

Humán transzmembrán proteom topológiájának meghatározása kombinált kísérleti és elméleti módszerekkel

Doktori értekezés

Langó Tamás

Szerkezeti Biokémia Program, Biológia Doktori Iskola
Eötvös Loránd Tudományegyetem, Természettudományi Kar

A doktori iskola vezetője: Prof. Erdei Anna, DSc, egyetemi tanár

A program vezetője: Prof. Kovács Mihály, DSc, egyetemi tanár

Témavezető:

Dr. Tusnády Gábor, PhD, kutatócsoport vezető

Membránfehérje Bioinformatika Kutatócsoport, Enzimológiai Intézet
Magyar Tudományos Akadémia, Természettudományi Kutatóközpont



Budapest

2018

Köszönetnyilvánítás

Szeretném megköszönni témavezetőmnek, Dr. Tusnady Gábornak, hogy kiváló szaktudásával, szemléletmódjának és logikájának átadásával segítette munkámat.

Külön köszönet illeti Dr. Róna Gergelyt és Kucsma Nórát, a proteomikai mintaelőkészítés kontroll kísérleteinek kivitelezésében és szakmai kérdésekben nyújtott segítségért. Köszönöm Dr. Várady Györgynek a FACS mérésekben nyújtott segítséget.

Köszönöm Dr. Özvegy-Laczka Csillának, Dr. Vértessy Beának, Dr. Sarkadi Baláznak és Dr. Szakács Gergelynek a szakmai tanácsokat, valamint, hogy lehetőséget biztosítottak, hogy a kísérleti munkámat a laborjaikban végezhessem.

Köszönöm Dr. Hunyadi-Gulyás Évának és Dr. Turiák Lillának a proteomikai mérésekben nyújtott segítséget, valamint a proteomikai laborok vezetőinek Dr. Medzihradzky Katalinnak és Dr. Drahos Lászlónak a projekt támogatását.

Köszönöm továbbá a Membránfehérje Bioinformatika Kutatócsoport korábbi és jelenlegi tagjainak az inspiráló és segítőkész munkakörnyezetet. Külön köszönöm Varga Juliának az MS adatok értékelésében nyújtott segítséget (és a nagymennyiségű kókuszcsemetét, amivel hozzájárult a kutatás sikeréhez) és Dobson Lászlónak a bioinformatikai számításokat.

Köszönöm Buday Lászlónak, az Enzimológiai Intézet igazgatójának, valamint Dr. Pokol Györgynek a Magyar Tudományos Akadémia Természettudományi Kutatóközpontjának főigazgatójának, hogy lehetőséget biztosítottak a munka kivitelezésére.

Köszönöm a MTA Lendület program, valamint az OTKA pályázatok nyújtotta anyagi támogatást.

Végül szeretném megköszönni családomnak, barátnőmnek és barátaimnak a sok-sok támogatást, valamint a téma iránti érdeklődő tekintetüket.

Tartalomjegyzék

1	Bevezetés	9
1.1	A biológiai membránok felépítése	9
1.2	A transzmembrán fehérjék szerkezete.....	10
1.3	Az α -helikális transzmembrán fehérjék szintézise és csoportosítása	12
1.4	Az α -helikális transzmembrán fehérjék sejten belüli útja és poszt-transzlációs módosulásai	14
1.5	Transzmembrán fehérjék funkciója.....	15
1.6	Transzmembrán fehérje 3D szerkezetek meghatározása.....	17
1.6.1	Röntgenkristallográfia	18
1.6.2	NMR.....	18
1.6.3	Krio-elektronmikroszkópia	19
1.7	Topográfia, topológia becslő bioinformatikai módszerek.....	20
1.8	Kísérleti topológia módszerek.....	21
1.8.1	A kódoló DNS szekvencia megváltoztatásával végzett technikák.....	22
1.8.2	A kódoló DNS szekvencia megváltoztatása nélküli technikák	27
1.9	Fehérjék és peptidek szekvenálása tömegspektrometriával	31
1.9.1	Tömegspektrometriáról általánosságban.....	31
1.9.2	Biomolekulák tömegspektrometriai vizsgálata	32
1.9.3	Tömegspektrometria a proteomikában, a peptidek azonosítása és szekvenálása... ..	34
1.9.4	Tömegspektrometriai spektrumok elemzése, feldolgozása.....	36
2	Célkitűzések	38
3	Anyagok és módszerek.....	40
3.1	Anyagok	40
3.1.1	Felhasznált sejtek/sejtkultúrák	40
3.1.2	Felhasznált anyagok	40
3.2	Felhasznált eszközök.....	40
3.3	Módszerek	40
3.3.1	A TOPDB adatbázisban levő adatok bővítése	40
3.3.2	Sejtizolálás, sejttenyésztés	42
3.3.3	Sejtfelszíni primer amin csoportok jelölése	43
3.3.4	Sejtek feltárása, membránpreparálás	46
3.3.5	Membránpreparátumok szolubilizálása, emésztése	50

3.3.6	A biotinált peptidok dúsítása, eluálása, C18 oszlopon való tisztítása	51
3.3.7	Peptid fragmensek nano UPLC-MS/MS analízise	52
3.3.8	A nanoLC-MS/MS adatok értékelése, a peptidok/fehérjék azonosítása	53
3.3.9	A módosított peptidok vizsgálata	55
4	Eredmények	58
4.1	A TOPDB adatbázis frissítése kísérletes topológiai adatok gyűjtésével.....	58
4.2	A TOPDB adatbázis és a CCTOP kapcsolata	60
4.3	A fejlesztendő módszer eredményeinek topológia becslésre gyakorolt hatásának bioinformatikai elemzése	62
4.4	A topológiai adatokat szolgáltató MSTOP kísérleti módszer fejlesztése	64
4.4.1	A jelölés maximalizálása ép sejtek felszínén, valamint a sejtintegritásra gyakorolt hatása.....	66
4.4.2	A Szulfo-NHS-SS-biotin membrán impermeabilitásának vizsgálata	70
4.4.3	Membránpreparátumok készítése, fehérje tartalom mérése, minőség ellenőrzés SDS-PAGE-gélelektroforézissel	72
4.4.4	A biotinált peptidok követésére fejlesztett dot-blot módszer	74
4.5	Eluálás körülményeinek analitikai eredményekre gyakorolt hatása	76
4.6	Vörösvértestekből készített jelölt membránpreparátumok analitikai eredményeiből szerzett tapasztalatok.....	77
4.7	Vörösvértest minták és további minták analízise, és jelölt fehérje és peptid kihozatalai	81
4.8	A kísérlet eredményeinek validálása.....	85
4.9	A jelölések szekvenciális környezetének elemzése.....	86
5	Diszkusszió	88
6	Irodalomjegyzék.....	95
7	Mellékletek.....	111
7.1	Felhasznált anyagok melléklete	111
7.2	Felhasznált eszközök melléklete	111
7.3	Alkalmazott módszerek melléklete	112
7.3.1	A felszíni jelölés eredményességének ellenőrzése.....	112
7.3.2	Jelölt és jelöletlen sejtek viabilitásának tesztelése	113
7.3.3	Fehérjemérés	114
7.4	Eredmények melléklete	115
7.4.1	Szulfo-NHS-SS-biotin jelölés nélküli sejtek vizsgálata.....	115

Rövidítések jegyzéke

- ABC:** ATP-binding cassette (ATP-kötő kazetta)
- ACN:** Acetonitril (acetonitril)
- ATCC:** American Type Culture Collection (amerikai sejtgyűjtemény)
- BSA:** Bovine Serum Albumin (Szarvasmarha szérum albumin)
- CID:** Collision-Induced Dissociation (ütközés által kiváltott disszociáció)
- CHO:** Chinese Hamster Ovary (kínai hörcsög petefészek)
- DAG:** Diacylglycerol (diacilglicerol)
- DNS:** Dezoxiribonukleinsav
- DTT:** Dithiothreitol (1,4-bisz-(szulfanil)-bután-2,3-diol)
- EDTA:** Ethylenediaminetetraacetic acid (etilén-diamin-tetraecetsav)
- EndoH:** Endoglycosidase H (endoglikozidáz H)
- ER:** Endoplasmic reticulum (endoplazmatikus retikulum)
- ESI:** Electrospray Ionization (electrospray ionizáció)
- ETD:** Electron-Transfer Dissociation (elektron-transzfer disszociáció)
- FA:** Formic Acid (hangyasav)
- FACS:** Fluorescence-activated cell sorting (fluoreszcencia-aktivált sejtválogatás)
- FDR:** False discovery rate (hamis találati arány)
- FITC:** Fluorescein isothiocyanate (fluoreszcens izotiocianát)
- GPCR:** G Protein Coupled Receptor (G-protein kapcsolt receptorok)
- HA:** Hemagglutinin
- His4C:** Catalytic domain of the His4p histidinol dehydrogenase (His4p hisztidinol dehidrogenáz katalitikus doménje)
- HPLC:** High-performance liquid chromatography (nagy hatékonyságú folyadékkromatográfia)
- HRP:** Horseradish peroxidase (tormaperoxidáz)
- HT:** High Throughput (nagy áteresztőképesség)
- IP3:** Inositol trisphosphate (inozitol trifoszfát)
- LC:** Liquid Chromatography (folyadékkromatográfia)
- MS:** Mass Spectrometry (tömegspektrometria)
- NMR:** Nuclear Magnetic Resonance (mágneses magrezonancia)
- PBS:** Phosphate-Buffered Saline (foszfát pufferes sóoldat)
- PCR:** Polymerase Chain Reaction (polimeráz láncreakció)
- PhoA:** Alkaline phosphatase (alkalikus foszfatáz)
- Ppm:** Parts per million (milliomod rész)
- PVDF:** Polyvinylidene difluoride (Polivinilidén-difluorid)
- RBC:** Red Blood Cell (vörösvértest)

SDS-PAGE: Sodium Dodecyl Sulphate Polyacrylamide Gelelectrophoresis (Na-dodecil-szulfátos poliakrilamid gélelektroforézis)

SRP: Signal recognition particle (szignál felismerő részecske)

Szulfo-NHS-SS-biotin: Sulfosuccinimidyl-20(biotinamido)ethyl-1,3-dithiopropionate (szulfoszukcinimidil-2-(biotinamido)etil-1,3-ditiopropionát)

TBS: Tris-Buffered Saline (Tris puffereelt sóoldat)

TCEP: Tris(2-carboxyethyl)phosphine (trisz(2-karboxietil)foszfin-hidroklorid)

TFA: Trifluoroacetic acid (trifluor ecetsav)

TM: Transmembrane (transzmembrán)

TMS: Transmembrane Segment (transzmembrán szegmens)

TOF: Time of flight (repülési idő)

Tris-HCl: Tris(hydroxymethyl)aminomethane hydrochloride (Trisz(hidroximetil)-amino-metán hidroklorid)

UPLC: Ultra Performance Liquid Chromatography (ultranagy hatékonyságú folyadékkromatográfia)

Kivonat

Az élő szervezetben minden sejtet, valamint sejtszervecskét kettős lipid réteg választ el a környezetétől. A transzmembrán (TM) fehérjék ebbe beágyazódva helyezkednek el és „kapuként” funkcionálnak a sejt vagy organellum külső környezete és belseje között. Emellett számos biológiai folyamatban fontos szerepet játszanak, úgymint külső ingerek érzékelése, iontranszport, energiatermelés. Egyes betegségek gyógyítása úgy lehetséges, ha ezeket a kapukat használjuk a hatóanyag célba juttatásához. Az új gyógyszerhatóanyag-fejlesztések elősegítéséhez elengedhetetlen, hogy minél többet tudjunk meg róluk; különös tekintettel a szerkezetükre, előfordulásukra és orientációjukra a biológiai membránokban.

Bár a különböző genomok által kódolt fehérjék megközelítőleg negyede TM fehérje, nagyon kevés szerkezetet oldottak meg speciális fiziko-kémiai tulajdonságaik miatt. A nagy felbontású szerkezet-meghatározás technikai korlátjai miatt előtérbe kerültek alternatív kísérleti és elméleti technikák fejlesztése, melyek elősegíthetik az alacsony felbontású szerkezeti modellezést.

Az alacsony felbontású szerkezet egyik leggyakrabban használt reprezentációja a topológia, amely megadja a fehérje TM szegmenseinek (TMS) a számát, valamint azok között levő régiók membránhoz viszonyított orientációját. Ennek részleges vagy teljes meghatározására számos kísérleti és bioinformatikai becslő módszer terjedt el. Egyes becslő algoritmus pontosságát megbízható kísérleti topológiaadatok segítségével jelentősen növelhetjük, megkönnyítve így a TM fehérjeszerkezetének modellezését.

2013 szeptemberében induló ELTE Biológia doktori képzésem alatt először az irodalomból a topológiai kísérleti technikákat és az azok által szolgáltatott adatokat gyűjtöttünk össze és ezekkel az adatokkal frissítettük a TOPDB adatbázist. A topológiai adatokat több forrásból gyűjtöttük: irodalmi adatokból a PubMed felhasználásával, a UniProt fehérjeszekvencia adatbázisból, valamint a megoldott fehérjeszerkezeteket tartalmazó PDB adatbázisból. A feldolgozott források egyedi fehérje vizsgálatokat (TM fehérjeszakaszok membránhoz való helyzetének jellemzése riporter enzimekkel vagy fluoreszcens fehérjékkel fuzionált konstrukciók elemzésével, posztranszlációs módosulások meghatározásával, proteáz enzim hozzáférhetőség vizsgálatokkal, immuno-lokalizációs tesztekkel, kémiai módosításokkal) és nagyskálás, több TM fehérjét egy időben jellemző vizsgálatokat (N- és O-glikozilációs adatok) is magukba foglaltak. A bővítés fő eredménye, hogy megközelítőleg háromszorosára növekedett mind a TM fehérjék, mind a kísérleti topológiai adatok száma az

adatbázisban a korábbi verzióhoz képest. A feldolgozott módszerek többsége azonban csak egy-egy TM fehérje topológiájáról nyújt információt; ráadásul az eredményük sokszor megkérdőjelezhető, mivel megváltoztathatják a vizsgált fehérje funkcióját vagy lokalizációját. A módszerek, adatok összegyűjtése után sok humán TM fehérjéről egyáltalán nem találtunk kísérleti topológiai információt. Ezért célul tűztük ki egy saját kísérleti módszer kidolgozását, amely felhasználásával gyorsan és költséghatékonyan tudunk megbízható topológiai adatokat generálni akár több száz TM fehérjéről egyszerre, tovább bővítve, kiegészítve a kísérleti topológiai adattal rendelkező TM fehérjék halmazát.

A fejlesztett és optimalizált MSTOP módszer gyorsan kivitelezhető (~2-3hét). A kidolgozott technika segítségével fiziológias körülmények között ép sejtek plazmamembrán TM fehérjeit tudjuk természetes környezetükben vizsgálni. A membrán külső oldalán lévő TM fehérje részekben elhelyezkedő lizin aminosav oldalláncot specifikusan, biotint hordozó kémiai reagenssel módosítjuk. A biotinon keresztül a módosított részeket avidin agarózzal töltött oszlopon dúsítjuk, a módosítás helyét és szekvenciális környezetét tömegspektrometriával azonosítjuk. Egy-egy TM fehérjéhez tartozó módosított aminosavnak szekvenciális pozíciója felhasználható az adott TM fehérje topológia becslése során egy általunk kidolgozott topológia becselő eljárásban, így növelve a megfelelő becslés pontosságát.

A kísérletet úgy optimalizáltuk, hogy mind a módosított lizin aminosavak, mind a jelölt TM fehérjék száma maximális legyen. A kísérleti módszert vörösvérteteken, valamint HL60 és K562 sejt vonalakon végeztük el és közel 200 TM fehérje több mint 700 lizin oldalláncát jelöltük. A kapott adatok felhasználásával 14%-kal lehetett növelni a CCTOP becslés pontosságát, amely képes figyelembe venni a különböző kísérleti adatokból származó adatokat, mint megszorításokat.

A TM fehérjéknek a membrán külső részén elhelyezkedő részeinek feltérképezése az aktuális fehérje pontosabb modellezésén túl alapul szolgálhat a fehérjével kölcsönható mediátor molekulák lehetséges kötődési helyeinek megtalálásához, valamint a megfelelő sejtre, organelumra specifikus antitest tervezéséhez.

Abstract

In the living organism all cells and their organelles are separated by lipid bilayer from the outside environment. Transmembrane (TM) proteins are embedded in the membrane and their polypeptide chain spans it at least once. TM proteins act as molecular „gate” in biological membranes: they play role in different signal transduction and metabolic processes, as well as in maintaining the dynamic equilibrium between the internal and external environment of cells. Treatment for many diseases involves these gates, as they can be used to deliver the active substances. According to the most recent studies, about 25% of the ORFs encode TM proteins. Despite their importance and abundance, only a handful of structures have been determined so far, due to their special physico-chemical properties. Given the persisting technical difficulties associated with high resolution structure determination, bioinformatic and other experimental approaches are widely used to characterize their low-resolution structure.

Topology is the most frequently used representation of TM protein structure, defining the number and location of TM segments, as well as their orientation relative to the membrane. In the recent decades several experimental methods and bioinformatic tools have been developed to determine the topology of TM proteins. The accuracy of some topology prediction algorithms can be significantly increased by incorporating reliable experimental data as constraint.

Therefore, the primary purpose of my work was to gather topology data from several resources, including literature using PubMed, protein sequence and structure databases such as the UniProt and the PDBTM. The processed resources included unique protein assays and high throughput experiments, labeling several proteins at the same time. The TOPDB database was updated using the collected data. Both the number of included TM proteins and the number of experimental data increased by threefold compared to the previous revision. Although collecting and organizing topology data yielded an enormous amount of topology information, we could not assign any experiment for ~50% of the human TM proteome. To overcome this and to produce experimental data for some of the uncharacterized proteins, we developed the MSTOP method, a rapid procedure generating topology data for hundreds TM proteins at the same time. The method was tested on three cell types: on red blood cell, on HL60 and on K562 cell lines. At this point the experiments yielded 730 extra-cytosolic positions for nearly 200 TM proteins. Topology data provided by MSTOP can increase the accuracy of CCTOP prediction method by 14%.

1 Bevezetés

1.1 A biológiai membránok felépítése

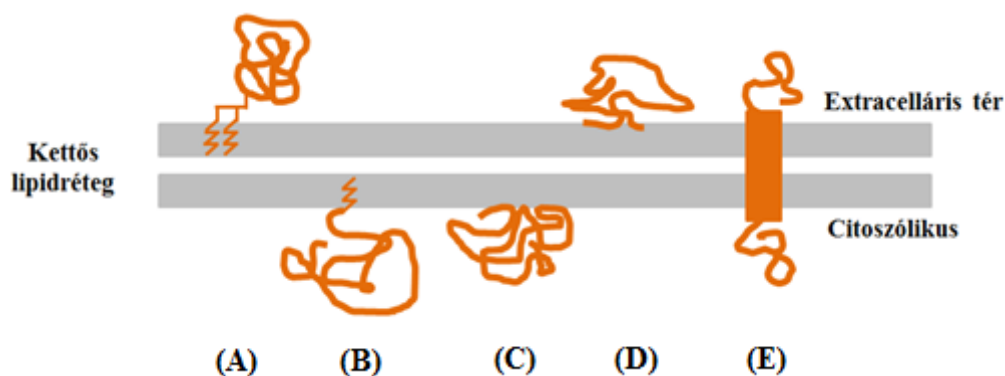
A sejteket és azok organelleumait kettős foszfolipid rétegek határolják, amelyek a külső és belső részek elválasztásán kívül szelektív permeabilitásuk révén szabályozzák, hogy mely anyagok képesek bejutni, vagy elhagyni az adott sejtet, illetve annak kompartmentjét. Az emlős szervezetek biológiai membránjaikban három fő alkotóelemet különböztetnek meg:

- A hidrofób farokkal és hidrofil poláris fejjel rendelkező lipideket, úgymint foszfolipidek, amelyek alapváza alapján megkülönböztetünk foszfoglicerideket és szfingomielineket. Ezen felül az emlősökre jellemző lipid még a koleszterol is, amely a membrán fluiditásában, valamint lipid raftokban jelentős szerepet játszó összetevő^{1,2}.
- A membránfehérjéket, melynek egy részhalmaza a dolgozat tárgyát képező TM fehérjék (részletesebben a felsorolás után).
- A harmadik az előző alkotóelemeket módosító szénhidrátok, így létrehozva glikolipideket és glikoproteineket. Ezek a szénhidrát módosulások a membrán extra-citoszólikus/lumenális oldalán következnek be mind a fehérjéken, mind a lipideken³.

A három fő komponensből a membránfehérjék a különböző biológiai membránokban eltérő mennyiségben vannak jelen. Az idegsejtek axon nyúlványainak szigetelését végző szaraz mielin membránnak kevesebb mint 30% a fehérjetartalma⁴, míg a legdúsabb fehérjetartalmú membránnak tartott mitokondriális belső membránnak a 60-70%-át teszik ki ezen fehérjék^{5,6}.

A membránokhoz kapcsolódó fehérjék kettős lipid réteghez való illeszkedésük alapján két fő egységre oszthatók: a perifériás fehérjékre, amelyek a membrán valamelyik oldalán horgonyoznak, valamint integráns fehérjékre, amelyek a membránba ágyazódnak. A perifériás fehérjék közvetve vagy közvetlenül is kapcsolódhatnak a kettős lipid réteghez. A közvetett kapcsolatra jellemző példa, a glikolipidek által membránhoz rögzített fehérjék, ezek az úgynevezett glikozil-foszfátidil-inozitol (GPI) kapcsolt fehérjék⁷ (1. ábra/A). Egyes perifériás

fehérjék közvetlen kapcsolódhatnak a membrán lipidekhez/zsírsavakhoz, ide sorolhatók a farnezilált (izoprén váz kapcsolódik a fehérje lánchoz, pl.: Ras, Rho, Rab fehérjék⁸ vagy mirisztoil-csoporttal módosított fehérjék (telített zsírsav kapcsolódik az N-terminális glicin α -aminocsoportjára, pl.: citokróm b5 reduktáz⁹) (1. ábra/B). A membrán két oldalán egyaránt előfordulnak olyan perifériás fehérjék is, melyek másodlagos kötőerők (hidrofób, ionos kölcsönhatások) révén csatlakozhatnak a lipidekhez vagy más membránfehérjékhez (pl.: epidermális növekedési faktor¹⁰, spektrin, ankirin fehérjék¹¹, 1. ábra/C, D). Az utolsó jelentős csoport az integráns fehérjék egységébe tartozó, legalább egy membránon áthaladó szakaszt tartalmazó TM fehérjék (1. ábra/E).



1. ábra: A membrán fehérjék membránhoz való asszociálódásának fő típusai. (A) GPI-kapcsolt fehérje, (B) Lipid-kapcsolt fehérje, (C, D) Másodlagos kötőerőkkel membrán-asszociált fehérjék, (E) TM fehérje

1.2 A transzmembrán fehérjék szerkezete

A biológiai membránok a vízoldható anyagok számára lényegében átjárhatatlanok, így a külső és belső részek közötti átjuttatásban speciális „molekuláris kapuk” vesznek részt, ezek az úgynevezett TM fehérjék. A teljes genomszekvenálási adatokat elemezve a nyitott leolvasási keretek (open reading frame, ORF) megközelítőleg 20-30%-a TM fehérjét kódol¹²⁻¹⁴. Ezek a fehérjék több csoportba sorolhatóak, egyrészt az átívelő szakaszok másodlagos szerkezete, illetve a TMS-eik száma és irányultsága alapján.

A membránon áthaladó másodlagos szerkezetek alapján megkülönböztetünk α -helikális, valamint β -hordó TM fehérjéket. Az α -hélixek önállóan vagy kötegekbe rendeződve fordulhatnak elő, míg a β -szálak önmagukba záródva egy hordó palástját alkotják. A hidrogéndonor és hidrogénakceptor atomok nem tudnak kedvező hidrogénhid kötést kialakítani az oldószerrel a kettős lipidrétegben, ezért a polipeptid lánc csak olyan szerkezetet

vehet fel, amelyben a lehető legtöbb hidrogéndonor és hidrogénakceptor atom a láncon belül alakít ki hidrogénhidat. Ez az oka, hogy a fentebb említett két membránon átívelő szerkezeti elem terjedt el, ahol α -helikális szerkezetben a szekvenciálisan 1-4 pozícióban levő karbonil-oxigén és az amid-hidrogén atomjai között, vagy két párhuzamosan futó (anti)parallel β -szál között alakulnak ki a hidrogénhidak^{15,16}. Míg az α -helikális fehérjék minden biológiai membránban megtalálhatók, addig a β -hordó fehérjék a Gram-negatív baktériumok külső membránjában, valamint az eukarióta sejtek mitokondrium és kloroplaszt membránjaiban fordulnak elő¹⁷.

A szerkezeti adatbázisok szerint közel 3000 α -helikális és 400 β -hordó TM fehérje térszerkezetét ismertjük^{18,19}, ami az összes ismert térszerkezet alig több mint 2%-a. Globuláris fehérjék esetében a lehetséges natív térszerkezetek (foldok) száma egyes becslések szerint nem több mint 10000^{20,21}. A TM fehérjékben a membránon keresztülhaladó α -hélixek általában merőlegesen a membránra és egymáshoz szorosan illeszkednek, ezért a TM fehérjék szerkezeti variabilitása kisebb, mint az azonos hosszúságú globuláris fehérjéké. A TM fehérjék azonban lényegesen nagyobbak a globuláris fehérjék többségénél, és mivel a lehetséges szerkezeti variációk száma exponenciálisan növekszik a fehérje méretével, így lényegesen többféle szerkezetet vehetnek fel. A lehetséges szerkezeti teret azonban szűkíti, hogy a különböző számú TM hélixet tartalmazó fehérjék eloszlása rendkívül inhomogén. Az egy TM hélixet tartalmazó fehérjék az összes TM fehérje közel 40%-át teszik ki, és a 7, illetve 12 TM hélixet tartalmazó fehérjék száma is kiugróan magas. Becslések alapján a TM fehérjék közel 80%-a 300 szerkezetbe, 90%-a pedig 500 szerkezetbe sorolható²². Jelenleg megközelítőleg 100 különböző szerkezetű TM fehérjét ismerünk, így a fenti becslés alapján számított nem redundáns szerkezeti térhez viszonyítva a lehetséges szerkezetek alig 20%-át ismerjük.

A megoldott 3D szerkezetek hiánya miatt, a TM fehérjék egy részéről mind kísérleti, mind bioinformatikai eszközökkel próbálnak topológiát meghatározni. A topológia megadja a TM fehérjék TMS-einek számát, azok szekvencián belüli helyét és membránhoz viszonyított orientációját. A „pozitív bent-szabály” a legismertebb szabály, amely befolyásolja a topológiát és az 1.3 fejezetben ismertetett α -helikális TM fehérjék csoportosítását is. Először a bakteriális fehérjékben fedezték fel, hogy a pozitívan töltött aminosavak négyszer nagyobb mennyiségben találhatóak a citoplazmás hurkokban, mint a periplazmatikus hurkokban²³. Ennek hátterében egyrészt az állhat, hogy a TM szegmensek membránba ágyazását végző transzlokon (heterotrimer fehérjekomplex) molekuláris gépezet α -alegysége negatívan töltött

aminosavat tartalmaz a riboszómális komplex kötésének környezetében. Továbbá a fejcsoportjában negatív töltésű lipidek a citoszólikus oldalon gyakrabban találhatóak, illetve az elektrokémiai gradiens jelenléte szintén hatással lehet a fehérje szintű pozitív töltés többletre a citoszólban²⁴⁻²⁶. Számos tanulmány vizsgálta, hogy ha a TM szegmensek citoszólikus szegélyező régióiban lévő pozitívan töltött aminosavakat kicserélik negatívan töltöttre, illetve az extra-citoszólikus szegélyező régióban lévő aminosavakat pozitív töltésűre, ezzel megzavarható a transzlokon molekuláris gépezet, a natívval ellentétes orientációt is el lehetett érni a megfelelő TM szegmens esetében²⁷⁻²⁹.

A dolgozat fő témája a humán eredetű sejtek plazmamembrán fehérjeinek vizsgálatáról készül, a továbbiakban az α -helikális fehérjéket tárgyalom részletesebben.

1.3 Az α -helikális transzmembrán fehérjék szintézise és csoportosítása

A TM fehérjék a fehérjék kódoló dezoxiribonukleinsav (DNS) szakaszokról átíródott hírvivő RNS érése után, a tradicionális fehérjeszintézisnek megfelelően, a riboszómán szintetizálódnak. Az α -helikális TM fehérjék csoportosítása a szintézis során keletkezett TM fehérje topológiája (TM hélixek száma és irányultsága) alapján történik.

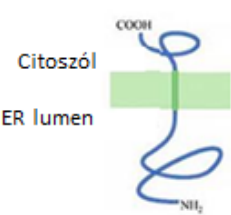
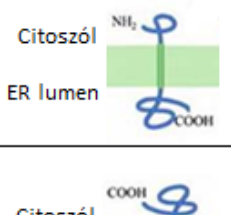
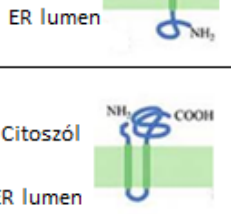
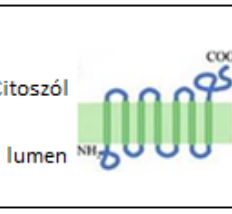
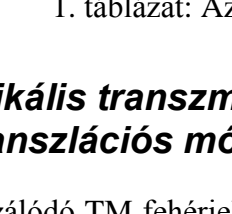
A TM fehérjék első három csoportját az egy TMS-sel (bitopikus) rendelkező fehérjék alkotják.

Az 1. csoportba tartoznak azok a TM fehérjék, amelyek a szolubilis fehérjékhez hasonlóan hordoznak egy endoplazmatikus retikulumba (Endoplasmic reticulum, ER) irányító szignál szekvenciát. Az adott fehérje mRNS-éről a riboszómán szintetizálódott szignál peptidet felismeri a szignál felismerő részecske (signal recognition particle, SRP), amely komplexet az ER membránban lokalizálódó SRP receptor megköt. Az ER felszínén az SRP receptor (röviden: SR) továbbítja az mRNS-riboszóma-SRP komplexet a transzlokon komplexre (ez a Sec61 komplex emlősökben³⁰). Ezt követően az SRP disszociál, miközben a szintetizálódó peptidlánc szignál peptid régiója áthalad a transzlokon pórusán, majd a transzlokon komplexből a kettős lipidréteg felé továbbítódik (laterális irányú továbbítás). Majd a szignál-peptidáz levágja a szignál peptidet az ER lumenében^{31,32}. Amikor a következő hidrofób szegmens/TMS szintén befűződik a transzlokon pórusába, akkor a transzlokáció megakad (szegmens neve stop-transzfer szekvencia, mivel onnantól kezdve, hogy a TMS stabilizálódik, a szintetizálódó szál többi része nem jut be a transzlokon pórusába), így a teljes

fehérje irányultsága a szintézis végeztével az N-terminális lumenális, C-terminális citoszólikus lesz.

A következő két csoport fehérjéi nem tartalmaznak szignál szekvenciát, így ezeket a fehérjéket a szintetizálódó TM régió rögzíti a transzláció során a transzlokonhoz. Annak függvényében, hogy a hidrofób szakaszt szegélyező N-vagy C-terminális részen (a „flanking” régióban) halmozódnak fel a pozitívan töltött aminosavak, a TM régió eltérő orientációban helyezkedhet el a membránban (lásd „pozitív-bent” szabály^{33,34}). A 2. csoportba tartoznak az N-terminális citoszólikus/C-terminális lumenális TM fehérjék (ennek TM régiója az úgynevezett start transzfer szekvencia, hiszen ettől a szekvenciától kezdve a szintetizálódó többi fehérje rész a lumenbe jut), míg a 3. csoportba tartozó TM fehérjéknek a TM régiója az előzőhöz képest fordított orientációjú (a TM régiója a stop transzfer szekvenciának felel meg, mint az 1. típusú TM fehérjéknél).

