

Bioinformatikai és proteomikai módszerek fejlesztése és alkalmazása a glikoproteomikai kutatásokban

Doktori Értekezés

Ozohanic Olivér

okleveles vegyész

Eötvös Loránd Tudományegyetem, Természettudományi Kar, Kémia Doktori Iskola
Szintetikus Kémia, Anyagtudomány és Biomolekuláris Kémia Doktori Program

Doktori Iskola vezetője: Dr. Inzelt György
Doktori Program vezetője: Dr. Perczel András

Témavezető:
Dr. Drahos László
tudományos főmunkatárs

MTA Természettudományi Kutatóközpont, Szerves Kémiai Intézet
Tömegspektrometriai Laboratórium



Budapest, 2012

Köszönetnyilvánítás

Elsőként szeretnék köszönetet mondani Dr. Drahos Lászlónak, témavezetőmnek, hogy szakmai tanácsaival és útmutatásaival segítette munkámat a tömegspektrometriai ismeretek elsajátításában és alkalmazásában.

Köszönettel tartozom Dr. Vékey Károlynak hasznos tanácsaiért, továbbá, hogy lehetővé tette, hogy kutatásaimat az MTA Természettudományi Kutatóközpont Tömegspektrometriai Osztályán végezzem.

Köszönöm Dr. Inzelt Györgynek, a doktori iskola vezetőjének és Dr. Perczel Andrásnak, a doktori program vezetőjének, hogy lehetővé tették doktori értekezésem elkészítését.

Ezúton mondok köszönetet Újszászy Kálmánnak útmutatásáért, a gyakorlati tömegspektrometriai tanításáért.

Köszönöm Dr. Ludányi Krisztinának a dolgozat javításában nyújtott felbecsülhetetlen segítségét.

Köszönetemet szeretném kifejezni dr. Turiák Lillának a kísérleti munkában nyújtott szívélyes segítségéért.

Köszönetet szeretnék mondani a Tömegspektrometriai Osztály mindazon munkatársainak, (Dr. Budai Líviának, Gömörly Ágnesnek, Fabio Marinonak, Dr. Imre Tímeának, Jekő Anita, Krenyácz Juditnak, Dr. Pollreisz Ferencnek, Sergio Indelicatonak, Grádné Szabó Ritának, Tóth Eszternek) akikkel együtt dolgozhattam. Tanácsaikkal, építő megjegyzéseikkel, kísérleti munkámban nyújtott segítségükkel hozzájárultak ismereteim bővítéséhez és kutatói munkám előrehaladásához.

A legnagyobb köszönet szüleimet és testvéreimet illeti, akik végtelen türelemmel és támogatással segítettek doktori munkám elkészítését.

Tartalomjegyzék

Köszönetnyilvánítás.....	1
Tartalomjegyzék.....	2
Rövidítés jegyzék.....	4
1 Bevezetés.....	6
2 Célkitűzés.....	8
3 Irodalmi összefoglaló.....	9
3.1 A tömegspektrometria.....	9
3.2 A proteomika.....	11
3.3 A tömegspektrometrián alapuló proteomika.....	13
3.3.1 Fehérjék azonosítása.....	14
3.3.2 Fehérjék kvantitatív analízise.....	16
3.3.3 Poszt-transzlációs módosítások meghatározása.....	17
3.4 A glikoziláció.....	18
3.5 A glikoproteomika.....	20
3.6 Glikoproteinek szerkezetjellemezése.....	21
3.7 Glikopeptidek tandem tömegspektrumát értékelő programok.....	25
3.8 A proteomikában szokásosan használt statisztikai, kemometriai módszerek.....	27
3.9 Az alfa-1 savas glikoprotein.....	29
3.10 Az atherosclerosis.....	31
4 A kísérletek és számítások részletei.....	33
4.1 Felhasznált vegyszerek, minták, alkalmazott készülékek, szoftverek.....	33
4.1.1 Vegyszerek.....	33
4.1.2 Minták.....	33
4.1.3 Műszerek.....	33
4.1.4 Szoftverek.....	33
4.2 Fehérjék proteolízisére használt protokollok.....	34
4.2.1 A GlycoMiner fejlesztés során az AGP triptikus emésztésére használt protokoll.....	34
4.2.2 Vérplazmából származó AGP triptikus emésztésére használt protokoll.....	34
4.2.3 Plazma minta fehérjéinek emésztése (dúsítás előtt).....	35
4.2.4 Plazma minta fehérjéinek emésztése (dúsítás után).....	35
4.3 Peptidek, glikopeptidek (fehérje emésztmények) nanoLC-MS(MS) vizsgálata.....	36

4.3.1	Nanoáramlásos folyadékkromatográfia és tömegspektrometriás vizsgálat	36
4.3.2	nanoLC-MS/MS adatokon alapuló fehérje azonosítás	37
4.3.3	(nano)LC-MS/MS adatokon alapuló glikopeptid azonosítás	37
4.4	Glikoproteinek/glikopeptidek dúsítására alkalmazott módszerek	38
4.4.1	Lektin affinitás kromatográfia	38
4.4.2	Fenilbórsavas (PBA) affinitás kromatográfia	38
4.4.3	Előre ekvilibrált anioncserélő SPE	38
4.5	Lineáris diszkriminancia analízis egészséges és beteg csoportok megkülönböztetésére	39
5	Eredmények és értékelésük	40
5.1	Helyspecifikus glikozilációs mintázat meghatározása	40
5.2	GlycoMiner: Glikopeptidek tandem tömegspektrumának automatikus értékelése	43
5.3	A GlycoMiner kezelőfelülete	50
5.4	A glikopeptid összetételt meghatározó algoritmus részletei	52
5.5	GlycoPattern: minor glikoformok azonosítása és a glikozilációs mintázat meghatározása.....	55
5.6	Példák glikozilációs mintázatok meghatározására	60
5.7	Glikopeptidek és glikoproteinek dúsítása	61
5.8	A helyspecifikus glikozilációs mintázat alkalmazása orvosi-diagnosztikai feladat megoldására: AGP glikozilációjának változása atherosclerosis esetén.....	66
6	Összefoglalás	72
7	Summary	75
8	Irodalomjegyzék	76

Rövidítés jegyzék

AGP	alfa-1-savas glikoprotein
AMRT	accurate mass measurement and corresponding retention time, pontostömeg mérés és hozzá tartozó retenciós idő
ASCII PKL	peak list, csúcs lista fájl formátum
BIRD	blackbody infrared radiation dissociation, feketetest infravörös sugárzás disszociáció
BIS-TRIS	bis(2-hidroxietyl)amino-tris(hidroxietyl)metán
CE	capillary electrophoresis, kapilláris elektroforézis
CI	chemical ionization, kémiai ionizáció
CID	collision induced dissociation, ütközéses aktiválással történő fragmentáció
Con A	konkanavalin A
CRM	charge residue model, „maradék töltés” modell
CT	computer tomography, komputertomográfia
Da	dalton
DDA	data dependent analysis, adatfüggő analízis
DNS	dezoxiribonukleinsav
DTT	1,4-ditio-DL-treitolt
ECD	electron capture dissociation, elektron befogásos disszociáció
EI	electron impact ionization, elektronionizáció
ELISA	enzyme linked immunosorbent assay, ezimhez kapcsolt immunszorbens vizsgálat
emPAI	exponentially modified protein abundance index, exponenciálisan módosult fehérje mennyiségi index
ESI	electrospray ionization, elektroporlasztásos ionizáció
ETD	electron transfer dissociation, elektron átadásos disszociáció
FA	formic acid, hangyasav
FDA	Food and Drug Administration, az Amerikai Egyesült Államok élelmezési és gyógyszerügyi hivatala
Fuc	fukóz
HILIC	hydrophilic interaction liquid chromatography, hidrofíl interakción alapuló folyadékkromatográfia
GlcNAc	<i>N</i> -acetyl-glükózamin
Glu-C	V8-endoproteináz
Hex	hexóz
Hex₃HexNAc₂	<i>N</i> -glikán egység
HexHexNAc	laktózamin
HexHexNAcNeuAc	szialil-laktózamin
HexNAc	<i>N</i> -acetyl-hexózamin
HexNeuAc	hexóz + szialil sav
ICAT	isotope coded affinity tags, izotóp kódolt affinitás jelölés
IEM	ion evaporation model, ionpárolgás modell

IRMPD	infrared multiphoton disszociation, több infravörös fotonnal kiváltott disszociáció
ITRAQ	isobaric tagging reagents, izobár reagensekkel történő jelölés
LC	liquid chromatography, folyadékkromatográfia; a dolgozatban nagynyomású és ultranagynyomású folyadékkromatográfiát is jelöl
LC-MS	liquid chromatography mass spectrometry, folyadékkromatográffal kapcsolt tömegspektrometria,
LC-MS/MS	liquid chromatography mass spectrometry, folyadékkromatográffal kapcsolt tandem tömegspektrometria
LDA	linear discriminant analysis, lineáris diszkriminancia analízis
LDL	low density lipoprotein, kis sűrűségű lipoprotein
Le^x	Lewis X glikán struktúra
m/z	tömeg/töltés
MALDI	matrix assisted laser desorption/ionization, mátrixszal segített lézer deszorpció
MGF	Mascot generic file, Mascot általános fájl formátum
MS/MS	tandem tömegspektrometriai mérés
MudPit	multidimensional protein identification technology, többdimenziós fehérje azonosítási technológia
nanoESI	nanoelectrospray, nanoliteres áramlásra optimált elektroporlasztásos ionforrás
nanoLC	nanoliquid chromatography, nanoáramlású folyadékkromatográf
NeuAc	<i>N</i> -acetyl-neuraminsav
NIH	National Institutes of Health, Amerikai nemzeti egészségügyi intézmények
ORM	orosomucoid, alfa-1-savas glikoprotein másik elnevezése
PAI	protein abundance index, fehérje mennyiségi index
PBA	fenilbórsav
PGC	porózus graphyte chromatography, porózus grafit kromatográfia
PLGS	ProteinLynx Global Server, adatbázis keresésen alapuló tandem tömegspektrum kiértékelő számítógépes program
PMF	peptide mass fingerprint, peptid tömeg ujjlenyomat
PNGaseF	peptid <i>N</i> -endoglikozidáz F
PPM	parts per million, rész a millióban
RP-C18	fordított fázisú, C18 láncokat tartalmazó kromatográfiás töltet
RPM	percenkénti fordulatszám
Sia	sziálsav
SILAC	stable isotope labeling by amino acids in cell culture, stabil izotópokkal történő jelölés sejt kultúrában aminosavakon keresztül
TOF (MALDI-TOF)	time of flight, repülési idő analizátor
UHPLC	ultrahigh pressure/performance liquid chromatography, ultranagy nyomású/teljesítőképességű folyadékkromatográfia
WGA	wheat germ agglutinin, búzacsira agglutinin
XML formátum	extensible markup language, bővíthető jelölő nyelv

1 Bevezetés

Az emberi génállomány feltérképezése, a genomika átütő sikere után a kutatók figyelme a fehérjék szekvenciájának, szerkezetének és funkcióinak megismerésének irányába fordult. Ez az új tudományterület proteomika néven vált ismertté. A proteomika, és az ezzel szoros kapcsolatban levő biomolekuláris tömegspektrometria világszerte a legdinamikusabban fejlődő kutatási irányok közé tartoznak; a kémiai-biokémiai kutatásokra fordított erőforrások jelentős és egyre növekvő hányada ebbe az irányba fókuszál. Az informatika és a műszerezettség párhuzamos fejlődése lehetővé teszi a tömegspektrometria rutinszerű felhasználását különböző biológiai minták vizsgálatára. Így lehetőség nyílt – többek között a legtöbbet vizsgált makromolekulák – a fehérjék mennyiségének, szekvenciájának és poszt-transzlációs módosulásainak meghatározására és jellemzésére.

A glikoziláció az emberi fehérjék egyik leggyakoribb poszt-transzlációs módosítása, amely számos biológiai folyamatban kulcsfontosságú szerepet játszik. A fehérjék glikozilációjának vizsgálata egyre nagyobb hangsúlyt kapott az elmúlt években: a glikoproteinek szerkezetének pontos ismerete hozzájárul számos biokémiai folyamat mélyebb megértéséhez. Ezek között szerepel a sejtek közötti kommunikáció és az immunfelismerés is, amelyek különböző betegségek kialakulásában meghatározó jelentőséggel bírnak. Az oligoszacharid struktúrák szerkezetében bekövetkező változások felismerése, azonosítása számottevő orvosi diagnosztikai jelentőséggel bír. Ezt jelzi az is, hogy az Egyesült Államok Élelmiszeri és Gyógyszerügyi Hivatala (FDA, Food and Drug Administration) által elfogadott tumor biomarkerek többsége glikoprotein. Mindemellett napjainkban egyre nagyobb számban jelennek meg bioszimiláris, fehérje alapú gyógyszer-molekulák, melyek a glikoziláció pontos jellemzése nélkül nem kerülhetnek forgalomba.

A genetikával és a proteomikával összehasonlítva a glikozilációval kapcsolatos ismereteink azonban még kezdeti stádiumban járnak. Ennek fő oka, hogy az oligoszacharid struktúrák szerkezetvizsgálata és analitikája a mai napig számos nehézséggel küzd. Az egyik legnagyobb probléma, hogy a glikoproteinek nagyszámú, de ugyanakkor kis koncentrációjú homológok, úgynevezett glikoformok formájában vannak jelen a szervezetben. Ezek szelektív dúsítására még nagy érzékenységgű tömegspektrometria alkalmazása esetén is szükség van. Ugyanezen okból, más szerkezetvizsgálati módszerek kis érzékenységgük miatt csak korlátozottan alkalmazhatók. Ezért fontos új dúsítási, mérési és értékelési módszerek kidolgozása. Bár a glikoziláció vizsgálatában a tömegspektrometria alapú módszerek

meghatározóak, továbbfejlesztésük elengedhetetlen ahhoz, hogy a glikoziláció biokémiai folyamatokban és az élő szervezetben betöltött szerepének feltárásában jelentős előrelépés történjen.

2 Célkitűzés

Doktori munkám kezdetekor a glikoziláció jellemzéséhez szükséges tömegspektrometriai módszerek csak részben álltak rendelkezésre. Mindez azonban soklépéses, hosszadalmas feladatot jelentett, még akkor is, ha a vizsgálandó glikoprotein nagy mennyiségben rendelkezésre állt. A tömegspektrometriás méréseket megfelelő számítógépes programok és (bio)informatikai háttér nélkül kellett értékelni, mely hosszú manuális munkát jelentett.

Munkám során célom volt egy olyan, tömegspektrometriás módszereken alapuló munkafolyamat kidolgozása, mely lehetővé teszi fehérjék helyspecifikus *N*-glikozilációjának vizsgálatát. A feladat része annak kidolgozása, hogy a módszer alkalmas legyen nagy áteresztőképességű vizsgálatokra – orvos-diagnosztikai feladatok megoldásánál ez ugyanis alapkövetelmény. Biológiai minták esetén komplex anyagkeverékekkel dolgozunk, ezért a folyamat minta-előkészítéssel kezdődik. Glikoproteinek minta-előkészítése esetén a cél részben ezek dúsítása, és több dúsítási módszer összehasonlításával a célra legmegfelelőbb technika kiválasztása.

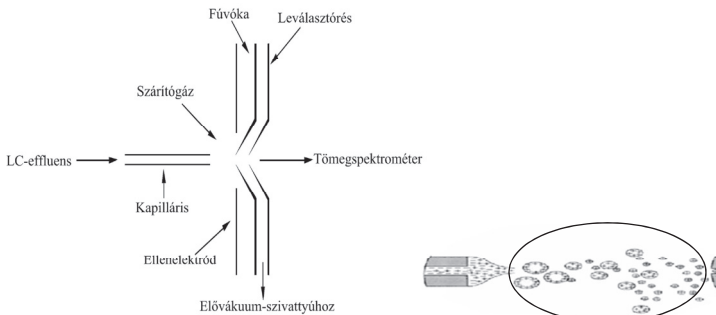
Ezt követően, munkám talán legfontosabb része a glikoproteinek szerkezetének jellemzése; ill. az erre szolgáló tandem tömegspektrometriás módszerek kidolgozása. A munka első lépése a vizsgált vegyületek fragmentációjának megértése. Ezt követően célunk volt az értékelés automatizálása; melyhez új számítógépes algoritmusok és programok kidolgozása szükséges. A fehérje glikoziláció heterogenitásának jellemzésére kombinált tömegspektrum/tandem tömegspektrum alapú munkafolyamat létrehozását terveztük, amely megbízható, pontos eredményt ad, és ugyanakkor széles körben elérhető.

Célom volt továbbá az így kidolgozott munkafolyamat segítségével a klinikai mintákból izolált fehérjeminták glikozilációjának jellemzése. Ennek segítségével lehetőség nyílik annak megállapítására, hogy van-e kapcsolat az adott betegség (érelmeszesedés) és a glikozilációs mintázat között.

3 Irodalmi összefoglaló

3.1 A tömegspektrometria

A tömegspektrométerben gázfázisú ionokat állítunk elő, amelyeket tömeg/töltés értékük szerint elválasztjuk, majd detektáljuk. Ennek eredményeként kapjuk a tömegspektrumot. A gázfázisú ionok előállításához különböző ionizációs technikák állnak rendelkezésre. Ezek közül a dolgozatban csak az elektroporlasztásos ionizációt (ESI, electrospray ionization) tárgyalom, mivel doktori munkám során ezt az ionizációs technikát használtam.



1. ábra: Electrospray ionforrás működése, valamint egy nagyított kapilláris és a folyadékcsseppek

Az ESI folyamat során (1. ábra) az „electrospray kapillárisra” pozitív potenciált kapcsolva az oldatban lévő pozitív töltésű ionok az oldat felszínén felhalmozódnak. Az ionok közötti taszítás, valamint a potenciálgradiens hatására a folyadék kúp alakot vesz fel (Taylor cone). Elegendően nagy potenciál esetén a felületi töltéssűrűségből származó elektrosztatikus taszítás meghaladja a felületi feszültséget, és többszörös töltésű cseppek válnak le a Taylor kúpról. A cseppek mérete függ az alkalmazott potenciáltól, az oldószertől és az áramlási sebességtől. A töltött cseppek az elektrosztatikus taszítás miatt eltávolodnak egymástól, így jön létre a porlasztás (spray). A cseppek a potenciálgradiens hatására az ellenlektrod és a belépő felé vándorolnak, miközben az oldószerpárolgás folyamatosan csökkenti méretüket. Az oldószér párolgását általában gázáram és intenzív fűtés is segíti, ez utóbbi a folyadék adiabatikus expanziója miatt bekövetkező hőmérséklet-csökkenést ellensúlyozza. A cseppek párolgása addig tart, míg a töltések közötti taszító erő meg nem haladja a folyadék felületi feszültségét (mely

összetartja a folyadékceppet), amit Rayleigh instabilitási határnak nevezünk. Ekkor az ún. Coulomb robbanás következik be, azaz számos kisebb méretű csepp keletkezik. A továbbra is sokszorosan töltött cseppekből két különböző mechanizmus szerint képződhetnek egyedi ionok:

A.) Charge residue model (CRM, „Maradék töltés” modell): A folyamatos oldószerpárolgás és a sorozatos Coulomb hasadások eredményeképpen olyan cseppek keletkeznek, melyek csak egyetlen (általában sokszorosan töltött) iont tartalmaznak. Ezekből további oldószer-evaporációval jönnek létre gázfázisú ionok.

B.) Ion evaporation model (IEM, Ionpárolgás): 10 nm átmérőnél kisebb, sokszorosan töltött cseppek esetén a Coulomb tasztítás következtében ion lökődhet ki a felületről.

E két folyamat általában párhuzamosan zajlik, azonban sok esetben az egyik válik dominánssá. Az eddigi ismeretek szerint a makromolekuláknál a CRM, a kis molekulatömegű vegyületek pedig IEM mechanizmus szerint játszódik le a folyamat.

Az ESI technika ionos, poláris és polarizálható vegyületek vizsgálatára kiválóan alkalmas, és egyszerűen kapcsolható kromatográfiai módszerekkel (LC, CE). A módszer páros elektronszámú kvázi molekulaionokat eredményez (pl. $[M + H]^+$, $[M + Na]^+$, $[M - H]^-$, $[M + Cl]^-$ stb), és a fragmentáció szinte elhanyagolható mértékű. Itt szeretném megemlíteni, hogy a tömegspektrumban észlelhetők időnként fragmensek, azonban ezek nem az electrospray folyamat eredményei, hanem később a készülékben bekövetkező fragmentáció termékei. Mindezek mellett az electrospray további előnye, hogy többszörös töltésű ionokat hoz létre (pl. $[M + nH]^{n+}$), lehetővé téve nagy molekulatömegű molekulák (proteinek, nukleinsavak, polimerek) vizsgálatát. A többszörös töltésnek köszönhetően a spektrumban az ionok kisebb m/z -nél láthatók, mint a molekulatömegük. Hátránya, hogy csak oldatok vizsgálatára alkalmas és nagy mennyiségben csak illékony oldószerek és pufferek használhatók. Az ESI módszert gyakran használják folyadékkromatográffal kapcsolva peptidek és fehérjék vizsgálatára. A gyakorlatban kis mintamennyiségek esetén igen nagy érzékenységű módszerre van szükség. Ilyenkor a kromatográfiát kis átmérőjű (kapilláris vagy nano) oszlopokon végzik, lassú eluens áramlás mellett. Az 1 μL /percnél kisebb áramlás mellett üzemelő folyadékkromatográfokat nanoáramlású folyadékkromatográfoknak nevezük (nanoLC). A nanoliteres áramlásra optimálalt elektroporlasztásos ionforrásokat nanoESI forrásnak nevezük.

A tömegspektrometriás vizsgálatok egyik fontos eszköze a tandem tömegspektrometria (MS/MS). Tandem tömegspektrometriának nevezzük azokat a tömegspektrometriás technikákat, amelyekben anyaion-leányion kapcsolat határozható meg. Egy tandem tömegspektrometriás kísérlet esetén a fragmentációra kiválasztott molekulát (iont) prekurzornak vagy anyaionnak nevezük. A fragmentáció eredményeként létrejövő fragmens ionokat leányionoknak nevezük.

A tandem tömegspektrometriai kísérletek egyik példája az ún. precursor ion scan, amely során kiválasztjuk a vizsgálni kívánt molekula iont, valamilyen módszerrel fragmentáltatjuk, majd a fragmenseket – tömeg/töltés szerinti elválasztás után – detektáljuk. A molekulák fragmentációját kiválthatjuk több módon is. Legelterjedtebb az ütközéses aktiválással történő fragmentáció (CID, [1]). Ennek során inert gázmolekulákkal (pl. Ar) ütköztetjük a célmolekulánkat. Az ütközések hatására a kinetikus energia belső energiává alakul, ami a molekulák fragmentációjához vezet. Fragmentációt kiválthatunk infravörös sugárással is (IRMPD, BIRD [2]). Az utóbbi tíz évben kifejlesztettek elektron befogással vagy leadással járó (electron capture dissociation [3] (ECD), electron transfer dissociation [4] (ETD)) aktivációs technikákat is. Ezek előnyösen használhatók többszörösen töltött molekulák vizsgálatánál, annál is inkább, mivel sok esetben a CID fragmentációval komplementer információt szolgáltatnak.

3.2 A proteomika

Marc Wilkins 1995-ben [5, 6] vezette be a proteom kifejezést. A proteom tradicionálisan a vizsgált organizmus által szintetizálható összes fehérjét, szűkebb értelemben az adott pillanatban az organizmusban jelen lévő saját fehérjéket jelenti. A különböző fehérjék száma sokkal nagyobb lehet, mint az őket kódoló géneké. A genommal ellentétben a proteome folyamatos változásnak van kitéve. A sejtben, a szövetekben, a szervezetben lévő fehérjeállomány (proteome) vizsgálatát foglalja össze a proteomika (proteins → proteome → proteomics [7]) kifejezés. A bővülő vizsgálati lehetőségek révén a proteomika eredeti jelentése több szempontból is tágabb értelmet nyert. A proteomika változó bonyolultságú fehérjeelemek vizsgálatát jelentheti. Az utóbbi időkben egyre növekvő hangsúlyt fektetnek a kvantitatív meghatározásra [8, 9] (fehérje-expresszió változásának vizsgálata). Vizsgálhatunk például olyan fehérjéket, amelyek bizonyos típusú komplexet képeznek (fehérje-DNS, fehérje-fehérje, vagy fehérje-kismolekula kölcsönhatás). A fehérjék poszt-transzlációs módosulásainak kiterjedt vizsgálata is lehetővé vált (diszulfidkötés kialakulása, oxidáció, glikoziláció, foszforiláció, ubiquitináció, acetiláció, metiláció, hidroxiláció, nitroziláció, szulfatálás). A proteomika céljai közé tartozik a sejt folyamatok, szabályozó folyamatok és betegségek mechanizmusának megismerése. További cél egyes gének és termékeik sejtbeli funkciójának meghatározása és megértése (funkcionális genomika) a fehérje szintjén (szekvencia, mennyiség, poszt-transzlációs módosításai, kölcsönhatásai, aktivitása, szubcelluláris eloszlása és szerkezete révén). Fontos proteomikai terület továbbá a gyógyszerfejlesztés és a biomarkerek keresése [10, 11]. A

proteomika, ezen belül a biomarkerek keresése az utóbbi tíz év egyik legdinamikusabban fejlődő kutatási területe.

A biomarker szót az 1999-es az amerikai NIH/FDA konferencián vezették be [12]. A biomarker a normális biológiai folyamatokat, patogén hatásokat és farmakológiai válaszokat jellemzi. Szűkebb értelemben egy organizmus (élőlény) biológiai és patológiai állapotára jellemző molekula vagy molekulák együttese. A proteomikai kutatások jelentős részének fókuszában olyan vizsgálatok állnak, amelyben a mintában található fehérjéket, mint biomarkerek nézik. A vizsgálatok a különböző patológiás egyénektől származó fehérjék közötti mennyiségi és/vagy minőségi különbségeket keresik patológiás esetekben. Ha sikertül patológiás állapottal összefüggő szignifikáns eltéréseket találni, akkor biomarker jelöltet találtak.

Az FDA által 2006-ig elfogadott és klinikumba bevezetett tumorbiomarkereket a 1. táblázat tartalmazza.

1. táblázat: FDA által elfogadott tumorbiomarkerek (2006)[13]

Biomarker	Betegség	Érzékenység	Szelektivitás
CEA	mellhártya (pleura) rosszindulatú daganatos megbetegedése	58%	79%
CEA	hashártya tumoros megbetegedése	76%	91%
Her-2/neu	IV stádiumú mellrák	40%	98%
Bladder Tumor Antigen	urothelialis sejt karcinóma	53%	70%
Thyro-globulin	pajzsmirigy tumor áttéte	75%	95%
Alpha-fetoprotein	hepatocellularis karcinóma	50%	70%
PSA	prosztata tumor	46%	91%
CA 125	nem kissejtes tüdőrák (NSCLC)	84%	80%
CA 19.9	hasnyálmirigy tumor	58%	96%
CA 15.3	mellrák	95%	95%
leptin, prolectin, osteopontin és IGF-II	ovárium tumor	95%	95%
CD98, fascin, sPIgR és 14-3-3 eta	tüdőrák	96%	77%

A táblázatban található biomarkerek érzékenysége és szelektivitása (lásd 27. oldal) tág határok között változik, emiatt szükséges további, még megbízhatóbb biomarkerek keresése. Jobb mutatókkal rendelkező biomarkerek kutatásának egyik fontos eleme lehet a megfelelő kísérlettervezés (kis mintaszám alapján néhány jelölt megtalálása, később nagy mintaszámú

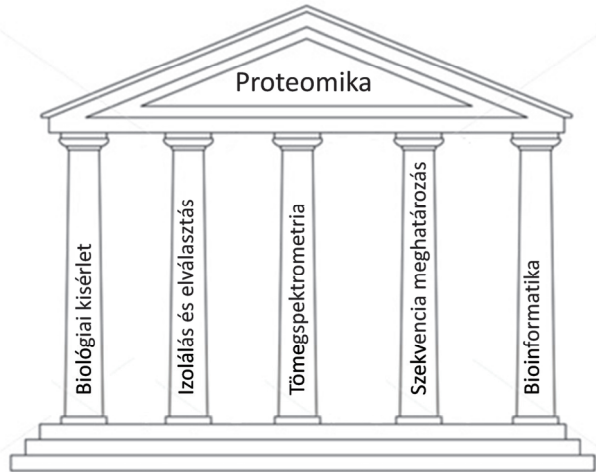
kísérletben a jelölt alkalmazhatóságának igazolása). Jelenleg FDA elfogadott protein biomarkerek többsége glikoprotein.

3.3 A tömegspektrometrián alapuló proteomika

A proteomikai munkafolyamat egy biológiai kísérletből, a fehérjék izolálásából és elválasztásából, azonosításból és/vagy mennyiségi meghatározásból áll. A proteomikai módszertan kezdetben főleg a tradicionális, biológiában elterjedt eszközökre támaszkodott. A gyakran vizsgált minták, mint a sejt lizátumok vagy a vér szinte kimeríthetetlen fehérje forrással szolgálnak. Anderson és Anderson rámutatott, hogy több mint tíz nagyságrend [14] mennyiségi eltérés is lehet egyes komponensek között. Tradicionálisan elektroforézis alapú (2D gél elektroforézis, kapilláris elektroforézis) módszereket használtak a fehérjék elválasztására. A módszerek fejlődésével egyre nagyobb teret nyertek a tömegspektrometriával kapcsolt folyadékkromatográfia alapú módszerek. Yates és munkatársai már 2001-ben közöltek 2D folyadékkromatográfián alapuló proteom [15] meghatározást. Bevezették a MudPit (multidimensional protein identification technology, többdimenziós fehérje azonosítási technológia) módszert, amely tandem tömegspektrometriát valamint adatbázis keresést használ a fehérjék azonosítására. A modern készülékek nagy érzékenységeknek, nagy felbontásának és tömegpontosságának köszönhetően a tömegspektrometria a proteomikai munkafolyamatok egyik leggyakrabban használt eszközévé vált.

