzihradszky, K. F., Kaasik, K., & Chalkley, R. J. (2015 (B)). Tissue-Specific Glycosylation at the Glycopeptide Level. *Molecular & Cellular proteomics: MCP*, 14(8), 2103–2110. <https://doi.org/10.1074/mcp.M115.050393>

Mikesh, L. M., Ueberheide, B., Chi, A., Coon, J. J., Syka, J. E., Shabanowitz, J., & Hunt, D. F. (2006). The utility of ETD mass spectrometry in proteomic analysis. *Biochimica et Biophysica Acta*, 1764(12), 1811–1822. <https://doi.org/10.1016/j.bbapap.2006.10.003>

Munkley J. (2016). The Role of Sialyl-Tn in Cancer. *International Journal of Molecular Sciences*, 17(3), 275. <https://doi.org/10.3390/ijms17030275>

Nagata, Y., & Burger, M. M. (1974). Wheat germ agglutinin. Molecular characteristics and specificity for sugar binding. *The Journal of Biological Chemistry*, 249(10), 3116–3122.

Nilsson, J., Rüetschi, Halim, U. A., Hesse, C., Carlsohn, E., Brinkmalm, G., and Larson, G. (2009) Enrichment of glycopeptides for glycan structure and attachment site identification. *Nat. Methods* 6, 809–811

Pan, Y., Yago, T., Fu, J., Herzog, B., McDaniel, J. M., Mehta-D'Souza, P., Cai, X., Ruan, C., McEver, R. P., West, C., Dai, K., Chen, H., & Xia, L. (2014). Podoplanin requires sialylated O-glycans for stable expression on lymphatic endothelial cells and for interaction with platelets. *Blood*, 124(24), 3656–3665. <https://doi.org/10.1182/blood-2014-04-572107>

Pap, A., Medzihradzky, K. F., & Darula, Z. (2017). Using "spectral families" to assess the reproducibility of glycopeptide enrichment: human serum O-glycosylation revisited. *Analytical and Bioanalytical Chemistry*, 409(2), 539–550. <https://doi.org/10.1007/s00216-016-9960-7>

Pap, A., Klement, E., Hunyadi-Gulyas, E., Darula, Z., & Medzihradzky, K. F. (2018). Status Report on the High-Throughput Characterization of Complex Intact O-Glycopeptide Mixtures. *Journal of the American Society for Mass Spectrometry*, 29(6), 1210–1220. <https://doi.org/10.1007/s13361-018-1945-7>

Pap, A., Tasnadi, E., Medzihradzky, K. F., & Darula, Z. (2020). Novel O-linked sialoglycan structures in human urinary glycoproteins. *Molecular Omics*, 16(2), 156–164. <https://doi.org/10.1039/c9mo00160c>

Park, G. W., Lee, J. W., Lee, H. K., Shin, J. H., Kim, J. Y., & Yoo, J. S. (2020). Classification of Mucin-Type O-Glycopeptides Using Higher-Energy Collisional Dissociation in Mass Spectrometry. *Analytical Chemistry*, 92(14), 9772–9781. <https://doi.org/10.1021/acs.analchem.0c01218>

Peng, J., Jiang, J., Wang, W., Qi, X., Sun, X. L., & Wu, Q. (2011). Glycosylation and processing of pro-B-type natriuretic peptide in cardiomyocytes. *Biochemical and Biophysical Research Communications*, 411(3), 593–598. <https://doi.org/10.1016/j.bbrc.2011.06.192>