A több TM régióval rendelkező fehérjék alkotják az utolsó két csoportot. A 4A csoportot alkotják azok a fehérjék, ahol a fehérje N-terminálisa a citoszól felé néz. A 4B csoportban lévő fehérjéknek ellentétes oldalon, azaz a lumenális/ extra-citoszólikus térben van az N-terminálisa^{26,29,35} (a csoportosítást az 1. táblázat szemlélteti).

Csoport	Sematikus ábra	Mechanizmus	Topológia	Példa
1	 <p>Citoszól ER lumen</p>	SRP/SR/Sec61+ szignál peptidáz	N-terminus_lumen/ C-terminus_citoszól	Glikoforin A, Alacsony sűrűségű lipoprotein (LDL) receptor
2	 <p>Citoszól ER lumen</p>	SRP/SR/Sec61	N-terminus_citoszól/ C-terminus_lumen	Transzferrin receptor, Galaktozil transzferáz
3	 <p>Citoszól ER lumen</p>	SRP/SR/Sec61	N-terminus_lumen/ C-terminus_citoszól	Szinaptotagmin 1, Neuregulin
4A	 <p>Citoszól ER lumen</p>	SRP/SR/Sec61	N-terminus_citoszól/ C-terminus_lumen vagy citoszól (páros vagy páratlan TMS szám függvényében)	Band 3 anion transzporter, Glükóz transzporter
4B	 <p>Citoszól ER lumen</p>	SRP/SR/Sec61	N-terminus_lumen/ C-terminus_citoszól vagy lumen (páros vagy páratlan TMS szám függvényében)	Rodopszin, Dopamin receptor

1. táblázat: Az α -helikális TM fehérjék alaptípusai

1.4 Az α -helikális transzmembrán fehérjék sejten belüli útja és poszt-transzlációs módosulásai

A szintetizálódó TM fehérjelánc megfelelő szegmenseinek transzlokon közvetített ER membránba épülése alatt, az ER lumenében, az egyes fehérjeszakaszokon glikoziláció történhet. Ennek két fajtáját különböztetjük meg: N- vagy O-glikoziláció. Az előbbi esetében az aszparagin aminosavakra kerülhet szénhidrátlánc, ha az aszparagin környezete NXS/T motívum, amelyben az X bármilyen aminosav, kivéve prolin. Ennek a motívumnak minimum 10-15 aminosav távolságra kell lennie a membrán határtól³⁶, hogy az oligoszacharid protein transzferáz át tudja helyezni az aszparagin amid nitrogénjére a dolicholon (poliizoprenoid típusú membránlipid) lévő oligoszacharid részt. Az O-glikozilációnál nincs konszenzus motívum, illetve membránhoz viszonyított távolsági megszorítás sem, a szerin és treonin

aminosavak hidroxil csoportjain következik be úgy, hogy az N-acetilgalaktózamin kötődik a fehérjékhez a GalNAc-transzferáz enzim segítségével³⁷.

Ezután a TM fehérjék az ER-ből a cisz-, mediális- és transz-Golgi ciszternákon haladnak keresztül, ahol a szacharid egységek tovább módosulhatnak (glikozidáz enzimek bonthatják a meglévő struktúrát, és különböző transzferázok kapcsolnak rájuk további aktivált monoszacharid egységeket), terminális lépésként fukóz és szialsavak kapcsolódnak a meglévő struktúrához, végül a fehérjék a rendeltetési helyükre (pl. endo lizoszomális rendszerbe, plazmamembránba stb.) kerülnek. Egyes esetekben még további poszttranszlációs módosításokon mehetnek keresztül - ilyenek például a foszforiláció, acetiláció - hogy az adott jelpályában betöltött szerepüket megfelelően elláthassák, szerkezetük stabilizálódjon³⁸⁻⁴⁰.

A plazmamembránban lokalizálódó TM fehérjék klatrin-függő vagy-független útvonalon internalizálódhatnak újra. Ezután vagy az endoszómális rendszeren keresztül juthatnak vissza a rendeltetési helyükre (reciklizáló endoszómából) vagy ubikvitinilációval (a korai és késői endoszómából) a lizoszómába⁴¹⁻⁴³.

1.5 Transzmembrán fehérjék funkciója

A TM fehérjék jelentős szerepet töltenek be a különböző extracelluláris és citoszólikus mediátorok felismerésével (receptor funkció) jelátviteli pályákban, transzport folyamatokban, enzimkatalizált reakciókban, energiatermelő folyamatokban, tápanyag felvételben, valamint a sejt/organelum külső és belső környezete között a komponensek dinamikus egyensúlyának fenntartásában, a sejt-sejt kapcsolatok kialakításában, valamint extracelluláris mátrixhoz való horgonyzásban.

A humán genomban a membránfehérjék legnépesebb receptor funkcióval rendelkező családja a G-protein kapcsolt receptorok (G Protein Coupled Receptor, GPCR), melyek száma 800 felett van⁴⁴⁻⁴⁶. A TM fehérjék jelátvitelben betöltött szerepét szemléltethetjük a GPCR-fehérjékkel. A GPCR-ek ligand kötés után a heterotrimer G-fehérjével hatnak kölcsön, amelynek hatására a G-fehérje α -alegység guanozin-difoszfátja cserélődik guanozin-trifoszfátra, így az α -alegység aktiválódik. Ez a foszfolipáz C-t (Phospholipase C, PLC) aktiválja, ami a foszfatidil-4,5biszfoszfátot (Phosphatidylinositol 4,5-bisphosphate, PIP2) inozitol-trifoszfátra (Inositol trisphosphate, IP3) és diacilglicerolra (diacilglicerol, DAG) bontja. Ezek a folyamatok a membrán közelében játszódnak le, majd a keletkezett kismolekulák a citoplazmába továbbítják a jelet. Következő lépésben az IP3 kötő receptor TM fehérje, amely egy intracelluláris ligandum hatására nyíló ioncsatorna, specifikusan köti az

IP3-at és ami endoplazmás retikulumból Ca^{2+} felszabadulást eredményez. A folyamat végén a Ca^{2+} -t köti a kalmodulin, és ez a komplex aktiválja a kalcium–kalmodulin-dependens protein kinázt (CaM kinázt), a DAG pedig aktiválja a protein kináz C-t⁴⁷. A jelátviteli pályán így elérkeztünk a kinázokhoz, amelyek foszforilálják a célfehérjét, így további jelátviteli útvonalakat aktiválhatnak/inaktiválhatnak⁴⁸.

A TM fehérjék egyes alcsoportjai passzív transzportot végeznek, mivel a transzporthoz nem szükséges energiaforrás - például az adenzin-trifoszfát (ATP) – hidrolízise, a folyamatot a koncentráció-gradiens hajtja. A GPCR jelpályában említett IP3 receptor által végzett Ca^{2+} transzport is passzív. Az IP3 kötése segíti elő a kalciumion transzportját az ER lumenből a citoszól felé, a folyamat az elektrokémiai gradiens irányába zajlik. Ilyen transzport mechanizmussal rendelkezik a glükóz transzporter (GLUT1)⁴⁹ is, és a nitrogén lebontás melléktermékét, az ureát transzportáló (UT-A, UT-B) fehérjék is^{50,51}.

Más TM fehérjék aktív transzporttal továbbítják az adott komponenst, ellentétes irányban, mint ahogy a koncentráció-gradiensből következne. Ezek további két csoportra oszthatóak. Az elsődleges aktív transzporterek csoportjába tartozó TM fehérjék esetében közvetlen energiafelhasználás, ATP-hidrolízise szükséges az adott komponens transzportálásához. A másodlagos aktív transzporter csoportba tartozó fehérjék esetén egy másik molekula energiagradiensének felhasználásával transzportálják az adott anyagot, azaz két komponens transzportálása történik egy időben. Ez utóbbi esetről az egyirányú transzportot szimpornak, míg az ellentétes irányút antiport-transzportnak nevezzük.

Az aktív transzporterek elsődleges csoportjába tartoznak az ATP-kötő kazetta (ATP-binding cassette, ABC) fehérje család tagjai, melyek baktériumokban, gombákban, növényekben, állatokban és az emberben is fellelhetőek⁵². Emberben 48 fehérjét azonosítottak a családból, amelyek 7 csoportba sorolhatóak (ABC A-G). Jellegzetességük a 6 TMS-t tartalmazó Transzmembrán Domén, amelyekből funkcionális ABC transzporter fehérjékben mindig kettő van (egy vagy több fehérjelánc alakíthatja ki). További jellegzetes egységük a nukleotid kötő domén (NBD), amelyet három konzervatív motívum (*Walker A/B*), valamint az *ABC signature* motívum jellemez. A citoszólikus NBD az ATP kötésért és a hidrolízisért felelős, az ebből származó energia fordítódik a transzporter működésére. Egyes tagjairól már azonosították, hogy milyen funkciókat látnak el a sejten belül. Az ABCA alcsoport tagjai részt vesznek a lipid és koleszterol transzportban, az ABCB1 multidrog rezisztenciát okozhat, míg ABCG2-t az emlőrák esetén emelték ki az alkalmazott kemoterápiás szerekkel szembeni rezisztencia kapcsán. A cisztás fibrózis transzmembrán konduktancia regulátor fehérje

(CFTR, ABCC7), egy kloridion csatorna, ami kivételt képez más ABC fehérjékkel szemben, mivel a transzportot nem ATP-hidrolízis segítségével végzi, hanem passzív transzporttal⁵³⁻⁵⁵.

Az aktív transzporterek másodlagos csoportjában az antiporterek közé tartozik például a Band 3 anion transzporter, amely az egyik leggyakoribb fehérje a vörösvértestek (Red Blood Cell, RBC) plazmamembránjában, és a kloridion és a hidrogénkarbonátion cseréjéért felelős^{56,57}. Szintén antiporter a $\text{Na}^+/\text{Ca}^{2+}$ cserélő fehérje, amely a transzport során 3 (Na^+):1(Ca^{2+}) arányban végzi az ionok kicserélését a membrán két oldala között⁵⁸. Szimporterek fehérjék jellegzetes képviselői a Na^+ és glükóz, illetve a Na^+ és aminosavak együttes szállítását végző transzporterek⁵⁹.

A membrán transzporter fehérjék csoportosítását az IUBMB (International Union of Biochemistry and Molecular Biology) által jóváhagyott TCDB (Transporter Classification Database) adatbázis tartalmazza. A transzporter csoportosítási rendszer hasonlít az Enzim Bizottság (Enzyme Commission, EC) rendszeréhez, amelyben az enzimeket csoportosítják mind funkció, mind filogenetikai információk alapján. Összesen 1277 transzporter családot rendszereztek az adatbázisban⁶⁰.

A funkcionális csoportosítások példáiból egyértelmű, hogy a TM fehérjék valóban széles spektrumú és biológiailag jelentős folyamatokban játszanak szerepet. Ezek alapján nem meglepő tény, hogy a jelenleg forgalomban lévő, Élelmiszer- és Gyógyszerügyi Hatóság (Food and Drug Administration, FDA) által engedélyezett gyógyszerhatóanyagok több mint fele, valamilyen TM fehérjét céloz meg⁶¹.

1.6 Transzmembrán fehérje 3D szerkezetek meghatározása

A biológiai célmolekulára (jelen esetben TM fehérjék) való gyógyszertervezés kiindulópontja, a célmolekula kiválasztása után, annak előállítása, majd tisztítása szerkezetkutatási célokra. A célmolekula és már ismert ligandja együtt úgynevezett bioaktív konformációt vehet fel, amely szerkezetének megismerése lehetővé teszi a kölcsönható felszínnek pontosabb azonosítását, ezáltal akár nagyobb affinitású, de szerkezetét tekintve analóg molekulák tervezését és további hozzáférhető molekuláris felszín megismerését is biztosítja. A fehérjék 3D szerkezetének (amit az atomi koordinátákkal írhatunk le) meghatározásra 3 fő módszert használnak jelenleg: a röntgenkristallográfiát, a mágneses magrezonancia spektroszkópiát (Nuclear Magnetic Resonance, NMR) és a krio-elektronmikroszkópiát. A következőkben ezeket tárgyalom általánosan, valamint TM fehérjékre vonatkozóan.

1.6.1 Röntgenkristallográfia

A röntgenkristallográfiával makromelekulák rendezett kristályai vizsgálhatók. A röntgensugárzás nagyenergiájú elektronmágneses sugárzás, amely a fehérjemolekulák által alkotott kristályrácsra elhajlást szenved (mivel kristályrácsponatok távolsága összemérhető a hullámhosszal). A detektált diffrakciós mintázat alapján Fourier-transzformáció segítségével megkapjuk a fehérje elektronsűrűség-térképét, amelyből a vizsgált fehérjeszekvenciájának ismeretében a 3D szerkezet modellezhető⁶². A módszerrel a túl flexibilis régiókat tartalmazó fehérjék nehezen vizsgálhatóak, mivel a szórt sugár az ilyen részokről változatosan interferál, ezért a diffrakciós mintázat annyira elmosódottá válik, hogy nem lehet az elektronsűrűség-térképet kiszámolni. A kristallográfia hátrányaként kezelik, hogy csak statikus információt ad az adott fehérjeszerkezetéről, a dinamikáját nem tudja jellemezni.

A membránfehérjék kristályosítását, más szolubilis fehérjékhez hasonlóan a fehérje overexpressziója előzi meg a sejtekben (akár specifikusan jelölve a tisztítás könnyítése céljából) és ezekből kell izolálni detergens felhasználásával a vizsgálni kívánt membránfehérjéket. Az alkalmas detergens megválasztása, amit még a megfelelő TM fehérje tolerál (ha nem sérül a szerkezet, feltételezhetően a funkció is megmarad) általában időigényes folyamat. Nem megfelelő detergenst, puffert, pH-t, hőmérsékletet választva a fehérje natív szerkezete is megváltozhat, a polipeptid láncok aggregálódhatnak, kicsapódhatnak, amelyek így ellehetetlenítik a kristályképződést^{63,64}. A legújabb technikák az izolált TM fehérjéket membránt mimikáló rendszerekbe (pl.: Lipidic Cubic Phase, LCP) ágyazzák és így vizsgálják a szerkezetüket^{65,66}. A kísérleti kihívások ellenére a legtöbb TM fehérjeszerkezetét röntgenkristallográfia segítségével határozták meg⁶⁷.

1.6.2 NMR

A másik elterjedt szerkezet-meghatározási módszer az NMR. Az eljárás azon az elven alapszik, hogy a neutron, vagy protont, vagy mindkettőt páratlan számban tartalmazó NMR-aktív atommagok információt adnak a környezetükről, szomszédos atom kapcsolataikról, elektronjaikról. Mikor elektromágneses térbe helyezzük azokat, gerjesztődnek, és az adott mag környezete miatti árnyékolásból kiszámolható az árnyékolatlan referencia anyaghoz viszonyított kémiai eltolódás, amelyből a vizsgált makromolekula szerkezetére következtethetünk. A technika előnye, hogy natív közegben és roncsolás nélkül végezhető a mérés, akár élő sejten belül is⁶⁸. Hátránya, hogy elsősorban kisebb szerkezetek vizsgálhatók (<25kDa), az érzékenységszűrés és a spektrális komplexitás miatt^{69,70}.

A TM fehérjék jelentős része az említett felső molekulaméretnél lényegesen nagyobb, így ezzel a technikával kevesebb szerkezetet oldottak meg^{71,72}. A krisztallográfiához hasonlóan további hátrány, hogy a célfehérjéből, például a tömegspektrometriához (Mass Spectrometry, MS) képest nagy mennyiség kell és gyakran költséges izotópjelzést is használni kell az NMR vizsgálat előtt.

1.6.3 Krio-elektronmikroszkópia

A krio-elektronmikroszkópia egy speciális mintaelőkészítés és az elektronmikroszkópia ötvözésével született. Az elektronmikroszkópok a normál fénymikroszkópokhoz hasonló elven működnek, de fény helyett gyorsított elektronokat bocsátanak a mintára vákuum alatt, a fókuszáláshoz pedig optikai lencsék helyett elektromágneses lencséket használnak. A hullámhossz csökkentésével a feloldóképesség csökken, azaz javul. Ahol nagy rendszámú elemek vannak a tárgylemezen, ott elektron eltérítést fogunk tapasztalni, azaz a detektor ernyőre kevés elektron jut, így sötét lesz a kép, míg ahol könnyű elemek vannak jelen, ott világos, hiszen azok az elektront kevésbé fogják eltéríteni. A krio-elektronmikroszkópiánál a „krio” szó a mintaelőkészítési körülményét jelenti, mely során a mintát nagyon gyorsan lefagyasztják. Ilyenkor a mintában lévő víz még nem tud kristályt képezni, vitrifikált amorf víz képződik, ezáltal nem is tudja roncsolni azt (folyékony nitrogénnel -196°C -ra előhűtött propánnal vagy etánnal végzik). A krio-elektronmikroszkópia lehetővé teszi az egyes makromolekulák szerkezeti és molekuladinamikai vizsgálatát natív állapotukban. A fagyasztott fehérje makromolekulák 2D vetületeiből rekonstruálják a molekula 3D szerkezetét^{73,74}.

A TM fehérjék tekintetében a jövőben jó kiegészítő módszere lesz az előző két technikának, hiszen ennél a módszernél nagyobb molekulák is vizsgálhatók natív körülmények között, ráadásul a konformáció változások is követhetőek egy adott ko-faktor hatására⁷⁵. Azonban az időigényes fehérje tisztítási lépéseket ennél a módszernél is el kell végezni.

A három felsorolt szerkezet-meghatározó módszer, habár az összes atom koordinátáját megadja az adott molekula szerkezetben, a fehérje membránon belüli orientációját és a membrán pontos elhelyezkedését nem, ezt utólag kell rekonstruálni bioinformatikai vagy kísérleti eszközökkel^{18,19,76}.

1.7 Topográfia, topológia becslő bioinformatikai módszerek

A TM fehérjék esetén a teljes szerkezet-meghatározása nagyon sok problémába ütközik, más kísérletek pedig időigényesek, illetve csak egyes fehérjeszegmensek helyzetét tudják meghatározni, a topológia meghatározására elterjedtek a különböző számítógépes predikciós módszerek.

Az első becslő eljárások azon a megfigyelésen alapultak, hogy a kettős lipidrétegben elhelyezkedő részek csak megadott hosszúságúak lehetnek és ezekben a membránon átmenő régiókban, a fehérje többi részéhez képest magasabb a hidrofób aminosavak aránya. Ez utóbbi mérésére számos skálát bevezettek, amelyek az egyes aminosavak a lipidből a vizes környezetbe való mozgásukhoz szükséges szabadentalpia változást írják le⁷⁷. A különböző hidrofóbicitás skálákat számos egyedi fehérje vizsgálatoknál használták a kísérleti adatok alátámasztására⁷⁸. A skála használatával becsülhető, hogy a szekvencián belül hol helyezkednek el a TM szegmensek (azaz megadja a topográfiát).

Később von Heijne és mtsai leírták a „pozitív-bent szabályt” (1.2 fejezetben részletesebben)⁷⁹ és ezzel nagyjából egyidőben elterjedtek a számítógépes topológia becslő módszerek is. Az első eljárások egyszerű csúszóablak átlagolással működtek⁸⁰. Később statisztikai eljárások is megjelentek, melyek jelentősen megnövelték a predikció pontosságát⁸¹. Az ismert szerkezetek számának folyamatos növekedésével elegendő adat vált elérhetővé, ami lehetővé tette a tanuló gépek eljárások használatát. Ezek nem csak a topológia becslés pontosságát növelték nagy mértékben, de lehetővé tették a szekvencia még részletesebb felosztását, így a köztes hurkok membránhoz közel és távol eső részeinek megkülönböztetését⁸². A következő lépést a homológ fehérjék tulajdonságainak figyelembevétele jelentette: mivel a szerkezet konzerváltabb, mint a szekvencia, azok a szekvenciák, melyeknek szekvenciális azonossága magas, nagy eséllyel ugyanabba a szerkezeti osztályba (foldba) tartoznak és ugyanazzal a topológiával fognak rendelkezni⁸³.

A topológia becslésére az egyik lehetséges gépi tanulási eljárás a rejtett Markov modell. A rejtett Markov modell legnagyobb előnye az általa biztosított architektúra, amellyel lehetséges figyelembe venni azokat a törvényszerűségeket (nyelvtani szabályokat), amelyek a TM fehérjéket jellemzik (pl citoszólikus és extra-citoszólikus szegmensek nem következhetnek egymás után; a membránrégió hossza legalább 14 aminosav, hogy átérje a lipidréteget, stb.). A rejtett Markov modellek másik nagy előnye, hogy képesek figyelembe

venni a különböző kísérleti eredményeket és ezeket kényszerfeltételként használni a becslés során⁸³.

A tanuló gép eljárásokat csoportosíthatjuk a tanítás módja alapján. A legtöbb módszer felügyelt tanításon alapult, azaz egy nagyobb halmaz elemeit felcímkézték és az algoritmus a címkéket figyelembe véve tanulja meg az egyes fehérjék jellemzőit⁸⁴. Nem felügyelt tanítás esetén ez a lépés elmarad: ilyenkor a különböző térrészekre jellemző aminosav eloszlások maximális különbségét keressük, és ez alapján osztjuk részekre^{83,85}. Mindkét módszernek vannak előnyei és hátrányai: felügyelt tanítás esetén a módszer alkalmas TM fehérjék szűrésére. A nem felügyelt tanításon alapuló módszereknél ez sok hibával járna, mivel az algoritmus egy globuláris vagy rendezetlen fehérjén is megtalálja a legkülönbözőbb részeket. A nem felügyelt tanítással működő módszerek viszont hosszútávon pontosabbak⁸⁶, mivel nem az aktuálisan ismert TM fehérjék halmazán lettek tanítva, hanem az egyes részek eltérő tulajdonságait használják ki.

A TM fehérjék becslése során fontos figyelembe venni a szignál peptid jelenlétét is. Ezek a szegmensek a TM régiókéhoz nagyon hasonló aminosav eloszlással rendelkeznek és ez sok predikciós módszert összezavar. Az újabb módszerek már képesek megkülönböztetni ezeket a részeket is. Régebbi eljárások használata esetén ezeket a részeket el kellett távolítani a szekvencia elejétől.

A legmodernebb módszerek (Consensus Constrained Topology, CCTOP⁸⁷, TOPCONS2⁸⁸ már ötvözik az itt felsorolt tulajdonságok összes előnyét: egyrészt több módszer eredményét is figyelembe veszik, és ez alapján hoznak létre egy konszenzus becslést. Ilyenkor az egyes módszerek egyedi hibái kiküszöbölődnek a többségi döntésnek köszönhetően. Másrészt ezek a módszerek már automatikusan figyelembe veszik az irodalomban fellelhető molekuláris biológiai kísérletek eredményét^{89,90}, illetve a szerkezet-meghatározásból származó adatokat (PDBTM¹⁸, OPM¹⁹), vagy a konzervatív módon, mindig a membrán azonos oldalán megjelenő doméneket és motívumokat (TOPDOM⁹¹).

1.8 Kísérleti topológia módszerek

A korábban ismertett TM fehérjék 3D szerkezet-meghatározás nehézségei miatt alternatív kísérleti módszerek alakultak ki a megoldatlan térszerkezetű fehérjék részbeni jellemzésére. Ezek a stratégiák a szubcelluláris lokalizáción és más biológiai információkon túl, az egyes TM fehérjék egy-egy aminosavának, szakaszának membránhoz viszonyított helyzetét is megadhatják (citoszólikus, extra-citoszólikus, membrán szegmens), azaz

topológiai adatot generálhatnak, amelyek az 1.7 fejezetben tárgyalt topológia becslő módszerek pontosabb eredményeit is elősegíthetik (amennyiben az adott algoritmusban megszorításként alkalmazzuk azokat). A kísérletek két nagy csoportba sorolhatók: a célfehérje kódoló DNS szakaszának változtatása alapján módosított, illetve nem módosított fehérjéket létrehozó technikákat különböztetünk meg.

1.8.1 A kódoló DNS szekvencia megváltoztatásával végzett technikák

A következőkben olyan topológiai adatot nyújtó technikákat tárgyalunk, ahol az adott TM fehérje kódoló DNS szekvenciáját a molekuláris biológia eszköztárával módosítjuk és a módosítások segítségével tudjuk az adott szegmens membránhoz viszonyított helyzetét jellemezni.

1.8.1.1 Fúziós fehérje konstrukciók

A topológiai adatszerzés egyik elterjedt módszere, hogy a vizsgálandó TM fehérjét kódoló DNS szakaszával fúziós fehérje konstrukciókat hoznak létre. A célfehérjére fuzionált enzim többek között lehet alkalikus foszfatáz (Alkaline phosphatase, PhoA)^{92,93}, β -galaktozidáz (β -galactosidase, LacZ)⁹⁴, β -laktamáz (β -lactamase, BLAM)⁹⁵. A fuzionált, riporterként használt enzim aktivitása eltér a membrán két oldalán, így következtetni lehet a célfehérje megfelelő részének (a fuzionált enzim szekvenciális helye a fehérjében) lokalizációjára (PhoA, BLAM periplazmatikus térben aktív, míg a β -galaktozidáz a citoszólikus térben). Szintén ide tartozó technika a fluoreszcens fehérjéssel történő fuzionálás, pl. zöld fluoreszcens fehérjét fuzionálnak az adott TM fehérje terminálására és annak membránhoz viszonyított helyzete antitesttel detektálható, membrán permeabilizált és nem permeabilizált körülmények között (az antitest normál körülmények között membránon nem tud áthatolni)⁹⁶. Az utóbbi opcionális módosítása, hogy a TM fehérje-fluoreszcens fehérje konstrukciót expresszáló sejteket fehérje degradáló enzimmel kezelik, amelyek antitestekhez hasonlóan membrán impermeabilisak, ezáltal a fuzionált fluorofor elveszítheti a fluoreszcenciáját és így meghatározható a terminális szubcelluláris lokalizációja (a módszer neve, fluoreszcencia proteáz védelem, Fluorescence protease protection, FPP)^{97,98}. További fejlesztése a fluoreszcens fehérje fúziós módszereknek, hogy redoxi-potenciál érzékeny fluoreszcens fehérje alkalmazásával a sejt permeabilizálása/roncsolása nélkül detektálható, hogy az adott fluoreszcens fehérje hol helyezkedik el, a membrán két oldala közötti elektromos potenciálkülönbség miatt (reduction-oxidation sensitive Green Fluorescent Protein, roGFP)⁹⁹.

A fúziós módszerek egy-egy célfehérjére is alkalmazhatók (lásd korábbiak), de akár teljes proteomokra is ki lehet terjeszteni azokat. Élesztőben a HA-Suc2-His4C (ahol a HA: az Influenza hemagglutinin epitóp, Suc2: az invertáz gén fragmentje, His4p hisztidinol dehidrogenáz katalitikus doménje) kiméria fehérjét fuzionálták 617 élesztő membránfehérje C-terminálisára. A terminálisok elhelyezkedését két módon vizsgálták. Először a kiméra fehérje egység glikozilációja alapján vizsgálták Endoglikozidáz H (endoglycosidase H, EndoH) enzim felhasználásával. A glikozilált fehérjékről a cukorláncok lehasadnak EndoH kezelés hatására, így a fehérje gélben futtatva gyorsabban vándorol, az enzimmel kezeletlen párjánál. Az eltolódás jelzi, hogy a kiméra glikozilálódott, azaz a kiméra elhelyezkedése lumenális/extra-citoszólikus, míg, ha nincs eltolódás a gélen, akkor citoszólikus volt. A másik vizsgálati mód hisztidin-mentes, de hisztinidolt tartalmazó lemezekben való növekedési teszt alkalmazása. A kiméra His4C része a citoplazmában aktív, így a hisztinidolt képesek felhasználni és tudnak növekedni a sejtek, viszont, ha extra-citoszólikus térbe kerül a domén, akkor sejtnövekedés nem figyelhető meg. Az élesztő TM fehérjék C-terminálisának membránhoz viszonyított helyzetének meghatározása, így egyféle fúziós egységgel, de kétféle kísérlettel történt¹⁰⁰. Az *Escherichia coli* belső membrán TM proteomját is fúziós fehérjékkel vizsgálták, 601 belső membránfehérje C-terminálisára fuzionáltak a membrán két oldalán ellentétes aktivitási profillal rendelkező fehérjéket. Alkalikus foszfátzt (aktivitása a terminális periplazmatikus helyzetét jelzi) vagy zöld fluoreszcens fehérjét helyeztek az adott TM fehérjék C-terminálisára (fluoreszcenciája a terminális citoplazmatikus lokalizációt jelzi)¹⁰¹.

Az ilyen jellegű topológiai adatok nem mindig egyértelműek, nehéz értelmezni a részleges enzimfunkciókat, valamint a célfehérjék méretéhez képest túl nagy módosítások, amelyek akár a fehérjeszerkezetére, ezáltal funkciójára is kihathatnak (különösen, ha nem csak a terminálisokra, hanem a célfehérje belső részeire helyezik őket). Az ABCG2 fehérje esete is megerősíti az előző állítást, a C-terminálisára helyezett fúziós fehérje konstrukció is problémát okozott: korlátozta a fehérje funkcióját és a fehérje intracelluláris retenciójára is hatással volt¹⁰².

1.8.1.2 Epitóp inzerció

A vizsgálandó fehérje méretével összemérhető fúziós fehérje konstrukciókhoz képest, ebben a fejezetben csak a néhány aminosav hosszúságú ismert antitest-epitóp célfehérje jelölési stratégiákról lesz szó. Az epitópok különböző fehérjékben előforduló lineáris vagy

konformációs antigén-determináns csoportok, amelyeket az immunrendszer fehérjéi képesek felismerni. Az ismert epitóp-célfehérje fúziós konstrukciók felhasználási lehetősége széles spektrumú: használhatóak a célfehérje tisztításához az epitópra specifikus antitestet hordozó affinitás oszlopok segítségével, a célfehérje Western blot módszerrel történő azonosításához vagy az epitóp membránhoz viszonyított elhelyezkedésének antitesttel való detektálása alapján célfehérje topológiai adatszerezésre is alkalmasak. Ilyen ismert epitópok:

- Influenza hemagglutinin (HA) epitóp, emberi influenza hemagglutinin, egy vírus felszíni glikoproteinjéből származó peptid szakasz, szekvenciája: YPYDVPDYA.
- Myc epitóp, c-myc gén termékéből származik, szekvenciája: EQKLISEEDL.
- FLAG epitóp mesterséges antigén, amelyhez nagy affinitású antitestet fejlesztettek, szekvenciája: DYKDDDDK, a felépítő aminosavak hidrofilitása miatt használják, mert így kevésbé denaturálhatják és ezáltal kevésbé inaktiválhatják a célfehérjét.