Megszületett az úgynevezett tömegspektrometrián alapuló proteomika [17-19] (mass spectrometry based proteomics). A tömegspektrometrián alapuló fehérjeazonosítást és jellemzést, valamint a kvantitatív meghatározást egyesek a proteomika szűkebb részének tekintik, és egyúttal azokat a kutatásokat, amelyekben meghatározó szerepe van a tömegspektrometriának, eléggé önkényes módon 5 fő csoportra osztják [20]

1. fehérjék azonosítása
2. fehérjék (relatív vagy abszolút) kvantitatív analízise
3. fehérje-fehérje kölcsönhatások tanulmányozása
4. poszt-transzlációs módosítások vizsgálata
5. strukturális proteomika



2. ábra: A tömegspektrometria alapú proteomika alapkövei: Biológiai kísérlet, fehérjék izolálása és elválasztása, tömegspektrometriai mérés, szekvencia meghatározás és bioinformatikai feldolgozás[16].

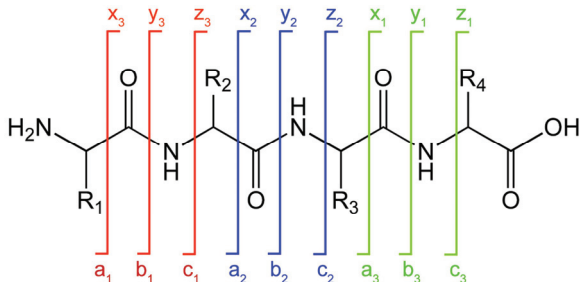
A tömegspektrometrián alapuló proteomika doktori munkámmal kapcsolatos csoportjait az alábbiakban röviden tárgyalom: ezek a fehérjék azonosítása, a relatív kvantitatív meghatározás és a poszt-transzlációs módosítások analízise.

3.3.1 Fehérjék azonosítása

Fehérjék azonosítására leggyakrabban az úgynevezett „bottom-up” (alulról-felfelé) technikát használják. A bottom-up technika lényege a fehérjeminta enzimatis, vagy kémiai úton történő hasítása („emésztése”) során keletkező peptidok meghatározása, majd ezek alapján a fehérje azonosítása. A peptidok azonosítására két tömegspektrometriás módszer terjedt el. Kezdetben a peptidokat főleg tömegük alapján azonosították az ún. PMF (peptide mass fingerprint, peptid tömeg ujjlenyomat) módszerrel. Egy meghatározott specificitású enzim hasítási termékeinek együttese jellemző az adott fehérjére. Ezt használja ki a PMF technika. Az egy fehérjéhez tartozó peptidok tömege egy ujjlenyomatot alkot, ami alapján adatbázis kereséssel meghatározható a fehérje. Az ujjlenyomat alapú azonosítás különösen többkomponensű minták vizsgálatánál nem mindig megbízható, sokszor elég nagy a hibás találatok aránya. A PMF vizsgálatokhoz tipikusan MALDI-TOF tömegspektrométert használtak. A módszer nagy

átteresztőképessége, érzékenysége és egyszerűsége miatt sokáig előszeretettel alkalmazott módszer volt a fehérjék azonosítására.

A másik – ma már – gyakrabban használt technika a peptid tandem tömegspektrometriás (MS/MS) fragmentációján alapuló azonosítása. E módszernél a peptid tömege mellett az aminosav szekvenciáról is nyerünk információt. A 3. ábrán peptid MS/MS spektrumában megfigyelhető lehetséges fragmensek láthatók. A peptidfragmentáció első nevezékantát P. Roepstorff [21] alkotta meg, amit Johnson és munkatársai továbbfejlesztettek [22]. Az, hogy melyik fragmensek képződnek az függ mind az aminosav szekvenciától, mind pedig a fragmentálás módjától (CID [1], ETD [4]). Ha a töltés az N-terminális felőli darabon marad, akkor az ionok a peptidkötés hasadási helyétől függően a, b, és c fragmensek képződhetnek. Ha a töltés a C-terminális oldalon lokalizálódik, x, y, és z fragmensek keletkezéséről beszélünk. Az MS/MS módszer alkalmas a peptid fragmenseiből annak szekvenciájának meghatározására, akár előzetes szekvencia információ nélkül is. Ha egy fehérjéhez hozzá tudunk rendelni a mintából azonosított peptideket, úgy tekintjük, hogy az adott fehérjét valamilyen valószínűséggel kimutattuk (azonosítottuk) az adott mintából. Az azonosítás megbízhatósága több tényezőtől függ, ezért általában minden fehérjéhez egy pontszámot társítanak, ami jellemzi az azonosítás megbízhatóságát.



3. ábra: Protonált peptid fragmentációja: Az MS/MS spektrumok esetében megfigyelhető lehetséges fragmensek: a, b, c fragmensek, esetén a töltés az N-terminális felőli darabon marad, illetve x, y, z fragmensek, ha a töltés a C-terminális oldalon lokalizálódik

A gyakorlatban proteolitikus emésztményekkel - peptid keverékekkel dolgozunk. A keverékek komponenseit a folyadékromatográfon szétválasztjuk, így már sok komponenset tartalmazó komplex minták vizsgálatára is lehetőség van. A fehérjekeverék proteolitikus emésztményének LC-MS/MS analízissel történő vizsgálatát „Shotgun Proteomics [23]”-nak nevezik.

Kromatográfiával kapcsolt mérések során számos komponensről lehetséges tandem tömegspektrum felvétele. Tipikusan valamilyen adatfüggő (DDA, data depend analysis) mérési módot választunk, amely a prekursor ion kiválasztásakor általában a legintenzívebb komponenseket részesítve előnyben. A fehérjék azonosítása a mérés során készült MS/MS spektrumok adatbázisban történő lekeresésén alapul. Ilyen adatbázis kereső programok (a teljesség igénye nélkül) például a Mascot [24], Protein Prospector [25], ProteinLynx Global Server, Sequest [26] és az X!Tandem [27]. Az adatbázis keresés felhasználja a peptid MS/MS spektrumát, amely alapján lehetséges aminosav szekvenciákat határoz meg, ezt összevetve ezt az adatbázisban lévő fehérjék szekvenciájával. Az adatbázis keresést megnehezíti a peptid nem várt kémiai (pl. oxidáció, ciklizáció) vagy poszt-transzlációs módosítása. További nehezítő tényező az adatbázisban nem szereplő izoformák jelenléte, valamint a használt proteolitikus enzim nem specifikus működése. Adatbázis keresésen alapuló peptid szekvencia meghatározáson kívül létezik még az úgynevezett „*de novo*” szekvenálás is. Ekkor kizárólag az MS/MS spektrumból nyerhető információkat használjuk fel a peptid szekvenciájának meghatározására. Ennek a technikának a hátránya, hogy nagyon jó minőségű MS/MS spektrum felvételét igényli [28, 29].

3.3.2 Fehérjék kvantitatív analízise

A fehérjék kvantitatív meghatározása feladattól függően elvégezhető abszolút vagy relatív módon is. A tömegspektrometria alapú proteomika esetén az abszolút mennyiségi meghatározás nehéz és csak standardok használatával oldható meg. Azonban relatív mennyiségi meghatározás standard nélkül is kivitelezhető, például izotóp jelzett komponensek egymáshoz viszonyított arányának meghatározásával. A továbbiakban csak a relatív mennyiségi meghatározás módszereit mutatom be.

A fehérjék analízise tipikusan proteolízis után történik. Mivel ez egy soklépéses feladat, a más területeken bevett gyakorlattal ellentétben a standardokat nem kalibrációs görbék kimérésére, hanem belső standardként alkalmazzák.

A relatív mennyiségi meghatározásra használt módszerek egy része izotóp jelölt standardokat használ. Ilyen módszerek például az ITRAQ (isobaric tagging reagents, izobár reagensekkel történő jelölés) [30], ICAT (isotope coded affinity tags, izotóp kódolt affinitás jelölés) [31] és a SILAC [32-35]. Az első két módszer izotóp kódolt reagenseket használ a peptid/fehérjék jelölésére. A SILAC sejt kultúrák jelölésére alkalmazható. Több sejt kultúra

esetén a használt táptalaj különböző izotópokkal jelölt aminosavakat tartalmaz. Ennek eredményeként a sejtek minden fehérjéje izotópkódolt lesz. Mennyiségi információt úgy kapunk, hogy egy kísérletben mérjük meg mind a jelölt és a jelöletlen mintákat. Ekkor a kapott intenzitás arányok azonosak a mintákban lévő komponensek mennyiségének arányával. Az izotópos jelölés legnagyobb hátránya a viszonylag költséges kivitelezés.

Egy másik megoldás a jelölés nélküli kvantitálás (ún. „label free” quantitation). A jelölés nélküli analízis egy vonzó alternatívája az izotópos jelölésnek. Legtöbbször relatív mennyiségi meghatározásra használják. Több változata is létezik aszerint, hogy LC-MS vagy LC-MS/MS spektrumok alapján végezzük a relatív összehasonlítást. Adatfüggő LC-MS/MS mérések alapján a fehérjék mennyiségének meghatározását leggyakrabban a spektrum számolást [36] (spectral count) használó módszerekkel oldják meg. Ezek a módszerek azon alapulnak, hogy egy bizonyos fehérjéhez hány spektrum tartozik [37, 38]. A különbözően expresszált fehérjék meghatározására a kvantitatív elemzést utólagosan statisztikai módszerekkel kombinálják [39].

LC-MS alapú méréseknél megfigyelték, hogy elektroporlasztásos ionizáció esetén a mért jelintenzitás lineárisan változik a koncentrációval [40, 41] (természetesen csak bizonyos koncentráció tartományokban, és akkor, ha reprodukálható vagy elhanyagolható mértékű az ion szuppresszió). LC-MS alapú módszereket dolgoztak ki komplex keverékek mennyiségi analízisére [42-44]. A fehérje mennyiségét a hozzátartozó peptidekhez tartozó intenzitások alapján becsülték. Az AMRT (accurate mass measurement and corresponding retention time, pontostömeg mérés és hozzá tartozó retenciós idő) módszer esetén az egyes peptidek retenciós idejét is figyelembe vették a mennyiségi meghatározásnál. Ez alapján relatív és abszolút mennyiségi meghatározási módszereket [45-47] is kidolgoztak. A fehérjék mennyiségét a fehérjéhez tartozó legintenzívebb jelet adó három peptid intenzitásának összegével jellemzik.

Munkám során a fehérjék mennyiségét a legintenzívebb jelet eredményező három peptid összegével jellemeztem.

3.3.3 Poszt-transzlációs módosítások meghatározása

Poszt-transzlációs módosítások jelenléte egy fehérjén a genomból mai tudásunk alapján nem jósolható meg. Az irodalomban közel 600 poszt-transzlációs módosítás ismert (UniMod, <http://www.unimod.org> (598), illetve RESID, <http://www.ebi.ac.uk/RESID> (572)) Az Unimod adatbázis kémiai és szekvencia módosulásokkal együtt, a dolgozat írásának időpontjában, 985 módosítást listáz. A leggyakoribb poszt-transzlációs módosítások a diszulfid-hidak, az oxidáció, az acetilezés, a metilezés, a foszforiláció, a szulfatálás, az ubiquitinálás és a glikoziláció. A

módosítások a fehérjék fizikai, kémiai és biológiai tulajdonságainak a megváltozásával járnak. Egy fehérje poszt-transzlációs módosításai jellemzőek lehetnek az adott sejt állapotára. Léteznek dinamikus és állandó módosulások. Az egyik legfontosabb dinamikus módosulás a foszforiláció, amely egyes enzimek aktivitását szabályozza. Sok kovalens módosítás a fehérjepopulációnak csak egy részét érinti (ilyen pl. a glikoziláció). A módosítások nagy száma, valamint dinamikus volta jelentősen megnehezíti a fehérjék vizsgálatát. Ezen kívül a fizikai tulajdonságok megváltozása miatt, módosulhatnak a peptidok kromatográfiás tulajdonságai, tömegspektrometriás detektálhatósága, illetve fragmentálódási tulajdonságai is. Ezek miatt az egyes poszt-transzlációs módosulások vizsgálata bonyolult feladat és az adott módosulásra speciálisan kidolgozott munkafolyamatot [48, 49] igényel. A továbbiakban a doktori munkámhoz kötődő glikozilációt és vizsgálatát tárgyalom részletesebben.

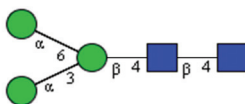
3.4 A glikoziláció

A glikoziláció az egyik leggyakoribb poszt-transzlációs módosítás, amely a humán fehérjék több mint 50%-at érinti [50]. A glikoziláció során a fehérjék peptidláncához oligoszacharid szerkezetek (glikánok) kötődnek. Háromféle glikozilációs módosítást különböztetünk meg, attól függően, hogy milyen aminosavhoz kötődnek a szacharidok: az *N*-, *O*- és *C*-glikozilációt. Ezek közül az *N*- és *O*- glikoziláció gyakrabban fordul elő, míg a *C*-glikoziláció ritka. Dolgozatomban részletesen az *N*-glikozilációval foglalkozom.

- *C*-glikoziláció esetén a glikánok a triptofán indolgyűrűjének 2-es helyzetű szénatomjához kapcsolódnak (például a komplement rendszer C6 fehérjéje [51]).
- *O*-glikoziláció során a fehérjék szerin, treonin és a tirozin aminosavainak hidroxil oldallánca szubsztituálódik szénhidrátokkal (egyik jellegzetes képviselője a mucinok [52]).
- *N*-glikoziláció esetén az aszparagin (ritkábban arginin) aminosavak oldallancának nitrogén atomjához kötődnek a glikánok (például sejtfelszíni receptor fehérjék esetén).

A glikoproteinek sajátossága a mikroheterogenitás jelensége. Ez glikoproteinek esetén azt jelenti, hogy egy fehérje különböző példányai, más és más oligoszacharid szerkezeteket hordoz(hat)nak.

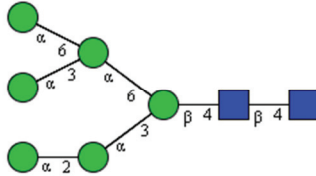
A humán glikoproteinek felépítésében leggyakrabban hexózok vesznek részt, pentózok ritkábban fordulnak elő. Előfordulnak ezek mellett a hexózok *N*-acetyl-származékai, szulfátjai és foszfátjai is. *N*-glikoziláció esetén a fehérjéhez kötődő oligoszacharidok szerkezete egy jól meghatározott mintát követ. A szerkezetek közös eleme egy öt szénhidrátból álló magstruktúra (core structure), amely az aminosavhoz kapcsolódik. Ez két *N*-acetyl-glükózamint és három mannózt tartalmaz meghatározott kötődési sorrendben (4. ábra). Amennyiben a molekula galaktózt tartalmaz, a galaktóz egy *N*-acetyl-glükózaminhoz kötődik, és együtt alkotnak egy laktózamin egységet. *N*-glikánok szíalsavakat (*N*-acetyl-neuramin sav) csak a láncok végén tartalmaznak.



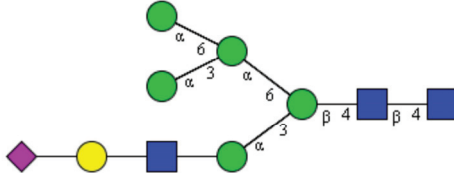
4. ábra: *N*-glikoziláció mag struktúra; a kék négyzetek *N*-acetyl-glükózamint, a zöld körök mannózt jelölnek.

Attól függően, hogy a magstruktúrához milyen egyéb szénhidrátok kötődnek, megkülönböztetünk „high mannose”, hibrid és komplex szerkezeteket (glikánokat). A „high mannose” glikánok a magstruktúrán kívül kizárólag mannózból állnak (5. ábra). A hibrid glikánok mind mannóz láncokat, mind laktózamin vagy szialil-laktózamin egységeket tartalmaznak (6. ábra). Amennyiben a glikán laktózamin antennákból épül fel, komplex glikánokról beszélünk (7. ábra).

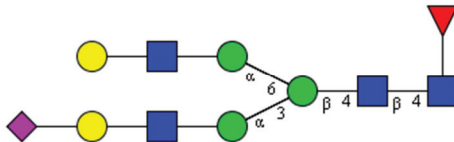
Egy fehérje *N*-glikozilációjának feltétele, a glikozilációs helyen egy úgynevezett konszenzus szekvencia megléte. A konszenzus szekvencia [53, 54] *N*-glikoziláció esetén *N*-X-S/T aminosavakból áll, ahol X nem lehet prolin. Egy aminosav szekvencia akár több konszenzus szekvenciát is tartalmazhat, azonban ezek nem mindegyikén történik glikoziláció, tehát nem glikozilációs helyek. Több glikozilációs hellyel rendelkező fehérjék esetén a glikán szerkezetek feltérképezésével jellemezhetjük a teljes glikozilációt. Az egyes glikozilációs helyekhez tartozó glikánok eloszlásának jellemzésével megkapjuk a helyspecifikus glikozilációt, tehát, hogy melyik glikozilációs helyhez milyen glikán kötődik, és milyen mennyiségben. A helyspecifikus glikoziláció meghatározása a glikoproteomika feladatai közé tartozik, és ez egyben a doktori értekezés fő témája is.



5. ábra: „High mannose” típusú glikán. A kék négyzetek *N*-acetil-glükózamint, a zöld körök mannózt jelölnek.



6. ábra: Hibrid típusú glikán. A kék négyzetek *N*-acetil-glükózamint, a zöld körök mannózt, a sárga körök galaktózt, a lila négyzetek szialsavat jelölnek.



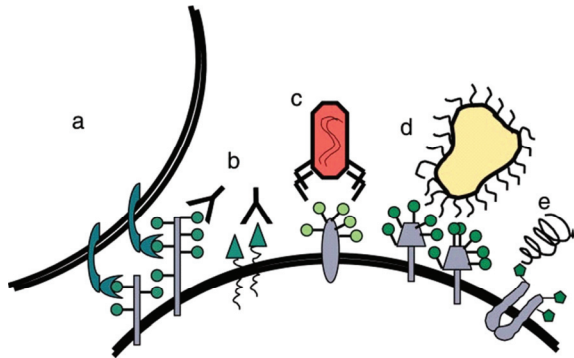
7. ábra: Komplex típusú glikán. A kék négyzetek *N*-acetil-glükózamint, a zöld körök mannózt, a sárga körök galaktózt, a lila négyzetek szialsavat a piros háromszög fukózt jelölnek.

3.5 A glikoproteomika

A glikoproteomika a proteomika azon része, amelynek célja a fehérjék glikozilációjának jellemzése. A glikoziláció jellemzése magában foglalja a glikánok szerkezeti meghatározást, kapcsolódási helyek felderítését, valamint az adott fehérje esetén glikoziláció funkciójának megértését. A fehérjék glikozilációjának sok különböző funkcióját mutatták ki. Attól függően, hogy melyik fehérjéről, és annak melyik glikozilációs helyéről van szó, a glikoziláció funkciója más és más lehet. Tehát egy adott funkció vizsgálatának csak egy adott fehérje esetén van értelme.

Az irodalomban számos publikáció foglalkozik a glikoziláció funkcióinak megértésével. E funkciók közül a legfontosabbak a következők:

- a fehérje végső térszerkezetének kialakítása (másodlagos és harmadlagos szerkezet), feltekeredés sebességének befolyásolása [55]
- a fehérje rigidításban és szerkezeti stabilitásában játszik szerepet [56]
- megváltoztatja az oldhatósági tulajdonságokat
- befolyásolja a fehérje szerkezetbeli biológiai felezési idejét [55, 57, 58]
- részt vesz a fehérje-fehérje kölcsönhatásban és felismerési folyamatokban: sejtek közötti felismerés és kommunikáció [59]
- alapvető szerepet tölt be az immunrendszer működésében és az immunválaszban (antigén) [60]



8. ábra: Glikoziláció funkciói: a) sejt-sejt interakció; b) antigén felismerés; c) és d) vírus illetve baktérium és gazdasejt receptor felismerés; e) toxin és receptor kölcsönhatás [59]

Helyspecifikus glikoziláció vizsgálata során egyrészt kimutatták, hogy esetenként a helyspecifikus glikoziláció gátolja a patogén T-sejt általi felismerését [61] (álcázás, saját anyagnak való tulajdonítás), valamint az autólízist [62] és a proteolízist is. Másrészt a helyspecifikus glikoziláció elősegíti a fehérje több lektinnel való kölcsönhatását [63].

3.6 Glikoproteinek szerkezetjellemezése

Glikoproteinek vizsgálatának céljai a fehérjék szerkezeti jellemzése, illetve biológiai szerepük megismerése. Emellett számos vizsgálat irányul a glikoproteinek biomarkerként való alkalmazhatóságára [11, 16, 64-66] is.

A vizsgálat első lépése a mintavétel, majd ezt követi a minták előkészítése. Leggyakrabban különböző betegségekben szenvedő egyének vér, vizelet és szövetmintáit vizsgálják, összefüggést keresve a glikoproteinek és a betegségek között. A minta-előkészítés a biológiai eredetű mintáknál természetesen elkerülhetetlen. Komplex biológiai mintákban jelen lévő glikoproteinek vizsgálatokor legtöbbször izolálási vagy dúsítási eljárásokat alkalmaznak [67]. Az izolálás során a vizsgálni kívánt glikoproteint elválasztjuk a többi komponenstől. Erre többféle lehetőség létezik: egy- vagy többdimenziós kromatográfiai technikák [68] (folyadék-kromatográfia, elektroforézis), immunaффinitáson [69] alapuló elválasztástechnikai módszerek. Dúsítás esetén a komponensek arányának megváltoztatásával érjük el, hogy a vizsgálni kívánt komponens a minta főkomponense legyen. A dúsítás lehetővé teszi csekély mennyiségben jelen lévő (minor) komponensek analitikai meghatározását. A dúsítás történhet lektin alapú módszerek [70, 71], gél-, HILIC [72] (hidrofil interakción alapuló folyadékkromatográfia), PGC [73, 74] (porózus grafit kromatográfia) kromatográfiai technikák alkalmazásával, TiO_2 [75] vagy ZrO_2 [76] affinitás segítségével és bórsav affinitással [77, 78]. A továbbiakban a dolgozatban is alkalmazott lektin affinitáson és a bórsav affinitáson alapuló módszereket tárgyalom részletesen.

A lektinek olyan fehérjék, amelyek szelektíven kötődnek bizonyos szénhidrátokhoz. A felismert szénhidrát struktúrák glikoproteinek vagy glikolipidek részei is lehetnek. A lektineket hosszú ideje használják a szénhidrátkémiában és a biokémiában különböző molekulák izolálására. Leggyakrabban alkalmazott lektinek a búzacsíra agglutinin (WGA, wheat germ agglutinin) és a konkanavalin. A (Con A) köszönhetően a széleskörű alkalmazhatóságuknak [79, 80]. A Con A főleg oligomannóz kisebb mértékben hibrid és kétantennás *N*-glikánokhoz kötődik. A WGA szelektíven kötődik *N*-acetyl-glükózaminhoz és szíálsavhoz (5-*N*-acetylneuraminsav). Mivel különböző lektineknek más és más a specificitásuk, minden glikánhoz található megfelelő lektin. A dúsítás technikai megvalósítása lehet kromatográfia, szilárd fázisú extrakció vagy mágneses részecskék felületére kikötött szorbens. Újabban Mechref és munkatársai bevezették az immobilizált lektinek használatát kvantitatív proteomikai vizsgálatokra [81]. Kimutatták, hogy az immobilizált lektinek jó reprodukálhatósággal használhatók kvantitatív vizsgálatokra. A bórsav bázikus körülmények között 1,2 vagy 1,3 ciszdiolokkal ciklikus észtereket képez. Bórsavas dúsítások esetén a glikánok kovalensen kötődnek az immobilizált bórsavhoz. A gyakorlatban a szilárdfázisú extrakció mellett a mágneses hordozóhoz kötött fenilbórsav terjedt el. A mágneses részecskékkel történő dúsítás kivitelezése egyszerűbb, de nagyobb mintatérfogatot igényel, mint a szilárd fázisú extrakció (ami már mikrométerben is kivitelezhető). A mágneses hordozónak megvan az az előnye is, hogy MALDI segítségével könnyen vizsgálhatóak [82]. A hátránya a bórsavas dúsításnak, hogy

kevésbé specifikus a lektinekhez képest [73, 83]. Nemrég Xu és munkatársai kifejlesztettek egy mezopórusos szilikahordozót [77] amellyel nagyobb hatékonyságot és egyszerű újra felhasználhatóságot értek el.

A minta-előkészítést követően a glikoproteinek vizsgálata történhet tömegspektrometrián (MS) alapuló módszerekkel, de egyéb eljárások is ismertek. Az egyéb módszerek közé soroljuk a lektinnel történő vizsgálatokat (lectin array [84]) és az ELISA-t (Enzyme Linked ImmunoSorbent Assay [85-87]). Ebbe a csoportba sorolhatók még a különböző elektroforézis alapú módszerek [88], mint például a kétdimenziós gélelektroforézis és az immunaffinitással kapcsolt kapilláris elektroforézis [89, 90]. A tömegspektrometriás analízist is magában foglaló technikák esetén dolgozhatunk intakt fehérjével (top-down módszer [91-93]) vagy a protein enzimatis, vagy kémiai hasítási termékeivel (bottom-up technikák). Gyakorlatban a bottom-up technikák sokkal elterjedtebbek. Esetenként a top-down és bottom-up módszereket együttesen alkalmazzák [94].

Bottom-up technikák során az enzimatis hasítást többféle módon végezhetjük, amikor is glikánokat vagy peptideket (és glikopeptideket) kapunk. A glikánok lehasítása történhet glikozidáz enzimekkel. Ezek közül *N*-glikoziláció esetén az egyik leggyakrabban alkalmazott enzim a peptid *N*-endoglikozidáz F (PNGase F) [95, 96] enzim, amely a peptid és a glikán közötti C-N kötést hidrolizálja. Kémiai hasításhoz trifluor-metánszulfonsavat, vagy hidrogén-fluoridot használnak. A peptid kötéseket proteolitikus enzimekkel hasítják („emésztik”). Leggyakrabban enzimatis hasításhoz tripszint, kimotripsint és Glu-C-t (V8-endoproteináz), kémiai hasításhoz CNBr-t alkalmaznak. A hasítást követi a tömegspektrometriás analízis.

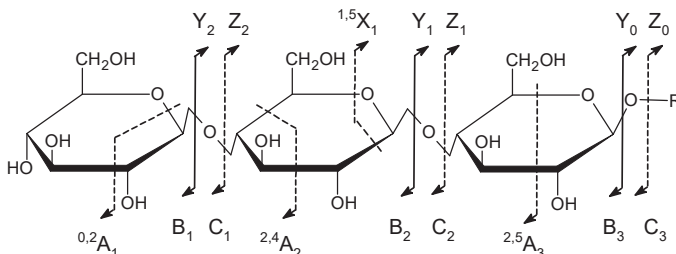
Lehasított glikánok tömegspektrometriás vizsgálata történhet MALDI, illetve ESI ionizáció használatával. Mindkét esetben főként a glikánok származékait vizsgálják, mivel ezek jobb érzékenységgel vizsgálhatók (Különösen jelentős az érzékenység növekedés a permetilezés esetén). Legelterjedtebb származékképzési eljárások a permetilezés [97, 98] és a 2-amino-benzoészav vagy savamid [99, 100] jelölés. A vizsgálatok során a glikánok a mért pontos tömegük alapján rendelhetők hozzá az elméleti összetételekhez. A permetilezett glikánok tandem tömegspektrometriás analízise során nagy gyakorisággal képződnek gyűrű kereszthasadási (cross-link) fragmensek. Ez fontos többletinformációt szolgáltat: lehetőség van a fragmensek alapján az egyes gyűrűk kapcsolódási helyének megállapítására. A lehasított glikánok vizsgálata nem szolgáltat információt arról, hogy a fehérjén hová kapcsolódott (azaz az ún. helyspecifikus glikozilációról). A mérési eredmények alapján a fehérjék átlagos glikozilációja határozható meg. A módszer hátránya, hogy több glikoproteint tartalmazó mintában, az összes glikoprotein együttes (átlagos) glikán mintázatát tudjuk csak meghatározni.