- Pinho, S. S., & Reis, C. A. (2015). Glycosylation in cancer: mechanisms and clinical implications. *Nature Reviews. Cancer*, 15(9), 540–555. <https://doi.org/10.1038/nrc3982>
- Pollex-Krüger, A., Meyer, B., Stuike-Prill, R., Sinnwell, V., Matta, K. L., & Brockhausen, I. (1993). Preferred conformations and dynamics of five core structures of mucin type O-glycans determined by NMR spectroscopy and force field calculations. *Glycoconjugate Journal*, 10(5), 365–380. <https://doi.org/10.1007/BF00731042>
- Powell, J. T., & Brew, K. (1974). The preparation and characterization of two forms of bovine galactosyl transferase. *European Journal of Biochemistry*, 48(1), 217–228. <https://doi.org/10.1111/j.1432-1033.1974.tb03759.x>
- Reisner, Y., Linker-Israeli, M., and Sharon, N. (1976) Separation of mouse thymocytes into two subpopulations by the use of peanut agglutinin. *Cell. Immunol.* 25, 129–134
- Ricciuto, J., Heimer, S. R., Gilmore, M. S., & Argüeso, P. (2008). Cell surface O-glycans limit *Staphylococcus aureus* adherence to corneal epithelial cells. *Infection and Immunity*, 76(11), 5215–5220. <https://doi.org/10.1128/IAI.00708-08>
- Rossez, Y., Maes, E., Lefebvre Darroman, T., Gosset, P., Ecobichon, C., Joncquel Chevalier Curt, M., Boneca, I. G., Michalski, J. C., & Robbe-Masselot, C. (2012). Almost all human gastric mucin O-glycans harbor blood group A, B or H antigens and are potential binding sites for *Helicobacter pylori*. *Glycobiology*, 22(9), 1193–1206. <https://doi.org/10.1093/glycob/cws072>
- Schwzyer, M., & Hill, R. L. (1977). Porcine A blood group-specific N-acetylgalactosaminyltransferase. *The Journal of Biological Chemistry*, 252(7), 2346–2355.
- M. Scigelova, A. Makarov: Orbitrap Mass Analyzer – Overview and Applications in Proteomics, Practical Proteomics 1-2/2006, DOI 10.1002/pmic.200600528
- Settineri, C. A., Medzihradzky, K. F., Masiarz, F. R., Burlingame, A. L., Chu, C., & George-Nascimento, C. (1990). Characterization of O-glycosylation sites in recombinant B-chain of platelet-derived growth factor expressed in yeast using liquid secondary ion mass spectrometry, tandem mass spectrometry and Edman sequence analysis. *Biomedical & Environmental Mass Spectrometry*, 19(11), 665–676. <https://doi.org/10.1002/bms.1200191106>

Smith, D. F., Podgorski, D. C., Rodgers, R. P., Blakney, G. T., & Hendrickson, C. L. (2018). 21 Tesla FT-ICR Mass Spectrometer for Ultrahigh-Resolution Analysis of Complex Organic Mixtures. *Analytical Chemistry*, 90(3), 2041–2047. <https://doi.org/10.1021/acs.analchem.7b04159>

Solecka, B. A., Weise, C., Laffan, M. A., & Kannicht, C. (2016). Site-specific analysis of von Willebrand factor O-glycosylation. *Journal of Thrombosis and Haemostasis: JTH*, 14(4), 733–746. <https://doi.org/10.1111/jth.13260>

Spiro, R. G., & Bhoyroo, V. D. (1974). Structure of the O-glycosidically linked carbohydrate units of fetuin. *The Journal of Biological Chemistry*, 249(18), 5704–5717.

Sricholpech, M., Perdivara, I., Yokoyama, M., Nagaoka, H., Terajima, M., Tomer, K. B., & Yamauchi, M. (2012). Lysyl hydroxylase 3-mediated glucosylation in type I collagen: molecular loci and biological significance. *The Journal of Biological Chemistry*, 287(27), 22998–23009. <https://doi.org/10.1074/jbc.M112.343954>

Stentoft, C., Vakhrushev, S. Y., Vester-Christensen, M. B., Schjoldager, K. T., Kong, Y., Bennett, E. P., Mandel, U., Wandall, H., Levery, S. B., & Clausen, H. (2011). Mining the O-glycoproteome using zinc-finger nuclease-glycoengineered SimpleCell lines. *Nature Methods*, 8(11), 977–982. <https://doi.org/10.1038/nmeth.1731>

Stults, J. T., & Watson, J. T. (1987). Identification of a new type of fragment ion in the collisional activation spectra of peptides allows leucine/isoleucine differentiation. *Biomedical & Environmental Mass Spectrometry*, 14(10), 583–586. <https://doi.org/10.1002/bms.1200141009>

Sukeno, T., Herp, A., & Pigman, W. (1972). Enzymic characterization of Golgi-rich fractions from rat submaxillary-sublingual glands. *European Journal of Biochemistry*, 27(3), 419–424. <https://doi.org/10.1111/j.1432-1033.1972.tb01854.x>