Az epitópokat kódoló nukleotid szekvenciát a célfehérjét kódoló DNS vizsgálni kívánt szakaszára (hurok régiók, terminálisok) építik be polimeráz lánreakció (Polymerase Chain Reaction, PCR) segítségével¹⁰³. A humán redukált folát karrier (RFC) fehérjénél HA epitóp inzerciót végeztek a becsült TMS-eket összekötő régiókba, valamint a fehérje terminálisokba. Az epitópok membránhoz viszonyított helyzetét, HA-epitóra specifikus rodamin-konjugált antitest jelöléssel vizsgálták permeabilizált, vagy nem permeabilizált körülmények között, fluoreszcens mikroszkóppal (ha permeabilizálás nélkül detektálták, akkor extra-citoszólikus, ellenkező esetben intracelluláris). Azokban az esetekben, ahol a fehérje funkció megmaradt, ott használható topológiai információt szolgáltatott, például a Q120-L121 aminosavak közötti epitóp (3/4 TMS közötti hurok) extracelláris, E226-L227 aminosavak közötti epitóp (6/7 TMS közötti hurok) citoszólikus stb¹⁰⁴. Hasonló elven a humán epesavtranszporter (ASBT) fehérje egyes szakaszait HA és FLAG epitóp inzercióval vizsgálták szaponinnal permeabilizált vagy nem permeabilizált sejteken¹⁰⁵, míg a humán szerves anion transzporter (OAT1) egyes szakaszainak elhelyezkedését Myc epitóp inzercióval vizsgálták¹⁰⁶. Az epitópok proteáz kezelés utáni antitesttel való detektálhatósága, egy további megközelítés az epitóp membránhoz viszonyított helyzetének meghatározására, így vizsgálták a Myc epitóppal jelölt humán Insig 1 fehérje terminálisait¹⁰⁷.

Ezek a célfehérje szekvenciájának hosszához képest relatív rövid módosítások, olykor szintén működés képtelen fehérjét eredményeztek, ezért még kisebb módosításokkal végzett módszereket is fejlesztettek topológiai adatszerzésre (1.8.1.3 és 1.8.1.4 fejezet).

1.8.1.3 Glikozilációs motívum vizsgálatán alapuló technikák

Az NXS/T motívumot az 1.4 fejezetben már megismertük. Az aszparagin oldallánc nitrogénjére helyeződhet a cukorlánc indító egység, ha a lineáris motívum az ER lumenében (extra-citoszólikus térben) helyezkedik el. Így ezen az aszparaginin bekövetkező glikoziláció szintén topológiai adatot szolgáltat, ami igazolja, hogy a célfehérje azon pozíciója extra-citoszólikus. A motívumot a fehérje kódoló szekvenciájának kívánt szakaszára megtervezett megfelelő NXS/T-t kódoló szakaszokkal 5' végen meghosszabbított PCR primerekkel végzik. Esetenként elég csak egy-egy nukleotid pontmutációval létrehozni a motívumot¹⁰⁴. A cukorlánccal ellátható pozíció membránhoz viszonyított helyzetét deglikozilálást végző enzimekkel vizsgálják, például Endo H enzimmel, ami a kitobióz (GlcNAc₂) magnál hasít^{100,108}, vagy PNGase F enzimmel, ami a legelső GlcNAc és az aszparagin között hasít¹⁰⁹. Továbbá a megfelelő aszparagin mutálása más aminosavra, szintén információt adhat annak glikoziláltságáról és ezáltal lokalizációjáról is^{110,111}. Bármely esetben, ha megváltozik a fehérje gélben mutatott vándorlási profilja, akkor arra lehet következtetni, hogy volt cukorlánc a fehérje aszparagin aminosaván.

Egy kísérletben az élesztő membránfehérjéinek C-terminálisára összesen 12db glikozilálható konszenzusmotívumot tartalmazó részt, valamint HA epitópot fuzionáltak. A mintákat Endo H enzimmel kezelték vagy nem kezelték, majd SDS-PAGE gélen megfuttatták. A HA epitópra specifikus antitesttel előhívott Western-blottal vizsgálták, hogy a futási profiljuk megegyezik-e, amiből lehet következtetni a glikoziláltságra¹⁰⁰, így jellemezni tudták a vizsgált élesztő TM fehérjék C-terminálisának membránhoz viszonyított helyzetét. Egy másik eljárás során a humán redukált folát karrier fehérjénél lokálisan egy-egy aminosav kodont mutáltak PCR-rel, hogy glikoziláláshoz szükséges motívum jöjjön létre, majd a már ismertetett módszerrel vizsgálták azok membránhoz viszonyított helyzetét¹⁰⁴.

Bár a glikoziláláshoz szükséges motívum létrehozása vagy elrontása kisebb módosítást jelent, mint az epitóp inszerció vagy fehérje fúzió, de ezekben a kísérletekben is előfordulhat, hogy az adott fehérje nem fejeződik ki vagy nem működőképes. A Western-blottal való előhíváshoz pedig mindig szükség van az adott fehérje elleni, vagy a fuzionált epitópra specifikus antitestre.

1.8.1.4 Aminosavak kémiai módosításán alapuló eljárások

Az 1.8.1.3 fejezetben is előfordult olyan konstrukció, amit egy aminosav módosításával elő lehetett állítani. Ebben a fejezetben olyan kísérletekről lesz szó, ahol mindig egy-egy aminosavat módosítanak egy TM fehérje megfelelő szakaszában. A vizsgálni kívánt aminosavat egy reaktív oldalláncot tartalmazó aminosavra cserélik, amelyekre kaphatók specifikus membrán permeábilis és membrán impermeábilis kémiai jelölő ágensek. A kémiai jelölők membrán permeabilitási tulajdonsága alapján tudják későbbiekben a vizsgált aminosav fehérjén belüli pozíciójának membránhoz viszonyított helyzetét jellemezni. Ezen kísérletek nagy hátránya, hogy mielőtt a célfehérje megfelelő pozícióiba bevezetnék a módosítandó aminosavat, a célfehérjéből létre kell hozni a módosítandó aminosavat nem tartalmazó konstrukciót, és ellenőrizni kell a módosított fehérje működését^{56,106,107,112}.

Az Insig-1 endoplazmatikus retikulum TM fehérjében egyedi ciszteint tartalmazó szekvenciákat akartak létrehozni, így először az összes endogén ciszteint mutálták szerinre, majd egyesével ciszteinre mutálták a vizsgálandó aminosavakat. Az egyedi cisztein tartalmú konstrukciókat HEK-293 sejtekben expresszálták, majd metántioszulfonát bromid (methanethiosulfonate bromide, MTSET), membrán impermeábilis szulfhidril reagenssel kezelték a sejteket, végül pedig hozzáadták a biotin-maleimid membrán permeábilis szulfhidril reagenst. Ezután azt vizsgálták, hogy a fehérje dúsítható-e avidin gyöngyön (biotin-kötésre specifikus oszlop), aminek eredményéről Western-blottal győződtek meg. Azok az egy-egy darab ciszteint tartalmazó fehérjék, amelyek biotinálása nem történt meg, így az avidin gyöngyön sem tudták dúsítani az MTSET blokkoló alkalmazásakor, csak a blokkoló nélküli kísérletben, a módosítással bevitt ciszteint citoszólikus térrészben tartalmazták. Azon cisztein tartalmú fehérjék, amelyek biotinálása mind blokkolóval, mind blokkoló nélkül megtörtént és így a gyöngyön dúsítható volt a fehérje, arról tanúskodott, hogy a bevitt cisztein pozíciója extra-citoszólikus. Számos egyedi cisztein tartalmú konstrukcióról, azonban nem lehetett ezzel a kísérlettel megállapítani, hogy hol lokalizálódott a ciszteinre módosított aminosav¹⁰⁷.

A humán szerves anion transzporter (hOAT1) esetében az egyetlen extra-citoszólikus térbe jósolt lizin aminosavat mutálták (K59-et) argininná, majd bizonyos aminosavakat lizinre módosítottak. Az egyes konstrukciók egyedi lizin aminosavának elhelyezkedését membrán impermeábilis szulfoszukcinimidil-2-(biotinamido)etil-1,3-ditiopropionát (Sulfosuccinimidyl-2(biotinamido)ethyl-1,3-dithiopropionate, Szulfo-NHS-SS-biotin) reagenssel tesztelték, avidin agaróz gyöngyön való dúsíthatóságát Western-blottal vizsgálták. Egyes konstrukciók

esetében megállapították, hogy nem hozzáférhető a reagens számára, így azok pontos membránhoz viszonyított helyzetét nem tudták meghatározni. Más esetekben nem működő fehérjét kaptak az adott aminosav mutálása által, így azokat nem vizsgálták tovább. Azonban 4 darab olyan konstrukciót tudtak létrehozni, amelyekben a módosított aminosav pozíciójának extra-citoszólikus lokalizációját sikerült megerősíteni¹⁰⁶.

1.8.2 A kódoló DNS szekvencia megváltoztatása nélküli technikák

Az 1.8.1 fejezetben felsorolt technikák, a kódoló szekvencia módosítása miatt, kihathatnak a célfehérje szerkezetére, funkciójára vagy a sejten belüli elhelyezkedésére, valamint sokszor nem egyértelmű vagy ellentmondó eredményt adnak, így további módszerek fejlesztésére volt szükség. A következőkben olyan technikákról lesz szó, ahol a vizsgálandó fehérjék kódoló DNS-ét nem módosítjuk. Ezen technikák közé tartozik a különböző proteázokkal való részleges emésztés által keletkezett fragmentek vizsgálata, vagy a célfehérje ellen termeltetett antitest kötődésének vizsgálata.

1.8.2.1 Endogén epitóp helyzetének vizsgálata

A célszekvencia módosítása nélküli technikák között szerepel az endogén epitóp ellen termeltetett antitest, amelynek topológiai térképezésre való felhasználása a korábbi epitóp inzerciós technikához hasonlóan történik. Az antitest fiziológias körülmények között nem membrán permeábilis, így csak a sejt felszínén lévő endogén epitópot képes felismerni. A detektálása többnyire fluoreszcens festékkel jelölt másodlagos antitest felhasználásával történik, ami nem a vizsgált fehérje epitópjára, hanem az ahhoz kötődő antitestre specifikus. Így kiküszöbölhető, hogy a jelzőmolekula esetleg befolyásolja az elsődleges antitest specificitását. A fluoreszcens jel feldolgozása fluoreszcencia-aktivált sejtválogatásos (Fluorescence-activated cell sorting, FACS) technikával, illetve fluoreszcens, konfokális, esetleg szuperrezolúciós mikroszkóppal történik.

Egy búza TM fehérje 6 darab endogén epitópjára ellen nyulakban termeltettek antitesteket. A fehérjét 293T sejtekben kifejeztették, majd Triton X-100-zal permeabilizált illetve nem permeabilizált körülmények között inkubálták az adott epitópra specifikus nyúl antitesttel. Végül Alexa Fluor 594 festékkel (AF594)-konjugált kecske anti-nyúl IgG antitestet használtak az elsődleges antitest előhívására, az AF594 jelző molekulát fluoreszcens mikroszkóppal detektálták¹¹³. Fontos megjegyezni, hogy a detergens hatására extra-

citoszólíkus térben natív körülmények között nem hozzáférhető epitóp is elérhetővé válhat az antitest számára, így a deterges használatával kapott eredmények nem mindig helyesek.

1.8.2.2 Parciális proteolízis vizsgálatok

Egy másik elterjedt technika a parciális proteolízis, amely során a célfehérjéket valamilyen sejtvonalban kifejeztetik és a sejteket specifikus vagy kevésbé specifikus hasítási hellyel rendelkező fehérje bontó enzimmel kezelik. A hasítás után kapott fragmenteket gélben megfuttatják, majd méretük az ismert tömegű marker molekulák felhasználásával számítható, majd az alkalmazott enzim hasítási specificitásának ismeretében következtethetünk, hogy hol mehetett végbe a hasítás a fehérjében. Az alkalmazott enzimek fiziológias körülmények között nem tudnak keresztülhatolni a membránon, így, ha sejt felszíni proteáz kezelést végeznek egyértelmű, hogy a fehérje megtalált hasító helye extra-citoszólíkus térben hozzáférhető volt. Egyes esetekben „inside-out” (kifordított) vezikulán végzik a kísérletet, így citoszólíkus hasítási helyeket térképezhetnek¹¹⁴. Homogén kifordított vezikula-populáció előállítása azonban nagy kihívás, a kontaminációk miatt kísérlet pontossága sérül.

A „antiproliferatív antitest célpont” (the target of an antiproliferative antibody, TAPA-1) fehérje topológiáját például sejtmentes transzlációs rendszerben vizsgálták mikroszóma frakcióval kiegészítve. Tripszin és proteináz K fehérje bontó enzimeket használtak, a fehérje gélen való futtatási profiljában jelentős eltérést figyeltek meg, ha az emésztés során triton X-100 permeabilizáló szert is alkalmaztak. Ez a feltételezett TMS-eket összekötő nagy hurkok lumenális/extra-citoszólíkus lokalizációját igazolja és indirekt módon a terminálisok citoszólíkus elhelyezkedését is¹¹⁵. Az interferon-indukálható transzmembrán fehérje család (interferon-inducible transmembrane proteins, IFITM) egyik tagjának C-terminálisára HA epitópot fuzionáltak, az ezeket expresszáló sejteket 5- 30 perces tripszines kezelésnek vetették alá. Gélen való megfuttatás után Western-blottal, a megfelelő specifikus antitesteket alkalmazva, vizsgálták egyrészt a fehérje, másrészt a HA epitóp jelenlétét. A fehérjére jellemző blot sávban, a hosszabb idejű enzim kezelés a HA epitóp detektálhatóságát csökkentette, tehát valószínűleg a fehérje C-terminálisán lévő „R” vagy „K” aminosavaknál hasítás következhetett be, ami fehérje C-terminálisa extra-citoszólíkus lokalizációjára utal. A HA epitóp detektálhatóságának csökkenését áramlási citometriával (FACS) is megerősítették¹¹⁶.

1.8.2.3 Glikoziláció azonosítása egyedi fehérjéken és teljes proteomokon

A glikozilációt nem csak mesterségesen létrehozott glikozilációs motívumot hordozó fehérjékben vizsgálják, hanem natív célfehérjékben is, az 1.8.1.3 fejezetnek megfelelően. Ha a vizsgált fehérje pontosan egy potenciális NXS/T motívumot tartalmaz, akkor a megfelelő cukorlánc rész eltávolítható a glikozidáz enzimek segítségével (PNGase F, Endo H), vagy glikozilációt blokkoló szerekkel (pl tunikamicin¹¹⁰) gátolható a kialakulásuk. Ezt követően a fehérje gélen való futtatásából következtetni lehet a fehérje glikozilációs állapotára és így a topológiára.

Az újabb tanulmányokban egyszerre több natív állapotú fehérje glikozilációját vizsgálják, komplex mintaelőkészítéssel, lektin affinitás kromatográfiával szelektíven dúsítva az adott glikocsoportokra. Többnyire még a dúsítás előtt eleméztik specifikus fehérje degradáló enzimmel/enzimekkel a komplex biológiai mintát és csak a glikocsoportokat tartalmazó fehérje peptideket dúsítják fel. A peptidek szekvenciáját tandem MS (MS/MS) technikával határozták meg, valamint a glikoziláció peptiden belüli pontos helyét és a peptiden lévő cukorláncok összetételét is azonosíthatták^{117,118}. Ezzel a technikával, valamint ennek alternatíváival, mind az O- glikozilációs¹¹⁹⁻¹²¹, mind az N-glikozilációs helyeket^{122,123} lehet vizsgálni.

A glikoproteomok részletes elemzésének a fő akadályá sokáig az volt, hogy nagy a heterogenitás mind a módosítási helyekben (az O-glikozilációnak nincs konszenzus motívuma), mind pedig az azokra kapcsolt glikán szerkezetekben. Az O-glikoziláció vizsgálatához egy tanulmányban a glikán szerkezet heterogenitását csökkentették, úgyhogy cink-ujj-nukleáz felhasználásával meggátolták az O-glikán szerkezeteket hosszabbító útvonalat. Így úgynevezett „SimpleCell” sejtvonalakat hoztak létre, amelyek homogén O-glikán szerkezetekkel rendelkeznek. A sejteket lizálták, emésztették, a megfelelő glikopeptideket lektin affinitás oszlopon kötötték ki, amelyeket végül tandem MS készülékkel vizsgáltak elektron-transzfer disszociációs (Electron-Transfer Dissociation, ETD) fragmentáció segítségével. Steentoft és mtsai 3 sejtvonalra több mint 100 O-glikoproteint találtak, több mint 350 O-glikán hellyel¹²⁰ ezzel a módszerrel. Később továbbfejlesztették a módszert szekretált fehérjék vizsgálatával, valamint MS analízis előtt az izolált peptideket izo-elektromos fókuszálásnak is alávetették, továbbá alternatív proteázt (kimotripszin) is használtak, így kibővítették a 3 sejtvonalra talált fehérjék és glikozilált peptidek listáját. Összesen 259 glikoproteint és 856 glikozilációs helyet azonosítottak¹²¹. A folyamatot 12 sejtvonalon elvégezve már közel 3000 glikozilációs helyet találtak, több mint 600

fehérjében¹¹⁹. Egy további tanulmányban a humán májból glikoproteineket dúsítottak HILIC és hidrazid oszlopon is, majd PNGaseF-fel eluálták az N-glikopeptideket és tandem MS készülékkal elemezték a mintákat. Ezzel a módszerrel összesen 2210 N-glikoprotein találtak 4783 N-glikozilációs helyet¹²². Végül 41 humán és 31 egér sejttípus N-glikopeptideit vizsgálva létrehoztak MS analízisből származtatott sejtfelszíni fehérje atlaszt is¹²⁴.

A fehérje halmazokban jelentős mennyiségű TM fehérjén azonosítottak a biológiailag jelentős N- és O-glikozilációs pozíciókat.

1.8.2.4 Aminosav specifikus jelölési stratégiák

A natív TM fehérje topológia vizsgálati technikák között megkülönböztethetünk még egy csoportot, amik az aminosav specifikus kémiai jelölő reagensekre épülnek. Az ezekkel történő módosításokat, az 1.8.2.3 fejezetben is alkalmazott tandem MS-sel mutathatjuk ki¹²⁵. Elsődlegesen az olyan reaktív oldalláncú aminosavakra fejlesztettek reagenseket, mint a cisztein szulfhidril csoportja, vagy mint a lizin primer amin csoportja. A szulfhidril csoportra specifikus reaktív csoport például a maleimid, piridildiszulfid, míg a primer amin csoporttal szobahőmérsékleten könnyen reagáló az N-hidroxi-szukcinimid (NHS)-észter, imido-észterek vagy 1,5-difluór-2,4-dinitrobenzén (DFDNB). A reaktív csoportok különböző távtartó karokkal és más dúsításra alkalmas egységgel ellátott reagensek szerves részei, amelyek eltérő membránpermeabilitási tulajdonságokkal rendelkeznek. A Szulfo-NHS-SS-biotin a dolgozatban leírt kísérletek fő komponense, amely tartalmaz egy primer amin specifikus NHS észtert, valamint egy avidin agarózzal töltött oszlopon való dúsításra alkalmas biotin részt. A két egység között egy 24,3Å hosszú távtartó kar helyezkedik el, a benne található diszulfid-híd redukálószerrel hasítható (pl.: (1,4-bisz-(szulfanil)-bután-2,3-diol) (Dithiothreitol, DTT) vagy trisz(2-karboxietil)foszfin-hidroklorid (tris(2-carboxyethyl)phosphine, TCEP), amely megkönnyíti az oszlopról való eluálást. Továbbá a reagens membrán impermeabilis tulajdonságú.

A nem csak egy aminosavra specifikus csoporttal rendelkező reagensek között léteznek homobifunkcionálisak (két ugyanolyan reaktív részt tartalmazó keresztkötő ágensek) és heterobifunkcionálisak (két eltérő), valamint trifunkcionálisak is (három különböző, vagy két ugyanolyan és egy eltérő). Az utóbbiak segítségével fehérjéken intra- és inter-molekuláris keresztkötéseket lehet végezni. A keresztkötések helyének azonosításával lehet következtetni arra, hogy az adott fehérjében mely aminosavak helyezkednek közel egymáshoz, hogyan nézhet ki a fehérje kölcsönható felszíne komplexben. Bár ezek a technikák limitált

felbontásúak, viszont tranziens kölcsönhatásokat is rögzíthetnek, így jó kiegészítői az 1.6 fejezetben tárgyalt szerkezet-meghatározásoknak, továbbá távolsági megszorításként használhatók a fehérjék modellezése során^{126–129}.

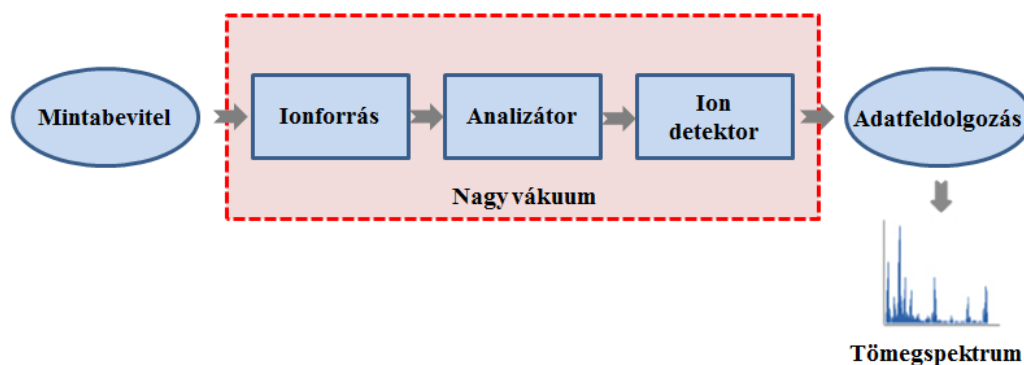
A fenti módosítások MS segítségével azonosítva szintén szolgáltathatnak topológiai adatot. A membrán impermeábilis, aminosav specifikus jelölőkkel csak a membrán egyik oldalán jelen lévő és hozzáférhető aminosavakat módosítják a kémiai reagenssel. Számos, úgynevezett sejtfelszíni fehérje izolálási technikánál használnak ilyen jelölőket, az ezeket leíró tanulmányok többségében a teljes fehérjét izolálják, mint például endotél sejtek¹³⁰, sztrómális őssejtek¹³¹, vagy a metasztázisban szerepet játszó melanóma sejtek¹³² felszíni proteomjának analízise során. Ezek munkák azt azonban kevésbé vizsgálják, hogy egy fehérjét, mely aminosav pozíciókon keresztül dúsítottak a vizsgált sejttípus felszínéről. Egyes tanulmányokban azonban a proteomikai peptid azonosításnál keresik az aminosav specifikus mesterséges módosításokat, például a hepatocita sejtvonalak (Lys-cell surface capture technology, Lys-CSC technika)¹³³, a Hodgkin és non-Hodgkin limfómák¹³⁴ vizsgálata során, valamint egy plazmamembrán fehérjék eltérő dúsítási módszertanát összevető tanulmányban is¹³⁵.

A korábbi tanulmányokban azonban nem optimalizálták a kísérleteket arra, hogy egy felszíni jelölésből minél több TM fehérjéről szerezzenek extra-citoszólikus topológiai adatot, így érdekesnek találtuk ennek a lehetőségnek a megvizsgálását (4.4 fejezet).

1.9 Fehérjék és peptidek szekvenálása tömegspektrometriával

1.9.1 Tömegspektrometriáról általánosságban

A tömegspektrometria egy olyan nagyműszeres analitikai módszer, ahol a vizsgálandó molekulát ionizáljuk, majd az egységnyi töltésre eső tömeget meghatározzuk. Kis molekulák és makromolekulák minőségi és mennyiségi meghatározására elterjedt módszer, nagyfokú érzékenysége és felbontóképessége miatt. A felbontóképesség (R) megadja, hogy két ion mekkora tömegkülönbség esetén tudunk egymástól megkülönböztetve detektálni (korszerű készülékekkel már akár 1 ppm (milliomod rész) felbontás is elérhető). Egy általános MS készülék mintabeviteli egységből, ionforrásból, tömeg analizátorból, detektorból, az előző háromban vákuumot biztosító rendszerből (az ionforrás olykor lehet légköri nyomáson is (Atmospheric Pressure Ionization, API) és végül a detektorban érzékelt jeleket feldolgozó egységből áll (2. ábra).



2. ábra: Tömegspektrométer általános felépítése

A mintabeviteli egység egyszerű minták esetén közvetlen, míg komplex minták esetén egy elválasztás-technikai elem van beépítve az ionforrás elé, például gázkromatográfia (gas chromatography, GC) vagy nagy hatékonyságú folyadékkromatográfia (High-performance liquid chromatography, HPLC). Az ionforrás segítségével állítják elő a vizsgálandó anyagból a gázhalmazállapotú molekulaiont, míg az analizátor egységen választják szét az ionokat tömeg/töltés (m/z) szerint térben és/vagy időben is, a detektorral pedig kvantálni lehet az adott ion mennyiségét. Az MS belső részeiben azért szükséges a légritkított tér, mert az ionok előállítására általában használt elektronok, valamint a képződő ionok nagyon reaktívak, ezért csökkenteni kell az ütközések valószínűségét. Végül a feldolgozó egység a detektorba beérkező jelekből spektrumot vesz fel, amely aztán tovább értékelhető. A tandem MS készülékekkel a kiválasztott molekulaion (prekursor ion) fragmentálható és az abból származó úgynevezett leányionok szintén meghatározhatók, amelyből akár vizsgált molekula szerkezetére is lehet következtetni¹³⁶ (1.9.2-1.9.4 fejezet).

1.9.2 Biomolekulák tömegspektrometriai vizsgálata

A biomolekulák vizsgálatát a lágy ionizációs technikák tették lehetővé, úgymint a mátrix segített lézer deszorpció/ionizáció (Matrix-Assisted Laser Desorption Ionization, MALDI) vagy electrospray ionizáció (Electrospray Ionization, ESI), amelyek előnye, hogy a vizsgált molekula alapváza nem vagy kevésbé fragmentálódik, helyette kvázi molekulaionok képződnek, $[MH_n]^{n+}$ formában (M: molekulaion, n: töltésszám). A többszörös töltések felvétele által a nagyméretű biomolekulák (nukleinsavak, fehérjék) is vizsgálhatóak, a molekulatömegükhöz képest kisebb m/z értéknél mutatkoznak. Az ESI kapcsolható közvetlenül kromatográfiával, így a komplex biológiai mintákat sok esetben HPLC-ESI-MSⁿ-

nel vizsgálják (n általában 2 vagy 3)^{137,138}. A dolgozat mintáit is ilyen készülékekkel analizáltuk, így a továbbiakban ennek alkotóelemeit tárgyalom.

A folyadékkromatográfia (Liquid Chromatography, LC) olyan elválasztási technika, melyben a mozgófázis folyadék, az állófázis pedig szilárd. A modern készülékeknél az állófázisnak olyan kicsi a szemcsemérete a nagy fajlagos felület eléréséhez (ami befolyásolja a vizsgálandó molekula állófázison való tartózkodási idejét), hogy emiatt nagy nyomáson (High Pressure - HP) kell az oldószert átpumpálni rajta. A komponensek elválasztásának alapja a komponensek megoszlási hányadosainak különbözősége. A komponensek eltérő ideig tartózkodnak az állófázisban, így elválnak. Az elválasztási célra használt HPLC vagy ultranagy hatékonyságú folyadékkromatográfiás (Ultra Performance Liquid Chromatography, UPLC) oszlopokat hosszúsággal (mm), belső oszlop átmérővel (mm), részecske töltet átmérőjével (mikrométer), valamint pórusmérettel jellemzik (Å). A vizsgálandó minták a megfelelő összetételű folyadékfázisban áramoltatva haladnak keresztül a kromatográfiás oszlopon.

A HPLC-ről a minta egyes elemei az ionforrásba jutnak elektroporlasztásos ionizáció során, amelyben folyadékkromatográfival szeparált folyadék fázisban lévő minta egy nagyfeszültségre kapcsolt kapillárison halad keresztül. A folyamat alatt a töltéssel rendelkező ionok az oldat felszínén felhalmozódnak, és a töltött részecskék közötti taszítás, valamint a potenciálgradiens hatására a kapillárisból kilépve kúp alakot vesz fel (Taylor kúp). Az ellenelektrod és a kúp között kialakult elektrosztatikus tér hatására töltött cseppek fűződnek le és haladnak az analizátorba vezető belépő rész felé. Közben a szárító/porlasztó gáz (N₂) és a hő hatására a lefűződő cseppek bepárolódnak, mely során a folyadékcseppek töltéssűrűsége addig tud növekedni, amíg a folyadék felületi feszültségét meghaladja az ionok taszítási ereje, és ezáltal a cseppek felrobbannak (Coulomb-robbanás). A vizsgált molekulákból ezen folyamat által lesz egyszeresen, vagy többszörösen töltött gáz fázisú molekulaion^{138,139}.

A molekulaion ezután kerül az analizátorba, amiben a tömeg/töltés (m/z) arány alapján szortírozhatjuk, illetve fragmentálás után a fragmentumokat hasonló analizátorokban elemezhetjük. A dolgozat szempontjából fontos analizátor típusok:

- kvadrupól-analizátorok: az analizátorba belépő ionok négy párhuzamos fémrúd között elektromos mező segítségével valamilyen – az m/z arány által meghatározott – trajektóriára állnak és így közülük csak egy adott m/z arányhoz tartozó ionok jutnak a kilépőreszre, így csak ezek hagyhatják el az

analizátort. Ha az m/z arányt meghatározó analizátor-paramétereket változtatjuk, a teljes vizsgált tömegszám-tartományt végigpásztázhatjuk (scannelünk).

- a repülési idő-analizátorok (Time of Flight, TOF) működése azon alapul, hogy azonos gyorsítótér esetén az ionok sebességét egyértelműen meghatározza m/z tulajdonságuk. Ezért az azonos út befutásához szükséges időből az m/z meghatározható.
- ioncsapda analizátorok: általánosságban egy belépő és egy kilépő és a köztük elhelyezkedő gyűrű-elektrodából állnak, utóbbira egyen-és váltófeszültséget kapcsolva az ionok oszcillálni kezdenek. Az m/z arány szerinti sorrendben engedik ki őket az analizátor belső teréből a detektor felé. Ide tartozik a nagyfelbontású Orbitrap elektrosztatikus ioncsapda, amelyben két hordó alakú elektróda között található egy orsó alakú elektróda, amely körül spirálisan keringenek (m/z -hez köthető körpályán) az ionok¹⁴⁰.

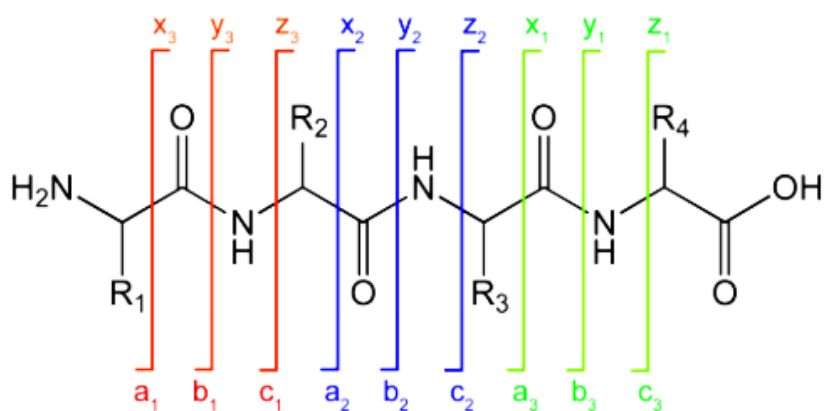
A dolgozatban használt két készülék egy Orbitrap Elite Hybrid készülék, amelyben egy érzékeny, de kis felbontású ioncsapda és egy nagy pontosságú és felbontású Orbitrap analizátor lett fuzionálva, valamint Bruker Maxis II ETD Q-TOF készülék, amelyben pedig egy kvadrupol és egy repülési idő analizátor helyezkedik el. A készülékekben lehetőség van a vizsgált molekula ionok fragmentálására és keletkezett fragmentek m/z szerinti elválasztására. Mindkét készülékben ütközés által kiváltott disszociációt (Collision-Induced Dissociation, CID) alkalmaztunk, valamint az Orbitrap készülékben egy szintén CID alapú módszert, a nagyenergiájú ütközéses fragmentációt (Higher-energy collisional dissociation, HCD). CID esetén inert ütközési gáz alkalmazásával fragmentálunk (pl.: argon), a fragment ionokból a biomolekula pontos szerkezetére következtethetünk. A detektorba érkező prekursor és fragmens ionoknak az m/z értékük és relatív intenzitásuk alapján létrejönnek a tömegspektrumok.