Glikoproteinek proteolitikus hasításakor peptid és glikopeptid keveréket kapunk. A proteolitikus emésztés során a glikopeptidek szénhidrát része nem változik meg, így lehetőség van ezek részletes vizsgálatára. A proteolízis során keletkező glikopeptidek tartalmazzák mind a fehérje glikozilációs helyre jellemző aminosavakat, mind a glikán részt. Az egyes glikopeptideket vizsgálva a glikozilációs helyek és helyspecifikus glikozilációs mintázat is meghatározható. Azonban nehezíti a vizsgálatokat, hogy a glikopeptidek a mintának csak egy kis részét teszik ki, továbbá a nagyfokú heterogenitás és a glikopeptidek kisebb tömegspektrometriás érzékenysége. A savas glikopeptidek általában, de a több szialsavat is tartalmazó glikopeptidek esetén még inkább igaz az, hogy kisebb az érzékenysége pozitív ionizációs módban (a proteomikában tipikusan alkalmazott kísérleti körülmények között). A gyakorlatban glikopeptidek vizsgálatára leggyakrabban folyadékkromatográffal kapcsolt tömegspektrometriát (LC-MS) és elektroporlasztásos ionizációt (ESI) használnak. Az LC-MS esetén egy töltött oszlopon elválasztjuk a vizsgálandó komponenseket. Az oszlop töltete alapján megkülönböztetünk fordított fázisú és normál fázisú módszereket. A fordított fázisú oszlop töltetek felülete leggyakrabban 18 szénatomot tartalmazó alkiláncokkal borított (RP-C18 oszlop). A fordított fázisú módszerek jól használhatóak mind peptid, mind glikopeptidek analízisére. RP-C18 oszlopok esetén a kromatográfiás retencióért főleg a peptid rész a felelős. A normál fázisú módszerek közül a glikopeptidek analízisére a HILIC (hidrofil interakción alapuló folyadékkromatográfia) módszer a legfontosabb. A HILIC kromatográfia során a peptid és glikopeptidek jelentősen szétválnak, így megkönnyítve a glikopeptidek azonosítását.

A folyadékkromatográfáról eluálódó glikopeptidekről szerkezeti információt tandem tömegspektrometriás kísérletekből nyerhetünk. A legtöbb kereskedelembe kapható tömegspektrométer esetén ütközései aktivációval (CID) lehet kiváltani a molekulák fragmentációját. A glikopeptidek fragmentálását [101] CID módszerrel végezve, a glikozid kötések hasadása nagyfokú, viszont a peptidkötések hasadása limitált [102]. E miatt CID módszerrel a peptid szekvencia meghatározása nehézkes, legtöbbször lehetetlen a tömegspektrum alapján. Ugyanakkor, a glikopeptidek CID alapú fragmentációja alkalmas a glikán rész szerkezetének a meghatározására. A tömegspektrometriai gyakorlatban a fragmentációs körülményeket a glikopeptidek vizsgálatára optimális körülményekre hangolják [103], hogy maximalizálják az MS/MS spektrum információ tartalmát.

A glikopeptidek CID fragmentációja nagymértékben hasonlít az oligoszacharidok fragmentációjára. Az oligoszacharidok fragmentációjának nevezékentát (9. ábra) Domon és Costello [104] alkotta meg. A fragmentáció során nagyfokú glikozidkötés hasadás figyelhető meg (B, C és Y, Z fragmensek). Glikopeptid CID spektrumban a B (oxónium) és Y fragmensek a

leggyakoribbak. Kisebb mértékben hasadhatnak a gyűrű kötése is, gyűrű keresztvasadási termékeket eredményezve (A és X fragmensek) [105, 106]. A keresztvasadási fragmensek fontos információt nyújthatnak a glikán egyes szacharidjainak kapcsolódási helyeiről.



9. ábra: Oligoszacharidok CID fragmentációja. A glikozid kötés hasadása B, C valamint Y és Z fragmenseket eredményez. A gyűrű kötéseinek hasadásakor A vagy X fragmensek képződnek

A glikopeptidek vizsgálatához újabban ECD vagy ETD fragmentációs technikák alkalmazására is van lehetőség [103, 107]. Ezen módszerek esetén a fragmentáció során a teljes glikán lehasad és a peptidre jellemző fragmensek képződnek, így lehetséges a peptid szekvencia meghatározása is. Az ECD és ETD technikák nem minden tömegspektrométeren érhetőek el. Probléma továbbá, hogy csak viszonylag kis tömeg/töltés tartományban működnek jól, és megadott töltéssel rendelkező prekuzort igényelnek [108, 109].

Glikoprotein LC-MS/MS méréseinek eredményeként akár több ezer MS/MS spektrum felvételére történik, amelyek manuális értékelése már ésszerűen nem kivitelezhető. Ezért a kutatások egyik igen fontos iránya automatizált rendszerek fejlesztése a glikopeptidek MS/MS és LC-MS/MS mérési adatainak értékelésére.

3.7 Glikopeptidek tandem tömegspektrumát értékelő programok

A glikopeptidek tandem tömegspektrumának kézi értékelése nehézkes és sok időt igényel: Először azt kell eldönteni, hogy a vizsgálni kívánt tandem tömegspektrum glikopeptidtől származik-e, vagy sem, ugyanis a triptikus emésztményben a legtöbb spektrum nem glikopeptidtől származik. Hosszú kromatográfiás mérések esetén ez hosszadalmas és bonyolult feladat, mivel az ilyen esetek több ezer spektrum kiértékelését igénylik. Míután sikerült találni egy olyan spektrumot, ami minden kétséget kizáróan glikopeptidtől származik, akkor következik a peptid és az oligoszacharid összetételének és szerkezetének megállapítása,

ami ugyancsak hosszadalmas. A legtöbb proteomikai programcsomag lényegesen megkönnyíti és lerövidíti az egyszerű proteomikai feladatok megoldását (fehérje azonosítás, kisebb mértékben *de novo* peptid szekvencia meghatározás). Azonban ezeknek a programoknak az alkalmazása poszt-transzlációs módosulások tanulmányozására korlátozott, ami különösen igaz, ha a feladat a glikozilációs módosulások azonosítása és szerkezet-meghatározása.

Figyelembe véve a glikoziláció kiemelkedő jelentőségét, számos szoftveres eszközt fejlesztettek ki a szerkezet-meghatározás nehézségeinek leküzdésére és a nagy áteresztőképességű munkában való segítségre [110-118]. Mivel a programok többsége egy célzott felhasználást/műszert szem előtt tartva készült, ezért legtöbbször csak egy bizonyos tudományos feladat megoldására alkalmas. Emellett sajnálatos módon a szoftverek közül csak egy érhető el szabadon (Glycomod [119]). A Glycomod programcsomag esetében azonban tesztleink nagy álpozitív találati arányt mutattak.

A különböző programok algoritmusai eltérő elképzelések alapján működnek. Mindegyik erősen támaszkodik arra, hogy az oligoszacharid rész összetételének változatossága aránylag kicsi, és ezért az elemi összetételek száma 1000 körül van. Ezt használja ki a Glycomod is, amely valószínűleg az első olyan program volt, ami a glikopeptidek MS/MS spektrumainak kiértékelését segítette. Ez a program proteolitikus fragmensek esetén képes kiszámítani egy adott tömeghez az összes lehetséges glikán és peptid kombinációt. Amennyiben a peptid rész már ismert, a glikán a listából kiválasztható. Különböző, egyéb programok is használják ezt az elvet ppm vagy ppm alatti tömegpontossággal párosítva. A glikopeptidek meghatározását kismértékben egyszerűsíti az a tény is, hogy a peptidek és a glikánok elemi összetétele között nagy a különbség. Ezeket a lehetőségeket jelenleg csupán két program használja ki [111, 118].

A glikopeptidek meghatározásának egy másik lehetősége a tandem MS-ből származó információ hasznosítása, a CID fragmentáció analízise. Ezen alapuló programok értelmezik a fragmentációs mintázatot, felhasználva mind a normál, mind a tandem tömegspektrumokból származó információkat [117, 118, 120]. A fragmentációs analízist értelemszerűen kiegészíti a glikopeptid pontos tömegmérése. A szoftver a peptid tömegét legtöbbször nem a mérések és a fragmentáció alapján határozza meg, hanem egy elméletileg lehetséges peptideket tartalmazó listából választja ki, majd ezt követően a peptidhez kapcsolódó glikán tömegét a mért értékek és a peptid különbségként számolja. Egy elméletileg lehetséges peptid tömeglista használatának legnagyobb problémája az, hogy fehérje keverékek esetén álpozitív eredményeket adhat. Tesztleink azt mutatják, hogy az álpozitív azonosítás előfordul már olyan keverékek esetén is, amelyekben mindössze 10 fehérje van.

Swamy [120] egy teljesen más megközelítést ír le, amely nem használ peptid listát, csak az MS/MS spektrumban jelen lévő szacharid fragmensek alapján dolgozik. Bár a megközelítés ígéretes, a használt algoritmus jó minőségű MS/MS spektrumok esetén is csak az esetek 67%-ban állapította meg helyesen a glikán összetételeket [120].

3.8 A proteomikában szokásosan használt statisztikai, kemometriai módszerek

A proteomikai vizsgálatok nagy részében az analízis célja két vagy több minta hasonlóságának vagy különbségeinek felderítése. Jó példa erre a biomarker keresés és a fehérje expressziós vizsgálatok. A feladat megoldására, a mérési adatokat csoportosítási vagy osztályozási eljárásokat használva, statisztikai elemzésének vetik alá. A csoportosítási (clustering) eljárások esetén hasonlóságot keresünk a vizsgált minták között úgy, hogy a csoportba tartozásukat illetően semmilyen előzetes ismeretünk vagy elvárásunk nincs. Az egyes minták csoportba sorolását csak a mért paramétereik által leírt „tulajdonságtérbeli” közelségük határozza meg. A csoportosítást nem felügyelt alakfelismerő (unsupervised pattern recognition) módszereknek is nevezzük. Ezzel szemben az osztályozási eljárások felügyelt alakfelismerő (supervised pattern recognition) módszerek. Osztályozás (classification) esetén azt vizsgáljuk, hogy a mintánk beletartozik-e valamelyik előre definiált csoportba (osztályba). Az osztályozás az egyes osztályok objektumainak egymás között hasonló tulajdonságain (proteomika esetén mért paraméterein) alapszik. Minden objektum osztályozása egy osztályozási szabály (függvény) alapján történik. Ez a szabály egy algoritmikusan megfogalmazott utasításkészlet, ami a tulajdonságtérben kijelöli az osztályhatárokat. Az osztályozási eljárásokat két csoportba sorolják: nem paraméteres és paraméteres vagy modellalapú osztályozás. Ezek közül a proteomikában gyakran használt lineáris diszkriminancia analízis (LDA) a második kategóriába tartozik.

A diszkriminancia analízis olyan adatelemzési módszer, amelyet kategóriába tartozás előrejelzésére lehet használni. A diszkriminancia analízis esetén kritériumváltozó kategóriákat jelöl, a független változók pedig számszerűek. Diszkriminancia analízis esetén az egyes osztályokba tartozó objektumokról feltételezik, hogy sokváltozós normális eloszlásfüggvénnyel jellemezhető sokdimenziós sokasághoz tartoznak. A modell egy sokváltozós eloszlás, illetve sűrűségfüggvény. Modellalapú osztályozáskor lényegében azt a valószínűséget számítjuk, amellyel egy objektum valamelyik eloszlásfüggvénnyel jellemzett osztályba kerülne. Ahhoz

hogy a diszkriminancia analízis alkalmazható legyen, a bemenő adatoknak meg kell felelniük a következőknek [121]:

- az adatok függetlenek legyenek, erre méréskor figyelni kell
- az osztályok mérete nagyjából ugyanakkora legyen
- a független változók normális eloszlásúak legyenek
- a lineáris diszkriminancia analízis esetében a különböző osztályokban a független változók varianciája hasonló (közel azonos) legyen (homoskedaszticitás)

A diszkriminancia analízis céljai többértékűek:

- olyan diszkriminancia függvények meghatározása, amely a független változók lineáris kombinációjaként a csoportokat a legjobban szétválasztja.
- annak meghatározása, mely becslő változók magyarázzák a legjobban a csoportok közötti különbséget.
- annak megállapítása, hogy van-e szignifikáns különbség a csoportok között a becslő változók alapján.
- az objektumok (proteomikában minták) valamely csoportba sorolása.

Az egyes osztálypárokhoz különböző diszkriminancia függvény tartozik. A diszkriminancia függvényt a független változók lineáris kombinációjaként keressük [122]. Az analízis kezdetén sem az eloszlásfüggvény paramétereit (modell paraméterek), sem a vizsgált objektumok tulajdonságai közötti kapcsolatot nem ismerjük. Ennek a problémának a megoldása úgynevezett „betanító halmaz” (training set) alkalmazása. A betanító halmaz olyan objektumokat tartalmaz, amelyekről biztosan tudott, hogy melyik osztályhoz tartoznak. Azok az objektumok, amelyeket azután osztályozni kell, az előrejelzési halmaz (test set) elemei. Ezeknek az elemeknek hovatarozása az osztályozó eljárás kimenő adata lesz. A betanító halmaz elemei alapján megalkotott osztályozási szabályt a gyakorlatban ellenőrzik. A leggyakrabban alkalmazott ellenőrzés során a betanító halmaz elemeit egyenként kihagyják, majd a nélküle megállapított modellbe visszahelyettesítik (leave-one-out cross-validation, egy-elem-kihagyásos keresztellenőrzés). Az eljárás a megállapított szabály szerint osztályozza az objektumokat. Ideális esetben minden elem saját, “valódi” osztályába jut, azaz abba, amelybe tartozónak azt a betanító halmazban megadták. Ha azonban nem a megfelelő módszert választották, vagy egyes objektumok kiugróak, a visszahelyettesítés bizonyos objektumokat eltérően sorol be. Az ekkor kapott eredmények az osztályozás előrejelzésbeli (*prediktív*) jóságát jellemzik [123]. Az osztályozás jóságát legegyszerűbben a helyesen besorolt objektumok arányával tudjuk jellemezni (l. egyenlet)

$$\text{Hiba arány}(\%) = \frac{n_{\text{helyes}}}{n_{\text{összes}}} \times 100$$

(1)

Orvosdiagnosztikai szempontból egy vizsgálatot többek között annak érzékenységeivel és specificitásával szokás jellemezni. Az érzékenység vagy szenzitivitás (2. egyenlet) annak a valószínűsége, hogy a diagnózis egy valódi beteg egyén esetén pozitív lesz-e. A specificitás (3. egyenlet) annak a valószínűsége, hogy egészséges egyén esetén a diagnózis negatív lesz-e.

$$\text{Érzékenység}(\%) = \frac{n_{\text{valós pozitív}}}{n_{\text{valós pozitív}} + n_{\text{álnegatív}}} \times 100 \quad (2)$$

$$\text{Specificitás}(\%) = \frac{n_{\text{valós negatív}}}{n_{\text{valós negatív}} + n_{\text{álpozitív}}} \times 100 \quad (3)$$

Biomarker keresés esetén a vizsgálat célja általában a diagnosztikai alkalmazás lehetőségének megállapítása. Fontos adat a módszer előrejelző képessége és specificitása, ezért minden esetben keresztellenőrzést kell végezni. A keresztellenőrzés hasznos módszer az osztályozás jóságának ellenőrzésére olyankor, amikor nem áll rendelkezésre elegendő tesztadat. A tömegspektrometrián alapuló proteomikai gyakorlatban az a tipikus, hogy nagyon sok független változónk (mérési pontunk) van, melyek akár korreláltak is lehetnek. Ugyanakkor a kutatás kezdeti stádiumában a mintaszám kicsi. Legtöbbször az egy csoportba eső minták száma a tízet sem éri el. Ilyen esetekben az első besorolás jósága és az egy-kihagyásos keresztellenőrzés besorolási jósága nagyon eltérhet egymástól. Az utóbbi a módszerünk előrejelző képességére utal, és sokkal kisebb lehet, mint amit nem keresztellenőrzött analízis adataiból számítunk. Ez azt jelzi, hogy a módszerünk nem robusztus.

3.9 Az alfa-1 savas glikoprotein

Proteomikai, glikoproteomikai módszerfejlesztéseink és klinikai vizsgálataink során az alfa-1-savas glikoprotein (AGP) helyspecifikus glikozilációját tanulmányoztuk. Ez a fiziológiailag fontos glikoprotein kiválóan alkalmas arra, hogy jelezze a szervezetben lejátszódó glikozilációs folyamatok megváltozását.

Az AGP, más néven orosomuroid (ORM), egy a humán plazmában található pozitív akut fázis fehérje. Fiziológiás körülmények között vérplazma koncentrációja kb. 1 mg/mL. Gyulladásos folyamatokban mennyisége akár tízszeresére is növekedhet. Az AGP legfontosabb

biológiai szerepe az immunrendszer működésének szabályozása [124], és a bázikus molekulák megkötése és szállítása [125, 126]. Az AGP az immunválasz során a gyulladásos reakciókat befolyásolja: stimulálja a limfocitákat, illetve citokinek expresszióját váltja ki. Az AGP expressziója főként a hepatocitákban, ritkábban epitél és endotél sejtekben, valamint makrofágokban történik.

Az AGP rendkívül heterogén fehérje, a humán plazmában - aminosav sorrend alapján - négy genetikai variánsa különböztethető meg (AGP F1, AGP F2, AGP S és az AGP A [127]). A genetikai variánsokat három különböző gén kódolja (AGP-A, AGP-B, AGP-B'): AGP-A gén az ORM1 lókuszon az AGP F1, AGP F2, valamint az AGP S genetikai variánsokat, AGP-B, ill. az AGP-B' gének pedig az ORM2 lókuszon az AGP-A variánst kódolják. A molekula heterogenitását jelentős mértékben növelik - az aminosav sorrend heterogenitása mellett - az öt *N*-glikozilációs helyhez kapcsolódó komplex oligoszacharid struktúrák. Az AGP oligoszacharid tartalma tömegének közel negyven százalékát teszi ki. Az oligoszacharid lánc változatosságának köszönhetően elméletileg több ezer különböző szerkezetű AGP izoform létezik.

Az oligoszacharidok egyszerűsített nevezéktanában [128] az első tag a laktózáminok számát mutatja (Bi-,Tri-,Tetra-). Az elnevezés ezt követő tagja a szialsav (S), ill. a fukóz (F) egységek jelenlétére utal, majd az azt követő szám ezen egységek számát mutatja. A BiS2 elnevezés például azt jelenti, hogy az oligoszacharid szerkezetében kettő laktózámin és kettő szialsav van jelen.

Az AGP glikozilációja (oligoszacharid antennák száma, monoszacharidok változatossága) számos kóros elváltozás esetén megváltozik [129-131]. A glikozilációs mintázat normálistól való eltérését figyelték meg akut gyulladásos állapotok, autoimmun betegségek és daganatos betegségek fennállása esetén. Ennek következtében a glikozilációs mintázat megváltozása bizonyos patológiás állapotok jelenlétére hívhatja fel a figyelmet, így fennáll a lehetőség, hogy az adott betegség esetében biomarkerként alkalmazható.

Különböző gyulladásos megbetegedések - így kardiovaszkuláris megbetegedések - esetén a plazma AGP szintje többszörösére növekszik. Spanyol együttműködő partnereink vizsgálták, hogy van-e kimutatható összefüggés az AGP glikozilációs mintázata és az atherosclerosis között [132]. UV detektálású kapilláris zóna elektroforézis segítségével a vérplazmából izolált AGP 13 izoformját választották el egymástól, és ezek mennyiségének változását vizsgálták egészséges, atherosclerosisban és hasi aorta aneurizmában (pozitív kontroll) szenvedő páciensek esetén. Az eredményeket (glikoformok relatív intenzitását/eloszlását) lineáris diszkriminancia analízis segítségével értékelték. Ennek alapján megállapították, hogy a beteg és egészséges egyénektől származó minták jól

megkülönböztethetők, azonban a beteg csoportok egymástól nem különíthetők el jól, mivel a módszer specificitása kicsi.

3.10 Az atherosclerosis

Nemzetközi együttműködésben egy fontos népbetegség, az érlelmeszesedés (atherosclerosis) korai kimutatására alkalmas biomarker azonosítását tűztük ki célul. Az atherosclerosis a fejlett országokban a vezető halálokok közé tartozik, kialakulásának hátterében az érfal tulajdonságainak megváltozása áll. Ennek fő oka lipid plakkok lerakódása az érfal mentén, amely gyulladással járó folyamatokat [133-135] indukál. A plakkok különböző típusú zsírokat, alacsony sűrűségű lipoproteint (Low Density Lipoprotein; LDL), koleszterint, kalciumot, illetve simaizom és kötőszöveti sejteket tartalmaznak, amelyek az ereket jelentősen beszűkíthetik. Ennek következtében az oxigéndús vér nem megfelelő mennyiségben jut el a szervekbe, valamint a plakkok leválása aktiválja az alvadási folyamatot, amely végső soron trombus (vérrög) kialakulásához vezethet. A trombus az ér elzáródását okozhatja (atherotrombózis), ennek eredményeként azok a szövetek, amelyeket ez az ér táplál, néhány percen belül elhalnak. Az érintett szövetek helyétől és nagyságától függően megkülönböztetünk többek között myocardialis infarktust (szív koszorúér elzáródása) ill. agyvérzést (agy erek elzáródása).

Egy másik fontos kardiovaszkuláris megbetegedés az atherosclerosis-sal hasonló patomechanizmussal (oxidatív stressz, gyulladás, vérröggképződés) rendelkező aneurizma. Az aneurizma egy vérrrel töltött gömbszerű kitüremkedés az érfalon, amely az érfal (betegség vagy sérülés miatti) elvékonyodása következtében alakul ki. Az egyre növekvő aneurizma miatt érfal sérülés veszélye áll fenn, amelynek következménye azonnali halál is lehet (pl. a hasi aorta aneurizma sérülés).

A klasszikus rizikófaktorokkal (magas vérnyomás és vérzsír szint) rendelkező betegek diagnózisa nem minden esetben egyszerű feladat. A nagyobb átmérőjű nyaki és hasi artériák vizsgálata ultrahanggal megvalósítható, ezzel szemben a kisebb koszorúerek elmeszesedéseinek vizsgálata csak angiográfiával vagy újabban komputertomográfiával (CT) lehetséges. Ez a két technika azonban nem megfelelő az érfalak állapotának szűrővizsgálati jellemzésére. Ezen kívül a komputertomográfia sugárterheléssel jár, az angiográfia pedig invazív eljárás. Így mindkettő előnytelen a betegek szempontjából.

A betegségek egyénenként változó lefolyásúak, sok esetben a betegek állapota hosszú időn át stabil marad (tünetmentesek). Ez nagyban megnehezíti a korai diagnózist és súlyos következményekkel járó történések előrejelzését és elkerülését. Ezért fontos kiegészítő diagnosztikai eljárások kutatása. Utóbbi időben egyre nagyobb teret nyer a vérplazma biomarkerek vizsgálata. Ezek segítségével előre jelezhetőek lehetnek az érfalban bekövetkező kóros elváltozások. Ez idáig azonban nincs még a klinikai gyakorlatban alkalmazott atherosclerosis vagy aneurizma biomarker.

4 A kísérletek és számítások részletei

4.1 Felhasznált vegyszerek, minták, alkalmazott készülékek, szoftverek

4.1.1 Vegyszerek

A folyadékkromatográfiához LC-MS-tisztaságú eluenseket és vegyszereket használtunk (acetonitril, víz, hangyasav). Az emésztéseket az alábbi reagensekkel végeztem: RapiGest (Waters, Milford, MA, USA), proteomikai tisztaságú sertés hasnyálmirigy tripszin (Promega, Madison, WI, USA), jódacetamidot és 1,4-ditio-DL-treitolt (DTT) (Sigma-Aldrich, Steinheim, Németország).

4.1.2 Minták

A humán plazma minták a Bajcsy-Zsilinszky Kórházból; az izolált AGP minták a madridi egyetemről (CSIC and Autonoma University) származtak. A glikoprotein standardokat (fetuin, haptoglobulin, transzferrin és alfa-1 savas glikoprotein) a Sigma-Aldrich-től vásároltuk.

4.1.3 Műszerek

A folyadékkromatográfiás vizsgálat előtt a mintákat Table Top Refrigerated Centrifuge Hermle Z300K (Hermle Labortechnik GmbH, Wehingen, Németország) készülékben centrifugáltuk.

A folyadékkromatográfiás elválasztást Waters C18 trap oszlopon (180 μm x 20 mm) történő sótalánítást követően 75 μm x 200 mm Waters BEH300 C18, 1,7 μm szemcseátmérővel rendelkező kapilláris kromatográfiás oszlopon végeztem, nanoAcquity UPLC rendszer alkalmazásával (Waters, Milford, MA, USA). A tömegspektrometriás vizsgálatok Waters Q-Tof Premier tömegspektrométeren (Waters, Milford, MA, USA) történtek.

4.1.4 Szoftverek

A kromatogramok és spektrumok felvételéhez és elemzéséhez a MassLynx szoftver 4.1 változatát (Waters, Milford, MA, USA) alkalmaztam. A peptidek azonosítását adatbázis keresés

alapján a Protein Lynx Global Server 2.3 verziójával (Waters) és Mascot Server 2.2.7 verzió (Matrixscience, USA) szoftverekkel végeztem. A glikopeptidek azonosításához és a glikozilációs mintázat meghatározásához az általam kidolgozott szoftvereket (GlycoMiner, GlycoPattern) alkalmaztam.

4.2 Fehérjék proteolízisére használt protokollok

4.2.1 A GlycoMiner fejlesztés során az AGP triptikus emésztésére használt protokoll

750 µg AGP fehérjét denaturáltam 37,5 µL 50mM-os NH_4HCO_3 oldatban oldott 0,2 w/v% RapiGest segítségével. A fehérje redukálásához 4,5 µL 45 mM-os DTT-t adtam a mintaoldathoz, 5mM-os végső DTT koncentrációt biztosítva az oldatban. Ezután a mintákat 60 °C-on 30 percig termosztáltam. A fehérjeoldat alkilálása céljából a mintához 7,5 µL 100 mM-os jódiacetamidot adtam, amely 15 mM-os végső jódiacetamid koncentrációnak felel meg. A jódiacetamidot is tartalmazó mintákat sötétben szobahőmérsékleten 30 percig inkubáltam. Ezt követően 37,5 µL frissen készített tripszinoldatot adtam a mintához. Az oldatot 2,00 mg tripszin 1,00 mL LC-MS tisztaságú vízben történő feloldásával készítettem. A tripszines emésztés 37 °C-on 60 percig történt. A RapiGest elbontásához és a proteolízis leállításához 7,5 µL 500 mM-os hangyasav oldatot adtam a mintához. A mintát ezt követően 30 percen keresztül 37 °C-on termosztáltam. A termosztálás után a mintát 10 percig centrifugáltam 17 000 g-nek megfelelő centrifugális erő alkalmazásával. A centrifugálást követően az oldhatatlan felülúszót eltávolítottam és a maradékot UHPLC vizsgálat céljából százszorosára hígítottam.

4.2.2 Vérplazmából származó AGP triptikus emésztésére használt protokoll

A vérplazmából származó AGP minták 1-5 pmol/µL AGP-t tartalmaztak. Az egyes mintákból 10 µL-t emésztettem az alábbi módon: az oldathoz 1,5 µL 0.5% RapiGest-et és 1µL 200 mM DTT-t adtam és a mintákat 60 °C-on 30 percig termosztáltam. A redukálást követően alkilálás céljából a mintához 2,5 µL 200 mM NH_4HCO_3 -ot és 1,5 µL 200 mM jódiacetamidot adtam és a mintákat szobahőmérsékleten 30 percig sötétben tároltam. A fehérjék peptidekre

hasítását tripszin (2 μL , 40 μM) hozzáadásával biztosítottam és a mintát 37 °C-on 3 óráig inkubáltam, amelyet követően a mintát kétszer 5 percig mikrohullámmal sugároztam be. Az emésztés leállítására és a RapiGest elbontására a mintához 1 μL hangyasavat adtam, majd 37 °C-on 30 percig inkubáltam. Ezt a minta 10 percig tartó centrifugálása követte 13 500 RPM (17 000 g) alkalmazásával. A centrifugálást követően a vízben oldhatatlan felülúszót eltávolítottam az oldatból. A mintát hígítás nélkül vizsgáltam nanoLC-MS módszerrel.

4.2.3 Plazma minta fehérjéinek emésztése (dúsítás előtt)

125 μL vizsgálati mintát tartalmazó oldathoz (2,5 μL plazma vízben) 3 μL 0.5% RapiGest-et és 2 μL 200 mM DTT-t adtam és a mintákat 60 °C-on 30 percig termosztáltam. A redukálást követően alkilálás céljából a mintához 15 μL 200 mM NH_4HCO_3 -ot és 2 μL 200 mM jódiacetamidot adtam és szobahőmérsékleten 30 percig sötétben tároltam. A fehérjék peptidekre hasítását tripszin (10 μL , 40 μM) hozzáadásával biztosítottam és a mintát 37 °C-on 3 óráig inkubáltam, melyet követően a mintát kétszer 5 percig mikrohullámmal sugároztam be. Az emésztés leállítására és a RapiGest elbontására a mintához 1 μL hangyasavat adtam és 37 °C-on 30 percig inkubáltam. Ezt követően a mintát 10 percig centrifugáltam 13 500 RPM (17 000 g) alkalmazásával. A centrifugálást követően a vízben oldhatatlan felülúszót eltávolítottam az oldatból. A glikopeptidek dúsítását 100 μL felülúszón hajtottam végre.