Sun, S., Yu, C., Qiao, Y., Lin, Y., Dong, G., Liu, C., Zhang, J., Zhang, Z., Cai, J., Zhang, H., & Bu, D. (2008). Deriving the probabilities of water loss and ammonia loss for amino acids from tandem mass spectra. *Journal of Proteome Research*, 7(1), 202–208. <https://doi.org/10.1021/pr070479v>

Sun, W., Liu, Y., Lajoie, G. A., Ma, B., & Zhang, K. (2019). An Improved Approach for N-Linked Glycan Structure Identification from HCD MS/MS Spectra. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 16(2), 388–395. <https://doi.org/10.1109/TCBB.2017.2701819>

Syka, J. E., Coon, J. J., Schroeder, M. J., Shabanowitz, J., & Hunt, D. F. (2004). Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proceedings of the National Academy of Sciences of the United States of America*, 101(26), 9528–9533. <https://doi.org/10.1073/pnas.0402700101>

Thanka Christlet, T. H., & Veluraja, K. (2001). Database analysis of O-glycosylation sites in proteins. *Biophysical Journal*, 80(2), 952–960. [https://doi.org/10.1016/s0006-3495\(01\)76074-2](https://doi.org/10.1016/s0006-3495(01)76074-2)

Thermo Fisher Scientific. (2017). Orbitrap Fusion Lumos Tribrid Mass Spectrometer: Breakthrough Gains for Quantitative Biology Sensitivity Transformed, Product Specifications, <https://planetorbitrap.com/orbitrap-fusion#tab:specifications>

Trinidad, J. C., Schoepfer, R., Burlingame, A. L., & Medzihradszky, K. F. (2013). N- and O-glycosylation in the murine synaptosome. *Molecular & Cellular Proteomics: MCP*, 12(12), 3474–3488. <https://doi.org/10.1074/mcp.M113.030007>

Vakhrushev, S. Y., Steentoft, C., Vester-Christensen, M. B., Bennett, E. P., Clausen, H., & Lavery, S. B. (2013). Enhanced mass spectrometric mapping of the human GalNAc-type O-glycoproteome with SimpleCells. *Molecular & Cellular proteomics: MCP*, 12(4), 932–944. <https://doi.org/10.1074/mcp.O112.021972>

Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., Darvill, A. G., Kinoshita, T., Packer, N. H., Prestegard, J. H., Schnaar, R. L., & Seeberger, P. H. (Eds.). (2015). *Essentials of Glycobiology*. (3rd ed.). Cold Spring Harbor Laboratory Press.

Vliegthart, J. F., & Casset, F. (1998). Novel forms of protein glycosylation. *Current Opinion in Structural Biology*, 8(5), 565–571. [https://doi.org/10.1016/s0959-440x\(98\)80145-0](https://doi.org/10.1016/s0959-440x(98)80145-0)

Wang, Y., Abernethy, J. L., Eckhardt, A. E., & Hill, R. L. (1992). Purification and characterization of a UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase specific for

glycosylation of threonine residues. *The Journal of Biological Chemistry*, 267(18), 12709–12716.

Wang, J., Bourne, P. E., & Bandeira, N. (2011). Peptide identification by database search of mixture tandem mass spectra. *Molecular & Cellular proteomics: MCP*, 10(12), M111.010017. <https://doi.org/10.1074/mcp.M111.010017>

Wen, D. X., Livingston, B. D., Medzihradzky, K. F., Kelm, S., Burlingame, A. L., & Paulson, J. C. (1992). Primary structure of Gal beta 1,3(4)GlcNAc alpha 2,3-sialyltransferase determined by mass spectrometry sequence analysis and molecular cloning. Evidence for a protein motif in the sialyltransferase gene family. *The Journal of Biological Chemistry*, 267(29), 21011–21019.

Wilson, I. B., Gavel, Y., & von Heijne, G. (1991). Amino acid distributions around O-linked glycosylation sites. *The Biochemical Journal*, 275 (Pt 2), 529–534. <https://doi.org/10.1042/bj2750529>

Wiśniewski, J., Zougman, A., Nagaraj, N. et al. Universal sample preparation method for proteome analysis. *Nat. Methods* 6, 359–362 (2009). <https://doi.org/10.1038/nmeth.1322>

Wu SW, Pu TH, Viner R, Khoo KH. Novel LC-MS² product dependent parallel data acquisition function and data analysis workflow for sequencing and identification of intact glycopeptides. *Anal. Chem.* 2014 Jun 3;86(11):5478-86.