1.9.3 Tömegspektrometria a proteomikában, a peptidok azonosítása és szekvenálása

A proteom alatt egy adott élőlény által szintetizált összes fehérjét értjük. A korszerű MS módszerek jelentős szerepet töltenek be a komplex fehérje elegyek vizsgálatában, valamint ezek posztranszlációs módosulásainak (pl.: foszforiláció, acetiláció, metiláció, glikoziláció) azonosításában. A tömegspektrometrián alapuló fehérje azonosításnak két fő

iránya van: a „bottom up” és a „top down” technika. Míg az előbbi a fehérjék proteolitikus hasítása után keletkező peptid fragmentumokat vizsgálja, addig az utóbbi a teljes fehérjét¹⁴¹.

A dolgozat *bottom up* proteomikára épül, amelyben a fehérjeminták izolálás után gélben vagy oldatban proteáz emésztésen mennek keresztül az MS analízis előtt. A keletkezett peptideket folyadék kromatográfiával szeparálják, amely on-line kapcsolható ESI ionforrással ellátott MS készülékhez. A keletkezett spektrumok azonosítása (peptidhez ill. fehérjéhez társítása) kétféle módon történhet. Az első a „peptide mass fingerprint” (PMF, peptid ujjlenyomat), a másik a tandem MS módszer. A PMF esetén a mintából nyert peptidek tömegét hasonlítjuk össze, az adott adatbázisban lévő fehérjék szekvenciáinak a kísérletünkben használt proteázzal *in silico* emésztett peptideinek tömegértékeihez. A tandem MS módszer kevésbé komplex mintáknál használatos, egy fehérje jelenlétének igazolásához több hozzátartozó PMF-et is meg kell találnunk. A másik megközelítésben nem csak a prekursor peptid ion tömegét tudjuk meg, hanem az ütközési cellában fragmentáljuk a peptideket. A fragmentumokból lehet következtetni a peptid aminosav-sorrendjére, azaz a peptid szekvenálása is megtörténik. Az ütközési cellában leggyakrabban használt CID fragmentálás esetén, ha a töltés a peptid N-terminálisán marad, akkor a peptidkötés hasadási helyétől függően „a” és „b” fragmens ionokat kapunk, a C-terminálison visszamaradó töltés esetén pedig „y” fragmentumokat^{142,143}. A fragmentumokat a megfelelő terminustól számozzuk. A 3. ábrán az előzőkön kívül további jellemző fragmentumok és jelölésük is látható. Az adatbázisokban lévő fehérjékből elméletileg keletkező peptidek, elméleti fragmentálódási mintázata összevethető a gyakorlatban tapasztalt MS/MS spektrummal, jó egyezés esetén a peptid azonosítható¹⁴⁴.



3. ábra: Az MS/MS spektrumok esetében megfigyelhető lehetséges fragmentumok és azok elnevezése, attól függően, hogy hol marad a töltés.

1.9.4 Tömegspektrometriai spektrumok elemzése, feldolgozása

A vizsgált peptidek prekursorionjairól és a hozzájuk tartozó fragmentumok ionjairól csúcslisták készülnek. A csúcslistákat a nyers fájllokból programokkal exportálják, különböző lekereső algoritmusok (mint a Mascot, Protein Prospector stb.) felhasználják ezt, mint a mintára jellemző gyakorlati spektrum lista^{145,146}.

A gyakorlati spektrumok azonosításához a lekereső algoritmusok a mintában lévő élőlény ismert fehérjeszekvenciáit használják fel a fehérjeszekvencia adatbázisokból (pl.: UniProt¹⁴⁷). A kísérlet elvégzéséhez használt emésztő enzimmel *in silico* emésztik a fehérjeszekvenciákat. A kísérletek során elképzelhető, hogy nem minden lehetséges hasítóhelyen történik emésztés, ezért megadható, hogy maximum hány hasítóhelyet hagyhat ki az *in silico* emésztés, így hosszabb peptidek is lekereshetők. Beállítható, hogy a lekeresendő peptidek terminálisai teljesen, vagy csak szemispecifikusak legyenek a használt emésztő enzimhez mérten (utóbbi esetében elég, ha csupán az egyik peptid terminális keletkezett az adott enzim hasítása által). Továbbá megadhatjuk a már felfedezett természetes módosulásokat is az adott aminosavakon opcionális módosulásként (ez a lekeresésnél lehetőséget ad az adott aminosavat tartalmazó peptidek módosult és módosulatlan változatának keresésére is). Ide tartozik például a szerin, treonin, tirozin foszforiláció (ha foszforilációs helyeket keresünk), vagy metionin oxidáció. Végül megadhatunk mesterségesen létrehozott módosításokat is, mint a cisztein szulfhidril csoportjainak alkilálása, általában fix módosításként (így csak az adott aminosavat tartalmazó peptidek alkilált változatával történik a keresés).

Általánosítva az előző bekezdésben felsorolt beállításokkal jön létre a használt adatbázis alapján egy elméleti peptid lista, amelyre jellemző prekursor- és leányionok (bizonyos ppm toleranciával) spektrumai keresendők a gyakorlati spektrumok között. Minden lekereső algoritmus másféle módon pontozza a találatokat, és kiválasztja a megbízhatónak ítélteteket. A szoftverek ún. E-value értéket számolnak minden peptidre, amelyik az adatbázis méretének, illetve egyéb paraméterek értékeinek felhasználásával megadja, hogy egy, legalább olyan jó pontszámmal rendelkező találat, mint a vizsgált, mennyire lehet a véletlen eredménye. Egy – a beállított konfidenciaintervallumnak megfelelő – küszöbérték meghúzásával végül megtörténik a megbízhatónak ítélt peptid találatok kiszűrése a listából. Az egymástól függetlenül azonosított peptideket végül fehérjékhez rendelik.

Könnyen belátható, hogy a kereséshez használt adatbázis méretének növelése, valamint az abban lévő szekvenciákon minél több természetes és mesterséges módosítás, valamint aspecifikus hasítás megengedése, az elméleti peptidlista jelentős bővülését eredményezi, ami növelheti a hamis pozitív, azaz véletlenszerű azonosítások lehetőségét. A véletlen találatok számának becslésére általában a „target-decoy” módszert használják, azaz randomizált (általában a megfordított fehérjeszekvenciákból képzett) peptidekkel is történik lekeresés. Ez megmutatja, hogy hány olyan peptidet talál meg az algoritmus a mintában, amelyik biztosan nincsen benne. Ebből becsülhető a hamis találati arány (decoy találatok száma/target találatok száma= false discovery rate, FDR)¹⁴⁸.

2 Célkitűzések

A TM fehérjék számos fontos biológiai folyamatban szerepet játszanak az élő sejtekben, ennek ellenére a meglévő TM szerkezetek száma rendkívül alacsony, mivel a 3D szerkezet-meghatározását lehetővé tevő technikák limitáltan és esetenként hosszú optimalizációs folyamatok után használhatók ezen fehérjékre. A szerkezeti tér jelentős része jelenleg még nem ismert, ezért alternatív topológia meghatározó elméleti és kísérleti technikák terjedtek el a TM fehérjék jellemzésére.

A bionformatikai megközelítésen alapuló topológia becslések univerzálisan használhatóak az összes TM fehérje jellemzésére, azonban egyes esetekben eltérő eredményeket szolgáltatnak a meglévő kísérleti eredményekkel. Léteznek azonban olyan becslő módszerek, amelyek képesek figyelembe venni a kísérleti topológiai megfigyeléseket, ezáltal pontosabb becslést szolgáltatni. Így az elsődleges célja a munkámnak:

- A tudományos irodalomban olyan kísérleti eljárások összegyűjtése, amelyek alkalmasak arra, hogy topológiai adatot szolgáltatassanak a TM fehérjéről (fúziós fehérjék, poszt-transzlációs módosítások, immuno-lokalizáció, proteáz-védelem stb.).
- Az általuk szolgáltatott adatok rendszerezése és a csoport korábbi topológiai adatokat összegyűjtő TOPDB adatbázisának frissítése.

A megismert topológia meghatározásához használható kísérleti eljárások sokszor csak egy-egy TM fehérje topológiájáról szolgáltatnak limitált információt, kivitelezésük időigényes, olykor a fehérje lokalizációját, funkcióját is megváltoztatják. Az egyes módszerek által hozott eredmények ráadásul nem mindig egyértelműek. A humán TM fehérjék több mint 50%-áról egyáltalán nem találtunk topológiai adatot a TOPDB adatbázis frissítése során. Így további célokat tűztem ki:

- Egy olyan kísérleti módszer kifejlesztését (továbbiakban: MSTOP), amely gyors és egyszerre akár több száz TM fehérje topológiájáról képes információt adni. Ezen belül:

- A módszerfejlesztés kezdeti szakaszában vörösvértestek felszíni fehérjéin a kísérlet egyes paramétereinek beállítását
- Az MSTOP módszer egyes szakaszainak kontroll kísérleteinek kidolgozását és alkalmazását
- A kísérlet egyes lépéseinek optimalizálását, úgy, hogy a kidolgozandó módszer minél több TM fehérjét jellemezzen
- Az MSTOP módszer kiterjesztését sejtvonalakra (HL60, K562)
- Az azonosított jelölt lizin pozíciók szekvenciális környezetének elemzését
- Az MSTOP módszer által szolgáltatott topológiai adatok pontosságának jellemzését (összehasonlítva a korábbi kísérleti adatokkal)
- A módszer által kapott adatoknak a topológia becslés pontosságára gyakorolt hatásának meghatározását

A végső cél, hogy a generált adatokkal a lehető legtöbb TM fehérjére pontosabb topológia álljon rendelkezésünkre. A pontosabb topológia ismerete megkönnyítheti a TM fehérjék 3D modellezését és így számos más kutatási területre segítséget fog nyújtani. Így többek között a hozzáférhető fehérjerégiók pontosabb meghatározása által az új kismolekulás hatóanyag-fejlesztéseket vagy a TM fehérjére specifikus antitest tervezéseket is segítheti.

3 Anyagok és módszerek

3.1 Anyagok

3.1.1 Felhasznált sejtek/sejtkultúrák

A dolgozatban bemutatott munkák során a következő sejteket és az Amerikai Sejtgyűjtemény (American Type Culture Collection, ATCC) egyes sejtkultúráit használtuk:

Vörösvértest (RBC), CHO-K1- kínai hörcsög petefészek (Chinese Hamster Ovary, CHO, ATCC[®] CCL-61TM), HL60 sejtvonal (ATCC[®] CCL-240TM), K562 sejtvonal (ATCC[®] CCL-243TM).

3.1.2 Felhasznált anyagok

A dolgozatban felhasznált anyagok 7.1-es fejezetben találhatóak, gyártók szerint csoportosítva.

3.2 Felhasznált eszközök

A dolgozatban felhasznált eszközök a 7.2-es fejezetben találhatóak, gyártók szerint csoportosítva.

3.3 Módszerek

3.3.1 A TOPDB adatbázisban levő adatok bővítése

Négy forrásból gyűjtöttük ki a TM fehérjék különböző szekvenciális pozícióinak membránhoz viszonyított lokalizációját:

- a Pubmed tudományos cikkek tároló adatbázis (2014. augusztusig megjelent cikkek),
- a UniProt annotált fehérjeszekvencia adatbázis (2014_07 kiadás),
- a PDB (Protein Data Bank) és PDBTM (Protein Data Bank of TransMembrane Proteins) fehérje 3D szerkezeti adatbázisok (mindkettő 2014. augusztus 22-i verzió).

A publikusan elérhető cikkek közül azokat vizsgáltuk át, amelyek tartalmazták a „topology” és „transmembrane” kulcsszavakat. Ezekből a kísérleti topológia adatokat kigyűjtöttük, és a megfelelő kategóriába soroltuk a TOPDB honlapján található

„*Documents/Experiments*” menü alatt tárgyaltak szerint. A legtöbb kísérleti adatot a riporter enzimmel, valamint fehérjével való fúziós technika általi lokalizáció vizsgálatok, poszttranszlációs módosulások meglétének és helyének meghatározása (ezek között mind egyedi fehérjére vonatkozó, mind nagy áteresztőképességű (high-throughput, HT) proteomikai vizsgálatokkal meghatározott poszttranszlációs módosulásokat is találtunk), proteázemésztésen alapuló topológiai kísérletek, immunolokalizációs kísérletek, valamint kémiai módosítások (ezek nagy részét részletesen tárgyaltuk a 1.8 fejezetben) szolgáltatták.

Az előzőken kívül a UniProt adatbázis bejegyzések közül az „RP” címkével kezdődő sorok TOPOLOGY kulcsszavát tartalmazó irodalmi forrásokat is átvizsgáltuk újabb topológia adatok megtalálása érdekében.

A TOPDB korábbi verziójának közlése óta több mint 1300 TM fehérjeszerkezetet oldottak meg. A PDB szerkezeti adatbázisba tárolt szerkezeteken, a TMDet algoritmust használva⁷⁶, topográfiai információkat nyertünk a TM fehérjéről. Ezek az eredmények a PDBTM fehérjeszerkezeti adatbázisba is bekerültek. A PDBTM adatbázis azonban nem tartalmaz topológiai adatot csak „Side1” és „Side2” bejegyzéseket, amely megkülönbözteti a membrán két oldalán túlnyúló részeket. Így a topológiai adatokat az eredeti szerkezet közlő cikkekből kerestük és manuálisan egészítettük ki a talált topológiai adatokkal a vizsgált fehérje TOPDB bejegyzését.

További adatokat kaptunk a UniProt, és a PDB kombinált vizsgálatával. A PDB adatbázisból TM fehérjékhez tartozó szolubilis fragmenteket tartalmazó bejegyzéseket és azok lokalizációját is kigyűjtöttük. Ehhez kiválogattuk azokat a UniProt bejegyzéseket, amelyek tartalmazzák az FT TRANSMEM sort vagy membrán szubcelluláris lokalizációt és van legalább egy referencia PDB bejegyzés a kereszthivatkozások között, de a TMDet algoritmus nem találta meg ezeket, mint TM fehérje. Ezeket a fehérjéket szintén kigyűjtöttük, publikációikat átvizsgáltuk.

Az összes összegyűjtött topológiai adatot a CCTOP topológia predikciós módszer használta fel megszorításként, a fehérjék pontosabb topológia becslésének elérése érdekében. A CCTOP az összegyűjtött topológiai adatokon kívül, további megszorításként használ 10 különböző topológia becselő módszert, valamint a konzervatív elhelyezkedésű doméneket a TOPDOM adatbázisból^{91,149} és a homológ fehérjékből meghatározott kísérletes adatokat.

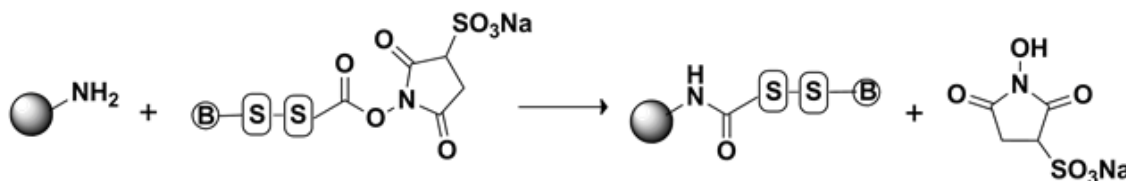
3.3.2 Sejtizolálás, sejtenyésztés

A RBC izolátumokon végzett kísérletekhez a vért mindig a kísérlet reggelén vették le egy önkéntestől az Országos Vérellátó Szolgálaton. A vért etilén-diamin-tetraecetsav (Ethylenediaminetetraacetic acid, EDTA) tri-kálium véralvadásgátlóval ellátott csövekben kaptuk meg, jégen tartottuk a RBC izolálásáig. A vért tartalmazó csövet kilengő rotorba helyeztük és 5 percen keresztül $500\times g$ -n $4^{\circ}C$ -on centrifugáltuk. Az RBC-k a cső alsó harmadába szeparálódtak, a fölöttük elhelyezkedő sárgás színű plazma felülúszót, valamint a plazma és az RBC-k határfázisán található „buffy coat” réteget is eltávolítottuk. Végül az RBC-ket kétszer $4^{\circ}C$ -os 10-10 ml foszfát pufferes sóoldattal (Phosphate Buffered Saline, PBS) átmostuk, minden mosás után $500\times g$ $4^{\circ}C$ -on 5 percig centrifugáltunk, ezután a felülúszót valamint a határfázist ismét eltávolítottuk. Az izolált RBC-ket jégen tároltuk a további kísérletekig.

A kísérletet további sejteken is tesztelni akartuk, ezért humán K562 krónikus és HL60 akut mieloid leukémia szuszpenziós sejt vonalakat tenyésztettünk. Továbbá a kontrasztosabb mikroszkópos képek készítéséhez CHO letapadó sejt vonalat használtunk. A tenyésztést erre alkalmas sejtenyésztő laboratóriumban, steril körülmények között végeztük. Az alkalmazott RPMI (HL60 és K562 sejtek táplálásához), valamint F12 HAM (CHO sejtek táplálásához) tenyésztő médiumokat lamináris fülke alatt (Streamline® Class II BSC) kiegészítettük $50\ \mu g/ml$ Penicillin-Sztreptomycin antibiotikumokkal és 10% főtális borjú szérummal (Fetal Bovine Serum, FBS). A folyamat után a tenyésztőmédiumokat egyszer használatos vákuumszűrő rendszeren szűrtük (Corning® bottle-top vákuumszűrő rendszer) a sterilitás biztosítása érdekében. A folyékony nitrogén alatt tárolt sejteket felvettük a sejtszámhoz mérten megfelelő mennyiségű médiumban, majd T25-ös vagy T75-ös flaskákban tenyésztettük azokat $37^{\circ}C$ -os termosztátban 5% CO_2 jelenlétében. A sejteket két-három naponként Bürker-kamrával számoltuk fénymikroszkóp alatt, majd passzáltuk a megfelelő médium felhasználásával, úgy, hogy a sejtszám a következő alkalomig az ATCC által megjelölt optimális tartományban maradjon. Az egy-egy kísérletekhez szükséges $\sim 10^7$ - 10^8 db sejtszámig növesztettük a sejteket. A RBC-k előkészítéséhez hasonlóan, a tenyésztett sejteket is kétszer 10-10 ml $4^{\circ}C$ -os PBS-sel mostuk, a mosások felülúszóit $300\times g$ $4^{\circ}C$ -on 3 percig végzett centrifugálással távolítottuk el.

3.3.3 Sejtfelszíni primer amin csoportok jelölése

A kereskedelmi forgalomban kapható „Sejtfelszíni fehérje izoláló kit” (Thermo Scientific) primer amin csoportokra specifikus reagensét, a Szulfo-NHS-SS-biotint használtuk az izolált sejtek jelölésére. A reagens membrán impermeábilis tulajdonsága miatt csak a sejtek külső felszínén hozzáférhető lizin ϵ -amino csoportot, valamint fehérje N-terminálisát képes kovalensen jelölni (4. ábra).



4. ábra: A fehérje primer amin csoport és a Szulfo-NHS-SS-biotin reagens reakciója

Mielőtt jelöltük az izolált sejteket a Szulfo-NHS-SS-biotin reagenssel egy harmadik 10 ml-es PBS mosást végeztünk, 20 percen keresztül 4 mM végkoncentrációjú jódcetamid alkilálószer jelenlétében. Az alkilálószer felhasználásával a szabad szulfhidril csoportokat blokkoltuk, ezzel elkerülve a jelölő reagensben jelenlévő diszulfid hídnak a jelölés fiziológiás körülménye alatt bekövetkező diszulfid izomerizációját. A sejtekről a mosó puffert eltávolítottuk 3-5 perces 300-500 \times g 4°C-on végzett centrifugálással, ezt követte a ~2 mM végkoncentrációjú Szulfo-NHS-SS-biotinnel történő jelölés, 20 percig 4°C-on fiziológiás (pH=7,4-es) PBS pufferben. A jelölést az internalizáció minimalizálása céljából végeztük alacsony hőmérsékleten. Az inkubációs idő végén a reakció leállításához hozzáadtuk a 10 ml 25 mM Tris (Tris(hydroxymethyl)aminomethane) primer amin tartalmú puffert (pH=7,2), majd ismét centrifugáltuk 300 \times g 4°C-on 3 percen keresztül és a felülúszót kidobtuk, végül ezt a lépést kétszer megismételtük a felesleges jelölő eltávolítása érdekében.

A jelölés sikerességét először CHO sejteken és RBC-ken teszteltük, valamint vizsgáltuk a membránpreparátumokban a jelölés jelenlétét, megmaradását (7.3.1. fejezet).

3.3.3.1 A sejtek felszíni biotinálásának optimalizálása

Az Szulfo-NHS-SS-biotin reagenst tekintélyes mennyiségű tudományos közleményben használták már korábban^{116,130,131}, amelyekben az alkalmazott reagens mennyisége változatos, olykor csak végkoncentrációkat ismertetnek, de pontos reakcióelegy-

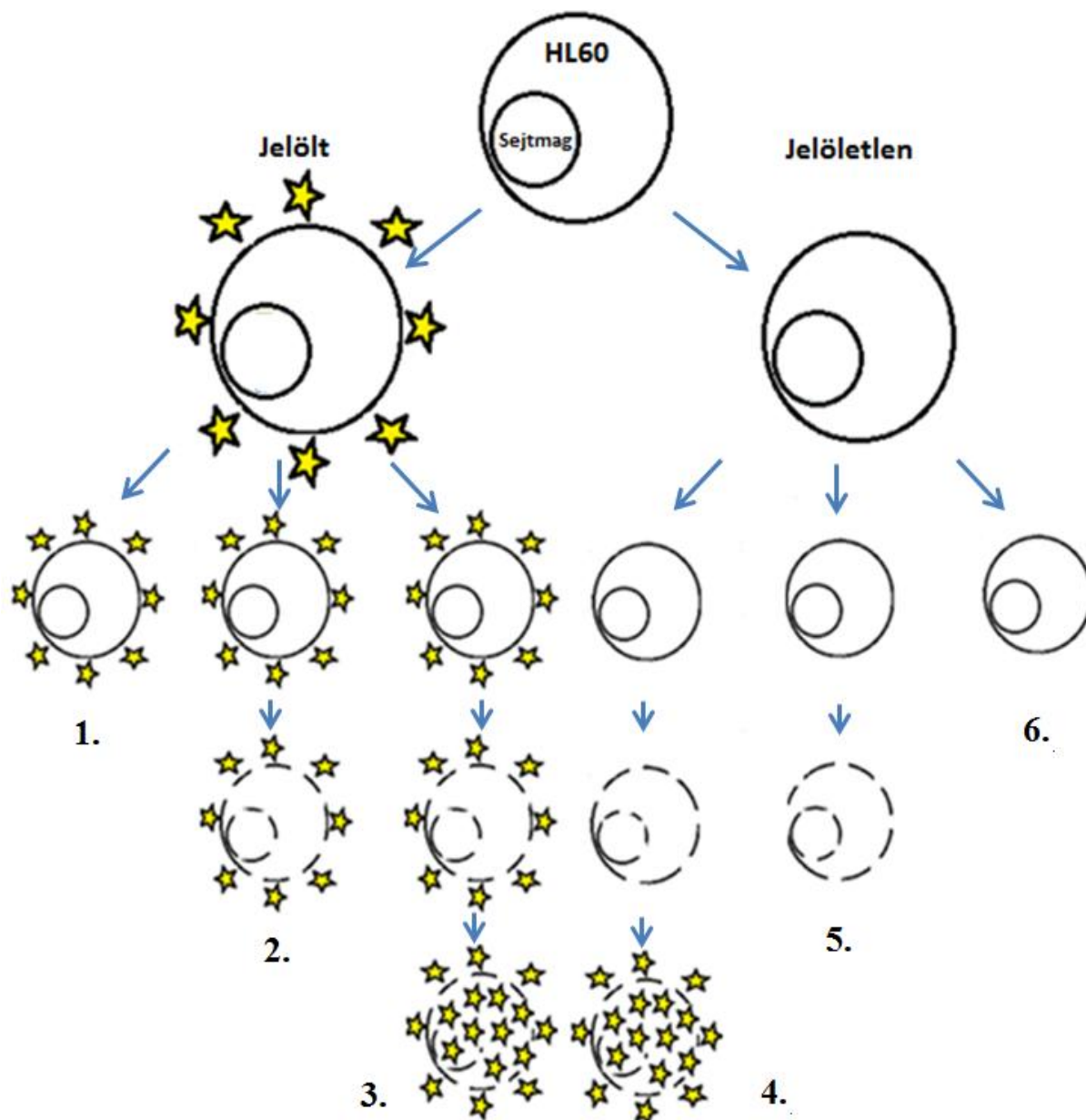
térfogatokat nem. A reakcióelegy másik fő összetevői a sejtek, amelyek száma szintén nagy variabilitást mutat (függ a sejttypustól, a kísérlet céljától). A különféle ismeretek miatt nem tudtunk egyértelmű döntést hozni, hogy a használt sejttypus mennyiségéhez mennyi jelölő reagensre van szükség, hogy a sejtek felszínét maximálisan jelölni tudjuk. A probléma megoldása érdekében FACS Attune® Acoustic Focusing Cytometer (Blue/Violet) készülékkel teszteltük a RBC-k felszínének biotinálási hatékonyságát a reagens végkoncentráció függvényében.

A kísérlethez az RBC-eket frissen izoláltuk (3.3.2-es fejezet), majd 16 db egyenlő részre osztottuk az izolátumot. Az első eppendorf csőbe nem került biotináló reagens (vak minta), míg a további 15 részhez növekvő koncentrációban (anyagmennyiséget nézve: 0,1; 1; 2,5; 5; 10; 25; 50; 100; 200; 500; 750; 1000; 2000; 5000 végül 10000 nmol reagens) adtuk hozzá a Szulfo-NHS-SS-biotin reagenst, ügyelve az egységes végtérfogatok tartására.

A jelölést tris pufferelt sóoldatos (Tris-buffered saline, TBS) mosással állítottuk le, majd 6%-os szarvasmarha szérum albumint (Bovine Serum Albumin, BSA) tartalmazó PBS-sel blokkoltuk a jelölési reakciót 4°C-on 10 percig (az aspecifikus antitest kötődések elkerülése végett). 2 µg/ml fluoreszcens izotiocianát (Fluorescein isothiocyanate, FITC) konjugált anti-biotin monoklonális antitesttel (szintén 6%-os BSA-t tartalmazó PBS-ben) inkubáltuk 1 órán keresztül sötétben enyhén rázatva mind a 16 mintát. A sejtek előre irányuló és oldalirányú fényszórásuk alapján áramlási citométerrel jellemeztük, a megfelelő szórási értékekkel rendelkező populációt a jelöletlen mintában kiválasztottuk („kapuztuk”), majd minden mintára az adott szórási érték tartományban („kapuban”) megjelenő sejtek átlagos fluoreszcencia értékeit vizsgáltuk. A kísérletünknel a sejteket a FITC flouorofór gerjesztéséhez szükséges hullámhosszúságú fényvel megvilágítottuk, majd az emittált fény sejtenkénti átlagos intenzitását detektáltuk. A sejtenkénti átlagos zöld (FITC) flureszcencia intenzitások alapján következtethettünk a sejt felszínen történő biotin felhalmozódásra, meghatározhattuk a minimális reagens koncentrációt, amely elegendő sejt felszín maximális biotinálásához.

3.3.3.2 Jelölő ágens membrán impermeabilitásának vizsgálata

A jelölő reagens a forgalmazó szerint a kettős foszfolipid rétegen nem képes áthatolni natív körülmények között. Ennek ellenőrzése a kidolgozandó módszer sikeréhez elengedhetetlen volt. A jelölő ágens membrán impermeabilis attribútumának tesztelését permeabilizált ill. nem permeabilizált HL60 sejt vonal felhasználásával készítettük el. A mintákat az 5. ábrán látható módon osztottuk fel és kezeltük.



5. ábra: A HL60 sejtek permeabilizált és nem permeabilizált körülmények közötti jelölésének folyamata. Szulfo-NHS-SS-biotin: sárga csillag. A számok a keletkezett 6 féle mintát jelölik, amelyeket FITC konjugált anti-biotin antitesttel és Hoechst 33342 DNS festékkel inkubáltuk, majd konfokális mikroszkóppal elemeztük (1. jelölt, nem permeabilizált, 2. jelölt, permeabilizált, 3. jelölt, permeabilizált és permeabilizálás után ismét jelölt, 4. jelöletlen, permeabilizált és permeabilizálás után jelölt, 5. jelöletlen, permeabilizált 6. jelöletlen)

A HL60 sejteket először két részre osztottuk, az egyik részt 20 percig 4°C-on 2 mM Szulfo-NHS-SS-biotin reagenssel jelöltük PBS pufferben, a másik részt az előzőhöz hasonlóan, csak jelölő ágens nélkül inkubáltuk PBS pufferben. A TBS puffer felhasználásával történő reakció leállítás közben 3-3 egyforma részre osztottuk a jelölt és jelöletlen mintákat

(5. ábra – 1-3: jelölt, 4-6: jelöletlen), majd fixáltuk azokat 4% PFA-val 4°C-on 10 percig. A fixáló szert kétszeri PBS pufferes mosással és centrifugálással (300×g 4°C-on 3 percig) távolítottuk el. Két-két jelölt és jelöletlen (5. ábra – 2-3: jelölt, 4-5: jelöletlen) sejtpopuláció membránját 0,1%-os TritonX-100 PBS pufferben permeabilizáltuk, 3 percen keresztül 4°C-on, majd a detergenses oldatot szintén 2-szeres PBS-es mosással távolítottuk el. Végül a permeabilizált mintából egyet-egyet Szulfo-NHS-SS-biotin jelölésnek vetettük alá (5. ábra: 3-4).

A keletkezett 6 féle mintát 6% BSA tartalmú PBS-ben blokkoltuk 10 percig 4°C-on a következő lépésben használt anti-biotin antitest aspecifikus kötődésének minimalizálása érdekében. A jelölő ágens biotin részére specifikus FITC konjugált anti-biotin monoklonális antitesttel (1:250-szeres hígításban PBS-ben, FITC gerjesztési és emissziós maximuma: 490/525 nm) inkubáltuk 4°C-on 30 percig az 5. ábrán bemutatott módon előkészített 6 sejtsoportot. Végül egy PBS-es mosás után, mind a 6 mintát 0,1 µg/ml Hoechst 33342 DNS festékkel inkubáltuk 2 percig.

A jelöletlen/jelölt és permeabilizált/nem permeabilizált mintákban a festékek (FITC, Hoechst) lokalizációját Zeiss LSCM 710 konfokális mikroszkóppal vizsgáltuk 63x NA=1.4 Plan Apo objektív felhasználásával. A festékeknek a sejtek adott síkjában lévő elhelyezkedését, a megfelelő hullámhosszúságú gerjesztő lézerek felhasználásával, valamint emissziós spektrumukra jellemző hullámhosszúságú fény kiszűrésére alkalmas szűrőkockákkal rögzítettük/fényképeztük.