4.2.4 Plazma minta fehérjéinek emésztése (dúsítás után)

A frakciókat vákumbepárló (SpeedVac) segítségével koncentráltam. A savat tartalmazó frakciók esetén VivaSpin (Sartorius AG, Goettingen, Germany) centrifuga csőben az oldószert savat nem tartalmazó vízre cseréltem a minták háromszori vizes mosásával. Minden frakció esetén a végtérfogat 100 μL lett, melyből 10 μL -t emésztettem az alábbi módon: az oldathoz 1,5 μL 0,5% RapiGest-et és 1 μL 200 mM DTT-t adtam és a mintát 60 °C-on 30 percig termosztáltam. Redukálást követően, alkilálás céljából a mintához 2,5 μL 200 mM NH_4HCO_3 -ot és 1,5 μL 200 mM jódiacetamidot adtam és szobahőmérsékleten 30 percig sötétben tároltam. A fehérjék peptidekre hasítását tripszin (2 μL , 40 μM) hozzáadásával biztosítottam, a mintát 37 °C-on 3 óráig inkubáltam. Ezt követően a mintát kétszer 5 percig mikrohullámmal sugároztam be. Az emésztés leállítására és a RapiGest elbontására a mintához 1 μL hangyasavat adtam és 37 °C-on 30 percig inkubáltam. Ezt a minta 10 percig tartó centrifugálása követte 13 500 RPM (17 000

g) alkalmazásával. A centrifugálást követően a vízben oldhatatlan felülúszót eltávolítottam az oldatból. A mintát hígítás nélkül vizsgáltam nanoLC-MS módszerrel.

4.3 Peptidek, glikopeptidek (fehérje emésztmények) nanoLC-MS(MS) vizsgálata

4.3.1 Nanoáramlásos folyadékkromatográfia és tömegspektrometriás vizsgálat

A plazmafehérjék peptidjeinek elválasztása a sóatlanítást követően fordított fázisú, C18-as nanooszlopon történt nanoAcquity UPLC System kromatográfias rendszer alkalmazásával. Eluensként A) 0,1 % hangyasavat tartalmazó LC-MS tisztaságú vizet és B) 0,1 % hangyasavat tartalmazó LC-MS tisztaságú acetonitrilt használtam. A folyadékkromatográfias mérés során alkalmazott gradiens paramétereit a 2. táblázatban láthatók.

2. táblázat: A nanoLC-MS mérések során használt gradiens

Idő (perc)	Áramlási sebesség (μL/perc)	A eluens aránya (%)	B eluens aránya (%)
0	0.4	90	10
70	0.4	50	50
72	0.4	15	85
80	0.4	15	85
82	0.4	90	10
100	0.4	90	10

Egy injektálás során 2,0-5,0 μL mintamennyiség került a nanooszlopra, amely közvetlenül kapcsolódott a tömegspektrométer nanoelektrospray ionforrásához.

Tömegspektrometriás paraméterek

- Ionizációs mód: pozitív ionizáció
- Kapilláris feszültség: 2,7 kV
- Cone feszültség: 15,0 V
- Ütközési energia: 3,3 eV
- Forrás hőmérséklete: 90 °C

A fehérjék azonosítását tandem tömegspektrometriával végeztem. A tömegspektrumok felvétele pozitív ionizációs módban 400-2000 m/z-ig terjedő tömegtartományban készült 7 s-os ciklusokat használva, mely ciklus egy teljes spektrumból és a 3 legintenzívebb ion 2 s hosszú MS/MS spektrumából állt. Az MS/MS kísérletek során Ar ütközőgázt alkalmaztam, az ütközési energiákat pedig a prekursor ion töltési állapota és tömege alapján változtattam 18-80 eV tartományban.

A „LockMass” referenciaoldat 900 pg/μL-es angiotenzin II oldat, 1:1 arányú acetonitril:víz elegyében oldva; az oldószer 0,1 %-ban hangyasavat tartalmazott.

4.3.2 nanoLC-MS/MS adatokon alapuló fehérje azonosítás

A kísérlet során nyert spektrumok feldolgozása a ProteinLynx (Waters) szoftverrel történt. A feldolgozott spektrumokat a Mascot (Matrix Science, London) programmal értékeltém ki.

A plazma fehérjék azonosítása esetén *Homo sapiens* taxonómiát és a SwissProt szekvencia adatbázist használtam. A keresés során egy kihagyott hasítási helyet engedélyeztem, fix módosulásként a ciszteinek jódcetamid származékát, míg lehetséges módosulásként a metionin aminosavak oxidációját állítottam be. Az anyaionok esetén 50 ppm toleranciát, míg a fragmensionok esetén 0.15 Da tömeg toleranciát engedélyeztem. A jódcetamid használatát karbamidometilcisztein rögzített módosulás használatával vettem figyelembe. Nem rögzített módosulásként a metionin oxidációját állítottam be. A peptidek közül a 95%-nál ($p < 0.5$) nagyobb valószínűséggel azonosítottakat fogadtam el [136]. A fehérjék esetén a 99%-nál ($p < 0.01$) nagyobb valószínűséggel azonosítottakat fogadtam el [137], ha legalább két hozzátartozó peptid fragmenssel is rendelkeztek.

4.3.3 (nano)LC-MS/MS adatokon alapuló glikopeptid azonosítás

A kísérletek során nyert spektrumok feldolgozása a ProteinLynx (Waters) szoftverrel történt. A feldolgozott spektrumokat az általam készített GlycoMiner programmal értékeltém ki.

A GlycoMiner kiértékelés során használt paraméterek a következők voltak:

- tömegpontosság: 15 ppm a prekursorra, 30 ppm a kis tömegű marker ionokra és 100 ppm a fragmens ionok közötti tömegkülönbségre.
- glikán összetétel: maximum 4 fukóz, 6 szialsav, 15 hexóz és 15 hexózamin.

- ponthatárok: marker pontszám esetén a 10 alatti találatokat elvetettük; a glycoscore esetén közepes (50%) megbízhatóságot követeltünk meg.
- a szekvencia adatbázis az AGP vizsgálatok esetén a négy AGP szekvenciavariánst tartalmazta

4.4 Glikoproteinek/glikopeptidek dúsítására alkalmazott módszerek

4.4.1 Lektin affinitás kromatográfia

A lektin affinitás kromatográfián alapuló dúsítást AffiSep WGA-SPE oszlopon (GALAB Technologies GmbH, Geesthacht, Németország) (0,6 mL / 4,0 x 50 mm) végeztem. Elsőként az oszlopot 3600 µL pufferrel (10 mM BIS-TRIS pH = 6,0, 15 mM NaCl, 1 mM CaCl₂, 1 mM MnCl₂) kondicionáltam. Következő lépésben történt a mintafelvitel (pufferrel tízszeresére hígított plazma minta vagy emésztett felülűszó) az oszlopra (1-es frakcióra). Ezt követően az oszlopot 1200 µL pufferrel mostam (2-es frakció). Végül a megkötődött glikoproteineket és glikopeptideket 1200 µL pufferrel (10 mM BIS-TRIS pH = 6,0, 15 mM NaCl, 1 mM CaCl₂, 1 mM MnCl₂, 150 mM *N*-acetil-glükózamin) oldottam le (3-as frakció).

4.4.2 Fenilbórsavas (PBA) affinitás kromatográfia

A PBA oszlopot (BondElut PBA, Agilent Technologies, Santa Clara CA, United States) 400 µL 90% acetonitril és 5% hangyasavat tartalmazó oldattal aktiváltam, majd 1 mL 200 mM ammónium hidrogén karbonát pufferrel (pH 10) mostam. Ezután vittem fel a mintát az oszlopra: 200 mM ammónium hidrogén karbonát pufferrel (pH 10) tízszeresére hígított plazma mintát vagy emésztett felülűszót (1-es frakció). A meg nem kötődött fehérjéket/peptideket 400 µL 200mM NH₄HCO₃ pH=10 (2-es frakció) oldattal, majd 600 µL 75% ACN + 25% 200mM NH₄HCO₃ pH=10 (2-es frakció) oldattal mostam. A megkötődött glikoproteineket és glikopeptideket 800 µL 50% ACN 5% FA oldattal oldottam le (3-as frakció).

4.4.3 Előre ekvilibrált anioncserélő SPE

Az anioncserélő oszlopot (ISOLUTE PEAX SPE 25mg töltet/ 1 mL SPE; Biotage AB, Uppsala, Svédország) 400 µL 99% ACN, 20 mM ammónium-acetát oldattal kondicionáltam. Ezt

követően vittem fel a mintát az oszlopra: 20 mM ammónium-acetáttal tízszeresére hígított plazma mintát vagy 100 μL emésztett felülúszót, amelyet 200 μL 10% ACN, 50 mM ammónium-acetáttal mostam (1-es frakció). A megkötődött glikoproteineket és glikopeptideket 300 μL 10% ACN, 5% hangyasav oldattal (2-es frakció) és 300 μL 10% ACN, 300 mM citromsav oldattal (3-as frakció) oldottam le.

4.5 Lineáris diszkriminancia analízis egészséges és beteg csoportok megkülönböztetésére

A lineáris diszkriminancia analízist (LDA) a Statistica számítógépes programcsomag 7-es verziójával végeztem (www.statsoft.com). A diszkriminancia analízishez a betanító halmaz az általam mért 28 darab, három kategóriába sorolható AGP minták helyspecifikus glikozilációs mintázatát tartalmazta. A glikozilációs mintázat összesen 156 különböző glikopeptidet tartalmazott. Ezek közül kiválasztottam 128 glikopeptidet (független változót), amelyek minden csoportban legalább egyszer nem zérus értékkel szerepeltek. A statisztikai modellhez hét paraméter használatát engedélyeztem.

A módszer érzékenységét és specificitását egy-elem-kihagyásos keresztellenőrzés (leave-one-out cross-validation) után számítottam. A keresztellenőrzés során kihagytam egy mintát, majd a diszkriminancia analízist a maradékkal, mint betanító halmazzal elvégeztem. A kihagyott mintát, mint teszt adatot használva megvizsgáltam, hogy a megfelelő osztályba lesz-e besorolva. Ezt minden egyes mintára elvégeztem.

5 Eredmények és értékelésük

A Doktori munkám fő célkitűzése és legfontosabb eredménye egy olyan, a helyspecifikus *N*-glikozilációs mintázat meghatározására alkalmas módszer kidolgozása, mely alkalmas orvosi, diagnosztikai problémák megoldására. Olyan folyadékkromatográfiával kapcsolt tömegspektrometriás módszert dolgoztam ki, amely

- 1) különböző glikoproteinek esetén is használható
- 2) nem csak izolált glikoproteinek, hanem glikoprotein keverékek vizsgálatára is alkalmas
- 3) elég érzékeny kis mintamennyiségek vizsgálatára (1-10 pmol), így tényleges biológiai minták is analizálhatók
- 4) nagy áteresztőképességű (több száz minta vizsgálatára alkalmas).

Kutatási tevékenységem első részében kidolgoztam a fent említett módszert, amelynek kulcsfontosságú része egy olyan algoritmus és szoftver (a GlycoMiner) fejlesztése, mely glikopeptidek tandem tömegspektrumainak automatikus elemzését teszi lehetővé.

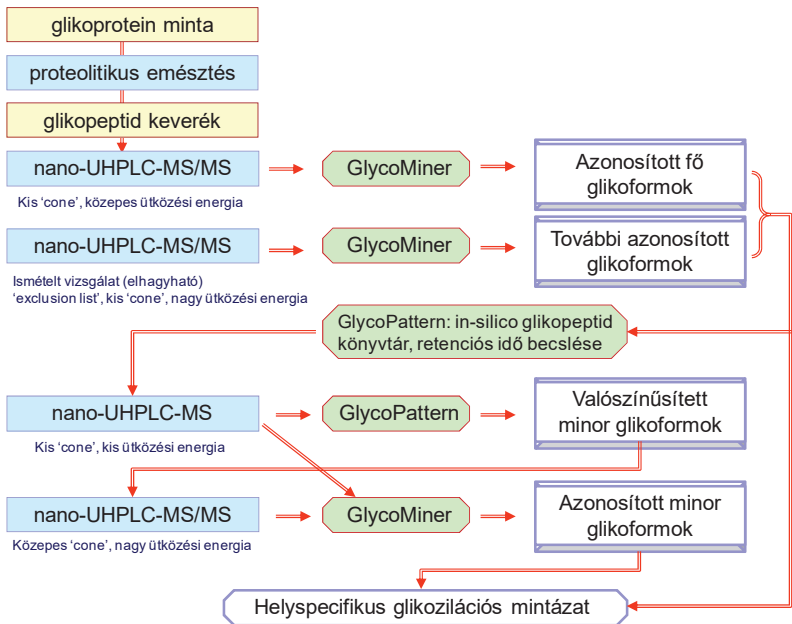
A módszer komplex mintákon történő alkalmazhatóságát a vérplazmában előforduló haptoglobin nevű fehérje glikozilációs mintázatának meghatározásával igazoltuk (60. oldal). A vizsgálatokat izolálás nélkül, közvetlenül albumin depletált plazmából végeztük. A módszer a gyógyszerkutatásban is jól alkalmazható, amit egy monoklonális antitest (influximab) glikozilációs mintázatának meghatározásán keresztül illusztrálunk (61. oldal). A kutatás következő lépése olyan dúsítási módszerek kidolgozása volt, amelyek kis koncentrációban jelen lévő glikoproteinek analizését teszi lehetővé.

Doktori dolgozatom utolsó részében a kidolgozott módszer gyakorlati alkalmazhatóságát mutatom be. Arra a kérdésre kerestük a választ, hogy az AGP glikozilációs mintázata alkalmazható-e orvosi-diagnosztikai feladatok megoldására. Példaként egy igen fontos népbetegség, az atherosclerosisot jellemző biomarkereket kerestünk, amelyek alkalmas lehetnek diagnosztikai célokra.

5.1 Helyspecifikus glikozilációs mintázat meghatározása

A helyspecifikus glikozilációs mintázat meghatározásának elvi folyamatábrája az 10. ábrán látható. A folyamat első lépése a vizsgálni kívánt glikoprotein (fehérjeminta) proteolitikus emésztése, amelynek eredményeként peptidek, illetve glikopeptidek képződnek. A komponensek

száma a fehérje méretétől, összetételétől, de leginkább a különböző glikoformáinak számától függően akár több ezer is lehet. Ezen komponensek azonosítására, összetételének (szerkezetének) meghatározására folyadékkromatográfiával kapcsolt (tandem) tömegspektrometria alkalmazható. A nagy érzékenység elérésének érdekében tipikusan nano-UHPLC-MS/MS technikát alkalmazunk. Itt szeretném megjegyezni, hogy a következőkben, mint LC-MS fogok a különböző folyadékkromatográfiai technikákra hivatkozni, hiszen a módszer nem csak nano-UHPLC-vel történő elválasztás esetén alkalmazható, hanem bármilyen folyadékkromatográfiai módszerrel.



10. ábra: A helyspecifikus glükozilációs mintázat meghatározásának folyamatábrája

Az enzimatikusan emésztett glükoproteinekről felvett tömegspektrumokban észlelhető csúcsok jelentős része peptidektől, kisebb hányada glükopeptidektől származik. Manuális értékelés a spektrumok nagy száma miatt nem megvalósítható. Ezért a doktori munkám során egy olyan algoritmust és számítógépes programot fejlesztettem ki, amely alkalmas több ezer glükopeptidről felvett tandem tömegspektrum értékelésére. Ennek a részleteit az 5.2. alfejezetekben tárgyalom. A folyamat első fázisában a fő (viszonylag nagy mennyiségben jelen

levő) glikopeptidek (glikoformok) azonosíthatók. A fő glikoformok azonosításához ún. adatfüggő mérési mód használata szükséges, ami azt jelenti, hogy a számítógépes algoritmus megadott szempontok (tipikusan a legnagyobb intenzitás) alapján választja ki a csúcsokat a tandem tömegspektrometriás felvételhez. Így a kis intenzitású csúcsok esetén, ill. abban az esetben, ha több glikopeptid párhuzamosan eluálódik, nem készül róluk MS/MS felvétel. Ennek a problémának a kiküszöbölésére újabb adatfüggő mérést kell elvégezni, ún. kizáró lista („exclusion list”) használatával. Ebben megadjuk azokat a csúcsokat, amelyekről már korábban készült felvétel, így az automatikus rendszer ezekről a csúcsokról a továbbiakban nem vesz fel újabb tandem tömegspektrumot. Ezen csúcsok helyett a rendszer más (kisebb intenzitású) csúcsot választ, így további glikopeptidek azonosítására nyílik lehetőség. Az így azonosított glikopeptidek retenciós idejének ill. tömegének segítségével egy glikopeptid könyvtárat készítünk.

Ezt követően azt vizsgáljuk, hogy egy adott glikozilációs helyen az összes elvileg lehetséges glikoform azonosításra került-e. A nem azonosított – de elméletileg lehetséges – glikoformokból *in silico* glikopeptid könyvtárat készítünk és egy általunk kidolgozott algoritmus és számítógépes program (GlycoPattern) segítségével *in silico* megbecsüljük ezen glikoformok várható retenciós idejét és csúcshélességet. A becslés alapját az adja, hogy a retenciós időt elsősorban a peptid szekvenciája határozza meg, a peptidhez kapcsolódó oligoszacharid ezt csak kis mértékben módosítja, és ez az eltolódás az oligoszacharid lánc összetétele (szialsav, fukóz, stb. tartalom) alapján előre jelezhető. Az algoritmus részleteit az 5.4. alfejezet tartalmazza.

A következő lépésben az így elkészített glikopeptid könyvtár segítségével (amely tartalmazza az összes lehetséges glikopeptid tömegét, retenciós idejét és csúcshélességet) további kis intenzitású glikopeptid jelöltek azonosíthatók. Fontos hangsúlyozni, hogy ezek a glikopeptidek csak lehetséges összetételek (szerkezetek), ugyanis tandem tömegspektrometriás szerkezetigazolás nem történt. Az összetétel meghatározásához egy újabb célzott tandem tömegspektrometriás vizsgálat elvégzése szükséges, amelyben a valószínűsített glikopeptidekről tandem tömegspektrometriás felvétel készül. Ezeknek a kis intenzitású csúcsoknak a teljes szerkezet igazolására nincs lehetőség, ezért nagy ütközési energiát használva a peptid részre jellemző fragmens megjelenését keressük. A fragmens megjelenése és a retenciós idővel való egyezés együttesen valószínűsíti az adott minor glikopeptid összetételét. Ezt a folyamatot az 5.5. alfejezetben tárgyalom részletesen. Az így nyert információkat összegezve, meghatározható a helyspecifikus glikozilációs mintázat. A módszer legnagyobb előnye, hogy összetett keverékek (azaz glikoprotein-keverékek) vizsgálatára is alkalmas.

5.2 GlycoMiner: Glikopeptidek tandem tömegspektrumának automatikus értékelése

A GlycoMiner algoritmust és számítógépes programot a glikopeptidek gyors azonosítására fejlesztettük ki. A munka során főként *N*-glikánokat tartalmazó glikoproteinek triptikus emésztményeit vizsgáltuk LC-MS/MS technikával, és a kapott tömegspektrumokat értékeltük. A glikoproteinek emésztése peptid – glikopeptid keveréket eredményez, amelyben a legtöbb glikopeptid kis intenzitású csúcsot ad. A szerkezet meghatározás során az első feladat azoknak a jeleknek az azonosítása, amelyek glikopeptidektől származnak. A feladatot legmegbízhatóbban MS/MS mérések alapján lehet megvalósítani, bár részleges információ az MS spektrumokból is nyerhető. A következő lépés a glikopeptid szerkezet-azonosítása, mely mind a glikán, mind pedig a peptid összetételének meghatározását jelenti. Keverékek esetén ez kizárólag MS/MS adatok alapján lehetséges. (Egy fehérjét tartalmazó minta esetén ez részben MS spektrumok alapján nagy pontosságú (1 ppm alatti) tömegméréssel is megoldható [111], amely azonban a gyakorlatban bonyolult minta-előkészítést igényel, és sokszor nem eléggé megbízható.)

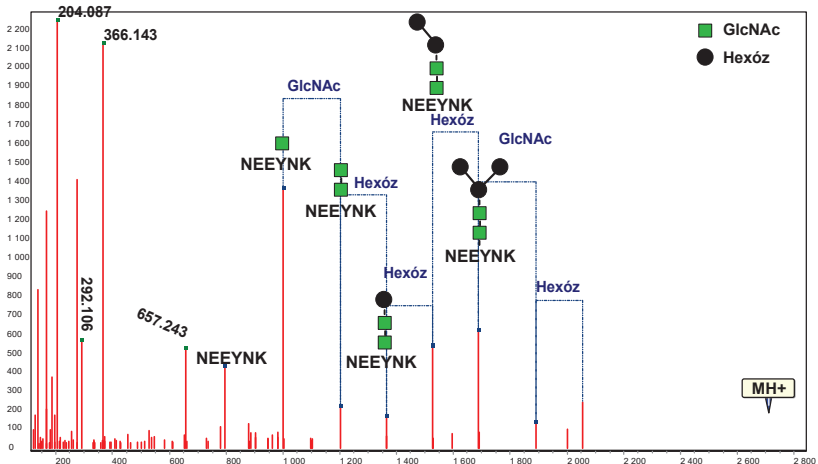
A GlycoMiner a tandem spektrumokból különböző információkat határoz meg és használ fel glikopeptidek azonosítására és összetételük/szerkezetük meghatározására. Ezek az információk, ill. az elvégzett műveletek a következők:

1. Egyszeres töltésű glikán fragmensek („B” típusú oxónium ionok, amelyek mono-, bi-, tri- és tetraszacharidoknak felelnek meg).
2. Közepes és nagy tömegtartományban észlelhető karakterisztikus tömegkülönbségek, amelyek nagy tömegű ionokból mono- és oligoszacharidok lehasadásával képződnek.
3. Az oligoszacharidokhoz kapcsolódó peptid tömege, amely a karakterisztikus ion sorozatok felhasználásával határozható meg.
4. A spektrum minőségének jellemzése, amely fontos lépés a valós jelek és zaj spektrumok elkülönítéséhez. A minőség jellemzéséhez egy kvantitatív jelzőt, úgynevezett „minőség pontszámot” (quality score, S_{quality}) vezetünk be.
5. A peptid összetétel keresése a peptid tömege alapján, figyelembe véve (1) csak azokat a fehérjéket, amelyek a mintában lehetnek, (2) a proteolitikus enzim specificitását és (3) azt, hogy a peptid szekvenciájában egy glikozilációs helynek megfelelő szekvencia darabnak kell lennie. A fehérje összetételt általában egy egyszerű proteomikai vizsgálatot követően állapítjuk meg, például ugyanazon minta LC-MS/MS felvételének adatbázis

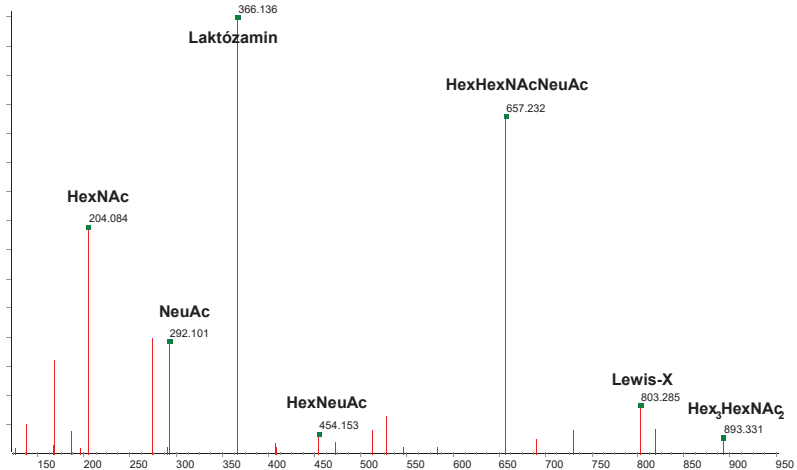
alapú (például *Mascott*al végzett) keresése alapján, de izolált fehérjék esetén a fehérje elméleti *in silico* emésztett adatai is használhatóak.

6. A glikopeptid összetétel meghatározása a fenti lépések és a prekursor tömege (ha elérhető a pontos tömege) alapján. Az összetétel meghatározás jóságának jellemzéséhez is bevezettünk egy pontszámot, az ún. „glycoscore”-t. A glycoscore két tapasztalati úton beállított részpontszámból áll, amely a glikán és peptid részek meghatározásának jóságát jellemzi (rangsorolja).

A prekursor ion tömegének pontos meghatározása rendkívül fontos a glikánok szerkezetének, ill. összetételének azonosításához. A GlycoMiner program alkalmazásakor az első lépés az általában többszörös töltésű ionokat tartalmazó ESI spektrumok egyszeres töltésűre konvertálása. Tömegspektrumok értékelésénél ez egy gyakran alkalmazott lépés. Jelenleg a Waters cég által forgalmazott ProteinLynx Global Server (PLGS) szoftver csomagot használjuk erre a célra, hasonló funkció a legtöbb tömegspektrométert gyártó cég szoftvere esetén szintén elérhető. Ezt követően történik az egyszeresen töltött tandem tömegspektrumok értékelése, amelynek során meghatározzuk a glikopeptid molekuláról oligoszacharid egységek lehasadásával képződő (viszonylag nagy tömegű) fragmenseket, valamint a kis tömegszámoknál megjelenő cukorláncokat (oxónium ionok). Az 11. és 12. ábrákon egy automatikusan értékelt tandem tömegspektrum látható.



11. ábra: NEEYNK-Bis2 glikopeptid MS/MS spektruma (azonosított fragmensekkel, automatikusan értékelve)



12. ábra: A NEEYNK-BiS2 glikopeptid MS/MS spektruma a kis tömegszám tartományban (oxónium ion típusú marker ionok)

A program első lépésben azonosítja azokat az MS/MS spektrumokat, amelyek glikopeptideknek felelnek meg. Az úgynevezett „*marker pontszám*” (marker score, S_{marker}) jelzi, hogy milyen megbízhatósággal azonosítható egy spektrum glikopeptidtől származónak. A program mind a kis tömegű oxónium ionok, mind a nagy tömegeknél előforduló karakterisztikus tömeg különbségeket figyelembe veszi. Az általunk vizsgált (humán) *N*-glikoproteinekből előforduló oxónium ionok listája az 3. táblázatban található. A lista szükség esetén (pl. nem humán eredetű minták esetén) bővíthető, ill. módosítható. A marker pontszám a spektrum minőségét is figyelembe veszi, amely nagyon zajos spektrumok esetén alacsony érték lesz. Ezzel megelőzhető, hogy nagyon zajos spektrumok esetén a háttérben véletlenszerűen észlelt csúcsokat a szoftver glikopeptidekhez rendelje. Ez a művelet az algoritmus fontos része, ami biztosítja, hogy a GlycoMiner használatakor az álpozitív találatokat szinte teljes mértékben elkerüljük. A marker pontszám értékei 0 és 100 között mozognak. Csak azok a spektrumok lesznek glikopeptidekhez hozzárendelve, amelyek marker pontszám értéke nagyobb, mint egy megadott határ (jelenleg 10), de ez a határ szükség szerint módosítható.

3. táblázat: A marker pontszám meghatározásánál figyelembe vett oxónium (B típusú) ionok

Tömeg	Szacharid (Rövidítés)
147.0757	Fukóz (Fuc)
163.0606	Hexóz (Hex)
204.0872	N-acetil-hexózamin (HexNAc vagy GlcNAc)
292.1032	Sziál sav vagy N-acetilneuramin sav (Sia, NeuAc)
366.1400	Laktózamin (HexHexNAc)
454.1561	Hexóz + Sziálsav (HexNeuAc)
657.2354	Szialis-laktózamin (HexHexNAcNeuAc)
803.2928	Szialis-Lewis X (Le ^X)
893.3251	N-glikán mag (Hex ₃ HexNAc ₂)

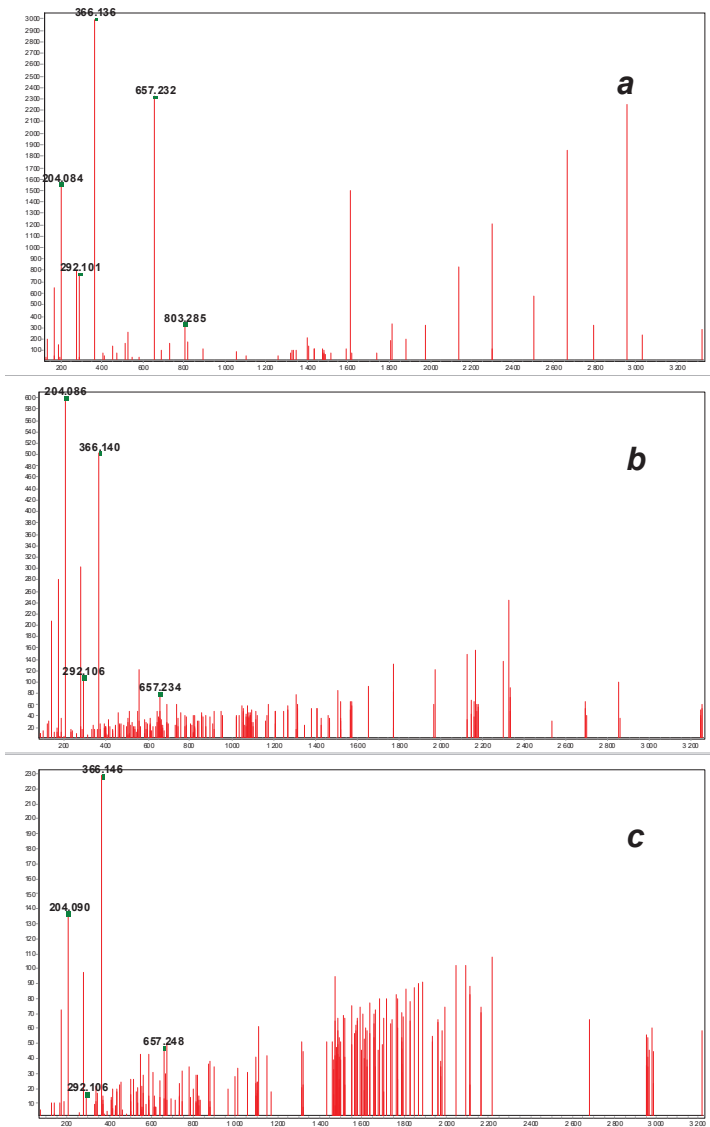
A spektrumok minőségét a *minőségi pontszámmal* jellemeztem (S_{quality}), amely 0 és 100 közötti értékeket vehet fel. A 13. ábrán három különböző minőségű spektrum látható: jó (A), közepes (B) és rossz (C) minőségű spektrumok. Az általunk felvett tandem tömegspektrumok mintegy fele volt „közepes” minőségű (S_{quality} 20 és 75 között), 30%-a jó (S_{quality} 70-100), 20%-a rossz minőségű (S_{quality} 20 alatt) volt. Ez a legtöbb proteomikai vizsgálat esetén tipikusnak mondható. Megjegyzem, ahhoz, hogy egy glikopeptid jelenlétét azonosítsuk, nem feltétlenül szükséges jó minőségű spektrum. A 13/C ábra bár rossz minőségű spektrumot mutat be, mégis megbízhatóan állítható róla, hogy a vizsgált minta glikopeptidet tartalmazott. Másrészt viszont a glikopeptid összetétel/szerkezet hozzárendeléséhez feltétlen jó, vagy legalább elfogadható minőségű spektrumra van szükség.