Yabu, M., Korekane, H., & Miyamoto, Y. (2014). Precise structural analysis of O-linked oligosaccharides in human serum. *Glycobiology*, 24(6), 542–553. <https://doi.org/10.1093/glycob/cwu022>

Yamada, K., Hyodo, S., Kinoshita, M., Hayakawa, T., and Takechi, K. (2010) Hyphenated technique for releasing and MALDI MS analysis of O-glycans in mucin-type glycoprotein samples. *Anal. Chem.* 82, 7436–7443

Zhang, H., Li, X. J., Martin, D. B., and Aebersold, R. (2003) Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat. Biotechnol.* 21, 660–666



Zhang, F., Ge, W., Ruan, G., Cai, X., & Guo, T. (2020). Data-Independent Acquisition Mass Spectrometry-Based Proteomics and Software Tools: A Glimpse in 2020. *Proteomics*, 20(17-18), e1900276. <https://doi.org/10.1002/pmic.201900276>

Zhao, P., Viner, R., Teo, C. F., Boons, G. J., Horn, D., & Wells, L. (2011). Combining high-energy C-trap dissociation and electron transfer dissociation for protein O-GlcNAc modification site assignment. *Journal of Proteome Research*, 10(9), 4088–4104. <https://doi.org/10.1021/pr2002726>

Zhokhov, S. S., Kovalyov, S. V., Samgina, T. Y., and Lebedev, A. T. (2017) An EThcD-Based Method for Discrimination of Leucine and Isoleucine Residues in Tryptic Peptides. *J. Am. Soc. Mass Spectrom.* 28, 1600–1611

9. RÖVIDÍTÉSEK JEGYZÉKE

Gyakori monoszacharidok neveinek rövidítései.

<i>Monoszacharid neve</i>	<i>Rövidítése</i>	<i>Egyezményes jele*</i>
Glükóz	Glc	
Galaktóz	Gal	
Mannóz	Man	
Fukóz	Fuc	
N-acetil-glükózamin	GlcNAc	
N-acetil-galaktózamin	GalNAc	
N-acetil-neuraminsav	NeuAc	

* SNFG szimbólumrendszer: <https://www.ncbi.nlm.nih.gov/glycans/snfg.html>

ABC – ammónium-bikarbonát

AGC – automatic gain control

CI – chemical ionization (kémiai ionizáció)

CID – collision-induced dissociation (ütközéses aktiválás)

DDA – data dependent acquisition (adatfüggő mérés)

DIA – data independent acquisition (adatfüggetlen mérés)

DTT – ditiotreitól (1,4-bisz-(szulfanil)-bután-2,3-diol)

EI – elektron ionizáció

ER – endoplazmatikus retikulum

ESI – electrospray ionization (elektroporlasztásos ionizáció)

ETD – elektron transzfer disszociáció

EThcD – electron transfer/higher-energy C-trap dissociation

FASP – filter-aided sample preparation

HCD (beam-type CID) – higher-energy C-trap dissociation

HILIC – hydrophilic interaction liquid chromatography (hidrofil interakciós folyadékkromatográfia)

HPLC – high-performance liquid chromatography (nagy teljesítőképességű folyadékkromatográfia)

IAM – jód-acetamid

LC-MS – folyadékkromatográfiával kapcsolt tömegspektrometria

LC-MS/MS – folyadékkromatográfiával kapcsolt tandem tömegspektrometria

MALDI-TOF - matrix-assisted laser desorption/ionization – time of flight

MS/MS – tandem tömegspektrometria

NCE – normalized collision energy (normalizált ütközési energia)

PTM – poszttranszlációs módosítás

SNFG – symbol nomenclature for glycans (glikán szimbólum-nevezéktan)

WGA – wheat germ agglutinin (búzacsíra agglutinin)

10. MELLÉKLETEK

1. melléklet 3.3.2 fejezetben vázolt N-glikopeptid keresés során alkalmazott N-glikán adatbázis.