3.3.4 Sejtek feltárása, membránpreparálás

A sejteken végzett Szulfo-NHS-SS-biotin jelölési kísérletek pozitív tapasztalatai után, a következő megoldandó probléma a minták komplexitása volt. A plazmamembránban előforduló fehérjék relatív alacsony száma a citoszólikus fehérjékhez képest, megnehezítette a használt reagens által módosított peptidek analitikai azonosítását. A minta komplexitásának csökkentése, valamint a plazmamembrán fehérjék dúsítása érdekében többféle membránpreparálási protokollt próbáltunk ki. Az eddigi tapasztalatok alapján analitikai szempontból leghatékonyabb (legtöbb jelölt TM fehérjét eredményezte) membránpreparálási módszert ismertetem.

A sejteket az izolálás és a megfelelő jelölés után hipotóniás lízis pufferbe (20 mM Tris-HCl (Tris(hydroxymethyl)aminomethane hydrochloride) pH=7,4, 10 mM KCl, 20 mM szukróz) helyeztük, amely tartalmazott 10 mM jódcetamidot, hogy a sejtek feltárása alatt

szabaddá váló szulfhidril csoportokat azonnal alkiláljuk (diszulfid-izomerizáció elkerülése érdekében).

Az izolált RBC-eket 10 percig kevertük jégen üvegbottal a hipotóniás lízis pufferben, a folyamat ideje alatt megközelítőleg az összes RBC feltáródott. A HL60 és K562 sejtek sejtizolátumok hatékonyabb sejtfeltárása érdekében először potteres eljárást alkalmaztunk (40 ciklusban mozgattuk le-fel a homogenizáló rudat), majd 26G-s tűn 1 ml-es fecskendő segítségével tovább homogenizáltuk (20 ciklusban vettük fel és injektáltuk vissza a homogenizátumot a fecskendő segítségével), majd a RBC-khez hasonlóan jégen 10 percig inkubáltuk a mintákat a lízis pufferben. A leukémiás sejtvonalak estén a feltárás után $1700\times g$ $4^{\circ}C$ -on 7 percig történő centrifugálással kiülepítettük a feltáratlan sejteket, sejtörmelékeket és sejtmagokat, majd a felülúszóval dolgoztunk tovább.

A lizátumokat speciális polikarbonát csövekbe (10,4 ml-es, 355603) töltöttük és $\sim 100000\times g$ 1 órán keresztül $4^{\circ}C$ -on 70.1 Ti fix rotor feltéttel L7-55 ultracentrifugában fugáltuk és ülepítettük ki a membrán pelleteket. Az RBC-eket kifehéredésig mostuk 10-szeresére hígított lízis pufferrel, amelynek nátrium-hidroxiddal állítottuk be a pH-ját, pH=7,7-re. Az RBC-knél 2-3, míg a sejtvonalaknál egy-egy mosási lépés volt, mosásonként szintén fugáltuk a mintákat $100000g$ -n, $4^{\circ}C$ egy órán keresztül.

A kifehéredett sejtpeletteket frissen készített 50 mM-os ammónium-bikarbonátban (NH_4HCO_3 , pH=8) vettük fel, majd azokat egy 2ml-es boroszilikát üveg homogenizáló csőbe pipettáztuk. A homogenizáló csőben a teflon (PTFE, Polytetrafluoroethylene) fejjel és rozsdamentes acél szárral ellátott homogenizáló rúddal a mintát jégen tartva potteres eljárásnak vetettük alá (25 ciklus le-felmozgatása a homogenizáló rúdnak). A keletkezett homogén opálos szuszpenziót egy nagyobb és egy kisebb részre osztottuk, a kisebb részt fehérje mérésre tettük félre, a nagyobb részt a fehérje mérés alapján megfelelő fehérjetartalmú részekre osztottuk. A fehérje mérést minden membránpreparátum minta esetén Lowry és mtsa¹⁵⁰ által kidolgozott módszerrel végeztük (7.3.3 fejezet).

3.3.4.1 SDS-PAGE

Az egységes fehérjetartalmú (5-20 μg) membránpreparátumok, minőségét, reprodukálhatóságát Na-dodecil-szulfátos poliakrilamid gélelektroforézissel (Sodium Dodecyl Sulphate Polyacrylamide Gelelectrophoresis, SDS-PAGE) teszteltük Coomassie Brilliant Blue (CBB) R-250 festék felhasználásával.

A redukálószeres (100 mM DTT) vagy redukálószermentes mintakoktélban (összetétele: 1,3 ml ultratiszta víz; 2,5 ml 0,5 M Tris-HCl; 2 ml glicerin; 4 ml 10% SDS; 200 µl 0,5% brómfenolkék) kétszeresére hígítottuk a mintáinkat, majd 5 percre 95°C-ra előmelegített termoblokk készülékbe helyeztük őket. Visszahűtés és centrifugálás után egységnyi térfogatú és azonos fehérjemennyiségű mintákat vittünk fel 4%-os tömörítő és 12%-os szeparáló SDS-Page gél zsebeibe. EuroClone Sharpmass markert használtunk a mintánként jellemző fehérjesávok méretének azonosítására.

A mintákat a tömörítő gélben futtatuk 10 percig, 140 V-on, a fehérjék elválasztását az elválasztógélben 60-90 percig végeztük, „PowerPac™ Universal” készülék segítségével. A gélt ezután kivettük az üveglapok közül, levágtuk a tömörítő részt, kétszer desztillált vízben átmostuk, majd 2mg/ml koncentrációjú Coomassie festékkel inkubáltuk 1-2 órán keresztül, vagy 16 órás időtartamig, a kontrasztosabb festés céljából. A festék elmosása után a szeparált fehérjék gélcsíkjait ChemiDoc XRS+ (Bio-Rad) készülékkel monitoroztuk.

Ezt a módszert használtuk a szolubilizált membránpreparátumok over night fehérjebontó enzimmel való kezelés (3.3.5 fejezet) utáni analizálására is.

3.3.4.2 Western-blot

Az SDS-PAGE gélfuttatást DTT mentes futtató pufferrel végezve, a biotin a jelölő ágenssel módosított fehérjéken marad (jelölő molekulában lévő diszulfid híd nem redukálódik, nem hasad le), így a membránpreparátum szolubilizált fehérjeinek biotináltsága vizsgálható a gél futtatása után (Coomassie festés nélkül) Western-blot technikával.

A komplex fehérje mintákat SDS-PAGE gélen elválasztottuk, majd polivinilidén-difluorid (Polyvinylidene difluoride, PVDF) membránra blottoltuk 70 V-on 200 mA-en, alul-felül 1-1 szivacs és 3-3 szűrőlap közé befogva a gélt és a membránt (membrán az anódhoz közelebb), transzfer pufferben (25 mM TRIS, 192 mM glicin, 15% MeOH), 1-2 órán keresztül végezve a folyamatot.

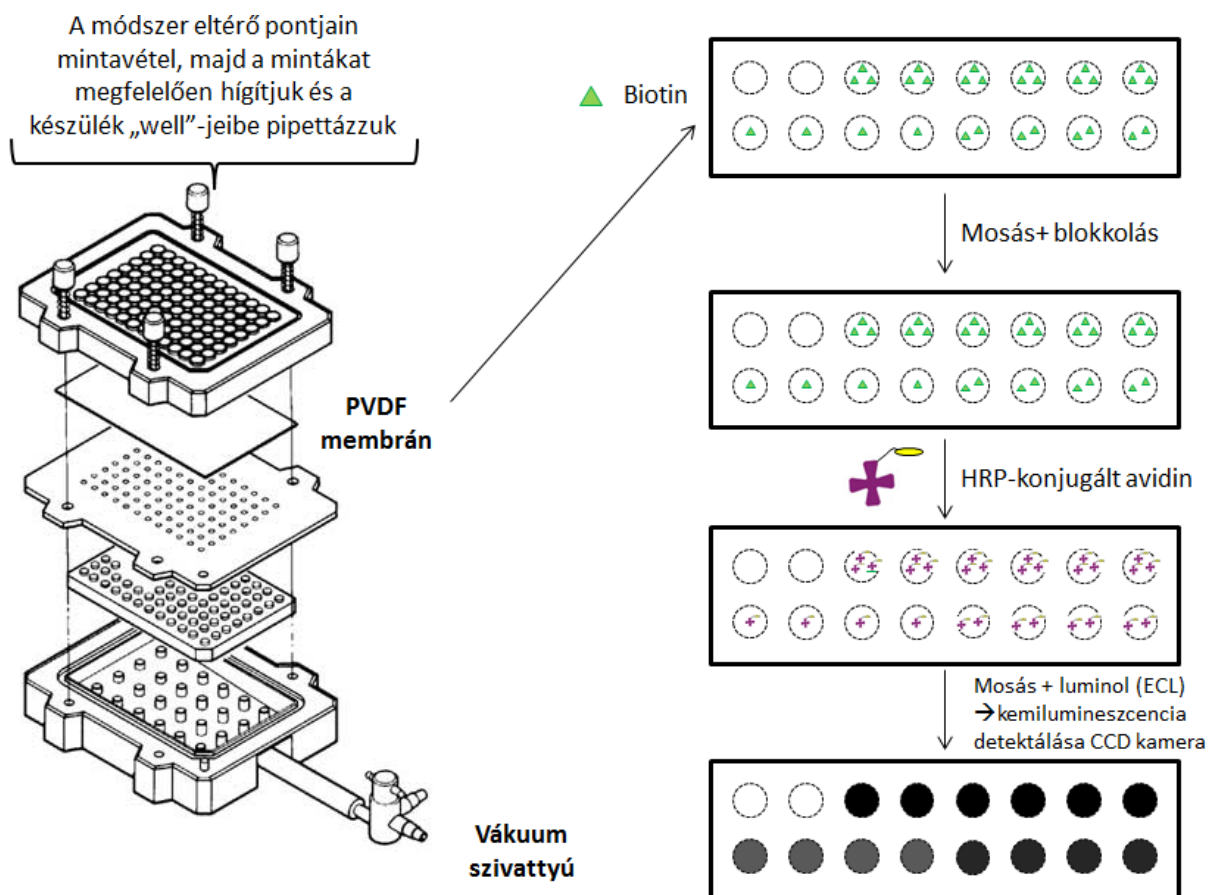
A blottolás után a PVDF membránt 5% BSA-TBS-T (TBS-T összetétele: 25 mM Tris-HCl, pH=7.4, 140 mM NaCl, 3 mM KCl, 0.05% Tween-20) pufferben blokkoltuk 45 percen keresztül. A következő lépésben a blokkoló pufferbe 0,02 µg/ml tormaperoxidáz (Horseradish peroxidase, HRP)-konjugált avidint adtunk és ezzel is 45 percig inkubáltuk a PVDF membránt. Végül 3-szor 10 perces TBS-T pufferes inkubálással elmostuk a felesleges jelölőt és blokkoló puffert. A membrán előhívásához HRP szubsztrátot (enhanced

chemiluminescence, ECL) használtunk, egy percig inkubáltuk. A szubsztrát peroxidáz általi oxidációja kemilumineszcens jelet eredményezett, amelyet ChemiDoc™ XRS+ készülékkel detektáltunk, különböző expozíciós időkkel felvételeket készíttve.

3.3.4.3 Dot-blot

Ezt az eljárást a membránpreparátumok biotináltsági szintjének szemikvantitatív elemzésén túl, a membránpreparátumok szolubilizált fehérjéinek elemésztése után az avidin agarózzal töltött oszlop (Pierce™ Spin Columns - Snap Cap) előtti és az oszlopon áteső frakciók biotin tartalmának jellemzésére használtuk. Az utóbbinál a cél, hogy meghatározzuk az avidin agaróz optimális mennyiségét, ami elégséges az összes biotinált komponens kikötésére a mintából.

Az avidin agarózzal töltött oszlop előtti és után frakciókat, a további membránpreparátum és mosó puffer mintákat 50 mM NH_4HCO_3 pufferrel hígítottuk. A mintákat előre eltervezett sorrendben vittük fel, a metanollal előaktivált PVDF membránt befogó összeállított Bio-Dot® mikrofiltrációs rendszer welljeibe (6. ábra bal oldala) 50 μl végtérfogatra kiegészítve. A mintákat vákuum szivattyú segítségével átáramoltattuk a wellék alá helyezett PVDF membránon, így blottolva azokat a biotinált komponenseikkel együtt. A membránt a blottoló berendezésből kivettük, majd a Western-blottnál bemutatott folyamatot végeztük el rajta (3.3.4.2-es fejezet), amelyet a 6. ábra jobb oldala szemléltet (membrán blokkolása, HRP-konjugált avidinnel inkubálás, majd a felesleges reagens eltávolítása, végül HRP szubsztrát hozzáadása, kemilumineszcens jelének detektálása).



6. ábra: Biotinált komponensek szemi-kvantitatív kimutatására fejlesztett dot-blot technika

3.3.5 Membránpreparátumok szolubilizálása, emésztése

A preparált membránok fehérjéinek bottom-up proteomikai vizsgálatához enzimatiskus emésztéssel kisebb poli- és oligopeptidekre hasítottuk a fehérjéket, úgy, hogy közben kerültek az olyan detergens szennyezők használatát, amelyek zavarhatják a peptidek MS analízisét.

A membránpreparátumokat a fehérjemérés alapján 100 µg, 200 µg és 500 µg fehérjetartalmú mintákra osztottuk. A minták szolubilizálásához 50 mM NH_4HCO_3 pufferben (pH=7,9) feloldott 0,5% (m/V-os) MS kompatibilis Rapigest detergens törzsoldatot és ultratiszta vízben feloldott 100 mM ditioetanol és 500 mM jódetamid törzsoldatot használtunk fel. A mintákhoz a következő végkoncentrációkban alkalmaztuk az előző három komponenst: 0,05-0,1% (m/V) Rapigest detergenst és 1-1 mM ditioetanol és jódetamidot. A reakció elegyet 5*1 percen keresztül vízfürdős szonikátorral kezeltük („degas” funkció az Elmasonic S 30 (H) készüléken), majd fél órán keresztül jégen inkubáltuk, 5 percenként vortexelve azokat.

A módszer hatékonyságának növelése céljából a szolubilizálás után 2 órán keresztül 37°C-on PNGaseF kezelésnek vetettük alá a mintákat, hogy az aszparaginokon lévő acetilglükózamin cukor egységeket lehasítsuk. Ezt követően 1:50/1:25 fehérje: proteáz (tripszin, MS Grade tripszin, Kimotripszin, Termolizin) arányban hozzáadtuk az enzimeket a mintákhoz és az adott enzim optimális hőmérsékletén (tripszin, MS Grade tripszin: 37°C, kimotripszin: 30°C; termolizin: 70°C) inkubáltuk 2, 4, 6, 8 vagy 16 órán keresztül. Az emésztések végén az enzimeket 10 percen keresztül 95°C-on hőinaktiváltuk, majd ezt követő jégen visszahűtöttük. A mintákhoz ezután a megfelelő proteáz inhibitorát adtuk a gyártók által előírt végkoncentrációban (Tripszin és MS Grade tripszin: N α -Tosyl-L-lysine chloromethyl ketone hydrochloride (TLCK), kimotripszin: N-p-Tosyl-L-phenylalanine chloromethyl ketone (TPCK), termolizin: EDTA). A biotinált és nem biotinált peptideket is tartalmazó oldatot -80°C-on tároltuk, vagy a folyamatban következő avidin agarózzal töltött oszlopra vittük fel (3.3.6 fejezet).

3.3.6 A biotinált peptidek dúsítása, eluálása, C18 oszlopon való tisztítása

A biotinált peptideket normál, vagy nagy kapacitású neutravidin agaróz gyöngyöt tartalmazó oszlopon tisztítottuk, hogy a jelölések proteomikai azonosítása a biológia minta komplexitásának további csökkentésével (jelöletlen peptid-háttér jelentős részét eltávolítottuk) egyszerűbbé váljon.

A szükséges neutravidin oszloptöltet mennyiségét dot-blot technikával optimalizáltuk (3.3.4.3 fejezet), hogy elkerüljük a biotinált komponensek elvesztését, azaz az oszlop utáni frakcióból ne detektáljunk biotin jelet. A megfelelő mennyiségű neutravidint tartalmazó oszlopon a 16 órán keresztül emésztett membránpreparátum mintákat (3.3.5 fejezet) szobahőmérsékleten 1 órán keresztül inkubáltuk enyhe ráztatás közben.

Az avidin agarózzal töltött oszlopon kikötődött aspecifikus komponensek lemosását különböző ionerejű és fiziológiás vagy magas pH-jú pufferekkel (50 mM NH₄HCO₃, pH=8; 5M NaCl; 100 mM NaHCO₃, pH=10; 60°-os 50 mM NH₄HCO₃, pH=8) végeztük. Az alkalmazott avidin agaróz gyöngy térfogatához képest a mosó pufferekből 10-20-szoros térfogatokat használtunk, 400-400 μ l-es mennyiséggel mostuk a gyöngyöt vákuum manifold segítségével.

A mosást követően a peptidek elúcióját 1 órán keresztül 37°C-on történő 10 mM TCEP vagy 50 mM DTT redukáló ágensek jelenlétében végeztük. A redukáló szerek a Szulfo-NHS-SS-biotin diszulfid kötéseit hasítják, így az inkubáció után 1000 \times g 4°C-on 1

perces centrifugálással nyertük vissza az oszlopon immobilizált, jelölt peptideket. A mintát az alkalmazott redukálószer végkoncentrációhoz képest 2,2-szeres végkoncentrációjú jódiacetammal alkiláltuk 45 percen keresztül, 37°C-on (ezzel meggátoltuk a szabad szulfhidril csoportok diszulfid-híd képzését), majd a mintákat vákuum alatt bepároltuk.

Az MS mérésre való mintaelőkészítés utolsó lépéseiként C18-as oszlopon tovább tisztítottuk és sómentesítettük. A C18 oszlopon való tisztítás menete:

- 2* 50% MeOH C18 oszlop kondicionálása az oszlop töltetének
- 2* 0,5% trifluor ecetsav (Trifluoroacetic acid, TFA) 5% acetonitril (ACN) mosás
- 2* 0,1% TFA mosás
- A vákuum bepárolt mintákat felvettük 50-50 µl 0,1% TFA-ban
- A mintákat az első három pontban írtak szerint előkészített C18 oszlopra helyeztük, centrifugálással (2000×g 4°C-on 1 percig) a C18 tölteten áteresztettük, majd az áteső frakciót visszahelyeztük az oszlopra és még egyszer centrifugáltuk
- Az oszlopon kötődött peptideket 2*0,1% TFA mosásnak vetettük alá
- Végül 50-50 µl 70% AcN és 0,1% TFA-val eluáltuk a peptideket kétszer, a keletkezett 100 µl-eket újra vákuum bepároltuk és 20-30 µl 2% AcN, 0,1% hangyasavban (Formic Acid, FA) vettük fel, majd az MS analízishez megfelelő teszt csövekbe töltöttük

3.3.7 Peptid fragmensek nano UPLC-MS/MS analízise

A tömegspektrometriai méréseket, valamint a spektrumok feldolgozását az MTA TTK Proteomikai Kutatócsoport és a Szegedi Biológiai Kutatóközpont Proteomikai Kutatócsoport végezte. A 3.3.6 fejezetben keletkezett komplex peptid elegy elválasztását és analízisét két nanoLC-MS/MS készülékkel végeztük:

Az első berendezésben a peptideket egy nanoAcquity UPLC rendszeren (Waters) választottuk el, majd azokat ionizálás után az LTQ-Orbitrap Elite (Thermo Scientific) tömegspektrométerrel elemeztük. A mérések során a komplex peptid elegyből 5 µl-es mennyiségeket injektáltunk a rendszerbe. A peptideket először egy Symmetry C18 nanoAcquity UPLC trap oszlopra kerültek (Waters, 5 µm, 100 Å, 0,18*20 mm) 10 µl/ perces áramlási sebességgel, 2 percig, majd az elválasztást BEH300C18 nanoAcquity UPLC oszlopon (Waters, 1,7 µm, 300 Å, 0,075*250 mm) végeztük. A használt lineáris gradiens 97-60% A eluens és 3-40% B eluens keveréke volt 40 vagy 100 percig (ahol A eluens: 0,1% FA-t

tartalmazó HPLC tisztaságú víz, B eluens: 5% DMSO-t és 0,1% FA-t tartalmazó acetonitril) 300 nl/perc áramlási sebességgel. Az MS spektrum felvétele az Orbitrap analizátorban (60000-es felbontás), majd az 5 vagy 10 leggyakoribb többszörösen töltött ion kiválasztása és fragmentációja a lineáris ioncsapdában történt, CID vagy HCD technikával (normalizált ütközési energia 35% volt és a dinamikus kizárás 30 másodpercre volt állítva).

A második berendezésben a peptidek elválasztása Dionex Ultimate 3000 NanoLC (Sunnyvale) kromatográfias rendszer használatával történt. CaptiveSpray nanoBooster ionforrással ionizáltuk, majd Bruker Maxis II ETD Q-TOF (Bremen) tömegspektrométerrel analizáltuk a peptideket. A mérések során a komplex peptid elegyből 1-6 µl-es mennyiségeket injektáltunk a berendezésbe. A peptidek először Acclaim™ PepMap™ 100 C18 Nano-Trap oszlopra (5 µm, 100 Å, 100 µm*20 mm, Thermo Scientific) kerültek, majd elválasztásuk 15cm Waters Peptide BEH C18 nanoACQUITY 1,7 µm részecske méretű UPLC oszlopon történt. Az alkalmazott nanoLC gradiens összetétele a következő volt: 97,5-75% A eluens és 2,5-25% B eluens keverékét 80 percen keresztül áramoltattuk, azután 20 percen keresztül futott a 75%-55% A eluens és 25-45% B eluens keveréke (ahol A eluens: 0,1% FA-t tartalmazó HPLC tisztaságú víz, B eluens: 0,1% FA-t tartalmazó acetonitril). Az MS mérésekhez pozitív ionizációs módot és 2,5 másodperc fix ciklus időt használtunk. A ciklus ideje alatt történt az MS spektrum feldolgozása 3 Hz-ig 150-2200 m/z tartományban, valamint az MS/MS spektrumok felvétele CID fragmentáció után 16 Hz-ig a gyakori prekursorok és 4 Hz-ig a ritkább prekursorok esetében.

3.3.8 A nanoLC-MS/MS adatok értékelése, a peptidek/fehérjék azonosítása

Az első berendezéssel mért nyers adatokból Proteome Discoverer (Thermo, v1.4) vagy Pava script¹⁵¹ segítségével generáltunk MS/MS csúcslistákat. A feldolgozott spektrumokat ProteinProspector v5.14.1 szoftverrel értékeltük ki. Két egymást követő lekeresést végeztünk. Az első lekeresés a teljes SwissProt fehérje adatbázison (SwissProt.06.10.2014 (545388/545388 bejegyzés) történt, amelynek az volt a célja, hogy a mintában lévő fehérjéket azonosítsuk. A második lekeresésnél a humán UniProt fehérje adatbázis (UniProtKB.2015.4.16 (145723/47262724 bejegyzés) és azok randomizált szekvenciái voltak összefűzve a SwissProt lekeresés fehérje találatainak szekvenciáival. A ProteinProspector v5.14.1 lekereső algoritmus további beállításai a következők:

- Egy vagy két, a hasító enzimre (a négyféle használt enzim: tripszin, MS Grade tripszin, kimotripszin, termolizin) nem specifikus hasítási vég megengedése, maximum két-három hiányzó hasítóhely peptidenként.
- Variábilis módosítások keresése:
 - karbamidometil a cisztein aminosavakon.
 - oxidáció a metionin aminosavakon.
 - deamidáció az aszparagin és glutamin aminosavakon.
 - tioaciláció vagy karbamidometiltio-propanoiláció a lizin aminosavakon és a fehérje N-terminálisokon (utóbbi a jódcetamiddal alkilált mintákban).
 - ha az avidin agarózról hangysavval eluáltuk a peptideket, akkor a lizin aminosavakon a tioacil-biotin (Thioacyl-B) módosulás keresése (mivel az ilyen elúciónál a biotináló reagens teljes egésze rajta maradt a módosított helyen).
 - ha a TCEP/DTT eluálás után jódcetamid helyett más reagenseket használtunk a szabaddá vált szulfhidril csoportok alkilálására: a brómetilamin esetében Thio(AE), az N-etilmaleimid alkilálás után a Thio(NEM) módosítások keresése a lizin aminosavakon.
- A három megengedett maximális módosítás peptidenként, 10 ppm prekursor tömeg tolerancia, a fragmentation tolerancia CID esetében 0,6 Da, HCD esetében 25 ppm.
- A Protein Prospector alapértelmezett elfogadási kritériumait (minimum 15 és 22 pontszám és 0,05 és 0,01 max E value peptidekre és fehérjékre) használtuk az értékeléshez, valamint a módosított lizineket tartalmazó peptideket manuálisan is ellenőriztük. Az FDR-t úgy határoztuk meg, hogy a kétszeres „decoy” peptid számot elosztottuk az összes azonosított spektrumok számával és így minden lekeresésnél kevesebb, mint 1%-ot találtunk.

A második berendezéssel mért nyers adatokat újra kalibráltuk a Bruker Compass DataAnalysis software 4.3 (Bruker Daltonik GmbH) segítségével a belső kalibráló anyagra. Az MS/MS csúcslista generálása ProteinScape 3.1 szoftverrel (Bruker Daltonik GmbH) történt. A feldolgozott spektrumokat a Mascot (Matrix Science) programmal értékeltük ki. A fehérjék/peptidek azonosításához a humán fehérjéket használtuk a SwissProt adatbázisból

(SwissProt 2014_08; 546238/546238 bejegyzés). A randomizált („decoy”) szekvenciákat a Mascot generálta. A Mascot lekereső algoritmus további beállításai a következők:

- Szemi-triptikus hasítási vég, maximum négy megengedett hiányzó hasítóhellyel.
- Fix módosításként: karbamidometil a cisztein aminosavakon.
- Variábilis módosításként:
 - oxidáció a metionin aminosavakon.
 - deamidáció az aszparagin és glutamin aminosavakon.
 - tioaciláció vagy karbamidometiltio-propanoiláció a lizin aminosavakon és a fehérje N-terminálisokon.
- Az anyaiionok esetén 7 ppm tolerancia, míg a fragmens ionok esetén 0,05 Da tömeg tolerancia.
- A Mascot iontalálatok közül a 95%-nál nagyobb valószínűséggel azonosítottak Mascot score értéke 13 volt, de a biztosabb módosított peptid azonosítás érdekében a találatokat 24-es pontszám fölött fogadtuk el. Az FDR minden lekeresés alatt 1% alatt volt.

3.3.9 A módosított peptidek vizsgálata

A komplex peptidmintákat a két tandem MS készülékkel analizáltuk, majd a keletkezett spektrumokat a 3.3.8 fejezetben tárgyaltak szerint különböző lekereső algoritmusokkal vizsgáltuk. Az analízisekből megbízhatóan azonosított peptid listát táblázatos formában kaptuk meg (táblázat egyes sorai tartalmazták az adott peptidet, a megfelelő analitikai tulajdonságaival együtt: m/z, retenciós idő, módosítások, tömeg pontosság stb.). A megkapott táblázatokból kiszűrtük a 2. táblázatban felsorolt módosulásokat.

Unimod ID	Módosítás rövidítése	Hozzáadott molekula neve és összegképlete	Hozzáadott molekula monoizotópos tömege	Szerkezeti képlet
126	Thioacyl	DSP, Thioacyl, 3-sulfanylpropanoyl: <i>C(3) H(4) O S</i>	87,9983	
293	CAMThio	3-(carbamidomethylthio)propanoyl: <i>C(5) H(7) N O(2) S</i>	145,0198	
-	Thio(NEM)	Thioacyl-N-ethylmaleimide: <i>C(9) H(11) N O(3) S</i>	213,0454	
-	Thio(AE)	Thioacyl-aminoethyl: <i>C(5) H(9) N O S</i>	131,0405	
1841	Thioacyl-B	Thioacyl-Biotin: <i>C(15) H(23) N(3) O(3) S(3)</i>	389,0902	

2. táblázat: Az alkalmazott fehérje N-terminális és lizin oldallánc módosítások jellemzői
(Unimod ID forrás: http://www.unimod.org/modifications_list.php¹⁵²)

A Szulfo-NHS-SS-biotin reagens membrán impermeabilitása miatt, a 2. táblázatban listázott módosítások az extra-citoszólikus térben elhelyezkedő és hozzáférhető lizin aminosavakon vagy fehérje N-terminálisokon következhetnek be. Az 1. sorban a membrán impermeabilis reagens redukálása után jelölt peptideken maradt módosítás látható. A 2-4. sorban a reagens redukálása után keletkezett szabad szulfhidril csoportok eltérő alkilálása miatt létrejött módosítások figyelhetők meg (sorrendben jóacetamid, N-etilmaleimid vagy brómetilamin alkilálószerrel végezve). Végül az 5. sorban a reagens redukálása nélkül előidézett módosítás látható, ekkor a biotináló reagens diszulfid híd utáni biotin felé eső része is a peptiden marad).

Mivel a peptidek értékeléséhez különböző kereső algoritmusokat, valamint adatbázisokat használtunk, az eredményeket egységesítettük, a megbízhatóan azonosított

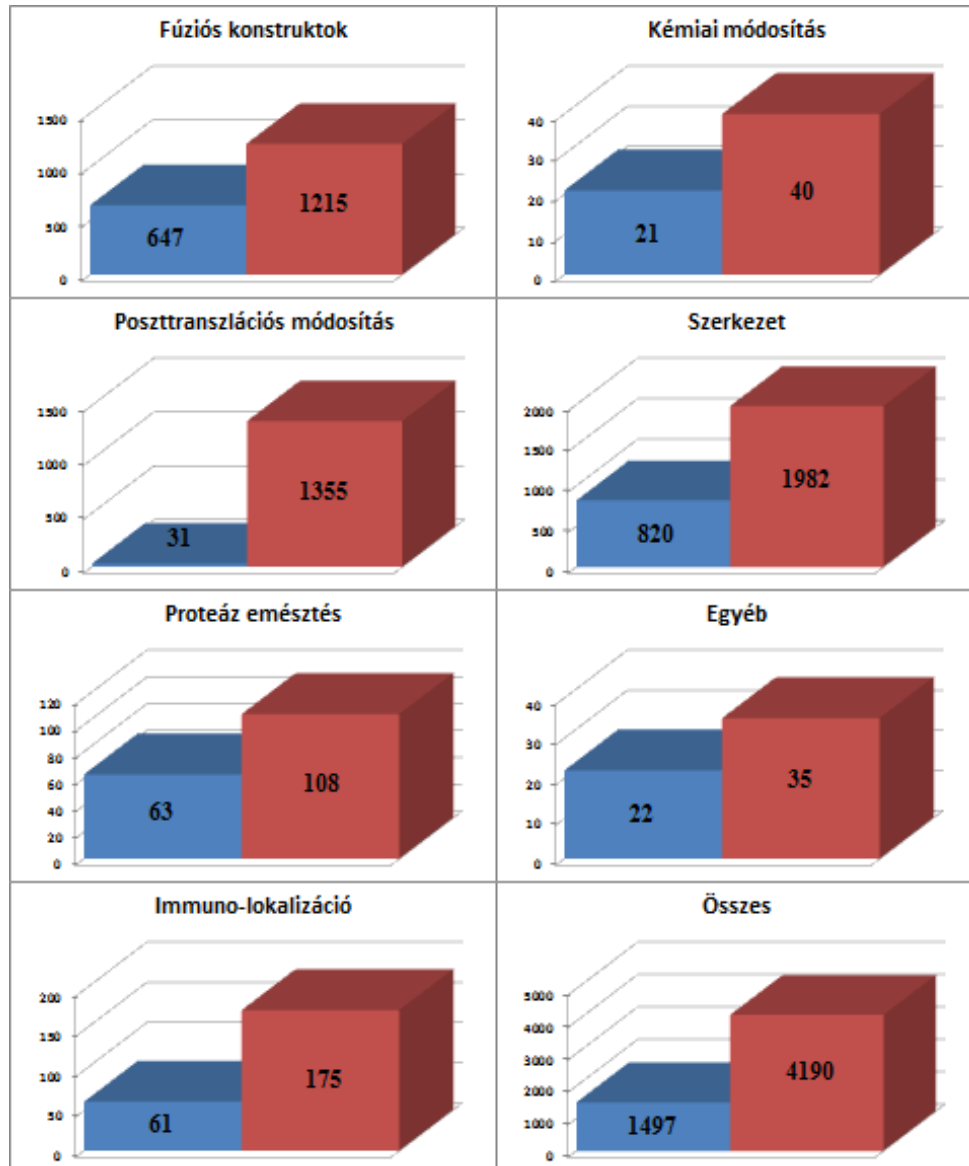
módosított peptideket visszaillesztettük a UniProt adatbázis humán szekvenciáira, majd a kapott fehérje halmazt szűrtük 95%-os szekvencia azonosságra. A CCTOP algoritmussal döntöttük el, hogy a keletkezett fehérje halmaz elemei közül, melyek TM fehérjék. A TM fehérjékhez tartozó, legalább három biológiai replikából meghatározott módosított pozíciókat tovább elemeztük.