Két különböző glikoprotein triptikus emésztményét vizsgálva teszteltük a programot, amelynek során összesen 3132 spektrumot értékeltünk a GlycoMiner-rel és manuálisan is. A spektrumok közül 338-at rendeltünk hozzá glikopeptidekhez. Manuális ellenőrzéssel megállapítottuk, hogy a program az összes jó és közepes minőségű spektrumot helyesen rendelte a glikopeptidekhez. Hibát a program csak a gyenge minőségű spektrumok esetén vétett. Ezek közül a GlycoMiner hibásan hármát glikopeptidekhez rendelt, míg három glikopeptidtől származó spektrumot nem talált meg. Ennek alapján a GlycoMiner álpozitív és álnegatív azonosítási aránya is 0.1%. Ez nemzetközi összehasonlításban rendkívül megbízhatónak mondható.

A glikopeptidek hozzárendelése után a következő lépés az összetételük meghatározása. A program a spektrum közepes és nagy tömegei között mono- és oligoszacharid különbségeket keres, amelyek alapján ion sorozatok azonosíthatók. (pl. a 11. ábrán, ahol számos hexóz és hexózamin különbség látható). Ezek az ionsorozatok szacharid egységekben térnek el egymástól

és a glikánok Y típusú fragmentációjakor jönnek létre, mely megfelel a glikozid kötés hasadásának. Szerencsés esetben ezek a sorozatok egészen az utolsó szacharid egység lehasadásáig folytatódhatnak. Az a fragmens ion, amely nem tartalmaz glikán részt a „csupasz” peptid ion. Ennek tömege alapján kiszámítható a peptid tömege. A peptid ion azonosítása az *N*-glikozileződés konzervatív struktúrájának meglétéen alapul. Ez a struktúra egy pentaszacharid, amelynek a fragmentációja egy hat ionból (tömegspektrometriás csúcspól) álló sorozatot eredményez. Az ionok monoszacharid tömeg távolságokra adnak csúcspontokat a spektrumban (Peptid + GlcNAc + GlcNAc + Hex + Hex + Hex), kedvező esetben mind a hat tömeg jelen van és egyértelműen azonosítható. Ezt az esetet mutatja a 11. ábra az NEEYNK szekvenciájú peptid esetén. A tömegkülönbség a peptid fragmens és a prekursor tömege között a glikán rész tömegének felel meg. Ezen tömegek ismerete feltétele az oligoszacharid összetétel meghatározásának. A lehetséges oligoszacharid összetételeket összehasonlítjuk, kiegészítjük az ion sorozatból nyert információval. Amennyiben rendelkezésre állnak fehérje szekvencia adatok, ezt is felhasználjuk a glikopeptid aminosav szekvenciájának meghatározására.

A meghatározott peptid szekvenciáknak tartalmazniuk kell legalább egy glikozilációs helyet (pl. *N*-glikopeptidek esetén ez az *N-X-S/T* szekvencia, ahol *X* nem *P*). A peptid szekvencia meghatározásánál a kísérletileg alkalmazott proteáz specificitását is figyelembe vesszük. Az algoritmus a peptid tömeget a fragmentáció alapján számolja, nem pedig fehérje adatbázisból nyeri. Így a peptid tömege akkor is meghatározható, ha a vizsgált fehérjéről más információ nem áll rendelkezésre. Amennyiben a peptid tömege adatbázis keresésből származna, túl nagy lenne az álpozitív arány, ezért nem ezt a megközelítést választottuk. A peptid szekvencia helyes meghatározása nem biztosítható minden esetben, mert tipikusan a tandem tömegspektrumban nem észlelhető elegendő peptid láncból származó fragmens. Ebben az esetben a peptid azonosítása csak pontos tömeg alapján történik. A peptid tömegének meghatározása után a program adatbázis keresést végez, és ha nem talál megfelelő peptid szekvenciát, a glikán összetételét és a peptid tömegét jeleníti meg. Ellenkező esetben a peptid szekvenciája és a glikán összetétele kerül kiírásra, és a kiválasztott ion nagyobb pontszámot kap. Természetesen ritka, hogy egyetlen MS/MS spektrumban minden fontosabb információ rendelkezésre álljon. Például a „csupasz” peptid legtöbbször nem látható, csak a peptid és egy *N*-acetyl-glikózamin kapcsolódásából képződő ion. Ez az ion tipikusan intenzív jelet ad, ezért az intenzitás eloszlást az ionok között a pontozó algoritmus figyelembe veszi. Általában nem látható mind a hat fragmens ion a Peptid+GlcNAc+GlcNAc+Hex+Hex+Hex sorozatból, ezért csökken az információ tartalom.



13. ábra: a.) Jó ($S_{\text{quality}} = 98$, $S_{\text{marker}} = 100$), b.) közepes ($S_{\text{quality}} = 41$, $S_{\text{marker}} = 48$) és c.) rossz ($S_{\text{quality}} = 9$, $S_{\text{marker}} = 45$) minőségű spektrumok

Az értékelés során több egyéb probléma is felmerülhet: 1) nincs meg a teljes információ az összetétel azonosításához, 2) nem minden adat valós (pl. zajcsúcsok), vagy félrevezető. Például a tömegkülönbség két ion között megfelelhet egy monoszacharid tömegének, amely akár véletlen is lehet. A proteolitikus enzim specificitását is figyelembe kell venni, de hibás és nem specifikus hasítások is előfordul(hat)nak. Konvencionális, gyakran azonosított glikán struktúrák valószínűségét is figyelembe kell venni, mint például az egy vagy két szialsavat tartalmazó kétantennás struktúrák gyakoriságát az AGP esetén. Bizonyos glikán struktúrák humán minták esetén kevésbé valószínűek és feltehetően hibás azonosítás eredményei (pl. kétantennás glikán három szialsav egységgel).

A GlycoMiner program a fent említett összes információt megfelelően súlyozva figyelembe veszi egy pontozó rendszer segítségével. (A pontozórendszer kialakítása egy soklépéses optimalás eredménye, ennek részletes tárgyalására itt nincs lehetőség.) A cél a lehető legtöbb valós glikopeptid azonosítása, a lehető legkevesebb álpozitív találat mellett.

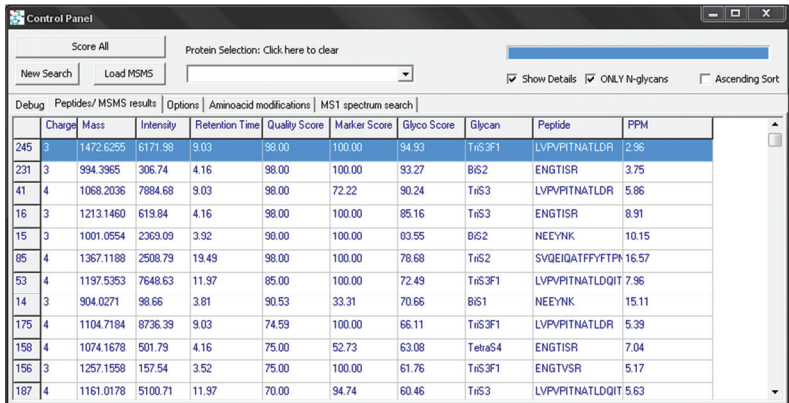
A korábban bemutatott példa esetében összesen 338 glikopeptid tandem tömegspektrumból - ezek manuális értékelésével - összesen 196 glikopeptid összetételét/szerkezetét határoztuk meg. A fennmaradó esetekben a kis csúcshintenzitások miatt a glikopeptidek összetételének meghatározására nem volt lehetőség. A GlycoMiner program szintén azonosította a manuálisan azonosított glikopeptideket, és ezen túlmenően még hét további glikopeptidet is meghatározott. Ezekről a későbbiekben újabb MS/MS vizsgálatok alapján, a retenciós időket és a glikoziláció biokémiai mechanizmusát is figyelembe véve megállapítottuk, hogy valós pozitív találatok. Hibás azonosítást egy esetben sem észleltünk. Az álpozitív találatok kiszűrésére szigorú minőségi követelményeket támasztunk. A program beállításainál lehetőség van a minőségi követelmények lazítására, amellyel további glikopeptid összetételeket lehetne meghatározni. Ezek nagy részénél azonban további, részletes vizsgálatok segítségével sem tudtuk igazolni, hogy ezek valós pozitív találatok.

Az eredmények azt mutatják, hogy a GlycoMiner hasonlóan, vagy jobban teljesít, mint az emberi szakértő. A markerionok alapján azonosított (azonosítható) spektrumok mintegy fele rendelkezik elégséges információval ahhoz, hogy a glikopeptid összetételét/szerkezetét meg lehessen határozni. Magas megbízhatósági szintet megkövetelve gyakorlatilag nincs álpozitív találat. A példában alkalmazott kísérleti körülmények hasonlóak a proteomikai gyakorlatban alkalmazottakhoz. Mivel az általunk azonosított spektrumok többségében közepes minőségűek voltak, így a teszt a mindennapi gyakorlatot modellezte. Meg kell jegyeznünk azonban, hogy a vizsgált glikoproteinek glikozilációja (viszonylag) jól ismert, ismeretlen glikoprotein esetén az emberi szakértő feltehetően rosszabbul teljesítene.

A GlycoMiner megbízhatóságát más, glikopeptidek MS/MS spektrumainak kiértékelését segítő programmal is összehasonlítottuk. Az irodalomban leírt változatos programok közül az ingyenesen elérhető Glycomod-ot [119] teszteltük az előzőleg megvizsgált spektrumok segítségével. Míg a GlycoMiner minden alkalommal helyesen azonosította a spektrumokat, a Glycomod esetén a várt eredmény csak az esetek 25%-ában szerepelt a kiadott listában. A Swamy [120] által készített program sikeres azonosítási aránya csupán 67% volt.

5.3 A GlycoMiner kezelőfelülete

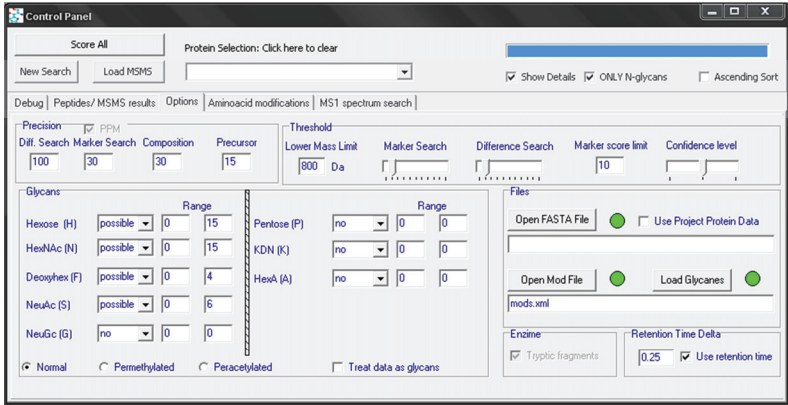
A GlycoMiner számítógépes program Windows operációs rendszer alatt működik. Fejlesztésekor nagy hangsúlyt fektettünk arra, hogy felhasználóbarát kezelőfelülete legyen. A felhasználói felület két részből áll: spektrum ablakból (16. ábra), ahol a spektrumok megjelenítése történik és egy vezérlőpultból, ahol a paraméterek beállítása és a számítások eredményei láthatók (14. és 15. ábrák).



The screenshot shows the 'Control Panel' window of GlycoMiner. It features a search bar for 'Protein Selection', a 'New Search' button, and a 'Load MSMS' button. There are checkboxes for 'Show Details', 'ONLY N-glycans', and 'Ascending Sort'. Below the search area is a table with columns: Charge, Mass, Intensity, Retention Time, Quality Score, Marker Score, Glyco Score, Glycan, Peptide, and PPM. The table contains 18 rows of data.

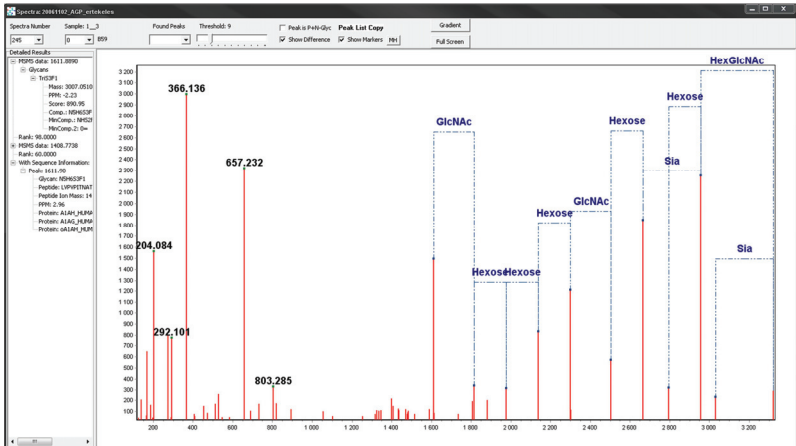
	Charge	Mass	Intensity	Retention Time	Quality Score	Marker Score	Glyco Score	Glycan	Peptide	PPM
245	3	1472.6255	6171.98	9.03	98.00	100.00	94.93	TriS3F1	LVPVPITNATLDR	2.96
231	3	994.3965	306.74	4.16	98.00	100.00	93.27	BIS2	ENGTISR	3.75
41	4	1068.2036	7884.68	9.03	98.00	72.22	90.24	TriS3	LVPVPITNATLDR	5.86
16	3	1213.1460	619.84	4.16	98.00	100.00	85.16	TriS3	ENGTISR	8.91
15	3	1001.0554	2369.09	3.92	98.00	100.00	83.55	BIS2	NEEYNK	10.15
85	4	1367.1188	2508.79	19.49	98.00	100.00	78.68	TriS2	SVDEIQATFFYFPH	16.57
53	4	1197.5353	7648.63	11.97	85.00	100.00	72.49	TriS3F1	LVPVPITNATLDQIT	7.96
14	3	904.0271	98.66	3.81	90.53	33.31	70.66	BIS1	NEEYNK	15.11
175	4	1104.7184	8736.39	9.03	74.59	100.00	66.11	TriS3F1	LVPVPITNATLDR	5.39
158	4	1074.1678	501.79	4.16	75.00	52.73	63.08	TetraS4	ENGTISR	7.04
156	3	1257.1558	157.54	3.52	75.00	100.00	61.76	TriS3F1	ENGTISR	5.17
187	4	1161.0178	5100.71	11.97	70.00	94.74	60.46	TriS3	LVPVPITNATLDQIT	5.63

14. ábra: Példa eredmény ablak, glikopeptid listával



15. ábra: GlycoMiner kezelőfelülete

A GlycoMiner egyszerűen töltésre konvertált bemeneti adatokkal dolgozik; az adatok formátuma lehet a ProteinLynx Global Server XML projekt formátuma, ASCII PKL (Peak List, csúcs lista) vagy MGF (Mascot Generic File, Mascot általános fájl) formátum, ami majdnem minden tömegspektrométer esetében elérhető. A GlycoMiner legfrissebb verziója letölthető a weblapunkról [138].



16. ábra: Spektrum kijelző ablak

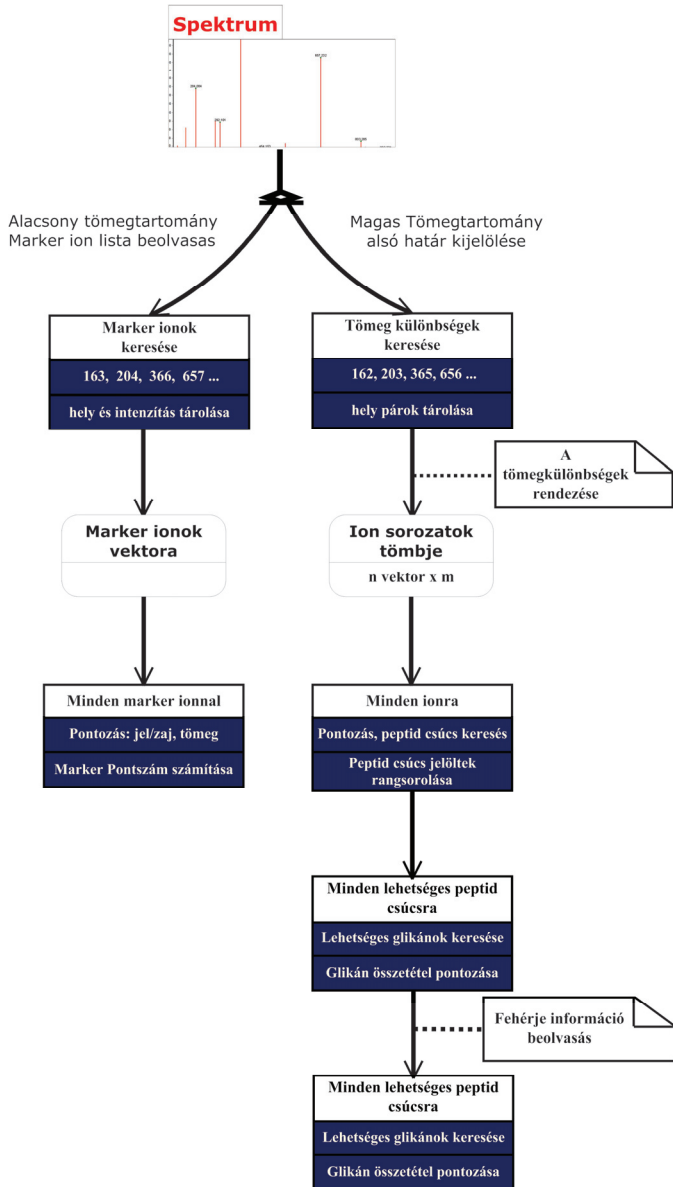
5.4 A glikopeptid összetételt meghatározó algoritmus részletei

A program működésének sematikus folyamatábrája a 17. ábrán látható. A pontozó rendszer által használt pontszámok jellemzőit a 4. táblázatban foglaltam össze. A spektrum minőségi pontszám (4. egyenlet) és a marker pontszám oxónium ionokat figyelembe vevő része kifejezhető analitikus formában (5. egyenlet). A marker pontszám erősen iteratív eljárások eredményét is tartalmazza. A spektrumokból nyert különféle adatokat a megfelelő megbízhatóság elérésének érdekében súlyozni kell. Az álpozitív azonosítások elkerülésére egy részletes pontozó rendszert fejlesztettünk ki. Abban az esetben például, ha egy triptikus hasítási helynél (lizin avagy arginin C terminálisa) nem történt meg a hasítás (pl. a glikán sztirikus gátlása miatt), a jelölt peptid kisebb pontszámot kap. A program a nem specifikus hasítási helyeket sem zárja ki, valamint az elméleti és a mért tömegek közötti nagyobb eltérést is megengedi, de ezek büntető pontokat vonnak maguk után. A büntetőpontok összeadódnak, és amennyiben az összegük nagyobb, mint egy előre beállított limit, a feltételezett összetételt az algoritmus elveti. A különféle pontszámok megfelelő optimalása elengedhetetlen a nagyarányú helyes azonosításhoz.

4. táblázat: A GlycoMiner által használt pontszámok összefoglalása

Pontszám	Leírás	Elfogadási alsó határ
Marker pontszám (S_{marker})	Azt mutatja, hogy egy spektrum glikopeptidről származik.	10
Spektrum minőségi pontszám (S_{quality})	Az MS/MS spektrumok minőségét (zajszint) és a peptidcsúcs megbízhatóságát jellemzi.	-
Glycoscore	A glikán összetétel meghatározásának megbízhatóságát jelzi, tartalmazza a peptid szekvencia pontszámát is.	15

A spektrum minőségi érték (S_{quality}) kvantitatívan jellemzi a spektrum minőségét, a kémiai és elektronikai zaj mennyiségét, és a glikán fragmentációból származó ionok spektrumbeli intenzitását. Az algoritmus a glikán részből származó veszteségek alapján kapott monoszacharid méretű tömegkülönbségek sorozatát is figyelembe veszi. Tapasztalataink szerint az eredmények jobbak, ha a spektrumban a zajt csak a közepes-nagy tömegtartományban vesszük figyelembe. Alapértelmezésként a középső tartomány alsó határa 800 m/z , ez a vezérlőpultban módosítható.



17. ábra: MS/MS alapú glikopeptid azonosítás folyamata

A spektrum minőségi érték az alábbi egyenlet alapján számítjuk:

$$S_{quality} = \frac{\sum_{i=0}^n Y[i]}{TIC} * \frac{n}{n_{spec}} * \frac{\sum_{i=0}^{n-1} (P[d[i]]^2 * d[i]^2)}{n * P_{max}^2 * d_{max}} * \frac{\prod_{i=0}^{nr} \left(2 - \frac{\delta[i]}{\delta_{max}}\right)}{\sum_{i=0}^{nr} \left(2 - \frac{\delta[i]}{\delta_{max}}\right)} * \rho * 100 \quad (4)$$

ahol n a fragmensként megjelölt ionok száma, n_{spec} a spektrum csúcsok számának összege, $Y[i]$ az i -edik csúcs intenzitása, TIC az összes csúcsintenzitás összege, $d[i]$ a csúcsok száma az i -edik és az előző fragmensként megjelölt csúcs között, d_{max} a figyelembe vett zaj csúcsok maximális száma, $\delta[i]$ az i -edik csúcs tömegpontossága, δ_{max} a legkisebb még elfogadható tömegpontosság, P a büntetőpontokat tartalmazó tömb, P_{max} a legnagyobb büntetőpont, és ρ a peptid tömegnek megjelölt csúcs megbízhatósági indexe (0.15-0.98 között).

A marker pontszám (S_{marker}) azt jelzi, hogy egy spektrum glikopeptidről származik-e. Ennek értékét nagyrészt a kis tömegeknél megjelenő oxónium ionok jelenléte határozza meg ($S_{marker,ox}$). Ezek az ionok tipikusan nagy intenzitásúak, ezért egyszerűen azonosíthatók még relatív zajos spektrumok esetén is (lásd 13. ábra). Abban az esetben, ha több oxóniumion észlelhető és a pontos tömegük megfelel az N -glikánok összetevőit alkotó szacharidoknak, nagy a valószínűsége, hogy a molekulában egy oligoszacharid szerkezet található. Ezek a kis tömegű oxónium ionok gyakran az MS spektrumokban is észlelhetők, N -glikopeptideken kívül más oligoszacharid részt tartalmazó molekulák esetén is észlelhetők. Ezért a marker pontszám például oligoszacharidok és O-glikopeptidek spektrumainak azonosítására is használható. Az S_{marker} és $S_{marker,ox}$ értéke 0 és 100 között változhat. Az $S_{marker,ox}$ analitikusan kifejezett képlete:

$$S_{marker,ox} = \frac{n}{N} * \frac{\sqrt{\sum_{i=1}^n P_i}}{n+1} * \frac{n}{n_{low} + 1} * 100 \quad (5)$$

$$P_i = \frac{Y[i]}{Y_{max}} * \left(1 - \frac{\delta[i]}{\delta_{max}}\right) * C \rightarrow \begin{cases} C=0, & S/N < 3 \\ C=1 & \end{cases}$$

ahol n a talált marker csúcsok száma, $Y[i]$ az i -edik marker ion intenzitása, C egy empirikus paraméter, amely 0 vagy 1 lehet attól függően, hogy mekkora a jel/zaj a csúcs 20 m/z körzetében, Y_{max} az alsó tömegtartomány legintenzívebb csúcsának intenzitása, $\delta[i]$ az i -edik csúcs tömegpontossága, δ_{max} a maximálisan megengedett tömegpontossági határ (amelyik csúcs tömege ennél nagyobb mértékben eltér az elméleti tömegtől, azt az algoritmus nem veszi

figyelembe), N az összes keresett marker ion darabszáma és n_{low} az alsó tömegtartományban lévő csúcsok száma.

Amint azt korábban ismertettem, a GlycoMiner grafikus felhatalmított felülettel rendelkezik. (14., 15. és 16. ábra). A program által meghatározott eredményeket a 14. ábra mutatja be. A listában láthatók, hogy mely spektrumok rendelhetők hozzá glikopeptidekhez (molekula tömeg, retenciós idő), a számított marker és minőségi pontszámok, és a peptid rész feltételezett tömege vagy a javasolt glikán összetétel (amennyiben a tandem tömegspektrumból meghatározható).

Az előbbieken bemutatott pontszámokon kívül bevezettünk egy harmadik pontszámot is, a *glikopontszámot* (glycoscore), amely azt jelzi, hogy mekkora a javasolt összetétel megbízhatósága. Eredményeink azt mutatják, hogy a 15-nél nagyobb glycoscore értékkel rendelkező struktúrák elfogadható azonosítást biztosítanak. A vezérlőpulton helyet foglaló opciós panel (15. ábra) lehetővé teszi a paraméterek (pl. elvárt tömegpontosság, a mintában található fehérjék szekvenciája, a figyelembe veendő monoszacharid egységek, stb.) változtatását. Ezek a paraméterek a mintára és a kísérletre vonatkoznak. A pontozó rendszer paraméterei a következők: a közepes tömeg tartomány kezdete, valamint a hozzátartozó intenzitás határok. A spektrum ablakban a spektrum és az analízis folyamatának néhány részlete látható.

5.5 GlycoPattern: minor glikoformok azonosítása és a glikozilációs mintázat meghatározása

LC-MS/MS mérések esetén részben a kromatográfiás futási idő, részben az érzékenység limitálja a felvehető tandem tömegspektrumok számát. Az automatikus rendszerekben jellemzően a spektrumok legintenzívebb 3-5 csúcsáról készül tandem tömegspektrometriás felvétel. Mivel a glikopeptidek intenzitása kisebb, mint a peptideké, így főleg peptidekről készül MS/MS spektrum. A glikopeptidekről felvett MS/MS spektrumok nem mindegyike értékelhető, azaz nem rendelhető hozzájuk összetétel/szerkezet. Ennek oka részben a rossz jel/zaj viszonyban, részben a mintában jelen lévő szennyezőkben, poszt-transzlációs módosulásokban keresendő. A fenti okok miatt a glikopeptidek többségéről nem áll rendelkezésünkre értékelhető tandem tömegspektrum, így az előzőekben ismertetett GlycoMiner sem használható. További probléma, hogy a glikozilációs mintázatok meghatározásához relatív mennyiségi meghatározás

is szükséges. Ez utóbbi azonban a tandem tömegspektrometriás mérésekből nem, vagy csak korlátozottan oldható meg.

A fent említett probléma normál (nem tandem) tömegspektrometria alkalmazásával oldható meg (kromatográfiával kapcsolva). Az LC-MS mérések a mintában lévő komponensek tömegéről, izotópeloszlásáról és egymáshoz viszonyított mennyiségéről adnak közvetlen információt, azonban a glikopeptidek fragmentációja (szerkezete) nem vizsgálható ezen módszerekkel.

A doktori munkám során célul tűztem ki egy olyan rendszer kidolgozását, amely a fent említett két technika lehetőségeit integrálja. Ehhez az LC-MS és LC-MS/MS mérések párhuzamos értékelése szükséges. A fent leírtak alapján egyrészt a LC-MS/MS (tandem tömegspektrometriás) mérések alkalmazása lehetővé teszi glikopeptidek szerkezetének (aminosav szekvenciájának és glikán összetételének) és retenciós idejének meghatározását, másrészt, ezen információkat felhasználva normál LC-MS mérés segítségével a glikofomákról kvantitatív adatok nyerhetők. Ez a koncepció az alapja a korábban ismertetett analitikai folyamatnak (10. ábra). A gyakorlati alkalmazáshoz egy újabb program kifejlesztése volt szükséges (GlycoPattern), melyet a későbbiekben a glikozlációs mintázatok meghatározására használtam. A munkafolyamat fő lépései a következők:

1) A mintában lévő fehérjék azonosítására a proteomikában szokásos bottom-up módszert alkalmazzuk. Adatfüggő LC-MS/MS mérés adatait adatbázis keresés segítségével értékeljük. Az így kapott fehérje listát felhasználjuk a következő lépésben.