HexNAc(1)	HexNAc(4)Hex(5)
HexNAc(2)	HexNAc(4)Hex(5)NeuAc(1)
HexNAc(1)Fuc(1)	HexNAc(4)Hex(5)NeuAc(2)
HexNAc(2)Fuc(1)	HexNAc(4)Hex(5)NeuAc(1)Na(1)
HexNAc(2)Hex(1)	HexNAc(4)Hex(5)NeuAc(2)Na(1)
HexNAc(2)Hex(2)	HexNAc(4)Hex(5)NeuAc(2)Na(2)
HexNAc(2)Hex(3)	HexNAc(4)Hex(5)Fuc(1)
HexNAc(2)Hex(3)Fuc(1)	HexNAc(4)Hex(5)Fuc(1)NeuAc(1)
HexNAc(2)Hex(4)	HexNAc(4)Hex(5)Fuc(1)NeuAc(2)
HexNAc(2)Hex(5)	HexNAc(4)Hex(5)Fuc(1)NeuAc(1)Na(1)
HexNAc(2)Hex(5)Fuc(1)	HexNAc(4)Hex(5)Fuc(1)NeuAc(2)Na(1)
HexNAc(2)Hex(6)	HexNAc(4)Hex(5)Fuc(1)NeuAc(2)Na(2)
HexNAc(2)Hex(7)	HexNAc(4)Hex(6)NeuAc(1)
HexNAc(2)Hex(8)	HexNAc(4)Hex(6)Fuc(1)NeuAc(1)
HexNAc(2)Hex(9)	HexNAc(4)Hex(6)Fuc(1)NeuAc(1)Na(1)
HexNAc(3)Hex(4)	HexNAc(5)Hex(3)
HexNAc(3)Hex(4)NeuAc(1)	HexNAc(5)Hex(4)
HexNAc(3)Hex(5)	HexNAc(5)Hex(5)
HexNAc(3)Hex(5)NeuAc(1)	HexNAc(5)Hex(5)Fuc(1)
HexNAc(3)Hex(6)	HexNAc(5)Hex(5)NeuAc(1)
HexNAc(3)Hex(6)NeuAc(1)	HexNAc(5)Hex(6)NeuAc(1)
HexNAc(3)Hex(6)Fuc(1)NeuAc(1)	HexNAc(5)Hex(6)NeuAc(2)
HexNAc(4)Hex(3)	HexNAc(6)Hex(3)
HexNAc(4)Hex(3)Fuc(1)	HexNAc(6)Hex(3)Fuc(1)
HexNAc(4)Hex(4)	HexNAc(6)Hex(3)Fuc(1)NeuAc(1)
HexNAc(4)Hex(4)Fuc(1)	HexNAc(6)Hex(3)Fuc(1)NeuAc(2)
HexNAc(4)Hex(4)NeuAc(1)	HexNAc(6)Hex(4)NeuAc(1)
HexNAc(4)Hex(4)NeuAc(1)Na(1)	HexNAc(6)Hex(4)NeuAc(2)

4. melléklet 3.3.4 fejezetben vázolt acetilált NeuAc-t tartalmazó O-glikánokkal módosított O-glikopeptidek azonosításához használt glikánadatbázis. Az ábrán szereplő szerkezeteket az EThcD spektrumok manuális kiértékelésével igazoltuk.