Vizsgáltuk ezeknek a jelölt pozícióknak a szekvenciális környezetét, a ± 10 pozícióban található aminosavak eloszlását, amelyek a várt&detektált lizinek környezetét jellemzik. Ezt összevetettük a várt&nem detektált lizinek környezetének aminosav eloszlásával ugyanezekben a pozíciókban. A várt (elméletileg jelölhető) lizinet hordozó peptidek tömegspektrometriai detektálhatóságának meghatározása következők szerint történt: a lizin extra-citoszólikus TM régióban legyen, szemi-triptikus peptid maximum négy módosítással, m/z 380-1600 közötti értéken legyen, és a töltés 2-4 között. Ezután a jelölhető pozíció körüli ± 10 aminosavat kivágtuk a szekvenciából, és a várt&detektált, valamint a várt&nem detektált lizinek előtti és utáni szekvenciák szekvencia logóit Seq2Logo szoftverrel készítettük el¹⁵³. A torzítás elkerülése érdekében a szekvencákat szűrtük 40% azonosságra CD-HIT¹⁵⁴ segítségével. Azon hurok/loop régióknak a hosszeloszlását is elemeztük, amelyben a jelölt pozíciók előfordulnak. A TOPDB adatbázisban megvizsgáltuk, hogy a megtalált jelölt szegmensekre léteznek-e már korábbi kísérletes topológiai adatok. Végül számításokat végeztünk arra, hogy a CCTOP becslő algoritmus pontosságát mennyire tudja növelni a módszer által generált adott arányú (megtalált extra-citoszólikus lizin aminosavak és az extra-citoszólikusán jószolt lizinek hányadosa) és minőségű (a korábbi kísérletes adatokkal való egyezőség mértéke) extra-citoszólikus lizin aminosav megszorítás.

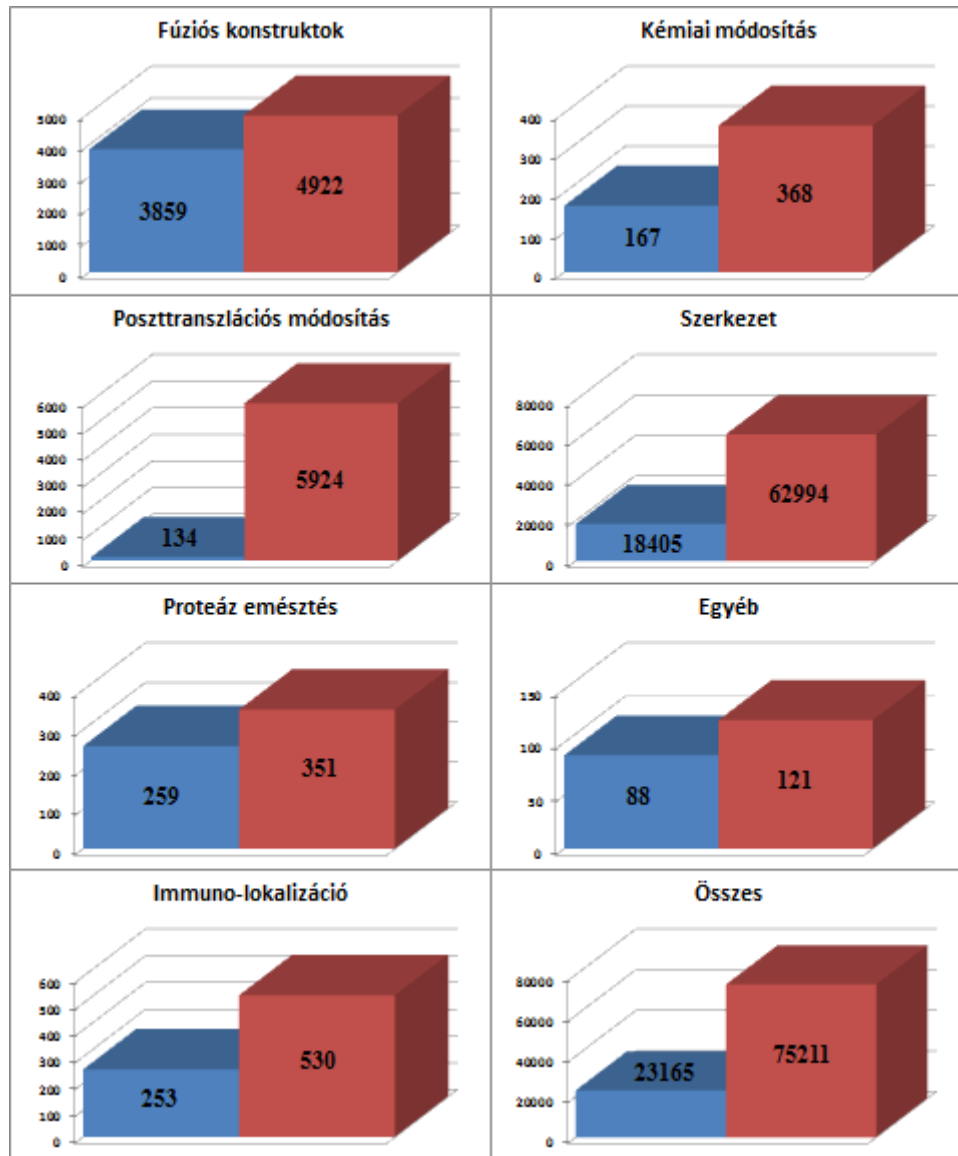
4 Eredmények

4.1 A TOPDB adatbázis frissítése kísérletes topológiai adatok gyűjtésével

A 3.3.1 fejezetben felsorolt forrásokat átvizsgáltuk, az azokban található topológiai adatokat kigyűjtöttük és kibővítettük a TOPDB adatbázist. Az 7. és 8. ábrán mutatjuk be a korábbi és a frissített adatbázis egyedi azonosító számainak és topológiai adat számainak kísérlet típusonkénti változását.



7. ábra: A korábbi (kék) és frissített (piros) adatbázis egyedi azonosító számának változása kísérlet típusonként



8. ábra: A korábbi (kék) és frissített adatbázis (piros) topológiai adat számának változása kísérlet típusonként

A korábbi és frissített adatbázis számaadatait arányaiban tekintve a legnagyobb növekedés a kísérleti módszerek közül a poszttranszlációs módosításoknál következett be, az N- és O- glikozilációs helyek meghatározásra fejlesztett HT proteomikai kísérletek által. Jelenleg a TOPDB 1400 TM fehérjéhez tartalmaz legalább egy HT módszerrel meghatározott topológiai adatot. A legnagyobb topológiai adat mennyiséget az új meghatározott 3D szerkezetek szolgáltatták, hiszen egy-egy TM fehérjeszerkezetének meghatározása, egy TM fehérje összes szegmensének helyzetét (citoszólikus, membrán, extra-citoszólikus) megadja. A TOPDB adatbázis legutóbbi frissítése óta sok szerkezetet határoztak meg. A TMDet algoritmus felhasználásával azonosítottuk, hogy a megoldott szerkezet közül, melyek TM fehérjék, és hogy a fehérjék szekvenciájában hol helyezkednek el a TM régiók. A

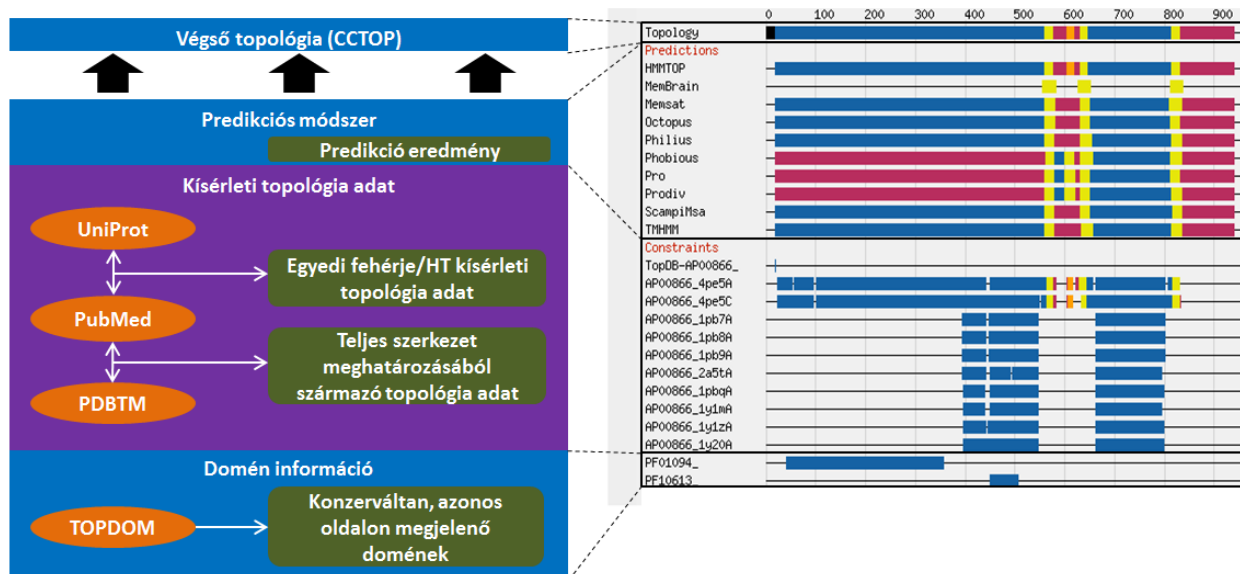
meghatározott szerkezeti információkkal szintén bővítettük a TOPDB adatbázist, így 1300-zal nőtt azon TM fehérjék száma, amelyekhez 3D szerkezeti adat áll rendelkezésre. A PDBTM adatbázisban ezeknek a fehérjéknek a membránhoz való irányultsága nincs meghatározva, ezért az eredeti szerkezetet közlő cikkekből gyűjtöttünk információkat az egyes fehérjeszegmensek membránhoz viszonyított helyzetéről és ez alapján egészítettük ki a pontos irányultságokkal a TOPDB adatbázist. A PDB adatbázisban egyes esetekben a TM fehérjékhez köthető citoszólikus és extra-citoszólikus domének 3D szerkezete is megtalálható. Ezeket a TMDET algoritmus nem tudja kiszűrni, mint TM fehérje. Ezek azonosítására a UniProt adatbázis összes olyan azonosítóját átvizsgáltuk, amely tartalmazta az 'FT transmem' sort vagy a membrán lokalizációt. A talált azonosítók közül, amelyeknél referenciaként egy vagy több PDB azonosító volt megjelölve, kerestük azokat a PDB azonosítókat, amelyeket a TMDET nem azonosított, majd manuálisan értékeltük az azokhoz tartozó cikkeket. Összesen 898 UniProt bejegyzéshez tartozott 2797 PDB referencia, melyekből a fentebbi szűrések által további 653 új TM fehérjével bővítettük a TOPDB adatbázist.

A 3D szerkezetmeghatározás mellett léteznek további népszerű módszerek, amelyek egyedi fehérjékről szolgáltatnak információt (ahogy az 7-8. ábrán is látható): a fúziós konstrukciók vizsgálatán alapuló módszerek, a poszttranszlációs módosulások meghatározása vagy megváltoztatására épülő módszerek, proteáz emésztéses módszerek, TM fehérjék TM szegmensei között levő régiók elhelyezkedését immunolokalizációval meghatározó technikák, a specifikus kémiai aminosav módosításokat használó módszerek stb. A felsorolt módszerekkel összesen 1438 új TM fehérjével bővítettük a TOPDB adatbázist, legalább egy topológiai adattal fehérjénként.

A TOPDB adatbázis korábbi adatmennyiségét háromszorosára növeltük a gyűjtött topológiai információkkal, ezáltal a kísérleti módszerekkel meghatározott topológiai adatok legnagyobb tárhelyévé vált.

4.2 A TOPDB adatbázis és a CCTOP kapcsolata

A TOPDB adatbázis olyan TM fehérjeszekvenciákat tartalmaz, amelyekhez legalább egy irodalmi topológiai adatot találtunk. A csoportunk által fejlesztett CCTOP algoritmus a meglévő topológiai ismeretek felhasználásával (becslő algoritmusok eredménye, TOPDB adatbázisban található topológiai adatokat, valamint TOPDOM adatbázisban található doméneket) becsli meg a TM fehérjék topológiáját. A TOPDB és a CCTOP kapcsolatát a 9. ábra bal oldala szemlélteti.



9. ábra: A CCTOP becsléshez felhasznált topológiai adatok forrásainak összesítő ábrája (bal oldal), valamint egy példa fehérje (Glutamate receptor ionotropic, NMDA 1, UniProt azonosítója: P35439) szekvenciáján a topológiai adatok megjelenítése (jobb oldal). Az ábra bal oldalán lila háttéren látható a TOPDB adatbázisba rendszerezett kísérleti topológiai adatok, kék háttéren pedig bioinformatikai módszerekkel meghatározott topológiai adatok. A narancssárga ellipszisek az adatbázisokat jelölik, ahonnan a topológiai adat származik, zöld lekerekített téglalapban pedig a származtatott topológiai információ figyelhető meg. Az ábra jobb oldalán látható egy példa fehérje a TOPDB-ben megjelenítve, a felső sorban a szekvenciájára a CCTOP algoritmus által meghatározott topológia becslés (a kézzel jelölt szakaszok az extra-citoszólíkus részek, a sárgával jelölt szakaszok a TM szegmens részek, végül a pirossal jelölt szakaszok a citoszólíkus részek). A következő sorokban látható a CCTOP által felhasznált különböző topológia predikációs algoritmusok (Predictions), valamint a példa fehérjeszekvenciájára adott topológia becsléseik, végül a TOPDB és a TOPDOM adatbázisból származó topológiai adatok általi megszorítások (Constraints). Forrás:

<http://topdb.enzim.hu/>

Egy példa fehérjéhez gyűjtött topológiaadatok és a fehérje CCTOP algoritmussal meghatározott topológiája a 9. ábrán jobb oldalt látható. A különböző kísérleti és bioinformatikai módszerek topológiai adatait felhasználva a CCTOP algoritmus a példa fehérje szekvenciájához becsült topológiáját a felső sor tartalmazza. Fentről-lefelé haladva a CCTOP predikció után először a 10 féle topológia becsülő algoritmus eredménye látható a

vizsgált fehérjeszekvenciára, majd a kísérleti (TOPDB) topológiaadatokat, valamint az azonos oldalon megjelenő doméneket (TOPDOM) figyelhetjük meg a „Constraints” feliratnál.

A TOPDB adatbázis a bemutatott példához hasonlóan összesen 1268 humán TM fehérjéhez tartalmaz legalább egy kísérleti vagy bioinformatikai úton meghatározott topológiai adatot. A jelenlegi ismereteink szerint ez a teljes humán TM proteomnak kevesebb, mint a negyede, amelyből következik, hogy a humán TM fehérjék jelentős hányadához még nem tudunk kísérletek alapján topológiaadatot rendelni. A további adatok szerzése érdekében új elméleti és gyakorlati módszerek fejlesztése szükséges, hogy minél több humán TM fehérjéhez létezzen legalább egy topológiaadat, amely jelentősen támogatná a teljes humán TM proteom topológia becslésének pontosságát.

4.3 A fejlesztendő módszer eredményeinek topológia becslésre gyakorolt hatásának bioinformatikai elemzése

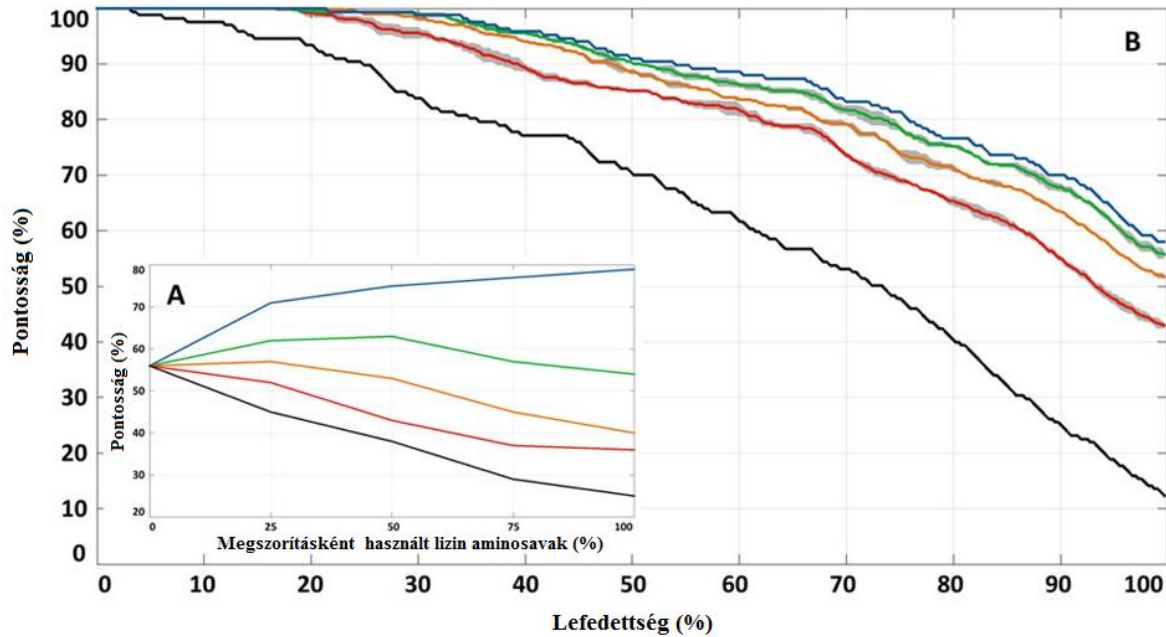
A TOPDB adatbázis frissítése alatt egyértelművé vált, hogy a jelenleg elérhető topológiai adatszerzésre is alkalmas módszerek megbízhatósága és érzékenysége nagy változékonyságot mutat, valamint a legtöbbjük nem alkalmas proteom szintű vizsgálatra. Továbbá a 4.2 fejezet végén tárgyaltuk, hogy a dolgozat tárgyát képező humán TM fehérjék jelentős részéhez még egyáltalán nincs kísérleti adat. Ezért humán TM fehérjékre topológiai adatot szolgáltató saját HT technika tervezését is elkezdtük (4.4. fejezet).

A módszer célja a TM fehérjék extra-citoszólikus hozzáférhető lizin aminosavainak specifikus jelölése és a jelölés segítségével az extracelluláris lizin pozíciók meghatározása. Mielőtt a kísérlet kivitelezésébe kezdtünk, megvizsgáltuk bioinformatikai elemzéssel, hogy ha ismernénk az extra-citoszólikus elhelyezkedő lizin aminosavak pozícióját, akkor javítanánk-e a topológia predikció pontosságán.

A megvalósítandó kísérletünk potenciális hatását a topológia becslésre egy emberi TM fehérje referencia-készleten analizáltuk, amely 333 nem redundáns TM fehérjeszekvenciát tartalmazott¹⁴. A fehérje halmaz 8099 extra-citoszólikus és 4892 intracelluláris lizint tartalmaz. A CCTOP algoritmus használatával elvégeztük a topológia predikciókat növekvő számú extra-citoszólikus lizineket megszorításként használva. Az esetleges torzítás elkerülése érdekében minden további bioinformatikai és kísérleti megszorítást elhanyagoltunk és csak a kiválasztott lizinek vettük figyelembe. Véletlenszerűen kiválasztottuk az extra-citoszólikus lizinek 25, 50 és 75%-át megszorításként, majd ennek függvényében összehasonlítottuk a becsült topológiák pontosságát a tesztzett TM fehérjéinek megállapított topológiájával. A 25,

50 és 75% extra-citoszólíkus lizin aminosav random kiválogatását megismételtük 50-szer, hogy kiszámoljuk a becslés pontosságának átlagát és szórását (10/B. ábra). A módszerünk elméleti korlátjának felméréséhez, olyan számításokat is végeztünk, amelyekben az összes extra-citoszólíkus lizint megszorításként használtuk (100% a 10/B. ábrán a kék vonal). A várakozásoknak megfelelően a topológia predikció pontossága szignifikánsan javult az extra-citoszólíkus lizin megszorításokkal. A számítások azt sugallják, hogy a becslés pontosságának a maximális növekedése 23% (56%-ról 79%-ra, 10/A. ábra kék vonal), amit az összes extra-citoszólíkus lizin megszorítással kaptunk meg. Az extra-citoszólíkus lizinek 20%-át megszorításként használva a topológia predikció pontossága 14%-kal növekedett (56%-ról 70%-ra).

Minden módszernek van hibája, így megvizsgáltuk, hogy mennyi hibát (úgy mint nem extra-citoszólíkus lizinek extra-citoszólíkusként azonosítása) lehet még tolerálni a fejlesztendő módszerrel szemben. A tévesen azonosított pozíciók hatását a becslés pontosságára úgy vizsgáltuk, hogy a véletlenszerűen kiválasztott 25, 50, 75 és 100% extra-citoszólíkus lizinek 4, 8, 12 és 16%-át random intracelluláris lizinekkal helyettesítettük. A 10/A. ábrán egyértelműen látható a predikció pontosságának drasztikus romlása már 4-8% citoszólíkus lizin extra-citoszólíkus megszorításként való használatával. Ezek alapján megállapítható, hogy a fejlesztendő kísérleti módszernek nagyon pontosnak kell lennie, hogy hozzájáruljon a becslés pontosságának növeléséhez.

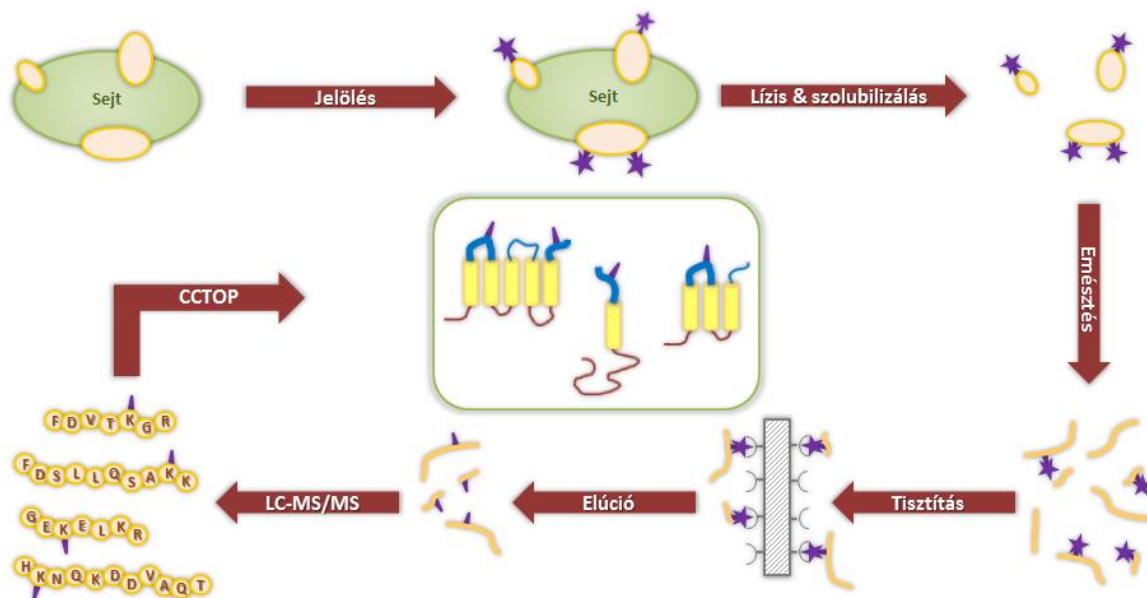


10. ábra: Az extra-citoszólikus lizinek felhasználásának a CCTOP topológia becslés pontosságára gyakorolt hatása a tesztszeten. A) A CCTOP predikció pontosságának változása a felhasznált extracelluláris lizin megszorítások százalékának függvényében. A megszorításként használt extra-citoszólikus lizineknek a 0, 4, 8, 12, és 16%-át lecseréltük intracelluláris lizinekre (kék, zöld, narancssárga, piros és fekete vonal), hogy vizsgálni tudjuk az esetleges hamis pozitív találatok hatását a becslésre. B) A topológia predikciókat sorbarendeztük a megbízhatósági értékük alapján és a legmegbízhatóbb fehérjehalmazokon mért pontosságot ábrázoltuk (a lefedettség az x tengelyen). A görbék színe jelzik a random kiválasztott extra-citoszólikus lizin arányát, amelyeket megszorításként használtunk a becslés során (kék, zöld, narancssárga, piros és fekete rendre a 100, 75, 50, 25 és 0% extra-citoszólikus lizin felhasználása). Az átlag értékeket a folytonos vonal, míg a szórást az árnyékolás jelzi.

4.4 A topológiai adatokat szolgáltató MSTOP kísérleti módszer fejlesztése

A módszerfejlesztés kezdeti kísérleteit vörösvértesteken végeztük. A Szulfo-NHS-SS-biotin reagenssel jelöltük a vörösvértesteket, majd a sejteket feltártuk, szolubilizáltuk és emésztettük, majd HPLC-MS/MS proteomikai elemzésnek vetettük alá. A komplex minta peptideinek azonosításánál az egyes fehérjék jelentős mennyiségi többsége (citoszólikus fehérjék: haemoglobin, aktin, tubulin stb.) megnehezítette az alacsonyabb kópiaszámmal rendelkező jelölt TM fehérjeszegmens azonosítását. Ezért a későbbiekben sejtek feltárása

után membránpreparátumot készítettünk, amelyet tisztítottunk, hogy a citoszólikus, más sejtszervecskékből származó fehérjék, valamint a membránpreparátumhoz csak asszociálódó fehérjék számát csökkentjük. A membránpreparátum elkészítése már jelentősen segítette a jelölt peptidek azonosítását, azonban még így is számos jelöletlen (a kísérlet szempontjából háttérként vagy jelölt peptidek azonosítást nehezítő faktoroként értelmezhető) peptid került azonosításra a mintákból. A membránpreparátum szolubilizálása és adott proteolitikus enzimmel történő emésztése után a jelöletlen peptid háttérrel a következő minták esetében tovább csökkentettük, úgy, hogy a jelölt peptideket a jelölő biotin részére specifikus neutravidin agarózzal töltött oszloppal dúsítottuk. Az oszlopon kikötött jelölt peptidek mellől különböző ionerejű és pH-jú pufferekkel (3.3.6. fejezet) távolítottuk el a jelöletlen peptidek jelentős részét, ezzel elősegítve a jelölt peptidek analitikai azonosítását.



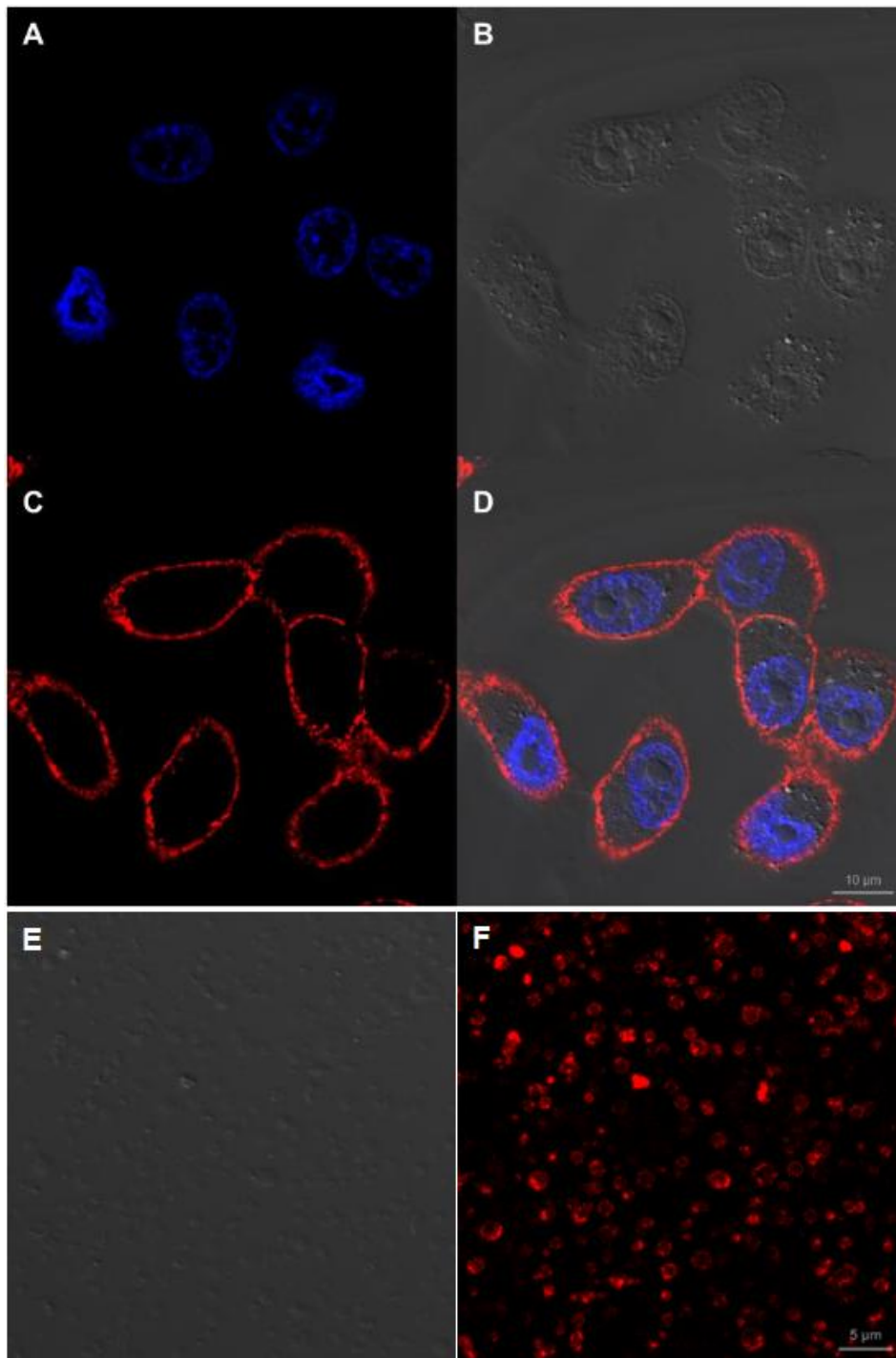
11. ábra: Az extra-citoszólikus térben hozzáférhető primer amin módosításra épülő HT topológiai adatot szolgáltató MSTOP módszer folyamatábrája. Az izolált sejteket, jelöljük membrán impermeabilis primer amin specifikus kémiai ágenssel (Szulfo-NHS-SS-biotin). A sejteket feltárjuk, a plazmamembrán frakciót tisztítjuk, majd szolubilizáljuk és emésztjük különböző proteolitikus enzimekkel. A módosított peptideket izoláljuk neutravidin agaróz gyöngyön, majd eluáljuk és szekvenáljuk tandem tömegspektrométerrel. A jelölt pozíciókat extra-citoszólikus megszorításként lehet használni a megfelelő TM fehérje szekvenciájára a CCTOP topológia becselő algoritmusban.