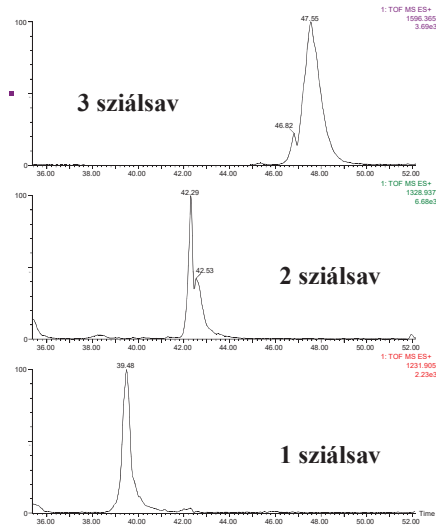
2) Glikopeptidek azonosításához is adatfüggő LC-MS/MS mérést végzünk. A glikopeptidek könnyebben fragmentálódnak, mint az azonos tömegű peptidek, ezért ezeket a méréseket alacsony ütközési energián végezzük [102]. A mérés tömegtartományát úgy állítjuk be, hogy lehetőleg minél kevesebb peptid és minél több glikopeptid essen bele. Jelen munkám során vizsgált glikopeptidek komplex glikánokat tartalmaztak, a legkisebb vizsgálandó tömeg is nagyobb volt 2400 Da-nál. Ezért a tömegtartományt m/z 800 és 2000 közé állítottuk be. Amennyiben csak kétszeresen és háromszorosan töltött ionokról veszünk fel spektrumot, akkor nagy eséllyel több glikopeptid ion kerül kiválasztásra, mint peptid. A tandem tömegspektrometriás mérés eredményeinek értékeléséhez a korábban bemutatott GlycoMiner programot használjuk. Így megkapjuk, a mintában található legintenzívebb glikopeptidek összetételét és retenciós idejét.

3) Az így kapott glikopeptid listában található peptid szekvenciák ismerete az alapja a következő lépésnek. Humán *N*-glikoproteinek esetén egy viszonylag kis számú (néhány száz) glikán összetétel lehetséges. A lehetséges glikán összetételekből elméleti listát állítunk össze,

amely alapján és a korábban azonosított glikopeptidek szekvenciájának felhasználásával kiszámíthatóak a keresendő glikopeptid tömegek. Fordított fázisú kromatográfia esetén elsősorban az aminosav szekvencia határozza meg a glikopeptid retenciós idejét, a peptidhez kapcsolódó cukrok csak kis mértékben módosítják azt (erről részletek a következő fejezetben találhatóak). Az elméletileg meghatározott tömegeket a szekvencia által meghatározott retenciós idő korlátok között keressük.

4) Ezt követően a glikozilációs mintázat meghatározásához normál LC-MS mérést végzünk alacsony ütközési energián. Ennek célja, hogy elkerüljük a glikopeptidek fragmentációját, és ez által egy torzított glikozilációs mintázatot nyerünk. A felvett spektrumokban megkeressük a listában szereplő molekula tömegeket több töltési állapottal, amelyekből ion kromatogramokat készítünk. Az ion kromatogramokat minden egyes tömeg esetén a megfelelő, előre meghatározott retenciós időablakra szűkítve hozzuk létre. Keresés közben nem csak a töltés állapotot, hanem az izotópeloszlást is figyelembe vesszük, mert az így kapott találatoknak sokkal kisebb az álpozitív aránya. Az egyes ion kromatogramokon csúcskeresést végzünk. A megtalált csúcsokra Gauss függvényeket illesztünk, majd kiszámoljuk ezek területét.

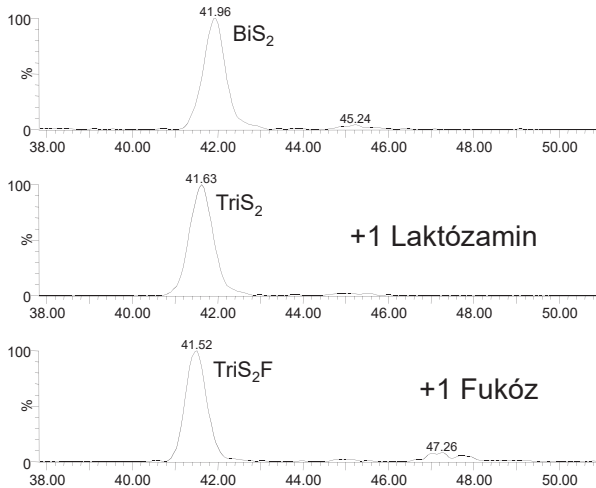
5) A mérések értékelésével kapott eredmények még mindig tartalmazhatnak véletlen találatokat. További problémát jelent, hogy a tömegspektrometria nagy szelektivitása ellenére is kaphatunk több kromatográfias csúcsot egy adott tömegre. Ezért az eredményeket szűrni kell, hogy minimalizáljuk a hibás találatok számát. Erre a szűrésre a kromatográfias relatív retenciós idők felhasználása ad lehetőséget. Fordított fázisú oszlopon a peptid szekvenciája határozza meg egy glikopeptid retenciós idejét. Ez a retenciós idő csökken, ha nő a molekula polaritása, vagyis ha nő a peptidhez kapcsolt glikán mérete. Ugyancsak a retenciós idő változását okozza a reziduális szilanol aktivitás is. Ez akkor fordul elő, amikor a fordított fázisú oszlop borítottsága nem tökéletes. Savas glikánok esetén a szilanol aktivitás megnövekedett retenciós időhöz és szélesebb csúcsalakhoz vezet, főleg a három vagy több szilolsavat tartalmazó glikánok esetén.[128, 139-141] (18. ábra) Ezek az effektusok felhasználhatóak az ugyanazon aminosav szekvenciával rendelkező glikopeptidek sorrendjének és relatív retenciós idejének a számítására. Amennyiben egy tömeghez több kromatográfias csúcs tartozik, a csúcshélesség és relatív retenciós idő alapján dönthető el, hogy melyik a keresett komponens.



18. ábra: Retenciósi idő és csúcsszélesség változása a sziálsav tartalommal

A relatív retenciósi idők előrejelzéséhez megvizsgáltuk három különböző típusú C18-as fázisú kromatográfiai oszlop retenciósi tulajdonságait. Először a laktózin egységek hatását a retenciósi időre gyakorolt hatását tanulmányoztuk az AGP glikopeptidjeinek segítségével. A vizsgált 32 glikopeptid pár esetén a laktózin egységgel nagyobb glikopeptid retenciósi ideje átlagosan 0.8%-al csökken (19. ábra). A retenciósi idő különbségek szórása 0.4%. A fukoziláció hatását 22 glikopeptid páron vizsgáltuk. Egy fukóz egység a retenciósi időt átlagosan 0.4%-kal csökkenti. A szórás ez esetben is 0.4% volt. Vizsgálataink alapján megállapítottuk, hogy a retenciósi idő változások additívak a laktózin és fukóz esetén. Egy laktózin és egy fukóz egység bevitel a molekulába $0.8 + 0.4 = 1.2\%$ -al csökkentette a retenciósi időt. A három vizsgált oszlop mind a laktózin, mind pedig egy fukóz egység különbség esetén hasonló eredményeket adott. A sziálsavak esetében az eredmény függ a vizsgált oszloptól. A legkisebb szilanolaktivással rendelkező oszlopon egy sziálsav átlagosan 0,3%-kal növeli a retenciósi időt, míg a legnagyobb esetben ez az érték 6% volt.

Az azonosított komponensek intenzitását a kromatográfiai csúcsra illesztett Gauss függvények paramétereinek alapján számoljuk. Az egy aminosav szekvenciához tartozó intenzitások leírják az adott glikopeptid glikozilációs mintázatát. Az egy fehérjéből származó összes glikopeptid relatív intenzitása az adott fehérje helyspecifikus glikozilációs mintázatát alkotja.



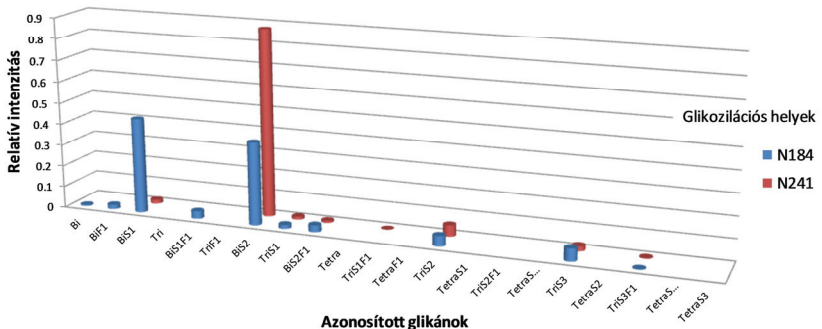
19. ábra: Retenciós idő változása antenna szám, fukóz tartalom függvényében.

Az eredményeket összefoglalva, kidolgoztunk egy LC-MS/MS és LC-MS méréseken alapuló munkafolyamatot, amely alkalmas glikoproteinek helyspecifikus glikozilációs mintázatának meghatározására. A munkafolyamat tartalmazza glikopeptidek főleg tandem tömegspektrometrián alapuló azonosítását, valamint az LC-MS alapú relatív mennyiségi meghatározását (10. ábra). Tandem tömegspektrometriás mérést végzünk a mintában található fehérjék azonosítására, hogy ez által szekvencia információhoz jussunk. Glikoproteinek glikozilációs helyeinek, vagyis a mintában található glikopeptidek aminosav szekvenciájának azonosítására is külön mérést végzünk. Ilyenkor más tömegtartományban veszünk fel spektrumokat, hogy maximalizáljuk a tandem tömegspektrometriára kiválasztott glikopeptidek arányát. A harmadik, normál tömegspektrometriai mérés során az azonosított aminosav szekvenciák és retenciós idők alapján meghatározzuk az elméletben lehetséges glikopeptid listából a kromatogramban jelen lévő glikopeptidek relatív intenzitását. A komponensek meghatározását izotópeloszlás és kromatográfiai paramétereken alapuló szűrők alkalmazásával pontosítjuk. A meghatározott intenzitások a mintában lévő glikoproteinek helyspecifikus glikozilációs mintázatát adják. A kidolgozott módszer (ill. a GlycoMiner program) nagy hatékonysággal képes glikopeptideket azonosítani. Mindössze 0,1% álpozitív és 0,1% álnegatív találati hibával dolgozik és még zajos spektrumok esetében is jól használható. A módszer további előnye, hogy proteinkeverékekben is használható, így a célfehérje előzetes izolálása nem

szükséges. Ez lehetővé teszi bonyolult keverékek, így pl. vérplazma és más biológiai minták közvetlen vizsgálatát.

5.6 Példák glikozilációs mintázatok meghatározására

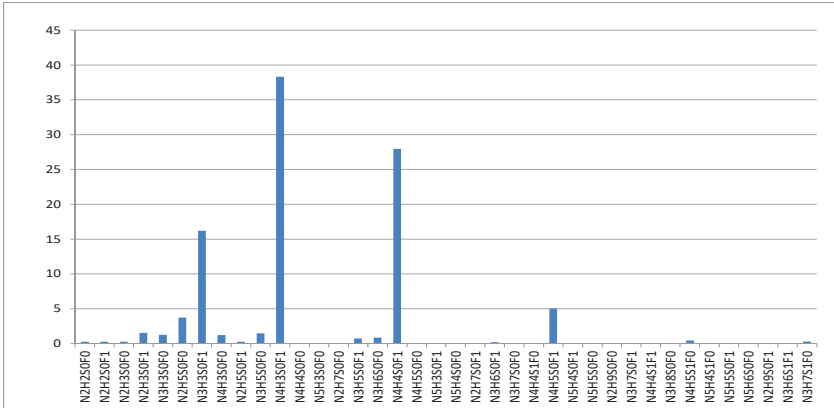
A kidolgozott módszer alkalmazhatóságát vizsgáltuk biológiai mintákon. Egyszerű minta-előkészítést követően (az albumin nagy részének az eltávolítása) ú.n. albumin depletált vérplazmát vizsgáltunk a fent leírt LC-MS módszerrel. Meghatároztuk a néhány leggyakoribb fehérje, mint például a haptogloblin glikozilációs mintázatát (20. ábra). Az ábrán a haptogloblin 184. és 241. aminosavaihoz kapcsolódó glikozilációs mintázata látható. Megfigyelhető, hogy a glikozilációs mintázat a két glikozilációs ponton jelentős mértékben eltér egymástól, pl. a BiS1 szerkezet csak a 184. aminosavhoz kapcsolódva jelenik meg, amely biológiai szempontból jelentőséggel bírhat.



20. ábra: Haptogloblin helyspecifikus glikozilációs mintázata: két glikozilációs helyen (^{184}N , ^{241}N) azonosított glikán összetételek és relatív intenzitásuk.

A módszer másik fontos alkalmazási területe a gyógyszeripar lehet. Például a terápiás fehérjék glikozilációjának vizsgálatában tölthet be kulcsfontosságú szerepet, mely a gyógyszer ellenőrzési folyamatok egyik legnehezebb lépése. Ennek illusztrálására egy monoklonális antitestet (infiximab) vizsgáltunk, amelynek glikozilációs mintázata a 21. ábrán látható. Itt szeretnénk megjegyezni, hogy az infiximab *N*-glikozilációs mintázata jelentősen eltér az emberi

fehérjéknél megszokottaktól, ezért a glikán szerkezetekről csak az összetétel meghatározása lehetséges, a részletes szerkezet nem.



21. ábra: Monoklonális antitest (infliximab) glikozilációs mintázata: azonosított glikán összetételek és relatív intenzitásuk. N – N-acetil-hexózamin, H – hexóz, S – szialsav, F – fukóz.

5.7 Glikopeptidek és glikoproteinek dúsítása

A glikoproteinek vizsgálatát nagyban nehezíti, hogy az ebből képződő glikopeptidek tömegspektrometriás érzékenysége kisebb, mint a peptidké, ezen túlmenően a glikopeptideket nagyfokú heterogenitás is jellemzi. Amennyiben tiszta glikoprotein minta áll rendelkezésünkre, a glikozilációs mintázat bár egy meglehetősen hosszadalmas eljárással (például a fentiekben ismertetett módon), de meghatározható. Amennyiben a vizsgálandó glikoprotein egy keverékben (pl. vérplazmában) van jelen, a vizsgálat érzékenysége és szelektivitása ideális esetben is csak néhány főkomponens glikozilációs mintázatának meghatározását teszi lehetővé. A feladat elvben egyszerűen, az adott protein izolálásával megoldható. Ez azonban általában hosszadalmas, bonyolult és költséges, ezért szükséges alternatív megoldások alkalmazása. Ennek egyik módja, ha például az izolálást az általában egyszerűbb és olcsóbb dúsítással helyettesítjük, amely glikoprotein keveréket eredményez. A haptogloblin vizsgálata esetén igazoltuk, hogy az ilyen keverékek vizsgálata az általunk kidolgozott LC-MS módszerrel megoldható.

Glikopeptidek, glikoproteinek dúsítására számos módszer található az irodalomban, de ezen dúsítási módszerek jellemzői (pl. szelektivitásuk, torzításmertességük) nem ismertek kellő részletességgel. A fent kidolgozott analitikai módszer egyik előnye, hogy a dúsítási eljárások jellemzésére is használható. A dúsítási folyamatot elvégezhetjük az intakt fehérjéken, amelyet proteolízis követ (dúsítás glikoprotein szinten). Alternatívaként proteolízis után a peptidkeveréken is elvégezhetjük a dúsítást (dúsítás glikopeptidek szintjén). Helyspecifikus glikoziláció vizsgálata tipikusan glikopeptidek jellemzését igényli, ezért az ilyen vizsgálatoknál a glikopeptidek dúsítása előnyösebb.

Munkám során cél volt különböző dúsítási módszerek összehasonlítása és tömegspektrometriai adatok alapján a dúsítási módszerek sikerességének kvantitatív jellemzésére alkalmas mutató kidolgozása. Egy másik fontos kérdés glikoproteinek dúsításakor, hogy megváltozik-e az eredeti glikoform arány (glikozilációs mintázat) a dúsított mintában. A glikoformák egymáshoz viszonyított arányának megváltozása a glikozilációs mintázat megváltozását jelenti és téves/torzított eredményekhez vezethet.

Számos lehetőség létezik glikoproteinek és glikopeptidek dúsítására. Dolgozatomban három, *N*-glikopeptidek dúsítására alkalmas szilárd fázisú extrakción (SPE) alapuló módszert hasonlítottam össze. A szilárd fázisú extrakció könnyen automatizálható, így használható nagy áteresztőképességű módszerekben. A munkánk során fenilbórsav affinitást, búzacásra agglutinin lektin affinitást és anioncserés módszereket használtunk. A fenilbórsav cisz-dioloikat képes kovalensen megkötni, míg a búzacásra agglutinin *N*-acetyl-glükózamin és szialsavat köt specifikusan. Más lektinek is használhatók glikopeptidek dúsítására, mint például a konkanavalin A, de a lektinek specifitásának összehasonlítása nem célja a dolgozatnak. Anioncsere estén a savas komponensek megkötésére és dúsítására nyílik lehetőség. Mivel a célkomponensek sok esetben komplex *N*-glikánok, amelyek több-kevesebb szialsavat tartalmaznak, anioncsere is alkalmazható a dúsításukra. A dúsításokat kezeletlen humán vérplazmán irodalmi módszerek alapján végeztem (a pontos eljárást a 4.4. fejezet tartalmazza). A dúsítási lépések után kapott mintákat nanoUHPLC-MS(MS) módszerrel vizsgáltuk, a glikopeptidek azonosítása tandem tömegspektrometriai adatok alapján a már bemutatott módon történt, a GlycoMiner program segítségével (5.2. fejezet).

A glikopeptidek dúsításának jellemzése nehéz feladat, ugyanis a tipikusan proteomikai módszerek a glikopeptideket nem ismerik fel, így róluk mennyiségi adatot sem szolgáltatnak. A gyakorlatban legtöbbször csak az azonosított glikoproteinek vagy peptidek számát használják e célra [72, 77, 142-146], amely nem szolgáltat kellően pontos eredményt. Ezért a glikopeptidek dúsulásának egyértelmű jellemzésére három különböző indexet dolgoztunk ki. Az első index (6.

egyenlet) az adott frakcióban található glikopeptidek mennyiségét jellemzi, a második (7. egyenlet) a nem glikozilált peptidek mennyiségét írja le, míg a harmadik index (8. egyenlet) a dúsítást jellemzi.

A peptidek mennyiségét a proteomikában általánosan alkalmazott analóg módon jellemeztük, a kromatográfias mérés során azonosított három legnagyobb intenzitású peptid intenzitás-összegét használva (6. egyenlet). Esetünkben ez jellemezte a minta peptid/fehérje mennyiségét. A peptidek jellemzésével analóg módon három intenzitás összegével jellemeztük a glikopeptidek mennyiségét is (7. egyenlet). Mivel a módszert alapvetően komplex *N*-glikánok vizsgálatára fejlesztettük ki, az *N*-glikánok leggyakoribb frgmenteiseit használtuk erre a célra. A glikán frgmenteiket 15 eV-ra növelt ütközési energiával nyertük. CID fragmentáció esetén a komplex glikánok spektrumában intenzív *m/z* 366,14, 512,19 és 657,23 csúcsoakat figyelhetünk meg: az első egy *N*-acetyl-laktózámin, a második ennek egy fukózzal növelt változata, míg a harmadik egy szialsavval szubsztituált *N*-acetyl-laktózámin. Mind a peptid, mind pedig a glikopeptid index esetén az egyes összegeket az összes frakció peptid illetve glikopeptid mennyiségére normáljuk.

$$\text{Peptid index az } i \text{ frakcióban (\%)} = \frac{\sum i \text{ frakció három legintenzívebb peptidje}}{\sum_{i=1}^n (\sum i \text{ frakció három legintenzívebb peptidje})} \times 100 \quad (6)$$

$$\text{Glikopeptid index az } i \text{ frakcióban (\%)} = \frac{\sum I_{366}, I_{512}, I_{657} \text{ az } i \text{ frakcióban}}{\sum_{i=1}^n \sum I_{366}, I_{512}, I_{657} \text{ az } i \text{ frakcióban}} \times 100 \quad (7)$$

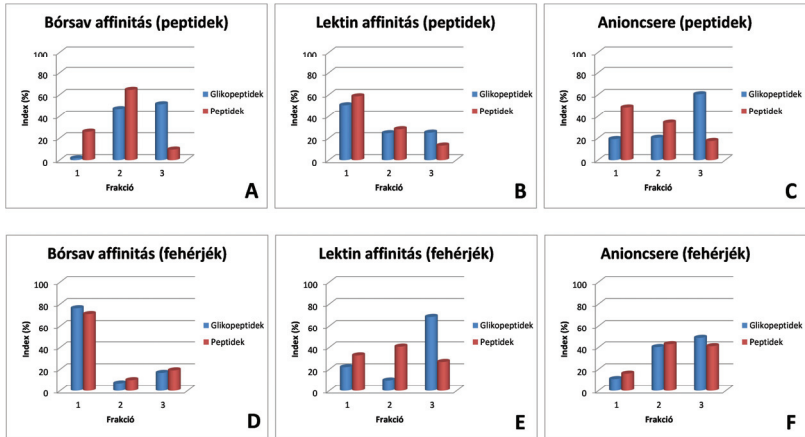
ahol *n* az összes frakció száma, I_{366} az *m/z* 366,14, I_{512} az *m/z* 512,19 és I_{657} az *m/z* 657,23 ion kivont ionkromatogramjának csúcs alatti területe.

A glikopeptid dúsítást a fenti egyenletekkel definiált két mennyiség arányaként vezetjük be (8. egyenlet).

$$\text{Glikopeptid dúsítási index az } i \text{ frakcióra} = \frac{\text{glikopeptid index az } i \text{ frakcióra}}{\text{peptid index az } i \text{ frakcióra}} \quad (8)$$

A kidolgozott indexek egyaránt használhatók LC-MS és MALDI felvételekből történő kiértékelésre is. A glikopeptid dúsítási index a glikoziláció mértékének jellemzésére is használható egy adott mintában. Itt szeretném megjegyezni, hogy az indexek abszolút értékben nem adnak pontos értéket, de különböző minták összehasonlítására nagyon kedvezőek.

A fent leírt indexeket alkalmaztuk három különböző dúsítási módszer összehasonlítására. Munkánk során vérplazmát dúsítottunk, amelynek nagy intenzitású komponensei többségében glikoziláltak. Mind fehérje, mind peptid szinten elvégeztük a dúsítást. Mindhárom módszer esetén három frakciót szedtünk (kísérleti részletek a 38. oldalon). A 22. ábrán láthatók az egyes módszerek esetén kapott dúsítási indexek. Fehérje szinten csak kismértékű dúsulást tapasztaltunk, amelynek legnagyobb mértéke (kb. 2x) a lektin a harmadik frakcióban észlelhető.

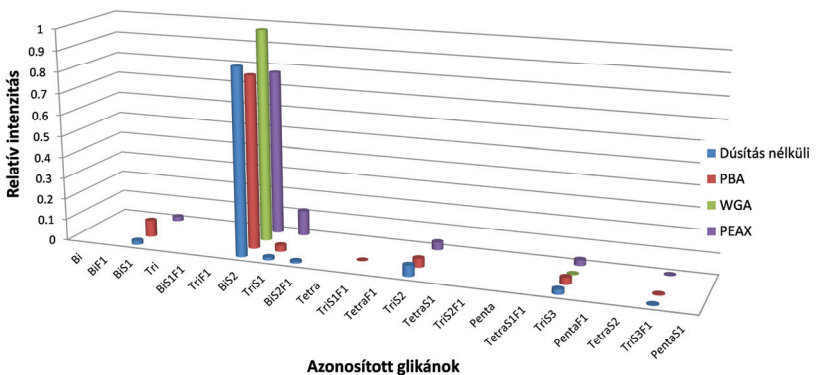


22. ábra: Glikopeptid és peptid index változása az egyes frakciókban: borsav affinitás peptidek esetén (A), fehérjék esetén (D); lektin affinitás peptidek esetén (B), fehérjék esetén (E); anioncsere peptidek esetén (C), fehérjék esetén (F)

Ezzel szemben a peptid szinten történő dúsítás sokkal eredményesebbnek bizonyult. A legnagyobb dúsítási indexet fehérjék esetében a lektin affinitás alkalmazásával nyertük ($\times 2,5$), míg a peptid szinten borsav affinitás alkalmazásával 5,3-szoros dúsulást tapasztaltunk. Megjegyzem, hogy ideális esetben a glikopeptid 100 %-a az utolsó frakcióban eluálódik és peptidek nincsenek jelen. A peptid szinten a lektin affinitás kromatográfiával nyert eredmények csak kis mértékű dúsulást eredményeztek (dúsítási index 1,9): nagyon sok glikopeptid eluálódott az első frakcióban, és sok peptid megkötődött az SPE oszlopon (ez ellentmondásban van a várakozásokkal). Az erős anion cserés és a borsavas dúsítás módszerek is ígéretesnek bizonyultak (dúsítási index 3,5, illetve 5,3). Mindkét esetben az utolsó frakció a glikopeptid több, mint a felét tartalmazta, míg a peptidek csupán átlagosan 15%-át. Az erős anion csere kb. négyszeres mértékű dúsítást eredményezett. A legjobb eredmények fenilborsavas SPE alkalmazásával nyerhetők: az utolsó frakcióban a glikopeptid fele található, és alig tartalmaz

peptideket (10%). A módszer további optimalizálásával, konkrétan a 2. és 3. frakció közötti határ jobb megválasztásával valószínűleg még jobb eredmények érhetők el.

A dúsítás mértékének meghatározásán kívül vizsgáltam továbbá, hogy a dúsítás során ezekben az esetekben a glikozilációs mintázat hogyan változik meg. A 23. ábrán peptid szinten történő dúsítás után a haptogloblin fehérje glikozilációs mintázata látható a legnagyobb mértékű dúsítást eredményező 3. frakciókban. Megállapítható, hogy a borsav affinitás és az anioncsere hatására az észlelt mintázat változása nem számottevő, míg a lektin affinitás során a csak a BiS2 szerkezetű glikán kötődik meg.



23. ábra: Haptogloblin VVLHPNYSQVDIGLIK glikopeptid glikozilációs mintázatának változása az alkalmazott három dúsítási eljárás során; a PBA – borsav affinitást, WGA – lektin affinitást, a PEAX – anioncsere jelöl.

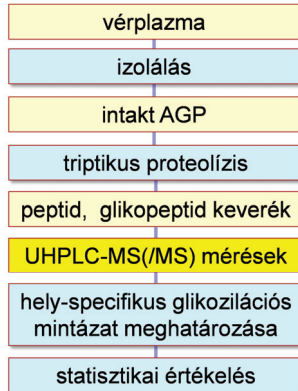
Az eredményeket összefoglalva, megvizsgáltunk három dúsítási eljárást mind peptid, mind fehérje szinten. Az egyes frakciók glikopeptid tartalmát egy dúsítási indexszel jellemeztük. Az eredmények alapján megállapítottuk, hogy a peptidok dúsítása hatékonyabb és előnyösebb, mintha ugyanezt a glikoproteinek szintjén végeznénk el. A legnagyobb mértékű dúsítás 5,3-szeres, míg fehérjéknél 2,5-szeres volt. A glikopeptidok dúsítására a borsav affinitás alapján működő SPE oszlop alkalmazása a legeredményesebb.

5.8 A helyspecifikus glikozilációs mintázat alkalmazása orvosi-diagnosztikai feladat megoldására: AGP glikozilációjának változása atherosclerosis esetén

Doktori munkám utolsó részében a glikozilációs mintázat meghatározására kidolgozott módszerünket egy fontos orvosi-diagnosztikai feladat megoldására használtuk. Azt feltételeztük, hogy a vérplazma molekuláris mintázata, ill. az ebben bekövetkező változás jellemezheti az egyes betegségeket (pato-fiziológiai folyamatokat). Ezért fontos kutatási irány az új „biomarkerek” keresése. Főleg technológiai okokból ezek a biomarkerek legtöbbször nukleinsavak (DNS, RNS), fehérjék vagy metabolitok jellemzésén alapulnak. Mivel a glikozilációt ezektől jelentősen eltérő mechanizmusok szabályozzák, feltételeztük, hogy a glikoziláció ismerete fontos további információt szolgáltathat egyes betegségek jellemzésére, melynek klinikai-diagnosztikai jelentősége is lehet. Spanyol együttműködés keretében egy fontos népbetegség, az atherosclerosis vizsgáltuk.