Összetétel	Monoizotópos additív tömeg	Szerkezet
HexNAcHex	365.132	
HexNAcHexNeuAc	656.228	
HexNAcHexNeuAcAc	698.238	
HexNAcHexNeuAc ₂	947.323	
HexNAcHexNeuAcNeuAcAc	989.334	
HexNAcHex(NeuAcAc) ₂	1031.344	
HexNAcHexNeuAc ₃	1238.418	
HexNAcHexNeuAc ₂ NeuAcAc	1280.429	
HexNAcHexNeuAc(NeuAcAc) ₂	1322.440	
HexNAcHex(NeuAcAc) ₃	1364.450	
HexNAcHexNeuAc ₄	1529.514	
HexNAcHexNeuAc ₃ NeuAcAc	1571.524	
HexNAcHexNeuAc ₂ (NeuAcAc) ₂	1613.535	
HexNAcHexNeuAc(NeuAcAc) ₃	1655.546	
HexNAcHex(NeuAcAc) ₄	1697.556	
HexNAc ₂ Hex ₂ NeuAcAc	1063.370	
HexNAc ₂ Hex ₂ NeuAc ₂	1312.455	
HexNAc ₂ Hex ₂ NeuAcNeuAcAc	1354.466	
HexNAc ₂ Hex ₂ NeuAcAc ₂	1396.476	
HexNAc ₂ Hex ₂ NeuAc ₃	1603.551	
HexNAc ₂ Hex ₂ NeuAc ₂ NeuAcAc	1645.561	
HexNAc ₂ Hex ₂ NeuAc(NeuAcAc) ₂	1687.572	
HexNAc ₂ Hex ₂ (NeuAcAc) ₃	1729.582	
HexNAc ₂ Hex ₂ NeuAc ₄	1894.646	
HexNAc ₂ Hex ₂ NeuAc ₃ NeuAcAc	1936.657	
HexNAc ₂ Hex ₂ NeuAc ₂ (NeuAcAc) ₂	1978.667	
HexNAc ₂ Hex ₂ NeuAc(NeuAcAc) ₃	2020.678	
HexNAc ₂ Hex ₂ NeuAcAc ₄	2062.688	

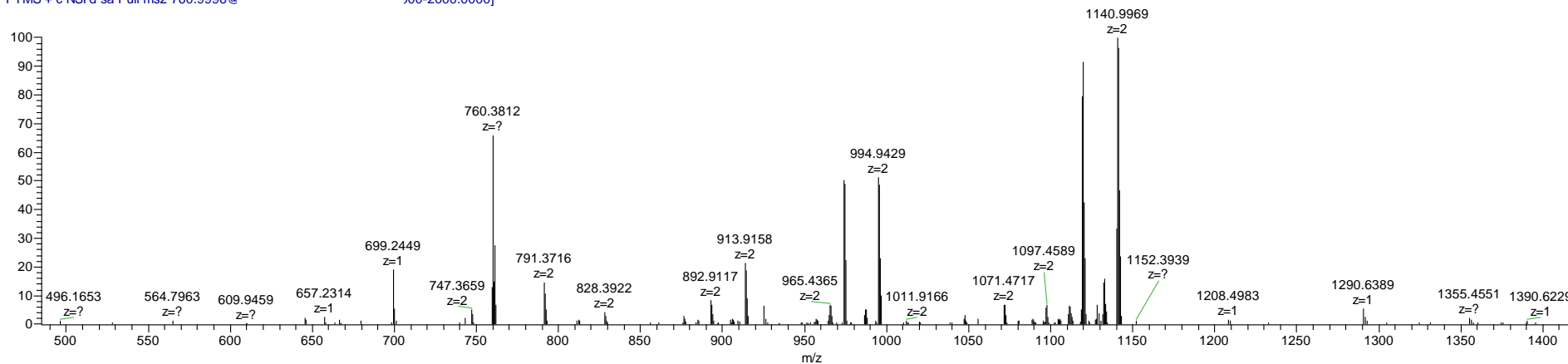
7. melléklet Összesítés a GF-Hunterrel azonosított O-glikán szerkezetekről.