Az előkísérletek tapasztalatait felhasználva kialakítottuk a fejlesztendő módszer alapjait, amely a sejtfelszíni fehérjék szelektív kémiai jelölésére támaszkodik, a Szulfo-NHS-SS-biotin reagenssel, amely membrán impermeábilis, ezáltal csak az extra-citoszólíkus térrészben előforduló fehérje amino-terminálist, valamint lizin-oldalláncot képes jelölni. Egyes cikkekben ezt a kovalens jelölést használják sejtfelszíni proteomok azonosítására^{130,131}, illetve egy adott fehérje topológiájának jellemzésére¹¹⁶. Az általunk közölt módszerben¹⁵⁵ viszont a sejtfelszínen kifejeződött endogén TM fehérjék extra-citoszólíkus szegmenseit jelöltük a humán TM proteom topológia becslésének támogatása érdekében. Az ép sejtek felszínének jelölési körülményeit úgy optimalizáltuk, hogy minimalizáljuk a hamis pozitív találatokat (úgy mint jelölt citoszólíkus TM fehérje régiók).

Az izolált sejtek jelölését követően (3.3.3 fejezet) az előkísérletek megállapításai alapján a jelölt sejteket feltártuk, membránt prepráltunk, a preprátumot pedig fehérje bontó enzimekkel emésztettük (3.3.4-3.3.5 fejezet). A biotinált peptideket neutravidin agarózzal töltött oszlopon dúsítottuk, majd az oszlopról a jelölt peptideket a rendukálószerrel (DTT, TCEP) eluáltuk. A proteomikai elemzés előtt C18 oszlopon is tisztítottuk a mintákat (3.3.6 fejezet). A peptideket tandem MS készülékekkel szekvenáltuk és a módosítás helyét meghatároztuk (3.3.7.-3.3.8. fejezet). A kapott eredmény listából a 3.3.9-es fejezetben ismertetett szerkezetekkel kovalensen módosított peptideket kiszűrtük, a megfelelő TM fehérjékhez tartozó módosított peptidek módosított pozícióját extra-citoszólíkus megszorításként tudtuk használni CCTOP topológia becselő algoritmusban. A fejlesztett módszer leegyszerűsített folyamat ábráját a 11. ábra szemlélteti, a technikát MSTOP eljárásnak neveztük el.

4.4.1 A jelölés maximalizálása ép sejtek felszínén, valamint a sejtingegritásra gyakorolt hatása

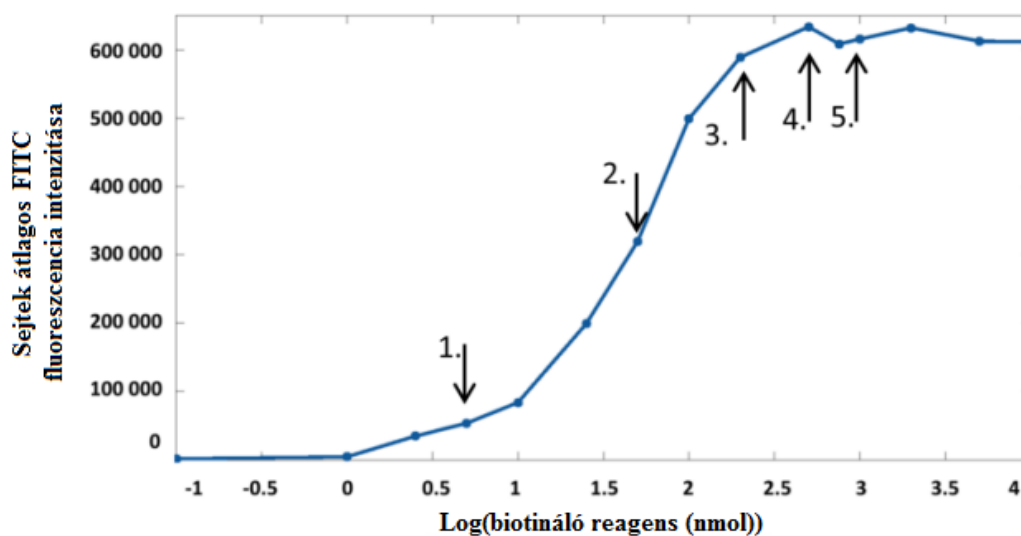
A sejtfelszíni biotinálás eredményességéről először letapadó kínai hörcsög petefészek sejteken győződünk meg konfokális mikroszkóp felhasználásával (12. ábra).



12. ábra: A CHO sejtek sejtfelszíni biotinálása Szulfo-NHS-SS-biotin reagenssel, majd TexasRed konjugált avidinnel a biotináltság tesztelése, DNS festése Hoechst 33342 reagenssel. Az A blokkban Hoechst (DNS marker), B blokkban differenciál interferencia kontraszt (DIC) vizsgálat, C blokkban TexasRed (jelölt extra-citoszólíkus rész) és a D blokkban egyesítettük az előző három blokk képeit. A vörösvértestekből készített membránpreparátum biotináltságának tesztelése szintén TexasRed konjugált avidinnel történt (F), az E blokkban a DIC látható.

A CHO sejteken elvégzett kísérlet igazolta, hogy a felszíni biotinálás működik, mivel a biotinra specifikus TexasRed konjugált avidin lokalizációját a jelölt sejtek felszínén figyeltük meg (piros szín, a C és D blokkban a 12. ábrán, az A blokkban a Hoechst DNS festék, B blokkban a DIC vizsgálat). A vak (jelöletlen) mintákban ellenben nem volt megfigyelhető a TexasRed festék felszíni felhalmozódása (7.4.1 fejezet, 28. ábra). A Szulfo-NHS-SS-biotin membránpreparálás utáni jelenlétét is megvizsgáltuk. Egy RBC izolátumot két részre osztottunk, az egyik részt 3.3.3 fejezet alapján jelöltük, a másik részt nem, majd mindkét részből a 3.3.4 fejezetben írtak szerint membránpreparátumot készítettünk a sejtek feltárása után. A keletkezett membránpreparátumokat szintén a biotinra specifikus TexasRed konjugált avidinnel inkubáltuk és konfokális mikroszkóppal vizsgáltuk. A jelölt minta esetében igazoltuk, hogy a membránpreparátumban is megfigyelhető a biotin jelenléte (piros szín a 12. ábra F blokkjában, E blokkban DIC), jelöletlen minta esetében viszont nem detektálható jelentős háttérfluoreszcencia.

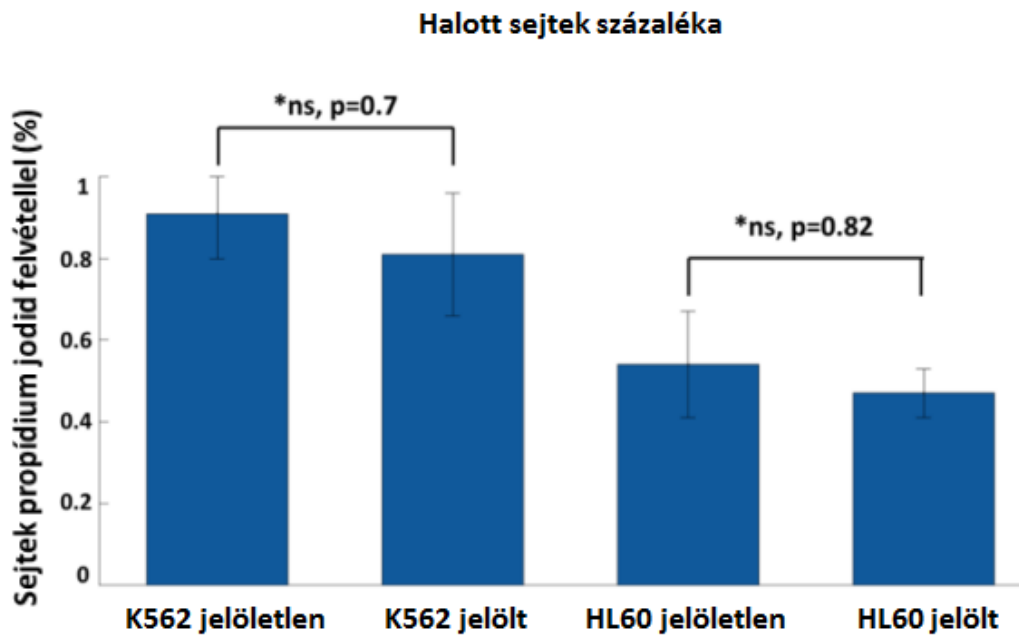
A céljaink érdekében az ép sejtek felszínének biotinálásának maximalizálásához szükséges Szulfo-NHS-SS-biotin reagens mennyiséget, végkoncentrációját is meg akartuk határozni a sejtszám függvényében. A maximális biotinálással növelhetjük annak az esélyét, hogy a hozzáférhető extra-citoszólikus szegmensek mindegyikét jelölni tudjuk és izolálásuk után megfelelő mennyiségben legyenek jelen a tömegspektrometriás kimutatáshoz és azonosításhoz. A 3.3.3.1 fejezetben ismertetett módon 1 jelöletlen és 15 különböző koncentrációjú jelölő ágenssel inkubáltunk egységes mennyiségű RBC-t. A sejteket FITC konjugált avidinnel inkubáltuk, így az átlagos sejtenkénti FITC intenzitását értékeltük a 16 sejtpopulációra áramlási citometriával, amely közvetett információt szolgáltatott a sejtek felszínének biotinálódásáról.



13. ábra: Sejtfelszíni biotinálás telítési görbéjének meghatározása áramlási citometriával, az alkalmazott biotináló reagens mennyiségének függvényében, FITC konjugált avidin felhasználásával. A nyilak (1-től-5-ig) jelzik azokat a jelölő koncentrációkat, amelyeknél membránpreparátumokat készítettünk a sejtekből és azokat dot-blottal elemeztük (18. ábra B2-től B6-ig).

A jelölő ~200-300 nmol anyagmennyiségben alkalmazva telítette az alkalmazott mennyiségű sejtek sejtfelszínét (13. ábra), amely a használt végtérfofogatokkal átszámolva ~1-2 mM végkoncentrációjú biotináló ágensnek felel meg, amely összhangban van a reagens más cikkekben használt végkoncentrációival¹⁵⁶.

Arról is meg akartunk győződni, hogy a megállapított maximális jelölő végkoncentráció a kezelt sejtek integritását befolyásolja-e a megválasztott reakciókörülmények között. Ezt három-három jelületlen és jelölt K562 és HL60 sejthalmaz propidium-jodid felvételével teszteltük (7.3.2. fejezet).

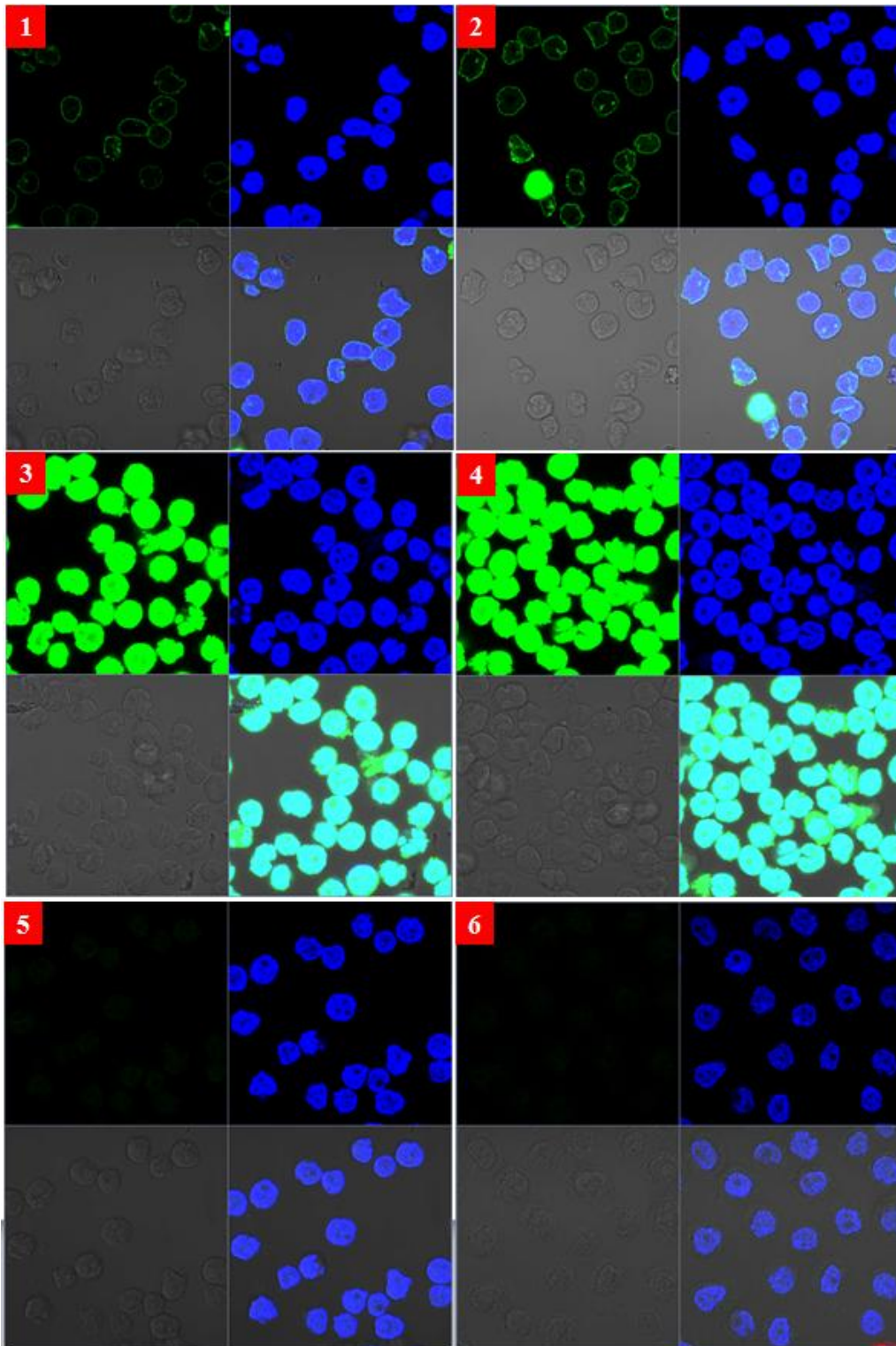


14. ábra: Jelöletlen és jelölt K562 és HL60 „halott” (PI pozitív) sejtek százaléka (* ns= nem szignifikáns a különbség a nem parametrikus Mann-Whitney teszttel)

Az áramlási citometriával megvizsgált sejtek propidium-jodid felvétele nem mutatott szignifikáns különbséget a jelölt és jelöletlen sejtek között (14. ábra). Kijelenthető, hogy a sejtek integritására nincs hatással a 2 mM Szulfo-NHS-SS-biotin, a PBS pufferben 4°C-on történő 20 perces inkubálás alatt.

4.4.2 A Szulfo-NHS-SS-biotin membrán impermeabilitásának vizsgálata

A primer amin specifikus jelölési technikával az extra-citoszólikus szegmensek szelektív jelölése volt a célunk. Mivel a jelölő ágens membrán impermeábilis, ezért elméletben csak permeabilizálás után juttatható be a citoszólba. A céljaink érdekében a reagens ezen attribútumát mindeképpen igazolni akartuk, hogy a folyamatot továbbvihessük. A 3.3.3.2 fejezetben részletezett módon, valamint az 5. ábra alapján vizsgáltuk a jelölő membrán impermeábilis tulajdonságát.



15. ábra: A jelölő ágens membrán impermeabilitás tesztje HL60 sejtvonalakon. A számok (1-től 6-ig) a 3.3.3.2 fejezetben ismertetett 5. ábra megfelelő minta sorszámainak felelnek meg.

Az ott leírt mintaelőkészítés után a biotinált részeket FITC konjugált anti-biotin antitesttel jelenítettük meg, valamint Hoechst DNS festékkel vizualizáltuk a sejtek magjait. Az egyes mintáknál a bal-felső sarokban a FITC, a jobb-felső sarokban a Hoechst festék, a bal-alsó a DIC kép, végül a jobb-alsó a kombinált (merge) kép látható.

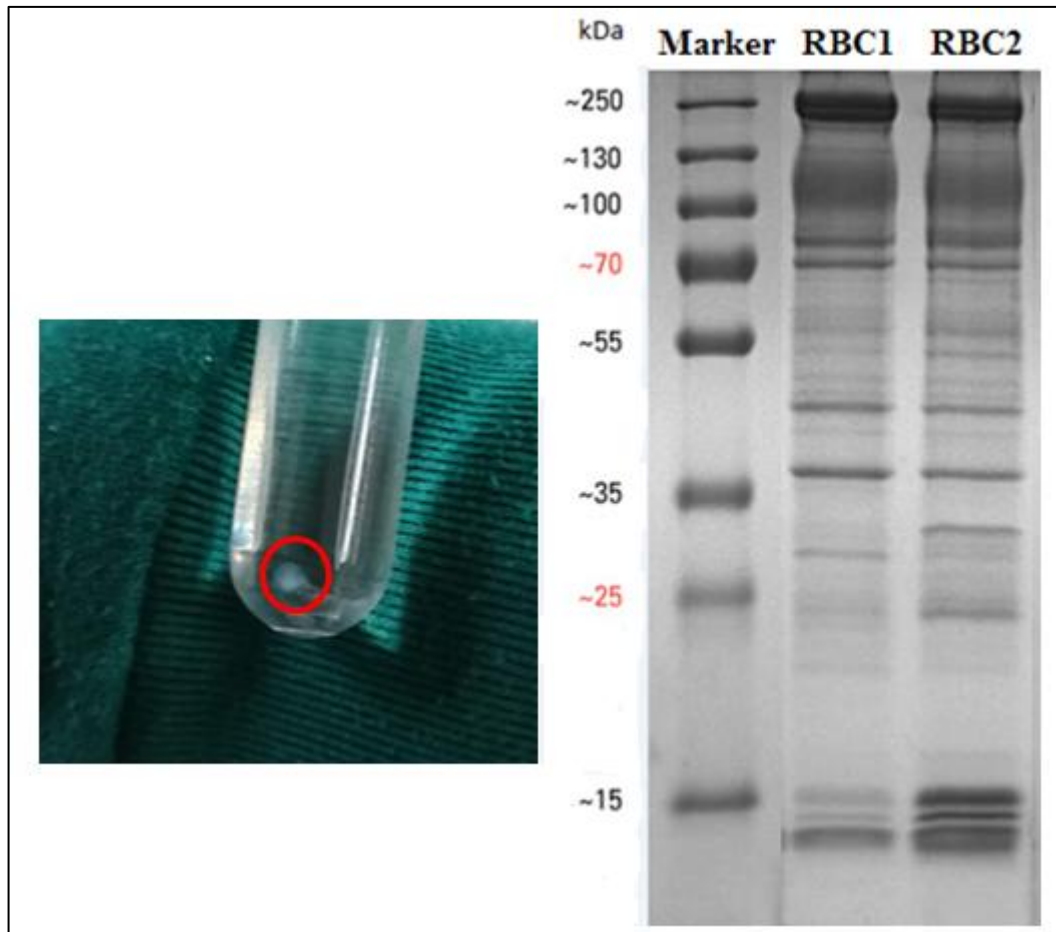
A FITC konjugált anti-biotin antitesttel a biotinnal jelölt egységek sejten belüli elhelyezkedését tudtuk vizsgálni, az eltérő mintaelőkészítési lépések függvényében (3.3.3.2 fejezet). A 15. ábrán a zöld színű fluoreszcens jelek mutatják meg a biotinált részek lokalizációját, míg a Hoechst festék a sejtek magját jelzi a kék színű fluoreszcenciával. Az 1. minta permeabilizálás nélkül jelölt, a FITC konjugált antitest a sejtfelületen felhalmozódik, míg a 6. jelöletlen és permeabilizálatlan mintánál az elvárásoknak megfelelően FITC jel nem látható. TritonX detergenssel kezeltük az egyes mintához hasonlóan jelölt 2-es és 3-as mintákat, valamint 6-os mintához hasonlóan jelöletlen 4-es és 5-ös mintákat, mely hatására az antitest és a biotin jelölő ágens számára átjárhatóvá vált a membrán. A négy minta közül a 3. (jelölt) és a 4. (jelöletlen) mintákat permeabilizálás után inkubáltuk a 2 mM Szulfo-NHS-SS-biotinnal, így lehetővé vált a sejten belüli biotinálás is. A FITC festék elhelyezkedése a 3. és 4. mintánál egyértelműen kiterjedt a konfokális mikroszkóppal rögzített sejtek síkszeleteinek teljes területére. Az 5. permeabilizált vak mintánál nem mutatkozott háttérfluoreszcencia (a FITC konjugált antitest az elvárásoknak megfelelően nem halmozódott fel a sejten belül). A 4. jelölt és azután permeabilizált mintánál az első mintához hasonlóan a plazmamembrán külső felületén halmozódott fel a FITC jel (egy sejt kivételével (15. ábra/2)), a sejtek belsejében nem mutatkozott zöld fluoreszcencia.

Az eredmények alapján egyértelműen kijelenthető, hogy a Szulfo-NHS-SS-biotin alkalmas extra-citoszólikus TM fehérjeszégmensek szelektív jelölésére, a használt reakció körülmények között a membrán impermeabilis maradt a jelölő molekula számára.

4.4.3 Membránpreparátumok készítése, fehérje tartalom mérése, minőség ellenőrzés SDS-PAGE-gélelektroforézissel

A sejteket az 4.4.1-es fejezetben tesztelt reakciókörülmények között, a maximális sejtfelületi jelölést biztosító koncentrációjú biotináló ágenssel jelöltük, majd a sejteket feltártuk hipotóniás lízis pufferben és ultracentrifuga felhasználásával membránpreparátumokat készítettünk a 3.3.4-es fejezetben részletezett módon. A preparátumok fehérjetartalmát Lowry és munkatársai¹⁵⁰ által fejlesztett módszerrel határoztuk meg, a 7.3.3-as fejezetben leírtak szerint. Az egy adott sejtípus membránpreparálási műveletének reprodukálhatóságáról SDS-PAGE gélelektroforézissel (3.3.4.1 fejezet) győződünk meg, gélzsebenként 10-20 µg fehérjetartalmú preparátumok felvitelével (16. ábra). Egy adott időpontban párhuzamosan és időben eltolva is elvégeztünk technikailag

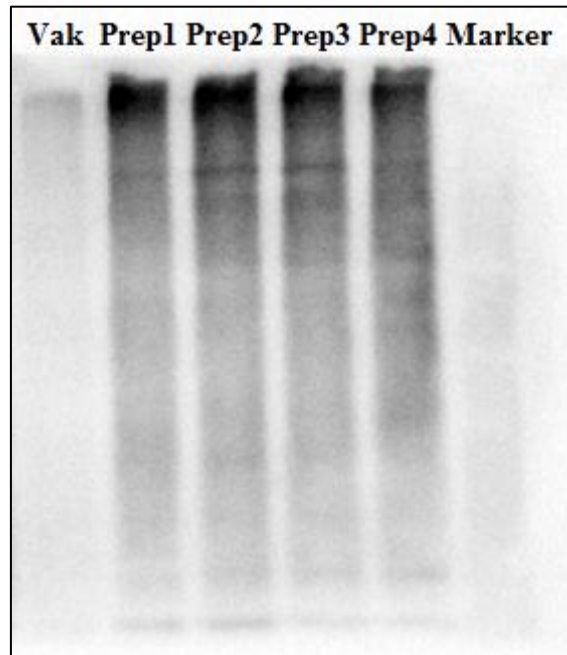
ugyanolyan membránpreparálásokat a megfelelő sejtekből és vizsgáltuk mennyire hasonló fehérje mintázatokat tapasztalunk a gélen Coomassie festés után. Így vizsgáltuk meg azt is, hogy ezek a jellemző sávok eltűnnek-e, ha fehérjebontó enzimekkel 16 órát kezeltük a szolubilizált membránpreparátumokat.



16. ábra: Membránpellet az ultracentrifugálás után, valamint két eltérő időpontban készített RBC (RBC1,2) membránpreparátum SDS-PAGE gélelektroforézise

A 16. ábrán bemutatott membránpelletből több mint 100 készült. Az azonos mintatípusból készített minták SDS-PAGE gélelektroforézise után látható fehérje sávok, a bemutatott példához nagy hasonlóságot mutattak. Ebből következik, hogy membránpreparálás folyamata ezzel a módszerrel ellenőrizve reprodukálhatónak mutatkozik. Ezt megerősíti, hogy ugyanannyi sejtől készített membránpellet teljes fehérjetartalma is jó egyezést mutatott. Végül a membránpreparátumokat szolubilizálva és 16 órán keresztül emésztve, majd SDS-PAGE-en futtatva a gélen emésztés nélkül látható éles fehérjesávok teljesen eltűntek, tehát az emésztés folyamata végbement.

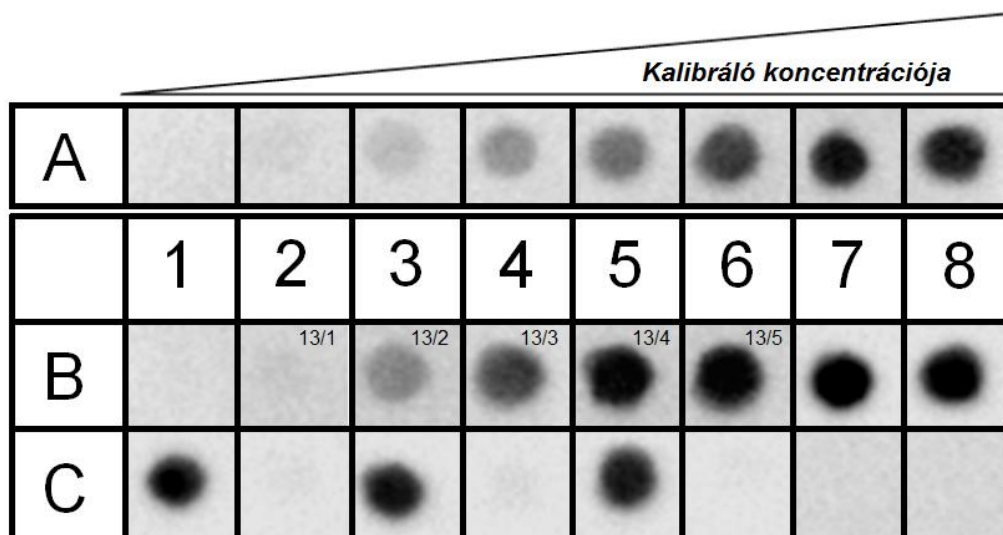
A redukálószer nélkül futatott emésztetlen mintákat Western-blottal is vizsgáltuk (3.3.4.2 fejezet) és az az egyes fehérjesávok biotináltságát HRP-konjugált avidint alkalmazva igazoltuk (17. ábra).



17. ábra: Redukálószerrel nem kezelt biotinálátlan (vak) és biotinált (Prep1-4) membránpreparátumok Western-blot analízise (előhívószer: HRP konjugált avidin)

4.4.4 A biotinált peptidek követésére fejlesztett dot-blot módszer

A sejtekből készített szolubilizált membránpreparátum 16 órán keresztül emésztése után keletkezett komplex peptidminták biotintartalmát normál vagy nagy kapacitású avidin agaróz gyöngyön (Thermo Scientific) izoláltuk. Az elméletileg biotinálható sejtfelszíni primer aminos csoportokra csak közelítő számítások állnak rendelkezésre, valamint azok hozzáférhetőségéről is limitált információk léteznek. Így a minták teljes biotintartalmának megkötéséhez szükséges avidin agaróz gyöngy-mennyiség megválasztása nem határozható meg egyértelműen. Ezért a dot-blot technikával optimalizáltuk, hogy a biotinált, és a Lowry módszerrel meghatározott fehérjetartalmú, szolubilizált és emésztett membránpreparátum biotin tartalmának kikötéséhez mennyi avidin gyöngy szükséges (3.3.4.3. fejezet, 6. ábra).



18. ábra: A biotin tartalom szemikvantitatív vizsgálatára alkalmas dot-blot módszer egyik eredménye. Az „A” sorban egy biotinált membránpreparátum hígítási sora látható (1%, 2%, 10%, 25%, 50%, 75%, 90%, 100%). B1, C7 és C8: negatív kontrollok, B2-től-B6-ig a 13.

ábrán az 1-től-5-ig nyilak által jelzett jelölő koncentrációval feljelölt sejtek membránpreparátumainak biotináltságának vizsgálata. B7 és B8: pozitív kontrollok, majd a megfelelő enzimmel emésztett membránpreparátumok avidin agarózzal töltött oszlop előtti és utáni frakcióinak biotintartalmának összevetése: C1-C2 az MS Grade tripszin, C3-C4 a kimotripszin, C5-C6 a termolizin emésztett minták oszlop előtti és utáni frakciója

A 18. ábrán bemutatott példa első sorában (A sor) egy 2 mM végkoncentrációjú biotináló reagenssel feljelölt membránpreparátum hígításával a különböző minták biotin tartalmának szemikvantáláshoz megfelelő tartományt értünk el, ami fontos volt a további minták elemzéséhez. A dot-blot módszerrel az áramlási citometria alapján mért felszíni jelöléssel azonos hatékonyságot mutattunk ki: a 13. ábrán nyilakkal jelzett 5 ponton készített membránpreparátumok egységes hígítása (a 18. ábra első sorában meghatározott lekisebb hígítással (A8)) után a kemilumineszcens jelintenzitások különbsége megfigyelhető (B2-B6 pont), kivéve a B5 és B6 esetében, ahol áramlási citometriával sem láttunk jelentős biotináltsági különbséget. A legfontosabb eredmény, hogy a dot-blot módszer segítségével minimalizáltuk a biotinált membránpreparátumok teljes biotintartalmának kikötéséhez szükséges avidin agaróz gyöngy mennyiségét, ezzel költséghatékonyabbá tettük a módszert, valamint biotinált komponensek mennyiségi veszteségét is meggátoltuk (C2, C4 és C6 minták avidin agarózzal töltött oszlopon áteső frakciójában nem volt detektálható biotin jel).

A dot-blot technika alkalmas volt továbbá az alkalmazott membránpreparálási stratégia egyes frakcióinak biotintartalmának elemzéséhez is, amely igazolta, hogy a biotinált komponensek a membránpelletben nagyságrendekkel jobban dúsulnak, valamint azt, hogy a pellet mosogatásakor nem veszünk jelentős mennyiségű biotint. Az avidin agarózzal töltött oszlopon dúsított jelölt peptidek mellől különböző összetételű mosópufferekkel mostuk el a jelöletlen komponenseket, az oszlopon átesett mosópufferek biotintartalmát szintén dot-blot módszerrel elemeztük. Az olyan mosópuffereket, amelyek az oszlopról dot-blottal detektálható biotinvesztést eredményeztek elhagytuk a rendszerből (például: 1% NP40 és 0,1% TritonX-100).