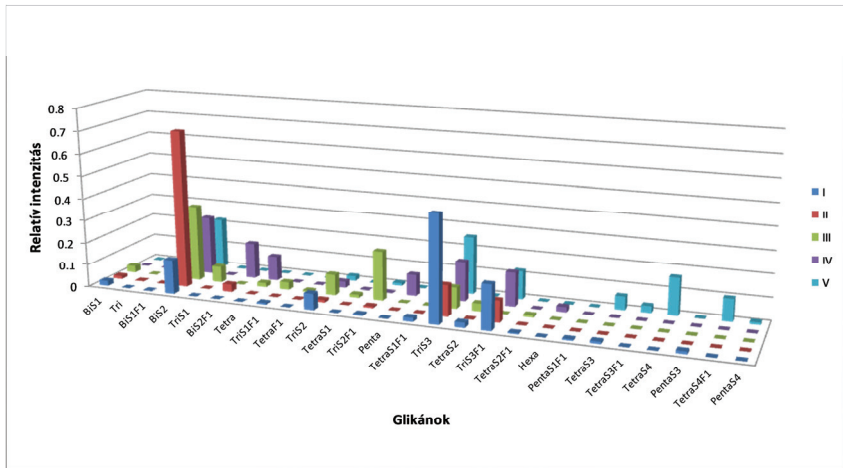
Az irodalomban először 2011-ben jelent meg tanulmány az AGP és az atherosclerosis kapcsolatáról [132]. A vizsgálat során kapilláris zónaelektroforézis segítségével 13 intakt AGP izoforma mennyiségi változásait követték egészséges és betegektől származó vérminták felhasználásával. A kapott eredményeket statisztikai analízisnek vetették alá, melynek eredményei ugyan az orvosi probléma megoldására nem voltak elegendőek, de biztatóak voltak. A tanulmány azonban csak az intakt AGP néhány (13) izoformjával foglalkozik, a helyspecifikus glikozilációs mintázatokat nem vizsgálták. Ezért a spanyol-magyar közös projectben megvizsgáltuk az AGP helyspecifikus glikozilációs mintázatát egészséges, atherosclerosisban és aneurizmában szenvedő betegektől származó mintákban. Célunk volt annak kiderítése, hogy az AGP biomarkerként alkalmazható-e atherosclerosisos megbetegedések esetén. A biomarker kutatásban mindig nehéz annak eldöntése, hogy a megtalált biomarker mennyire specifikus az adott betegségre vagy – rossz esetben – csak egy olyan biomarkert sikerült azonosítani, amely az általános rossz egészségügyi állapotra jellemző (pl. azt mutatja, hogy valaki hetekig az ágyban fektött, vagy csak azt, hogy gyulladásos folyamat van a szervezetében). Ezért nagyon fontos a megfelelő kontroll csoportok kiválasztása, azaz az egészséges csoport mellett szükséges egy másik érbetegségben szenvedő csoport kiválasztása is. Jelen esetben az együttműködő orvos kollégák választása az aneurizmára esett (részletek a 31. oldalon található). Összesen 28, három csoportba tartozó, izolált AGP mintát vizsgáltunk. A minták eloszlása a következő volt: 9 egészséges, 10 atherosclerosisban szenvedő beteg és 9 aneurizmás (pozitív kontroll).

Az elvégzett munkafolyamat a 24. ábrán látható:



24. ábra: A vizsgálatok munkafolyamata

Az első lépés az AGP izolálása véralplazmából, amit a spanyol együttműködő partnerek végeztek [132]. Az így izolált AGP minták helyspecifikus glikozilációs mintázatának meghatározására az előző fejezetben ismertetett módszert alkalmaztuk: a mintákat először tripszines emésztésnek vetettük alá a 34. oldalon ismertetett módon. Az AGP legintenzívebb glikopeptidjeinek szerkezeti azonosítását a fő komponensek adatfüggő nanoUHPLC-MS/MS mérésekkel végeztük, amelyeket a már ismertetett GlycoMiner segítségével értékeltünk ki. Így nyertük a mintában előforduló glikopeptidek aminosav szekvenciáját és retenciós idejét. Ezt követően minden egyes mintát megvizsgáltunk nanoUHPLC-MS-sel. A nanoLC-MS mérési adatok és a nanoLC-MS/MS-sel meghatározott aminosav szekvenciák és retenciós idők alapján meghatároztuk az egyes mintákban jelen lévő AGP helyspecifikus glikozilációs mintázatát az előző fejezetben ismertetett módon (GlycoPattern program segítségével). Peptidenként (9 peptid) negyvennégy glikoformot, azaz összesen 396 glikopeptidet kerestünk. Ebből a nanoUHPLC-MS mérések során pontos tömeg alapján 156 glikopeptidet azonosítottunk. Ezt követően az összes azonosított glikopeptidről MS/MS felvételt készítettünk. Az összes azonosított 156 glikopeptidből 83-at sikerült MS/MS spektrumok alapján is igazolni, a többi csúcsonak nem volt kellő intenzitása a tandem tömegspektrometriás azonosításhoz. A tandem tömegspektrumok alapján is azonosított glikopeptideket az 5. táblázat tartalmazza. Egy atherosclerosisban szenvedő beteg esetén a helyspecifikus mintázat az 25. ábrán látható.

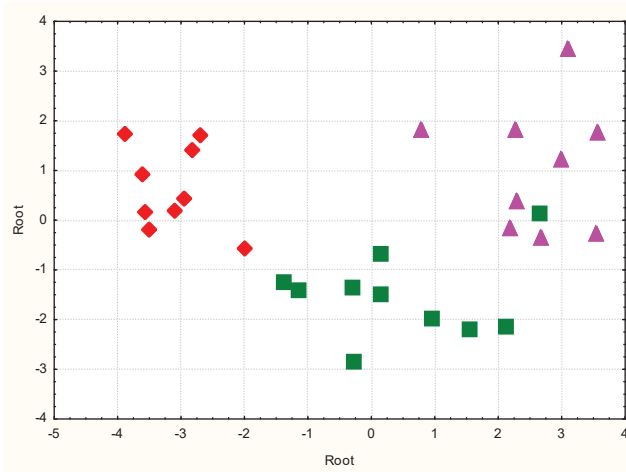


25. ábrán: Egy atherosclerosisban szenvedő beteg véréből izolált AGP helyspecifikus glikozilációs mintázata. (a különböző színek a különböző glikozilációs helyeket jelölik)

A fenti analitikai módszer reprodukálhatóságát kereskedelmi AGP mintákon vizsgáltuk, figyelembe véve a proteolízist, a tömegspektrometriás mérés és a kiértékelésből adódó hibákat is. Összesen 9 kereskedelmi AGP mintát vizsgáltunk. Az AGP proteolízise ugyanazon a napon történt. A nanoLC-MS mérések alapján meghatározott 128 glikopeptid mintázata 12% átlagos szórást mutatott. Amennyiben a kis intenzitású (<100 cps) glikopeptideket kihagyjuk, az átlagos szórás kisebb, mint 7%.

A 28 minta vizsgálata után a mérési adatokat statisztikai elemzésnek vetettük alá. Az egyes csoportok megkülönböztetésére lineáris diszkriminancia analízist végeztünk a Statistica számítógépes programcsomag segítségével. A betanítóhalmaz az összes mérési eredményt tartalmazta, tehát 28 esetből (minták) és 128 változóból (azonosított glikopeptidek) állt. Az osztályozás eredményeit egy-elem-kihagyásos keresztellenőrzéssel (cross validation) ellenőriztük. Az első osztályozás átlagosan 96%-ban helyes osztályba sorolta be a betanítóhalmaz elemeit (26. ábra). A keresztellenőrzött eredmények 89%-a helyes besorolási aránnyal rendelkezik. Az egyes csoportok megkülönböztetése nagyon jó (egészséges 89%, atherosclerosis 80%, aneurizma 100%). Itt szeretném hangsúlyozni, hogy nemcsak az egészséges és beteg, hanem az atherosclerosis és az aneurizma csoportok között is képes a módszer különbséget tenni. Vizsgálataink során az aneurizma csoportot pozitív kontrollként használtuk, amely a vizsgálat specifikusságának növelését szolgálja. Véleményünk szerint az, hogy a két beteg csoport egymástól megkülönböztethető sokkal fontosabb eredmény, mint önmagában az a

tény, hogy a beteg és az egészséges csoportokat meg tudtuk egymástól különböztetni. Ez arra utal, hogy a glikozilációs mintázat specifikus a betegségek megkülönböztetésére. Ez a vizsgálat specifikusságának a növelését szolgálja. A módszer érzékenysége atherosclerosisra nézve 80%, specificitása 94%. A statisztikai értékelés eredményei azt jelzik, hogy az AGP helyspecifikus glikozilációjának változása potenciális biomarker atherosclerosis esetén.



26. ábra: AGP minták lineáris diszkriminancia analízise: piros rombusz – aneurizma, zöld négyzet – atherosclerosis, lila háromszög- egészséges

Végül az eredményeinket összehasonlítottuk az irodalomban Angel Puerta és munkatársai által közölt eredményekkel [132]. Puerta és munkatársai összesen 13 intakt AGP izoformát különböztettek meg 28 mintában; vizsgálatuk helyspecifikus glikozilációra nem tért ki. Statisztikai analízis alapján a módszerük atherosclerosis esetén 33.3% érzékenységgel és 73.7% szenzitivitással jellemezhető (keresztellenőrzött értékek). A jelen dolgozatban leírt helyspecifikus glikozilációs adatok statisztikai analízise alapján a módszer lényegesen jobb szenzitivitással rendelkezik (80%, illetve 94%). Ez alapján kijelenthető, hogy az AGP helyspecifikus glikozilációjának vizsgálata előnyösebb az intakt AGP izoformok vizsgálatánál, vagyis atherosclerosis esetén ez jobb biomarker jelölt.

5. táblázat: Azonosított glikopeptidok; +++ az MS/MS azonosított intenzív glikopeptid, ++ MS azonosított, később MS/MS igazolt glikopeptid, + MS azonosított, - nem azonosított

Glikánok (in silico)	Glikozilációs hely				
	I	N33	II N56	III N72	IV N93
Bi	+	-	+	-	-
BiF1	-	-	-	-	-
BiS1	+++	+++	++	+	-
BiS1F1	-	+	+	+	-
BiS2	+++	+++	+++	++	+++
BiS2F1	-	+++	+	+	-
Tri	-	-	-	-	-
TriF1	-	-	-	-	-
TriS1	+++	++	++	++	+
TriS1F1	++	+	-	-	-
TriS2	+++	+++	+++	+++	+++
TriS2F1	+++	+++	++	+++	++
TriS3	+++	+++	+++	+++	+++
TriS3F1	+++	+++	++	+++	+++
Tetra	-	-	-	-	-
TetraF1	-	-	-	-	-
TetraS1	++	-	++	++	-
TetraS1F1	-	-	-	+	-
TetraS2	++	-	+++	+++	++
TetraS2F1	+	-	++	++	++
TetraS3	++	-	+	+++	+++
TetraS3F1	++	-	+	+	+++
TetraS4	++	-	+	+++	+++
TetraS4F1	-	-	+	+	+++
Penta	-	-	-	-	-
PentaF1	-	-	-	-	-
PentaS1	-	-	-	+	-
PentaS1F1	-	-	-	-	-
PentaS2	-	-	-	++	-
PentaS2F1	-	-	-	-	-
PentaS3	+	-	+	+	++
PentaS3F1	-	-	-	+	-
PentaS4	-	-	-	+	++
PentaS4F1	-	-	-	+	+
Hexa	-	-	-	-	-
HexaF1	-	-	-	-	++
HexaS1	-	-	-	-	-
HexaS1F1	-	-	-	-	-
HexaS2	-	-	-	-	-
HexaS2F1	-	-	-	-	-
HexaS3	-	-	-	-	++
HexaS3F1	-	-	-	-	-
HexaS4	-	-	-	-	++
HexaS4F1	-	-	-	-	-

Az eredményeket összefoglalva az AGP helyspecifikus glikozilációs mintázatának vizsgálatával az atherosclerosis megbetegedés előrejelzésére egy igen jó, 89%-os predikciós mutatóval rendelkező biomarkert találtunk. Ez alátámasztja, hogy az általunk kidolgozott minta-előkészítési és analitikai módszerek, valamint a GlycoMiner program alkalmazásának nemcsak elméleti, hanem komoly gyakorlati jelentősége is van. Segítségükkel kis mennyiségű fehérjemintából is (AGP esetén 10-20 pmol) lehetőség van a különböző glikopeptidek azonosítására, és a glikozilációs mintázat különbségeinek meghatározására. A glikopeptid mintázat sok esetben alkalmas arra, hogy egészséges és beteg csoportok, valamint pozitív kontrollcsoportok között különbséget tudjunk tenni, azaz különböző patológias állapotokat előre tudjunk jelezni. A kezdeti, igen pozitív vizsgálati eredményeink alapján az Országos Kardiológiai Központtal együttműködésben a kutatás folytatását tervezzük. Célunk a kezdeti eredmények jelentősen nagyobb betegszám esetén történő igazolása (validálása). Ezt követően a módszer szűrővizsgálatként történő bevezetése is elképzelhető.

6 Összefoglalás

Doktori munkám során helyspecifikus *N*-glikozilációs mintázat meghatározására alkalmas módszert dolgoztam ki és alkalmaztam különböző orvosi-diagnosztikai és analitikai problémák megoldására. A kidolgozott módszert atherosclerosis glikozilációs biomarkerek, valamint vérplazma fehérjék glikozilációs mintázatának meghatározására használtam, amelyet közvetlenül albumin depletált plazmából mértünk. Kidolgoztam továbbá a glikoproteinek dúsításának jellemzésére alkalmas mérőszámokat és vizsgáltam a glikozilációs mintázat dúsítás közben bekövetkező változását. Az elért tudományos eredményeket az alábbiakban foglalom össze:

- 1) Kidolgoztam egy nano-UHPLC-MS(/MS) alapú munkafolyamatot, amely alkalmas glikoproteinek helyspecifikus glikozilációjának meghatározására. A módszer legnagyobb előnye, hogy i) nem csak izolált glikoproteinek, hanem glikoprotein keverékek vizsgálatára is alkalmas, ii) kellően érzékeny kis mintamennyiségek vizsgálatára (1-10 pmol), így tényleges biológiai minták vizsgálatára használható, iii) nagy áteresztőképességű (azaz több száz minta vizsgálatára is alkalmazható). A munkafolyamat több – egymásra épülő – mérési folyamat és értékelés elvégzésén alapul. Tandem tömegspektrometriás mérést végeztem a mintában található fehérjék azonosítására (ezáltal szekvencia információkat nyertem), glikoproteinek glikozilációs helyeinek, vagyis a mintában található glikopeptidok aminosav szekvenciájának meghatározására és normál tömegspektrometriai mérés segítségével nyertem a glikopeptidok pontos arányát.
- 2) Kidolgoztam egy olyan számítógépes algoritmust és programot (GlycoPattern), amely az azonosított aminosav szekvenciák és retenciós idők alapján *in silico* összeállítja az elméletben lehetséges glikopeptidok listáját. Ezt a program a kromatogramban jelen lévő glikopeptidok relatív intenzitásarányának meghatározására használja fel. Az álpozitív találatok számát izotópeloszlás és kromatográfias paramétereken alapuló szűrők alkalmazásával csökkenttem. A meghatározott intenzitások a mintában lévő glikoproteinek helyspecifikus glikozilációs mintázatát adják.
- 3) A glikopeptidok szerkezetének ill. összetételének tandem tömegspektrumuk alapján történő azonosítására algoritmust és számítógépes programot fejlesztettem ki (GlycoMiner). A tesztek alapján a program jól használható a glikopeptidok azonosítására,

mivel a GlycoMiner mindössze 0,1% álpozitív és 0,1% álnegatív hibával dolgozik, és zajos spektrumok esetében is jól használható. A GlycoMiner 100% pontossággal határozza meg a szerkezeteket azokban az esetekben, amelyekben a manuális kiértékelés is lehetséges. A program az MTA TTK SZKI honlapjáról ingyenesen elérhető (<http://www.chemres.hu/ms/glycominer>). A dolgozat megírásáig több mint 150 felhasználó töltötte le.

- 4) A kidolgozott algoritmust és munkafolyamatot alkalmaztam glikoprotein standardok és vérplazma glikoproteinek glikozilációs mintázatának meghatározására. Módszerünk az első, amely a glikozilációs mintázat meghatározásakor figyelembe veszi nemcsak az izotópeloszlást, de a fordított fázisú kromatográfiából nyerhető összes információt (retenciós idő és kromatográfiás csúcs szélessége). A módszer nem igényel izolált fehérjéket, keverékek vizsgálatára is alkalmas. A módszer alkalmazhatóságát vérplazmából izolált fehérjék, pl. a haptogloblin, AGP glikozilációs mintázatának meghatározásával bizonyítottam.
- 5) Doktori munkám során megvizsgáltam három különböző dúsítási eljárást (fenilbórsav affinitást, búzacásra agglutinin lektin affinitást és anioncserés módszereket) mind peptid, mind fehérje szinten. Az egyes frakciók glikopeptid tartalmának jellemzésére egy dúsítási indexet dolgoztam ki és alkalmaztam mindhárom dúsítási módszer frakcióira. Megállapítottam, hogy a peptidek dúsítása előnyösebb az alkalmazott protokollok esetén, valamint a három vizsgált módszer közül bórsav affinitás a legjobb glikopeptidek dúsítására.
- 6) A kidolgozott algoritmust alkalmaztam atherosclerosis biomarkerek meghatározására. Munkám során egészséges, valamint atherosclerosisban szenvedő betegek esetében meghatároztam az AGP helyspecifikus glikozilációs mintázatát. Pozitív kontrollként aneurizmában szenvedő betegek glikozilációs mintázatát használtam. Megállapítottam, hogy az AGP helyspecifikus glikozilációs mintázatának vizsgálatával egy igen jó, 89%-os predikciós mutatóval rendelkező potenciális biomarkert azonosítottam, amely alkalmas az atherosclerosisos megbetegedés előrejelzésére. Ez alátámasztja, hogy az általam kidolgozott minta-előkészítési és analitikai módszerek, valamint a GlycoMiner program alkalmazásának nemcsak elméleti, hanem komoly gyakorlati jelentősége is van. Segítségükkel kis mennyiségű fehérjemintából (pl. 10 pmol AGP esetén) is lehetőség van különböző glikopeptidek azonosítására, és a glikozilációs mintázat különbségeinek meghatározására. A glikopeptid mintázat sok esetben alkalmas arra, hogy egészséges és

beteg csoportok, valamint pozitív kontrollcsoportok között különbséget tudjunk tenni, azaz különböző patológiás állapotokat előre tudjunk jelezni.

7 Summary

Glycosylation is one of the most frequent post-translational modifications of proteins. Its analysis is very difficult, the lack of suitable analytical methodologies form a bottleneck for understanding glycochemistry and glycobiology. This prompted us to develop an integrated workflow for glycosylation analysis, which is capable to efficiently determine *N*-glycosylation patterns.

The developed workflow starts with proteolytic sample digestion resulting in glycopeptides, which are subsequently analyzed. Glycopeptides retain information on the glycosylation site (so both the protein and the respective glycosylation site may be determined). Analysis is based on a combined use of LC-MS and LC-MS/MS analysis, integrated with software-based data evaluation. A major advantage using glycopeptide analysis is that glycoprotein mixtures can also be analyzed – which simplifies the required sample preparation enormously. The key, novel aspect of the workflow presented is automatic spectra evaluation. To identify glycopeptide structures based on tandem mass spectra we have developed the algorithm and software named GlycoMiner. This works automatically (does not require manual control) and has a very low false positive and false negative rate (both 0.1%). To analyze glycosylation patterns automatically, we have developed another software (GlycoPattern). This works in conjunction with GlycoMiner. Beside quantitation, it is capable of identifying minor glycoforms, although with less reliability (5 %).

We have applied the developed workflow to analyze various glycosylation-related problems. It was shown that the protocol can identify glycosylation patterns of glycoproteins in complex mixtures (blood plasma); and can be used to monitor changes in glycosylation of therapeutic antibodies. We have compared various enrichment methodologies, and compared their efficiencies. Probably most important, that we have analyzed the glycosylation pattern of a number of AGP samples, isolated from the sera of healthy individuals and patients suffering from atherosclerosis and aneurism. Statistical analysis showed significant differences among the glycosylation patterns characteristic for each of the subject groups. Cross-validation showed 89% predictive power for identifying atherosclerosis. The glycosylation pattern of AGP is therefore a very promising biomarker candidate.

The developed workflow and the accompanying software is therefore a promising new tool in glycobiology. Preliminary results indicate that it can be used advantageously in several new directions.

8 Irodalomjegyzék

1. Wells, J. M., and McLuckey, S. A. (2005) Collision-induced dissociation (CID) of peptides and proteins. *Methods in enzymology* 402, 148-185.
2. Sleno, L., and Volmer, D. A. (2004) Ion activation methods for tandem mass spectrometry. *J. Mass Spectrom.* 39, 1091-1112.
3. Cooper, H. J., Håkansson, K., and Marshall, A. G. (2005) The role of electron capture dissociation in biomolecular analysis. *Mass Spectrometry Reviews* 24, 201-222.
4. Syka, J. E. P., Coon, J. J., Schroeder, M. J., Shabanowitz, J., and 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, 9528-9533.
5. Dunn, M. J. (1995) 2d Electrophoresis - From Protein Maps To Genomes - Proceedings Of The International Meeting - Siena, September 5-7, 1994. *Electrophoresis* 16, U3-U4.
6. Dunn, M. J. (1997) From protein maps to genomes - Proceedings of the Second Siena Two-Dimensional Electrophoresis Meeting held in Siena, September 16-18, 1996. *Electrophoresis* 18, U3-U4.
7. Anderson, N. L., and Anderson, N. G. (1998) Proteome and proteomics: New technologies, new concepts, and new words. *Electrophoresis* 19, 1853-1861.
8. Ong, S. E., and Mann, M. (2005) Mass spectrometry-based proteomics turns quantitative. *Nat. Chem. Biol.* 1, 252-262.
9. Steen, H., and Pandey, A. (2002) Proteomics goes quantitative: measuring protein abundance. *Trends in Biotechnology* 20, 361-364.
10. Reyzer, M. L., and Caprioli, R. M. (2005) MALDI mass spectrometry for direct tissue analysis: a new tool for biomarker discovery. *Journal of Proteome Research* 4, 1138-1142.
11. Zhou, W., Petricoin, E. F., and Longo, C. (2012) Mass spectrometry-based biomarker discovery. *Methods in molecular biology (Clifton, N.J.)* 823, 251-264.
12. Downing, G. J., and Conference, N.-F. (2000) *Biomarkers and surrogate endpoints : clinical research and applications ; proceedings of the NIH-FDA Conference held on 15-16 April 1999 in Bethesda, Maryland, USA*, Elsevier, Amsterdam [u.a.].
13. Polanski, M., and Anderson, N. L. (2007) A list of candidate cancer biomarkers for targeted proteomics. *Biomarker insights* 1, 1-48.
14. Anderson, N. L., and Anderson, N. G. (2002) The human plasma proteome - History, character, and diagnostic prospects. *Molecular & Cellular Proteomics* 1, 845-867.
15. Washburn, M. P., Wolters, D., and Yates, J. R. (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature Biotechnology* 19, 242-247.
16. Marchetti-Deschmann, M., and Allmaier, G. (2011) Mass spectrometry - One of the pillars of proteomics. *Journal of Proteomics* 74, 915-919.
17. Gygi, S. P., and Aebersold, R. (2000) Mass spectrometry and proteomics. *Current Opinion in Chemical Biology* 4, 489-494.

18. Aebersold, R., and Goodlett, D. R. (2001) Mass spectrometry in proteomics. *Chemical Reviews* 101, 269-295.
19. Derrick, P. J., and Patterson, S. D. (2001) Mass spectrometry of proteomics. *European Journal of Mass Spectrometry* 7, 79-80.
20. Stults, J. T., and Arnott, D. (2005) Proteomics. In: Burlingame, A. L., ed. *Biological Mass Spectrometry*, pp. 245-289, Elsevier Academic Press Inc, San Diego.
21. Roepstorff, P., and Fohlman, J. (1984) Proposal For A Common Nomenclature For Sequence Ions In Mass-Spectra Of Peptides. *Biomedical Mass Spectrometry* 11, 601-601.
22. Johnson, R. S., Martin, S. A., Biemann, K., Stults, J. T., and 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, 2621-2625.
23. Wolters, D. A., Washburn, M. P., and Yates, J. R. (2001) An automated multidimensional protein identification technology for shotgun proteomics. *Analytical Chemistry* 73, 5683-5690.
24. Perkins, D. N., Pappin, D. J. C., Creasy, D. M., and Cottrell, J. S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20, 3551-3567.
25. 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. *Analytical Chemistry* 71, 2871-2882.
26. Eng, J. K., McCormack, A. L., and Yates, J. R. (1994) An Approach To Correlate Tandem Mass-Spectral Data Of Peptides With Amino-Acid-Sequences In A Protein Database. *Journal of the American Society for Mass Spectrometry* 5, 976-989.
27. Fenyo, D., and Beavis, R. C. (2003) A method for assessing the statistical significance of mass spectrometry-based protein identifications using general scoring schemes. *Analytical Chemistry* 75, 768-774.
28. Tannu, N., and Hemby, S. (2007) De novo protein sequence analysis of *Macaca mulatta*. *BMC Genomics* 8, 270.
29. Dancik, V., Addona, T. A., Clauser, K. R., Vath, J. E., and Pevzner, P. A. (1999) De novo peptide sequencing via tandem mass spectrometry. *Journal of computational biology : a journal of computational molecular cell biology* 6, 327-342.
30. Ross, P. L., Huang, Y. L. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A., and Pappin, D. J. (2004) Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Molecular & Cellular Proteomics* 3, 1154-1169.
31. Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., and Aebersold, R. (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotechnology* 17, 994-999.
32. Berger, S. J., Lee, S. W., Anderson, G. A., Pasa-Tolic, L., Tolic, N., Shen, Y. F., Zhao, R., and Smith, R. D. (2002) High-throughput global peptide proteomic analysis by combining stable isotope amino acid labeling and data-dependent multiplexed-MS/MS. *Analytical Chemistry* 74, 4994-5000.

33. Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., and Mann, M. (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Molecular & Cellular Proteomics* 1, 376-386.
34. Soufi, B., Kumar, C., Gnad, F., Mann, M., Mijakovic, I., and Macek, B. (2010) Stable Isotope Labeling by Amino Acids in Cell Culture (SILAC) Applied to Quantitative Proteomics of *Bacillus subtilis*. *Journal of Proteome Research* 9, 3638-3646.
35. Geiger, T., Wisniewski, J. R., Cox, J., Zanivan, S., Kruger, M., Ishihama, Y., and Mann, M. (2011) Use of stable isotope labeling by amino acids in cell culture as a spike-in standard in quantitative proteomics. *Nature Protocols* 6, 147-157.
36. Zhu, W., Smith, J. W., and Huang, C. M. (2010) Mass spectrometry-based label-free quantitative proteomics. *Journal of biomedicine & biotechnology* 2010, 840518.
37. Allet, N., Barrillat, N., Baussant, T., Boiteau, C., Botti, P., Bougueleret, L., Budin, N., Canet, D., Carraud, S., Chiappe, D., Christmann, N., Colinge, J., Cusin, I., Dafflon, N., Depresle, B., Fasso, I., Frauchiger, P., Gaertner, H., Gleizes, A., Gonzalez-Couto, E., Jeandenans, C., Karmime, A., Kowall, T., Lagache, S., Mahe, E., Masselot, A., Mattou, H., Moniatte, M., Niknejad, A., Paolini, M., Perret, F., Pinaud, N., Ranno, F., Raimondi, S., Reffas, S., Regamey, P. O., Rey, P. A., Rodriguez-Tome, P., Rose, K., Rossellat, G., Saudrais, C., Schmidt, C., Villain, M., and Zwahlen, C. (2004) In vitro and in silico processes to identify differentially expressed proteins. *Proteomics* 4, 2333-2351.
38. Liu, H. B., Sadygov, R. G., and Yates, J. R. (2004) A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Analytical Chemistry* 76, 4193-4201.
39. Colinge, J., Chiappe, D., Lagache, S., Moniatte, M., and Bougueleret, L. (2005) Differential proteomics via probabilistic peptide identification scores. *Analytical Chemistry* 77, 596-606.
40. Chelius, D., and Bondarenko, P. V. (2002) Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry. *Journal of Proteome Research* 1, 317-323.
41. Muller, C., Schafer, P., Stortzel, M., Vogt, S., and Weinmann, W. (2002) Ion suppression effects in liquid chromatography-electrospray-ionisation transport-region collision induced dissociation mass spectrometry with different serum extraction methods for systematic toxicological analysis with mass spectra libraries. *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences* 773, 47-52.
42. Wang, W., Zhou, H., Lin, H., Roy, S., Shaler, T. A., Hill, L. R., Norton, S., Kumar, P., Anderle, M., and Becker, C. H. (2003) Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards. *Analytical Chemistry* 75, 4818-4826.
43. Radulovic, D., Jelveh, S., Ryu, S., Hamilton, T. G., Foss, E., Mao, Y., and Emili, A. (2004) Informatics platform for global proteomic profiling and biomarker discovery using liquid chromatography-tandem mass spectrometry. *Molecular & cellular proteomics : MCP* 3, 984-997.
44. Silva, J. C., Denny, R., Dorschel, C. A., Gorenstein, M., Kass, I. J., Li, G. Z., McKenna, T., Nold, M. J., Richardson, K., Young, P., and Geromanos, S. (2005) Quantitative

- proteomic analysis by accurate mass retention time pairs. *Analytical Chemistry* 77, 2187-2200.
45. Silva, J. C., Gorenstein, M. V., Li, G. Z., Vissers, J. P., and Geromanos, S. J. (2006) Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. *Molecular & cellular proteomics : MCP* 5, 144-156.
 46. Silva, J. C., Denny, R., Dorschel, C., Gorenstein, M. V., Li, G. Z., Richardson, K., Wall, D., and Geromanos, S. J. (2006) Simultaneous qualitative and quantitative analysis of the *Escherichia coli* proteome: a sweet tale. *Molecular & cellular proteomics : MCP* 5, 589-607.
 47. Cheng, F. Y., Blackburn, K., Lin, Y. M., Goshe, M. B., and Williamson, J. D. (2009) Absolute protein quantification by LC/MS(E) for global analysis of salicylic acid-induced plant protein secretion responses. *Journal of Proteome Research* 8, 82-93.
 48. Gesellchen, F., Bertinetti, O., and Herberg, F. W. (2006) Analysis of posttranslational modifications exemplified using protein kinase A. *BBA-Proteins Proteomics* 1764, 1788-1800.
 49. Larsen, M. R., Trelle, M. B., Thingholm, T. E., and Jensen, O. N. (2006) Analysis of posttranslational modifications of proteins by tandem mass spectrometry. *Biotechniques* 40, 790-798.
 50. Spiro, R. G. (2002) Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology* 12, 43R-56R.
 51. Hofsteenge, J., Blommers, M., Hess, D., Furmanek, A., and Miroshnichenko, O. (1999) The four terminal components of the complement system are C-mannosylated on multiple tryptophan residues. *The Journal of biological chemistry* 274, 32786-32794.
 52. Corfield, A. P., and Warren, B. F. (1996) Mucus glycoproteins and their role in colorectal disease. *The Journal of pathology* 180, 8-17.
 53. Kowarik, M., Young, N. M., Numao, S., Schulz, B. L., Hug, I., Callewaert, N., Mills, D. C., Watson, D. C., Hernandez, M., Kelly, J. F., Wacker, M., and Aebi, M. (2006) Definition of the bacterial *N*-glycosylation site consensus sequence. *The EMBO journal* 25, 1957-1966.
 54. Nita-Lazar, M., Wacker, M., Schegg, B., Amber, S., and Aebi, M. (2005) The N-X-S/T consensus sequence is required but not sufficient for bacterial *N*-linked protein glycosylation. *Glycobiology* 15, 361-367.
 55. Helenius, A., and Aebi, M. (2004) Roles of *N*-linked glycans in the endoplasmic reticulum. *Annual review of biochemistry* 73, 1019-1049.
 56. Sears, P., and Wong, C. H. (1998) Enzyme action in glycoprotein synthesis. *Cellular and molecular life sciences : CMLS* 54, 223-252.
 57. Caramelo, J. J., and Parodi, A. J. (2007) How sugars convey information on protein conformation in the endoplasmic reticulum. *Semin. Cell Dev. Biol.* 18, 732-742.
 58. Hammond, C., Braakman, I., and Helenius, A. (1994) Role of *N*-Linked Oligosaccharide Recognition, Glucose Trimming, and Calnexin in Glycoprotein Folding and Quality Control
- 10.1073/pnas.91.3.913. *Proceedings of the National Academy of Sciences* 91, 913-917.