Ssz.	Összetétel	Additív tömeg	Ssz.	Összetétel	Additív tömeg
1	HexNAcHex	365,1322	30	HexNAc ₃ Hex ₂ FucNeuAc	1370,497
2	HexNAcNeuAc	494,1478	31	HexNAc ₂ Hex ₂ NeuAcNeuAcAcNa	1376,448
3	HexNAcHexNeuAc	656,2276	32	HexNAc ₃ Hex ₃ NeuAc	1386,492
4	HexNAcHexNeuAcNa	678,2096	33	HexNAc ₂ Hex ₂ NeuAc ₂ Sulpho	1392,415
5	HexNAc ₂ NeuAc	697,2542	34	HexNAc ₂ Hex ₂ (NeuAcAc) ₂	1396,476
6	HexNAcHexNeuAcAc	698,2382	35	HexNAc ₂ Hex ₂ NeuAc ₂ SulphNa	1414,394
7	HexNAcHexNeuAcAcNa	720,2202	36	HexNAc ₂ Hex ₂ (NeuAcAc) ₂ Na	1418,458
8	HexNAc ₂ HexNeuAc	859,307	37	HexNAc ₂ Hex ₂ FucNeuAc ₂	1458,513
9	HexNAc ₂ HexNeuAcAc	901,3176	38	HexNAc ₂ Hex ₂ FucNeuAc ₂ Na	1480,495
10	HexNAcHexNeuAc ₂	947,323	39	HexNAc ₂ Hex ₂ FucNeuAcNeuAcAc	1500,524
11	HexNAcHexNeuAc ₂ Na	979,305	40	HexNAc ₃ Hex ₂ NeuAc ₂	1515,535
12	HexNAcHexNeuAcNeuAcAc	989,3336	41	HexNAc ₂ Hex ₂ FucNeuAcNeuAcAcNa	1522,506
13	HexNAcHexNeuAcNeuAcAcNa	1011,316	42	HexNAc ₂ Hex ₂ Fuc(NeuAcAc) ₂	1542,534
14	HexNAc ₂ Hex ₂ NeuAc	1021,36	43	HexNAc ₂ Hex ₂ NeuAc ₂ NeuAcAc	1645,561
15	HexNAcHex(NeuAcAc) ₂	1031,344	44	HexNAc ₃ Hex ₃ NeuAc ₂	1677,587
16	HexNAc ₃ HexNeuAc	1062,386	45	HexNAc ₂ Hex ₂ NeuAc(NeuAcAc) ₂	1687,574
17	HexNAc ₂ Hex ₂ NeuAcAc	1063,37	46	HexNAc ₃ Hex ₃ NeuAc ₂ Na	1699,569
18	HexNAc ₂ Hex ₂ FucNeuAc	1167,418	47	HexNAc ₄ Hex ₃ FucNeuAc	1735,629
19	HexNAc ₃ Hex ₂ NeuAc	1224,439	48	HexNAc ₄ Hex ₄ NeuAc	1751,624
20	HexNAcHexNeuAc ₃	1238,418	49	HexNAc ₃ Hex ₃ FucNeuAc ₂	1823,645
21	HexNAcHexNeuAc ₂ NeuAcAc	1280,429	50	HexNAc ₃ Hex ₃ FucNeuAc ₂ Na	1845,627
22	HexNAcHexNeuAc ₂ NeuAcAcNa	1302,411	51	HexNAc ₃ Hex ₃ Fuc ₂ NeuAc ₂	1969,703
23	HexNAc ₂ Hex ₂ NeuAc ₂	1312,455	52	HexNAc ₃ Hex ₃ Fuc ₂ NeuAc ₂ Na	1991,685
24	HexNAcHexNeuAc(NeuAcAc) ₂	1322,44	53	HexNAc ₄ Hex ₄ NeuAc ₂	2042,72
25	HexNAc ₂ Hex ₃ FucNeuAc	1329,471	54	HexNAc ₃ Hex ₃ NeuAc	2116,756
26	HexNAc ₂ Hex ₂ NeuAc ₂ Na	1334,437	55	HexNAc ₄ Hex ₄ NeuAc ₃	2333,815
27	HexNAcHexNeuAc(NeuAcAc) ₂ Na	1344,422	56	HexNAc ₃ Hex ₄ FucNeuAc ₂	2391,857
28	HexNAc ₂ Hex ₂ NeuAcNeuAcAc	1354,466	57	HexNAc ₃ Hex ₅ NeuAc ₂	2407,852
29	HexNAcHexNeuAcNeuAcAcNeuAcAcAc	1364,45			

8. melléklet A 21. ábra kiterjesztett m/z skálájú EThcD spektrumai, amelyeken láthatóak az m/z 496,166, illetve az m/z 1355,46 ionok.