59. Saxon, E., and Bertozzi, C. R. (2001) Chemical and biological strategies for engineering cell surface glycosylation. *Annual review of cell and developmental biology* 17, 1-23.
60. Rudd, P. M., Elliott, T., Cresswell, P., Wilson, I. A., and Dwek, R. A. (2001) Glycosylation and the immune system. *Science (New York, N.Y.)* 291, 2370-2376.
61. Gavrillov, B. K., Rogers, K., Fernandez-Sainz, I. J., Holinka, L. G., Borca, M. V., and Risatti, G. R. (2011) Effects of glycosylation on antigenicity and immunogenicity of classical swine fever virus envelope proteins. *Virology* 420, 135-145.
62. Russell, D., Oldham, N. J., and Davis, B. G. (2009) Site-selective chemical protein glycosylation protects from autolysis and proteolytic degradation. *Carbohydr. Res.* 344, 1508-1514.
63. Brooks, S. A., Dwek, M. V., and U., S. (2002) *Functional & Molecular Glycobiology*, Bios scientific publishers LTD.
64. Davis, M. T., Auger, P., Spahr, C., and Patterson, S. D. (2007) Cancer biomarker discovery via low molecular weight serum proteome profiling - Where is the tumor? *Proteom. Clin. Appl.* 1, 1545-1558.
65. Holland, M., Yagi, H., Takahashi, N., Kato, K., Savage, C. O. S., Goodall, D. M., and Jefferis, R. (2006) Differential glycosylation of polyclonal IgG, IgG-Fc and IgG-Fab isolated from the sera of patients with ANCA-associated systemic vasculitis. *Biochim. Biophys. Acta-Gen. Subj.* 1760, 669-677.
66. Leiserowitz, G. S., Lebrilla, C., Miyamoto, S., An, H. J., Duong, H., Kirmiz, C., Li, B., Liu, H., and Lam, K. S. (2008) Glycomics analysis of serum: a potential new biomarker for ovarian cancer? *Int. J. Gynecol. Cancer* 18, 470-475.
67. Budnik, B. A., Lee, R. S., and Steen, J. A. J. (2006) Global methods for protein glycosylation analysis by mass spectrometry. *BBA-Proteins Proteomics* 1764, 1870-1880.
68. Cao, J., Nie, A. Y., Chen, Y. H., Wang, S., Lu, H. J., and Yang, P. Y. (2009) Separation and Enrichment of Glycoproteins/Glycopeptides. *Prog. Chem.* 21, 1888-1894.
69. Williams, T. I., Saggese, D. A., and Muddiman, D. C. (2008) Studying O-linked protein glycosylations in human plasma. *Journal of Proteome Research* 7, 2562-2568.
70. Wang, L., Li, F., Sun, W., Wu, S., Wang, X., Zhang, L., Zheng, D., Wang, J., and Gao, Y. (2006) Concanavalin A-captured glycoproteins in healthy human urine. *Molecular & cellular proteomics : MCP* 5, 560-562.
71. Wang, Y. H., Wu, S. L., and Hancock, W. S. (2006) Approaches to the study of N-linked glycoproteins in human plasma using lectin affinity chromatography and nano-HPLC coupled to electrospray linear ion trap-Fourier transform mass spectrometry. *Glycobiology* 16, 514-523.
72. Calvano, C. D., Zamboni, C. G., and Jensen, O. N. (2008) Assessment of lectin and HILIC based enrichment protocols for characterization of serum glycoproteins by mass spectrometry. *Journal of proteomics* 71, 304-317.
73. Zhang, L., Lu, H., and Yang, P. (2010) Specific enrichment methods for glycoproteome research. *Anal. Bioanal. Chem.* 396, 199-203.
74. Sumer-Bayraktar, Z., Kolarich, D., Campbell, M. P., Ali, S., Packer, N. H., and Thaysen-Andersen, M. (2011) N-glycans modulate the function of human corticosteroid-binding globulin. *Molecular & Cellular Proteomics.*

75. Yan, J., Li, X., Yu, L., Jin, Y., Zhang, X., Xue, X., Ke, Y., and Liang, X. (2010) Selective enrichment of glycopeptides/phosphopeptides using porous titania microspheres. *Chemical communications (Cambridge, England)* 46, 5488-5490.
76. Wan, H., Yan, J., Yu, L., Sheng, Q., Zhang, X., Xue, X., Li, X., and Liang, X. (2011) Zirconia layer coated mesoporous silica microspheres as HILIC SPE materials for selective glycopeptide enrichment. *Analyst* 136, 4422-4430.
77. Xu, Y., Wu, Z., Zhang, L., Lu, H., Yang, P., Webley, P. A., and Zhao, D. (2008) Highly Specific Enrichment of Glycopeptides Using Boronic Acid-Functionalized Mesoporous Silica. *Analytical Chemistry* 81, 503-508.
78. Lee, J. H., Kim, Y., Ha, M. Y., Lee, E. K., and Choo, J. (2005) Immobilization of Aminophenylboronic Acid on Magnetic Beads for the Direct Determination of Glycoproteins by Matrix Assisted Laser Desorption Ionization Mass Spectrometry. *Journal of the American Society for Mass Spectrometry* 16, 1456-1460.
79. Kaji, H., Saito, H., Yamauchi, Y., Shinkawa, T., Taoka, M., Hirabayashi, J., Kasai, K.-i., Takahashi, N., and Isobe, T. (2003) Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify *N*-linked glycoproteins. *Nat Biotech* 21, 667-672.
80. Vercoutter-Edouart, A.-S., Slomianny, M.-C., Dekeyser-Beseme, O., Haeuw, J.-F., and Michalski, J.-C. (2008) Glycoproteomics and glycomics investigation of membrane *N*-glycosylproteins from human colon carcinoma cells. *Proteomics* 8, 3236-3256.
81. Madera, M., Mann, B., Mechref, Y., and Novotny, M. V. (2008) Efficacy of glycoprotein enrichment by microscale lectin affinity chromatography. *Journal of Separation Science* 31, 2722-2732.
82. Qi, D., Zhang, H., Tang, J., Deng, C., and Zhang, X. (2010) Facile Synthesis of Mercaptophenylboronic Acid-Functionalized Core-Shell Structure Fe₃O₄@C@Au Magnetic Microspheres for Selective Enrichment of Glycopeptides and Glycoproteins. *The Journal of Physical Chemistry C* 114, 9221-9226.
83. Sparbier, K., Asperger, A., Resemann, A., Kessler, I., Koch, S., Wenzel, T., Stein, G., Vorverg, L., Suckau, D., and Kostrzewa, M. (2007) Analysis of glycoproteins in human serum by means of glycospecific magnetic bead separation and LC-MALDI-TOF/TOF analysis with automated glycopeptide detection. *Journal of biomolecular techniques : JBT* 18, 252-258.
84. Krishnamoorthy, L., and Mahal, L. K. (2011) Lectin Microarrays: Simple Tools for the Analysis of Complex Glycans Functional and Structural Proteomics of Glycoproteins. In: Owens, R., and Nettleship, J., eds., pp. 91-102, Springer Netherlands.
85. Smith, E. R., Ford, M. L., Tomlinson, L. A., Rocks, B. F., Rajkumar, C., and Holt, S. G. (2010) Poor agreement between commercial ELISAs for plasma fetuin-A: An effect of protein glycosylation? *Clinica Chimica Acta* 411, 1367-1370.
86. Chua, C.-C., Lim, M.-L., and Wong, B.-S. (2010) Altered apolipoprotein E glycosylation is associated with A β (42) accumulation in an animal model of Niemann-Pick Type C disease. *Journal of Neurochemistry* 112, 1619-1626.
87. Stork, R., Zettlitz, K. A., Muller, D., Rether, M., Hanisch, F. G., and Kontermann, R. E. (2008) *N*-glycosylation as novel strategy to improve pharmacokinetic properties of bispecific single-chain diabodies. *The Journal of biological chemistry* 283, 7804-7812.
88. Zhou, H. J., Liu, Y. K., Chui, J. F., Guo, K., Shun, Q. L., Lu, W. J., Jin, H., Wei, L. M., and Yang, P. Y. (2007) Investigation on glycosylation patterns of proteins from human

- liver cancer cell lines based on the multiplexed proteomics technology. *Archives of Biochemistry and Biophysics* 459, 70-78.
89. Guzman, N. A., Blanc, T., and Phillips, T. M. (2008) Immunoaffinity capillary electrophoresis as a powerful strategy for the quantification of low-abundance biomarkers, drugs, and metabolites in biological matrices. *Electrophoresis* 29, 3259-3278.
 90. Guzman, N. A., and Phillips, T. M. (2011) Immunoaffinity capillary electrophoresis: A new versatile tool for determining protein biomarkers in inflammatory processes. *Electrophoresis* 32, 1565-1578.
 91. Hanisch, F.-G. (2011) Top-Down Sequencing of O-Glycoproteins by In-Source Decay Matrix-Assisted Laser Desorption Ionization Mass Spectrometry for Glycosylation Site Analysis. *Analytical Chemistry* 83, 4829-4837.
 92. Resemann, A., Wunderlich, D., Rothbauer, U., Warscheid, B., Leonhardt, H., Fuchser, J., Kuhlmann, K., and Suckau, D. (2010) Top-Down de Novo Protein Sequencing of a 13.6 kDa Camelid Single Heavy Chain Antibody by Matrix-Assisted Laser Desorption Ionization-Time-of-Flight/Time-of-Flight Mass Spectrometry. *Analytical Chemistry* 82, 3283-3292.
 93. Mazur, M. T., Cardasis, H. L., Spellman, D. S., Liaw, A., Yates, N. A., and Hendrickson, R. C. (2010) Quantitative analysis of intact apolipoproteins in human HDL by top-down differential mass spectrometry. *Proceedings of the National Academy of Sciences* 107, 7728-7733.
 94. Pflieger, D., Przybylski, C., Gonnet, F., Le Caer, J.-P., Lunardi, T., Arlaud, G. J., and Daniel, R. (2010) Analysis of Human C1q by Combined Bottom-up and Top-down Mass Spectrometry. *Molecular & Cellular Proteomics* 9, 593-610.
 95. Kaji, H., Yamauchi, Y., Takahashi, N., and Isobe, T. (2007) Mass spectrometric identification of *N*-linked glycopeptides using lectin-mediated affinity capture and glycosylation site-specific stable isotope tagging. *Nat. Protocols* 1, 3019-3027.
 96. Kotwal, G. (2010) Influence of glycosylation and oligomerization of vaccinia virus complement control protein on level and pattern of functional activity and immunogenicity. *Protein & Cell* 1, 1084-1092.
 97. Kang, P., Mechref, Y., Klouckova, I., and Novotny, M. V. (2005) Solid-phase permethylation of glycans for mass spectrometric analysis. *Rapid Commun. Mass Spectrom.* 19, 3421-3428.
 98. Bleckmann, C., Geyer, H., and Geyer, R. (2011) Nanoelectrospray-MS(n) of native and permethylated glycans. *Methods in molecular biology (Clifton, N.J.)* 790, 71-85.
 99. Prien, J. M., Prater, B. D., Qin, Q., and Cockrill, S. L. (2010) Mass Spectrometric-Based Stable Isotopic 2-Aminobenzoic Acid Glycan Mapping for Rapid Glycan Screening of Biotherapeutics. *Analytical Chemistry* 82, 1498-1508.
 100. Ahn, J., Bones, J., Yu, Y. Q., Rudd, P. M., and Gilar, M. (2010) Separation of 2-aminobenzamide labeled glycans using hydrophilic interaction chromatography columns packed with 1.7 μ m sorbent. *Journal of Chromatography B* 878, 403-408.
 101. Wuhrer, M., Catalina, M. I., Deelder, A. M., and Hokke, C. H. (2007) Glycoproteomics based on tandem mass spectrometry of glycopeptides. *Journal of Chromatography B* 849, 115-128.

102. Krenyacz, J., Drahos, L., and Vekey, K. (2009) Collision energy and cone voltage optimisation for glycopeptide analysis. *European Journal of Mass Spectrometry* 15, 361-365.
103. Froehlich, J. W., Barboza, M., Chu, C., Lerno, L. A., Clowers, B. H., Zivkovic, A. M., German, J. B., and Lebrilla, C. B. (2011) Nano-LC-MS/MS of Glycopeptides Produced by Nonspecific Proteolysis Enables Rapid and Extensive Site-Specific Glycosylation Determination. *Analytical Chemistry* 83, 5541-5547.
104. Domon, B., and Costello, C. E. (1988) A Systematic Nomenclature for Carbohydrate Fragmentations in Fab-MS Ms Spectra of Glycoconjugates. *Glycoconjugate J.* 5, 397-409.
105. Harvey, D. J. (2005) Proteomic analysis of glycosylation: structural determination of N- and O-linked glycans by mass spectrometry. *Expert Rev. Proteomics* 2, 87-101.
106. Harvey, D. J., Bateman, R. H., and Green, M. R. (1997) High-energy collision-induced fragmentation of complex oligosaccharides ionized by matrix-assisted laser desorption/ionization mass spectrometry. *J. Mass Spectrom.* 32, 167-187.
107. Halim, A., Nilsson, J., Rüetschi, U., Hesse, C., and Larson, G. (2011) Human urinary glycoproteomics; attachment site specific analysis of N- and O-linked glycosylations by CID and ECD. *Molecular & Cellular Proteomics*.
108. Chen, Y. H., Zhang, Y., and Yang, P. Y. (2010) Glycosite Analysis in Glycoproteomics by Mass Spectrometry. *Curr. Proteomics* 7, 158-167.
109. Catalina, M. I., Koeleman, C. A. M., Deelder, A. M., and Wührer, M. (2007) Electron transfer dissociation of N-glycopeptides: loss of the entire N-glycosylated asparagine side chain. *Rapid Commun. Mass Spectrom.* 21, 1053-1061.
110. Campbell, M. P., Royle, L., Radcliffe, C. M., Dwek, R. A., and Rudd, P. M. (2008) GlycoBase and autoGU: tools for HPLC-based glycan analysis. *Bioinformatics (Oxford, England)* 24, 1214-1216.
111. An, H. J., Tillinghast, J. S., Woodruff, D. L., Rocke, D. M., and Lebrilla, C. B. (2006) A new computer program (GlycoX) to determine simultaneously the glycosylation sites and oligosaccharide heterogeneity of glycoproteins. *Journal of Proteome Research* 5, 2800-2808.
112. Ren, J. M., Rejtar, T., Li, L. Y., and Karger, B. L. (2007) N-glycan structure annotation of glycopeptides using a linearized glycan structure database (GlyDB). *Journal of Proteome Research* 6, 3162-3173.
113. Joenvaara, S., Ritamo, I., Peltoniemi, H., and Renkonen, R. (2008) N-Glycoproteomics - An automated workflow approach. *Glycobiology* 18, 339-349.
114. von der Lieth, C. W., Lutteke, T., and Frank, M. (2006) The role of informatics in glycobiology research with special emphasis on automatic interpretation of MS spectra. *Biochim. Biophys. Acta-Gen. Subj.* 1760, 568-577.
115. Gaucher, S. P., Morrow, J., and Leary, J. A. (2000) STAT: A saccharide topology analysis tool used in combination with tandem mass spectrometry. *Analytical Chemistry* 72, 2331-2336.
116. Clerens, S., Van den Ende, W., Verhaert, P., Geenen, L., and Arckens, L. (2004) Sweet Substitute: A software tool for in silico fragmentation of peptide-linked N-glycans. *Proteomics* 4, 629-632.

117. Goldberg, D., Bern, M., Parry, S., Sutton-Smith, M., Panico, M., Morris, H. R., and Dell, A. (2007) Automated *N*-glycopeptide identification using a combination of single- and tandem-MS. *Journal of Proteome Research* 6, 3995-4005.
118. Goldberg, D., Sutton-Smith, M., Paulson, J., and Dell, A. (2005) Automatic annotation of matrix-assisted laser desorption/ionization *N*-glycan spectra. *Proteomics* 5, 865-875.
119. Cooper, C. A., Gasteiger, E., and Packer, N. H. (2001) GlycoMod - A software tool for determining glycosylation compositions from mass spectrometric data. *Proteomics* 1, 340-349.
120. Swamy, S. (2004) The Automation of Glycopeptide Discovery in High Throughput MS/MS Data. *Master of Mathematics in Computer Science Thesis, University of Waterloo, Canada.*
121. Fukunaga, K. (1990) *Introduction to Statistical Pattern Recognition*, Second Ed., Academic Press, San Diego, California.
122. Venables, W. N., and Ripley, B. D. (2002) *Modern Applied Statistics with S*, Fourth Ed., Springer Verlag.
123. Taylor, D. L. (1991) *Evaluating the Sample Specificity of Discriminant Analysis Results Using the Jackknife Statistic [microform]* / Dianne L. Taylor, Distributed by ERIC Clearinghouse, [Washington, D.C.] .:
124. Shiyan, S. D., and Bovin, N. V. (1997) Carbohydrate composition and immunomodulatory activity of different glycoforms of alpha1-acid glycoprotein. *Glycoconjugate J.* 14, 631-638.
125. Fitos, I., Visy, J., Zsila, F., Mady, G., and Simonyi, M. (2006) Selective binding of imatinib to the genetic variants of human alpha1-acid glycoprotein. *Biochimica et biophysica acta* 1760, 1704-1712.
126. Schonfeld, D. L., Ravelli, R. B., Mueller, U., and Skerra, A. (2008) The 1.8-Å crystal structure of alpha1-acid glycoprotein (Orosomucoid) solved by UV RIP reveals the broad drug-binding activity of this human plasma lipocalin. *Journal of molecular biology* 384, 393-405.
127. Dente, L., Pizza, M. G., Metspalu, A., and Cortese, R. (1987) Structure and expression of the genes coding for human alpha 1-acid glycoprotein. *The EMBO journal* 6, 2289-2296.
128. Imre, T., Schlosser, G., Pocsfalvi, G., Siciliano, R., Molnar-Szollosi, E., Kremmer, T., Malorni, A., and Vékey, K. (2005) Glycosylation site analysis of human alpha-1-acid glycoprotein (AGP) by capillary liquid chromatography-electrospray mass spectrometry. *J. Mass Spectrom.* 40, 1472-1483.
129. Saldova, R., Wormald, M. R., Dwek, R. A., and Rudd, P. M. (2008) Glycosylation changes on serum glycoproteins in ovarian cancer may contribute to disease pathogenesis. *Disease Markers* 25, 219-232.
130. Imre, T., Kremmer, T., Héberger, K., Molnár-Szöllosi, É., Ludányi, K., Pócsfalvi, G., Malorni, A., Drahos, L., and Vékey, K. (2008) Mass spectrometric and linear discriminant analysis of *N*-glycans of human serum alpha-1-acid glycoprotein in cancer patients and healthy individuals. *Journal of Proteomics* 71, 186-197.
131. Saroha, A., Biswas, S., Chatterjee, B. P., and Das, H. R. (2011) Altered glycosylation and expression of plasma alpha-1-acid glycoprotein and haptoglobin in rheumatoid arthritis. *Journal of Chromatography B* 879, 1839-1843.

132. Puerta, A., Diez-Masa, J. C., Martin-Alvarez, P. J., Martin-Ventura, J. L., Barbas, C., Tunon, J., Egido, J., and de Frutos, M. (2011) Study of the capillary electrophoresis profile of intact [small alpha]-1-acid glycoprotein isoforms as a biomarker of atherothrombosis. *Analyst* 136, 816-822.
133. Fuster, V., and Moreno, P. R. (2005) Atherothrombosis as a systemic, often silent, disease. *Nat Clin Pract Cardiovasc Med* 2, 431-431.
134. Drouet, L. (2002) Atherothrombosis as a systemic disease. *Cerebrovascular diseases (Basel, Switzerland)* 13 Suppl 1, 1-6.
135. Diehm, N., and Baumgartner, I. (2009) Determinants of Aneurysmal Aortic Disease. *Circulation* 119, 2134-2135.
136. Keller, A., Nesvizhskii, A. I., Kolker, E., and Aebersold, R. (2002) Empirical Statistical Model To Estimate the Accuracy of Peptide Identifications Made by MS/MS and Database Search. *Analytical Chemistry* 74, 5383-5392.
137. Nesvizhskii, A. I., Keller, A., Kolker, E., and Aebersold, R. (2003) A statistical model for identifying proteins by tandem mass spectrometry. *Analytical Chemistry* 75, 4646-4658.
138. www.chemres.hu/ms/glycominer (2008) GlycoMiner.
139. Blanken, W. M., Bergh, M. L. E., Koppen, P. L., and van den Eijnden, D. H. (1985) High-pressure liquid chromatography of neutral oligosaccharides: Effects of structural parameters. *Anal. Biochem.* 145, 322-330.
140. Chen, X., and Flynn, G. C. (2007) Analysis of N-glycans from recombinant immunoglobulin G by on-line reversed-phase high-performance liquid chromatography/mass spectrometry. *Anal. Biochem.* 370, 147-161.
141. Krokhin, O. V., Craig, R., Spicer, V., Ens, W., Standing, K. G., Beavis, R. C., and Wilkins, J. A. (2004) An improved model for prediction of retention times of tryptic peptides in ion pair reversed-phase HPLC: its application to protein peptide mapping by off-line HPLC-MALDI MS. *Molecular & cellular proteomics : MCP* 3, 908-919.
142. Imami, K., Ishihama, Y., and Terabe, S. (2008) On-line selective enrichment and ion-pair reaction for structural determination of sulfated glycopeptides by capillary electrophoresis-mass spectrometry. *J. Chromatogr. A* 1194, 237-242.
143. Qiu, J., Zhang, Y., Lu, H. J., and Yang, P. Y. (2011) Small Size 4-Mercaptophenylboronic Acid Functionalized Gold Nanoparticles for Selective Enrichment and Mass Spectrometric Identification of Glycoproteins in Rat Liver. *Acta Chimica Sinica* 69, 2123-2129.
144. Tang, J., Liu, Y., Yin, P., Yao, G. P., Yan, G. Q., Deng, C. H., and Zhang, X. M. (2010) Concanavalin A-immobilized magnetic nanoparticles for selective enrichment of glycoproteins and application to glycoproteomics in hepatocellular carcinoma cell line. *Proteomics* 10, 2000-2014.
145. Yu, L., Li, X. L., Guo, Z. M., Zhang, X. L., and Liang, X. M. (2009) Hydrophilic Interaction Chromatography Based Enrichment of Glycopeptides by Using Click Maltose: A Matrix with High Selectivity and Glycosylation Heterogeneity Coverage. *Chemistry-a European Journal* 15, 12618-12626.
146. Zhang, L. J., Xu, Y. W., Yao, H. L., Xie, L. Q., Yao, J., Lu, H. J., and Yang, P. Y. (2009) Boronic Acid Functionalized Core-Satellite Composite Nanoparticles for Advanced

Enrichment of Glycopeptides and Glycoproteins. *Chemistry-a European Journal* 15, 10158-10166.

Publikációk

A Ph.D. értekezésben felhasznált, referált tudományos folyóiratokban megjelent közlemények

1. Ozohanics, O., Krenyacz, J., Ludanyi, K., Pollreis, F., Vekey, K., and Drahos, L. (2008) GlycoMiner: a new software tool to elucidate glycopeptide composition. *Rapid Communications in Mass Spectrometry* 22, 3245-3254. IF: 2.846
2. Ozohanics, O., Turiak, L., Drahos, L., and Vekey, K. (2012) Comparison of glycopeptide/glycoprotein enrichment techniques. *Rapid communications in mass spectrometry* 26, 215-217. IF: 2.846
3. Ozohanics, O., Turiak, L., Puerta, A., Drahos, L., and Vékey, K. (2012) HPLC-MS methodology for analyzing site-specific glycosylation patterns. *Journal of Chromatography A*. IF: 4.194 (közlésre elküldve)

Referált tudományos folyóiratokban megjelent egyéb közlemények

1. Turiak, L., Ozohanics, O., Marino, F., Drahos, L., and Vekey, K. (2011) Digestion protocol for small protein amounts for nano-HPLC-MS(MS) analysis. *Journal of Proteomics* 74, 942-947. IF: 5.074
2. Turiak, L., Misjak, P., Szabo, T. G., Aradi, B., Palocz, K., Ozohanics, O., Drahos, L., Kittel, A., Falus, A., Buzas, E. I., and Vekey, K. (2011) Proteomic characterization of thymocyte-derived microvesicles and apoptotic bodies in BALB/c mice. *Journal of Proteomics* 74, 2025-2033. IF: 5.074
3. Ambrus, A., Torocsik, B., Tretter, L., Ozohanics, O., and Adam-Vizi, V. (2011) Stimulation of reactive oxygen species generation by disease-causing mutations of lipamide dehydrogenase. *Human Molecular Genetics* 20, 2984-2995. IF: 8.058

4. Zsoldos-Mady, V., Ozohanics, O., Csampai, A., Kudar, V., Frigyes, D., and Sohar, P. (2009) Ferrocenyl pyrazolines: Preparation, structure, redox properties and DFT study on regioselective ring-closure. *Journal of Organometallic Chemistry* 694, 4185-4195. IF: 2.205
5. Budai, L., Ozohanics, O., Ludanyi, K., Drahos, L., Kremmer, T., Krenyacz, J., and Vekey, K. (2009) Investigation of genetic variants of alpha-1 acid glycoprotein by ultra-performance liquid chromatography-mass spectrometry. *Analytical and Bioanalytical Chemistry* 393, 991-998. IF: 3.841
6. Budai, L., Pollreisz, F., Ozohanics, O., Ludanyi, K., Drahos, L., and Vekey, K. (2008) Analysis of complex oligosaccharides using graphitized carbon liquid chromatography/mass spectrometry. *European Journal of Mass Spectrometry* 14, 419-422. IF: 1.34

Tudományos konferencián tartott előadások

1. Ozohanics, O., Krenyacz, J., Pollreisz, F., Vékey, K., Drahos, L.: *Automatic identification of saccharides and glycopeptides in MS/MS*, 25th Informal Meeting on Mass Spectrometry; Nyíregyháza-Sóstó, 2007. május 6-10.
2. Ozohanics, O., Krenyacz, J., Budai, L., Ludányi, K., Kremmer, T., Vékey, K., Drahos, L.: *Analysis of genetic variants of α -1 acid glycoprotein in cancer*, Second Central and Eastern European Proteomic Conference; Jéna, 2008. október 12-15. (plenáris előadás)
3. Ozohanics, O., Krenyacz, J., Budai, L., Ludányi, K., Kremmer, T., Vékey, K., Drahos, L.: *Az alfa-1 savas glikoprotein genetikai variánsainak vizsgálata*, Elvlasztástudományi Vándorgyűlés 2008; Sárovar, 2008. november 5-7.
4. Ozohanics, O., Drahos, L., Vékey, K.: *Glycosylation, mass spectrometry and informatics*, 3rd Central and Eastern European Proteomics Conference; Budapest, 2009. október 6-9.

5. Ozohanics, O., Memboeuf, A., Indelicato, S., Drahos, L., Vékey, K.: *Size dependence of fragmentation – predictable or mysterious?*, 28th Informal Meeting on Mass Spectrometry; Kőszeg, 2010. május 2-6.
6. Ozohanics, O., Turiák, L., Lengyel, Á., Vékey, K., Drahos, L.: *Glycosylation Patterns and Glycopeptide Biomarkers*, 5th Central and Eastern European Proteomic Conference; Prága, 2011. szeptember 19-22.
7. Ozohanics, O., Turiák, L., Lengyel, Á., Vékey, K., Drahos, L.: *A Challenge for nanoUPLC-MS/MS: Determination of Glycosylation Patterns*, HPLC 2011; Budapest, 2011. június 19-23.