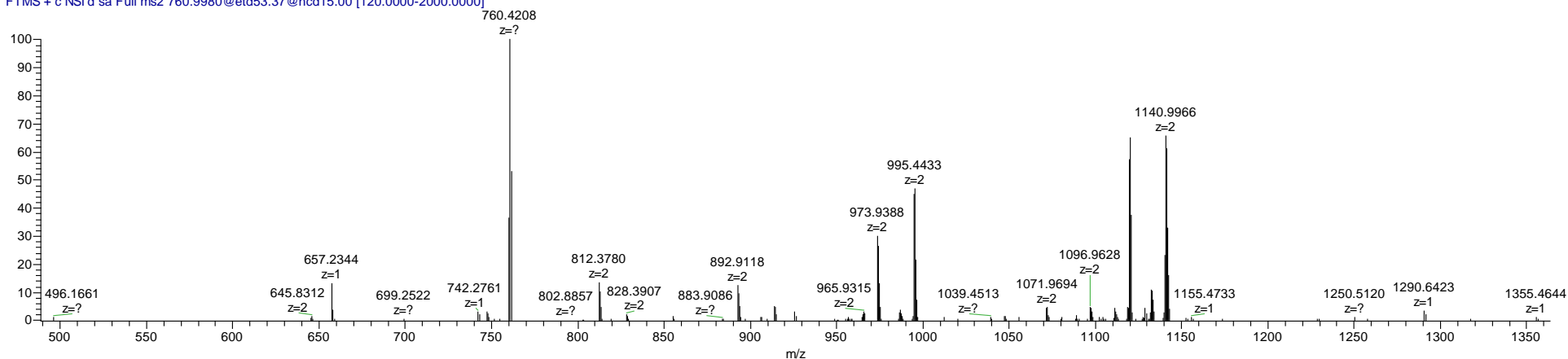
170919_08 #4177 RT: 14.95 AV: 1 NL: 1.00E5
 T: FTMS + c NSI d sa Full ms2 760.9996@

100-2000.0000]



170919_08 #5406 RT: 18.29 AV: 1 NL: 6.69E5
 T: FTMS + c NSI d sa Full ms2 760.9980@etd53.37@hcd15.00 [120.0000-2000.0000]

100-2000.0000]



11. SAJÁT KÖZLEMÉNYEK LISTÁJA

MTMT azonosító: 10058600

Összesített impakt faktor: 18.593

A dolgozat alapjául szolgáló közlemények:

Pap, A., Klement, E., Hunyadi-Gulyas, E., Darula, Z., & Medzihradzsky, K. (2018). Status Report on the High-Throughput Characterization of Complex Intact O-Glycopeptide Mixtures. *JOURNAL OF THE AMERICAN SOCIETY FOR MASS SPECTROMETRY*, 29(6), 1210–1220. <http://doi.org/10.1007/s13361-018-1945-7>

IF: 3.255 (2019)

Darula, Z., **Pap, A.**, & Medzihradzsky, K. F. (2019). Extended Sialylated O-Glycan Repertoire of Human Urinary Glycoproteins Discovered and Characterized Using Electron-Transfer/Higher-Energy Collision Dissociation. *JOURNAL OF PROTEOME RESEARCH*, 18(1), 280–291. <http://doi.org/10.1021/acs.jproteome.8b00587>

IF: 4.074 (2019)

Pap, A., Tasnadi, E., Medzihradzsky, K. F., & Darula, Z. (2020). Novel O-linked sialoglycan structures in human urinary glycoproteins. *MOLECULAR OMICS*, 16(2), 156–164. <http://doi.org/10.1039/c9mo00160c>

IF: 2.273 (2019)

További közlemények:

Pap, A., Medzihradzsky, K., & Darula, Z. (2017). Using “spectral families” to assess the reproducibility of glycopeptide enrichment: human serum O-glycosylation revisited. *ANALYTICAL AND BIOANALYTICAL CHEMISTRY*, 409(2), 539–550. <http://doi.org/10.1007/s00216-016-9960-7>

IF: 3.637 (2019)

Pap, A., Prakash, A., Medzihradzsky, K. F., & Darula, Z. (2018). Assessing the reproducibility of an O-glycopeptide enrichment method with a novel software, Pinnacle. *ELECTROPHORESIS*, 39(24), 3142–3147. <http://doi.org/10.1002/elps.201800223>

IF: 3.081 (2019)

Chalkley, R. J., Medzihradzsky, K. F., Darula, Z., **Pap, A.**, & Baker, P. R. (2020). The effectiveness of filtering glycopeptide peak list files for Y ions. *MOLECULAR OMICS*, 16(2), 147–155. <http://doi.org/10.1039/c9mo00178f>

IF: 2.273 (2019)