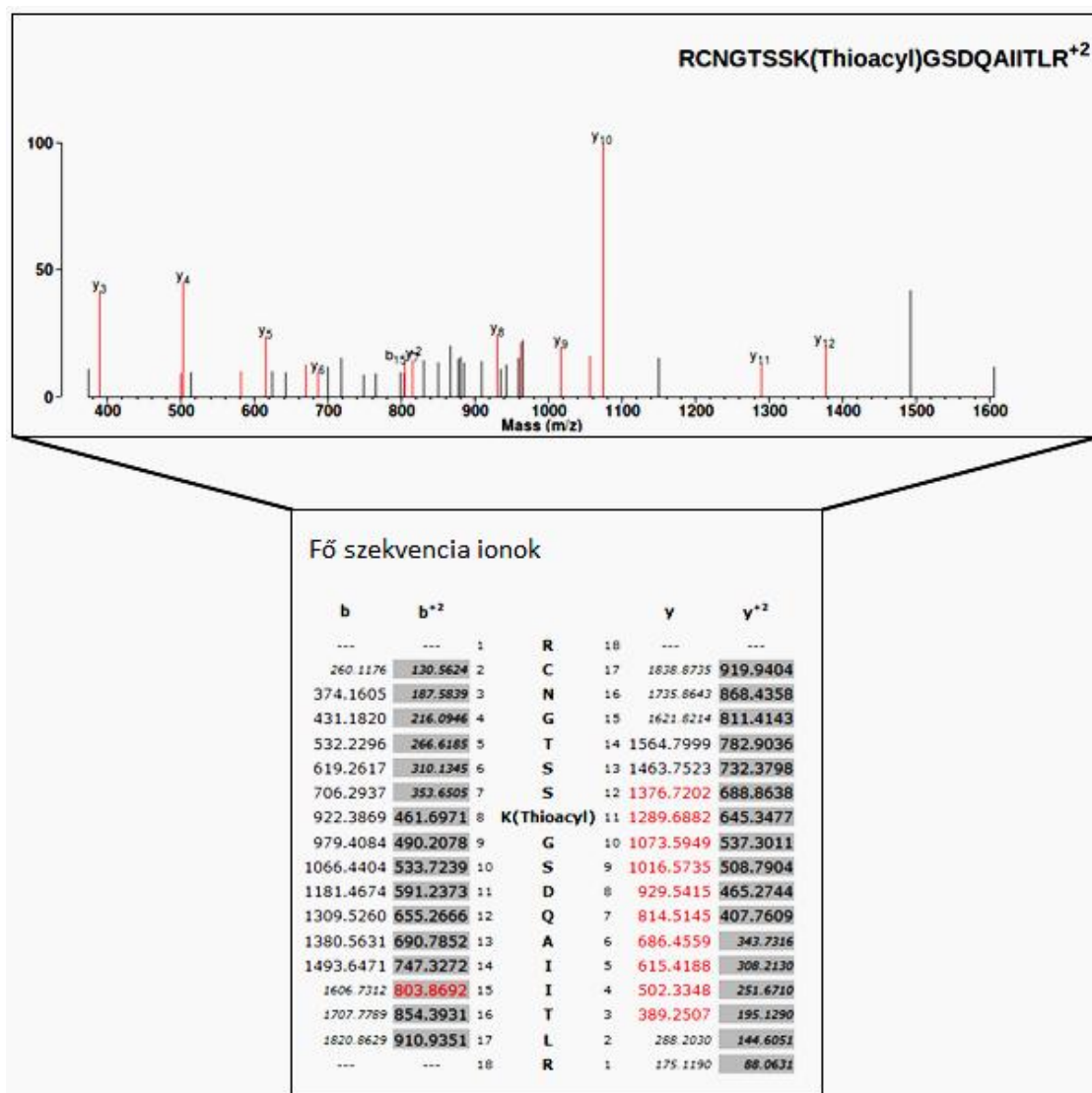
4.5 Eluálás körülményeinek analitikai eredményekre gyakorolt hatása

Az avidin agaróz gyöngyön dúsított jelölt peptideket a jelölő ágens diszulfid-hídjának redukálásával eluáltuk a gyöngyről. Kétféle redukálószer, TCEP-et, vagy DTT-t alkalmaztunk erre a célra a kezdeti RBC kísérleteknél. A TCEP redukálószer alkalmazásakor a 3.3.9-es fejezetben tárgyalt módosulások mellett, a tömegspektrometriai szekvenálás során megfigyeltünk egy nem várt alternatív módosulást a jelölt peptideken, amit „propionil” módosulásnak nevezünk. A jelölő redukálása után jelölt amino-csoportokon visszamaradó „tioacil” csoport, valamilyen mellékreakció folytán szulfhidril-csoportját elveszti, továbbá a jelenség kiterjedt néhány cisztein tartalmú peptidre, amelyekben cisztein aminosav hasonlóan módosult. A DTT redukálószer alkalmazásakor, ilyen jellegű módosulásokat az izolált peptideken analitikailag nem azonosítottunk, így a tömegspektrometriai lekeresésénél nem kellett növelnünk a megengedett módosulások számát (nem növekedett az elméletileg lehetséges peptid találatok száma), a további kísérletekben ezért csak DTT-vel dolgoztunk.

Az avidin agaróz gyöngyről többlépcsős eluálást is kipróbáltunk. Maximum háromszor eluálva a jelölt peptideket ugyanarról az avidin agarózyöngyről azt tapasztaltuk, hogy minden esetben az első eluálás eredményezett a legtöbb jelölt peptidet a proteomikai azonosítás során, a következő két eluátum analíziséből csak olyan jelölt peptideket kaptunk, amelyeket az első eluátumban is azonosítottunk. A további kísérletekben ezért csak egy eluálást végeztünk, az eluátumot 3.3.6-as fejezetben tárgyaltak szerint tisztítottuk, majd elemeztük tandem MS készülékekkel.

4.6 Vörösvértestekből készített jelölt membránpreparátumok analitikai eredményeiből szerzett tapasztalatok

A 4.4-es fejezetben ismertetett MSTOP módszert legelőször RBC-ken alkalmaztuk. A legelső mintákból kapott eredmények kevés jelölt peptidet eredményeztek, ráadásul nem TM fehérjékből, hanem globuláris fehérjékből (pl. fibrinogén, haemoglobin) kaptunk jelölt és jelöletlen peptideket. Az előző fejezetekben ismertetett optimalizálási lépések, valamint az avidin agarózzal töltött oszlop megfelelő mosási lépéseinek közbeiktatásával (3.3.6 fejezet) jelentős mennyiségű fehérjét azonosítottunk tömegspektrometriával (közel 50 db) és a megtalált fehérjék jelentős hányadához (~50%) legalább egy jelölt peptid társult (példa a 19. ábrán).

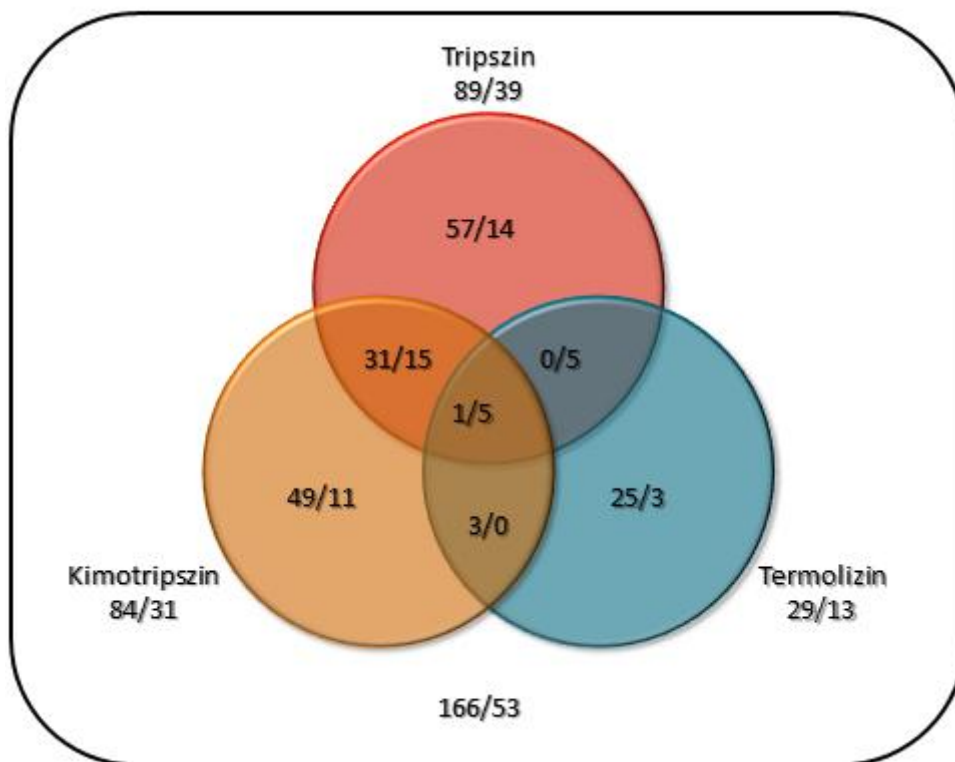


19. ábra: A Basigin fehérje jelölt peptidjének MS/MS spektruma. Pirossal kiemeltem a peptid megtalált fragmentationjai mind a spektrumban, mind a jellemző szekvencia ionok közül. A jobb felső sarokban a peptid szekvenciája és a módosításunk (Thioacyl) peptidszekvencián belüli helye látható.

A legelső minták 16 órán keresztül tripszinnel emésztett jelölt membránpreparátumok voltak, amely időintervallum jelentősen növelte a teljes folyamat kivitelezésének idejét. Meg akartuk tudni, hogy a 16 órás emésztés helyett, rövidebb idejű emésztésekkel növelhető-e az analitikailag azonosítható jelölt fehérje pozíciók száma. A membránpreparátumok 100-100 µg fehérjetartalmú részeit 2, 4, 6 és 8 órán keresztül emésztettük tripszinnel, majd az emésztést leállítottuk hőinaktiválással és inhibitor hozzáadásával, majd a keletkezett peptideket a 3.3.6-os fejezet alapján készítettük elő MS analízisre. Az analitikai eredmények jelölt peptid kihozatala alapján nem keletkeztek további találatokat a 16 órás emésztéshez

képest, sőt a 2 órás emésztés közel fele annyi jelölt pozíciót eredményezett. A továbbiakban a 16 órás emésztéseket alkalmaztuk a minták készítésekor, hogy az emésztési folyamat lejátszódásához elegendő időt biztosítsunk.

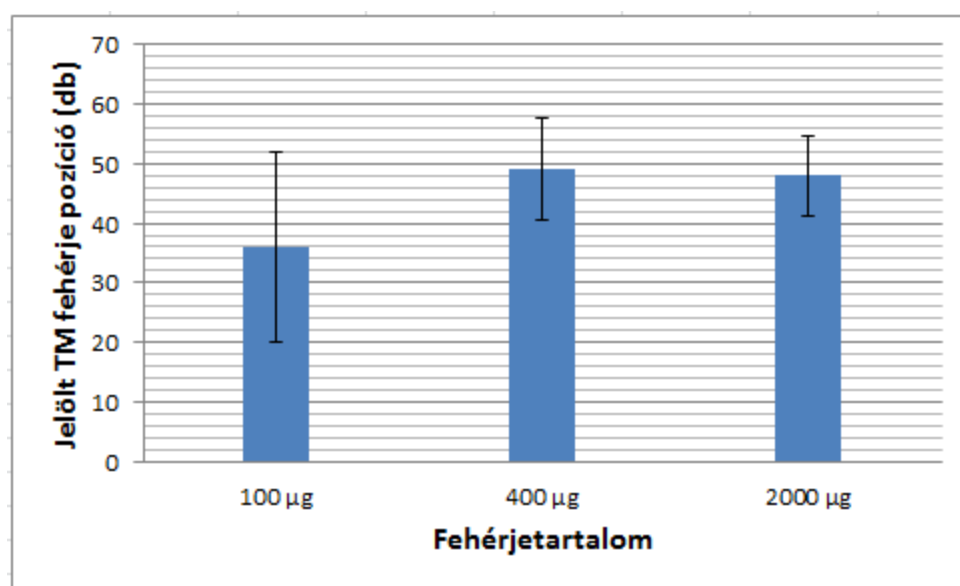
Az emésztési folyamat idejének változtatásán kívül más paraméterek változtatásával, például más fehérjebontó enzimekkel is teszteltük a módszert: párhuzamosan készítettünk tripszin, kimotripszin és termolizin enzimekkel emésztett mintákat. A három enzimből a tripszinnel készített minta produkálta a legtöbb jelölt TM fehérjét és hozzájuk tartozó jelölt peptidet. A kimotripszin és a termolizin enzimek általánosságban a tripszin enzimmel készített minta TM fehérjeihez szolgáltatott újabb jelölt peptideket, esetenként újabb jelölt pozíciókat is (20. ábra). Az eredmények alapján a tripszinnel készítettük a további mintákat, de a másik két enzimmel kapott eredményeket az eredmények összesítésénél felhasználtuk.



20. ábra: Az RBC membránpreparátum három féle fehérjebontó enzimmel (tripszin, kimotripszin, termolizin) történő kezelésének jelölt peptid/jelölt pozíció kihozatalának összevetése

Ezután teszteltük, hogy a módszerben alkalmazott szolubilizálandó membránpreparátumok fehérjetartalmának növelése segíti-e az MS készülékkel történő jelölt peptid azonosítást, növeli-e a megbízhatóan azonosított jelölt peptidek mennyiségét. A jelölt membránpreparátumokból - a Lowry módszerrel mért fehérjetartalmuk alapján - 10, 100, 400

μg és 2 mg fehérjetartalmú mennyiséget emésztettük az emésztési idő optimalizálásánál tapasztaltak alapján 16 órán keresztül tripszinnel, majd dúsítottuk arányosan növelve az avidin agaróz gyöngy mennyiségét a fehérjetartalmakhoz mérten, ugyanígy a mosási térfogatok mennyiségét is növeltük, a folyamat további része változatlan volt (3.3.6 fejezet). A 10 μg fehérjetartalmú preparátumok esetén alig figyeltünk meg jelölt peptidet, míg a 100 μg és nagyobb fehérjetartalmú preparátumokból kiinduló minták MS elemzése nem eredményezett többszörös mennyiségű jelölt peptidet, egyfajta telítés volt megfigyelhető, valószínűleg a dinamikus tartományba eső peptidek változatlansága miatt (mindenféle peptidből több lett, a háttérből is) (21. ábra). Így a továbbiakban 100-200 μg fehérjetartalmú preparátumokat alkalmaztunk egy-egy kísérlet elvégzéséhez.



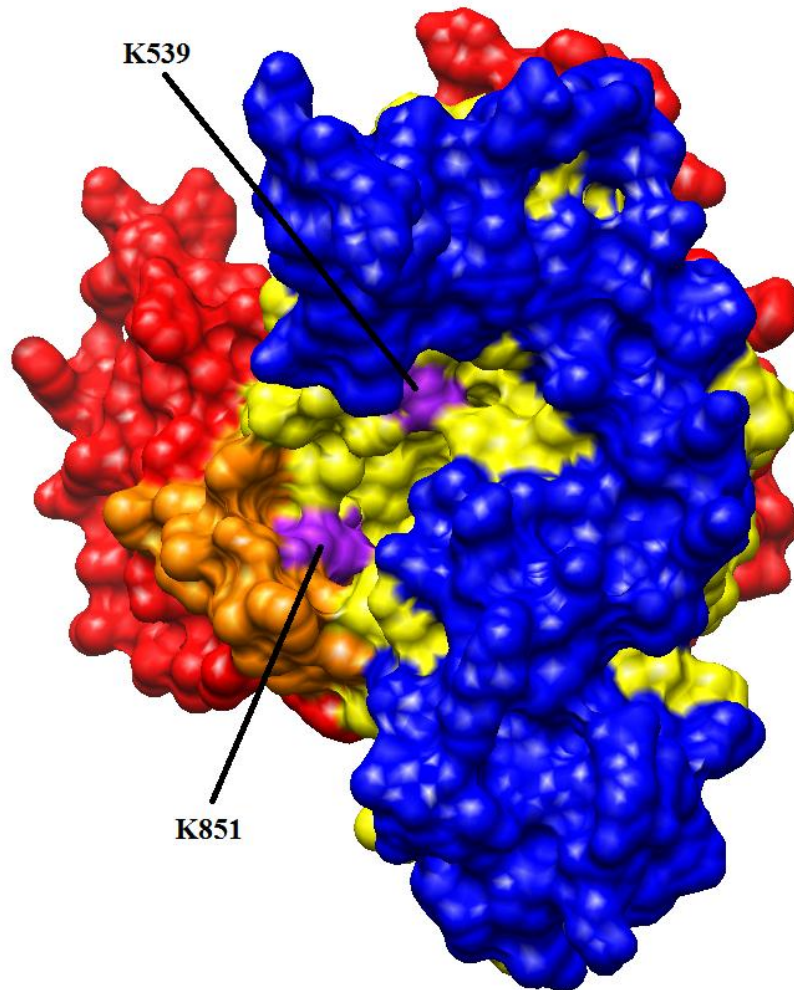
21. ábra: A különböző fehérjetartalmú membránpreparátumokból készített mintákból kijött TM fehérjékhez tartozó jelölt pozíciók kihozatalának összevetése

Az extra-citoszólikus térben a legjellemzőbb fehérje poszttranszlációs módosulás a glikoziláció¹⁵⁷, amely a feltételezhetően jelölhető extra-citoszólikus szegmensek tripszines emésztését nehezítheti, valamint a jelölt peptid izolálására és a MS azonosítására, szekvenálhatóságára is hatással lehet. Így a 16 órán keresztüli tripszines emésztés elé beépítettünk egy 2 órás PNGase F enzim kezelését a folyamatba, az aszparaginokon lévő acetilglükózamin cukor egységek lehasítása céljából. Az RBC-k plazmamembránjában legnagyobb kópiaszámban előforduló Band 3 anion transzporter esetén, -amelynek már az előző kísérletekből is jelentős jelölt peptid kihozatala volt- így növeltük az azonosított jelölt pozícióinak a számát. A 631K és 639K pozíciókkal bővült a módszerünk által extra-citoszólikus azonosított lizinek listája a Band 3 anion transzporter fehérjéből, amelyek az

irodalomban már igazolt 642-es pozícióban glikozilált aszparagin mellett találhatóak^{56,158,159}. A PNGase F kezelést, így a későbbi mintaelőkészítésnél mindig alkalmaztuk.

4.7 Vörösvértest minták és további minták analízisa, és jelölt fehérje és peptid kihozatalai

Az RBC mintákból összesen 65 nano UPLC-LTQ-Orbitrap Elite futtatást végeztük el, amely összesen 48 TM fehérjéről szolgáltatott 155 különböző jelölt pozíciót. Az eredmények között voltak olyan jelölt RBC TM fehérjék, amelyeknek térszerkezetét már valamelyik 3D szerkezet-meghatározó módszerrel (1.6 fejezet) megoldották, de találtunk olyanokat is, amelyek egyáltalán nem rendelkeztek még topológiaadattal. A 3D szerkezettel rendelkezőre példa a Band 3 anion transzporter, amelyből 8 db lizint sikerült megjelölni (lizinek szekvenciális pozíciója: 430, 539, 542, 551, 562, 631, 639, 851). Ezeket rávetítve a topológia becslésre, amely már figyelembe veszi a megoldott térszerkezetet és további topológiai adatokat, valamennyi jelölt pozíció extra-citoszólikus elhelyezkedésű volt (23. ábra/A). Az 539 és a 851 pozíció elfogadása azonban nem volt egyértelmű. Az 539-es lizin aminosav pozíció az 5. TM hélix extra-citoszólikus határán található, valószínűleg a reagensünk számára a membrán fluiditásából adódóan olykor hozzáférhető. A 851-es pozícióban lévő lizin aminosav egy citoszólikus oldalra visszahurkolódó régió közepén található. A szegmens környékét egy cisztein mutagenézisre épülő módszerrel vizsgálták korábban. Egy membrán impermeábilis, ciszteinre specifikus reagens jelölni tudta 852-től-855-ig terjedő aminosav pozíciókba bevezetett egyedi cisztein aminosavakat⁵⁶. A 3D szerkezetet elemezve és a membránfluiditást jelenségét figyelembe véve előfordulhat, hogy ez a szakasz extra-citoszólikus oldalról a jelölt molekula számára elérhetővé válik (22. ábra). A Band 3 anion transzporter fehérje eredményei megerősítik a módszer által szolgáltatott adatok minőségét, hiszen gyakorlati és elméleti eredményekkel is szinkronban voltak.

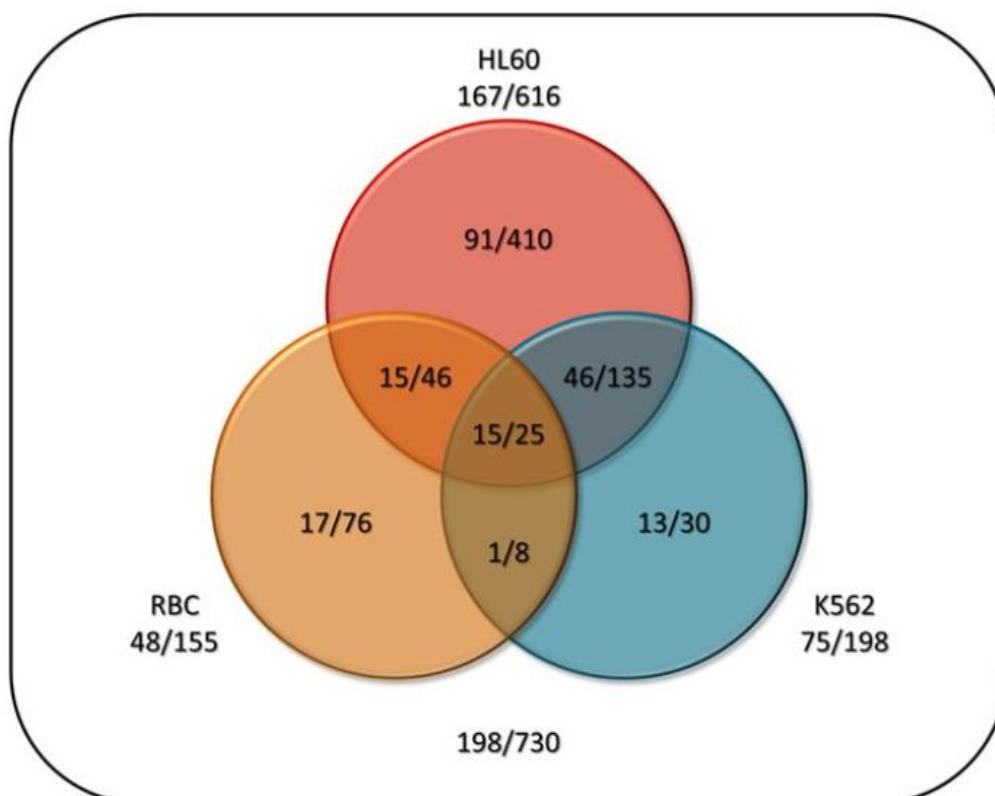


22. ábra: A Band 3 anion transzporter 539. és 851. lizinje a 3D felszíni szerkezeten (PDB: 4YZF, piros: citoszólikus tér, sárga: transzmembrán régió, narancs: visszahurkolódó régió, kék: extracelluláris tér)

Az ammónium transzporter RhA fehérjéről még nem közöltek topológiai és térszerkezeti adatot sem, azonban a módszerünk szolgáltatott négy jelölt lizin pozíciót, mint lehetséges extra-citoszólikus topológiai adatot (lizinek szekvenciális pozíciója: 40, 107, 164 és 235). A fehérjeszekvenciájára futtatott CCTOP konszenzus topológia becslő algoritmus eredményét elemezve, kijelenthető, hogy a jelölt négy pozíció a predikció alapján is extra-citoszólikus (23. ábra/B). Tehát egy korábban topológiai adattal nem rendelkező fehérjéről is olyan adatot szolgáltatott a módszer, ami szinkronban van a bioinformatikai topológia becslés eredményével.

A RBC mintáknál tapasztalt pozitív eredmények után, a bizonyos részeken optimalizált, topológiai adat szerzésre fejlesztett módszert további sejteken is vizsgálni akartuk. Olyan sejtvonalat kerestünk, ami nem letepadó és már megszekvenálták a transzkriptomját, a választás a humán K562 krónikus és a HL60 akut mieloid leukémia szuszpenziós sejtvonalakra esett. A HL60 sejtvonalból 115, míg a K562 sejtvonalból 14 tandem MS futtatást készítettünk, a kísérletek kezdeti fázisában nanoUPLC-LTQ-Orbitrap Elite készülékkel végeztük a méréseket, majd nanoLC-Bruker Maxis II ETD Q-TOF készülékre is kiterjesztettük az egyes minták analizálását. Az újabb tandem MS segítségével vizsgálni tudtuk, hogy eltérő készülékek, valamint a hozzájuk megvásárolt kereső algoritmusok hasonló jó eredményeket szolgáltatnak-e számunkra, amely a módszer robusztusságáról is információt szolgáltatott. A két készülék a tapasztalatok alapján azonos mintára hasonló mennyiségű, de nem teljesen azonos jelölt TM fehérjét és jelölt peptidet eredményezett. A különbség elsődleges oka lehetett az eltérő készülék összetétel, a lekereső szoftverek közötti különbség, valamint az eltérő kromatográfias beállítások.

A három sejtvonalra közel 200 tandem MS vizsgálat során 17493 jelölt peptidet kaptunk, legalább egy kovalensen módosított pozícióval. Az ezeket a peptideket támogató MS/MS spektrumok a ProteinProspector publikus weboldalán közzétettük (prospector.ucsf.edu, hozzáférési kulcsok: „Supplementary Table 2/Online Repository Links”¹⁵⁵). A különböző lekeresésekből származó jelölt peptideket egyedi TM fehérjékhez társítottuk. Azokat a jelölt pozíciókat fogadtuk el, amelyeket legalább 3 független jelölt peptiddel azonosítottunk. Ezzel a szűrővel 730 (47%) pozíciót azonosítottunk 198 (38%) TM fehérjében, 593 (38%) pozíciót 212 (40%) citoszólikus, valamint 250 (16%) pozíciót 114 (22%) extra-citoszólikus fehérjében. A jelentős mennyiségű citoszólikus fehérje-háttér, a jelölési stratégiát alkalmazó korábbi tudományos cikkekben is megfigyelhető¹³³. Az alkalmazott modell sejtekre külön-külön lebontottuk a különböző jelölt TM fehérje és jelölt pozíció kihozatalokat: HL60 sejtvonalból 167 TM fehérjét 616 jelölt pozícióval, a K562 sejtvonalból 75 TM fehérjét 198 jelölt pozícióval és az RBC-ékből 48 TM fehérjét 155 jelölt pozícióval azonosítottunk. A három sejtvonalból kapott jelölt TM fehérjék és pozíciók közötti átfedéseket a 24. ábrán látható Venn-diagram szemlélteti.



24. ábra: A jelölt TM fehérjék/pozíciók összesítő ábrája a különböző sejtvonalakból

Az módszer által szolgáltatott topológiaadatok összesen 198 TM fehérjét 730 jelölt pozícióval jellemezték. A homológ TM fehérjéket is figyelembe véve, a módszerünk által szolgáltatott jelölt peptidok, a UniProt adatbázison vizsgálva, 2776 humán TM fehérjére adtak legalább egy topológiai adatot.

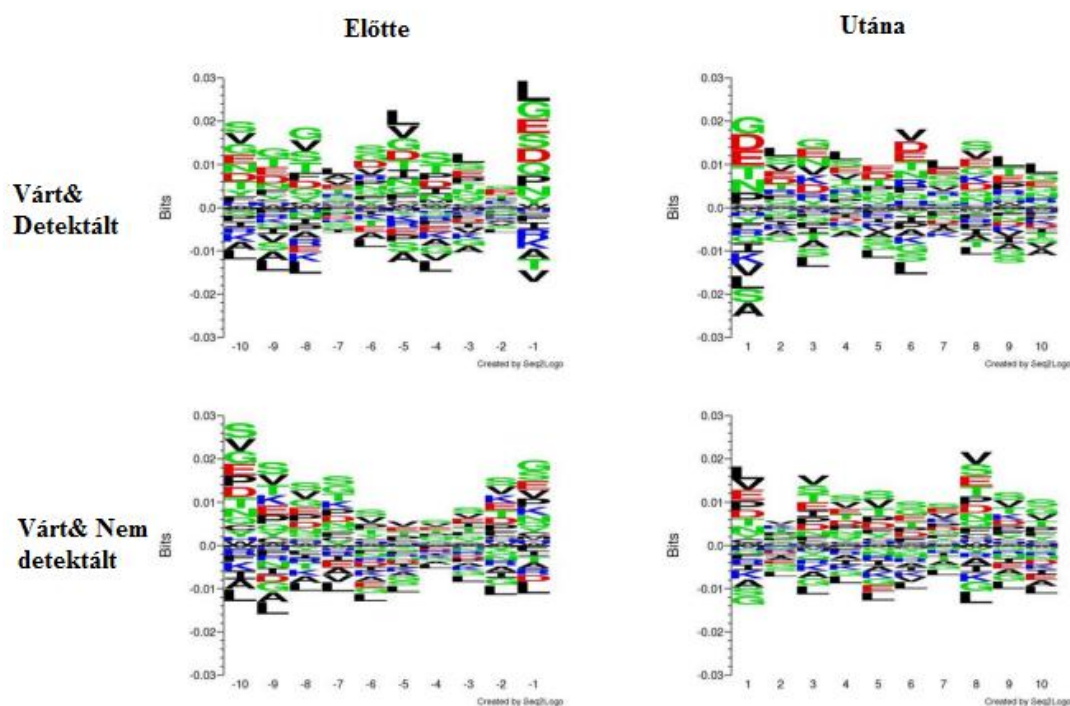
4.8 A kísérlet eredményeinek validálása

A módszer által szolgáltatott topológiaadatok (jelölt szekvenciális pozíciók) validálásának érdekében összehasonlítottuk az eredményeinket a jelölt lizin aminosavak membránhoz viszonyított helyzetét meghatározó független kísérleti eredményekkel, melyeket tudományos cikkekből gyűjtöttünk össze a TOPDB adatbázisba (4.1 fejezet). A redundáns adatokból következő torzulás elkerülése érdekében, csak nem homológ fehérjéket vettünk figyelembe. A vizsgálat folyamán 85 TM fehérjére nem találtunk kísérleti módszerrel meghatározott topológiaadatot a jelölt pozíciók szekvenciális helyein, azonban a 730 jelölt pozícióból 450 pozícióhoz (113 fehérjében) találtunk független kísérleteket, amelyek megerősíthetik azok extra-citoszólikus lokalizációját. A 450 pozíció 98,7%-a összhangban van a korábban más kísérleti módszerekkel meghatározott topológiai adatokkal, másképpen

megfogalmazva, a 450 pozíciónak 98,7%-a független kísérletek által is extra-citoszólikus oldalon volt megtalálható.

4.9 A jelölések szekvenciális környezetének elemzése

Annak érdekében, hogy betekintést nyerjünk a jelölési módszer hatékonyságát befolyásolható szekvenciális tényezőkbe, összehasonlítottuk a TM fehérjéből azonosított jelölt lizin pozíciók (25. ábra, felül: Várt&Detektált), valamint a TM fehérjéken elméletileg jelölhető, de a kísérletünkből nem detektált pozíciók (25. ábra, alul: Várt&Nem detektált) körüli ± 10 aminosav hosszúságú szekvenciák aminosav összetételét. Szekvencialogók felhasználásával vizsgáltuk az egyes pozíciókban előforduló aminosavak gyakoriságát is (25. ábra).

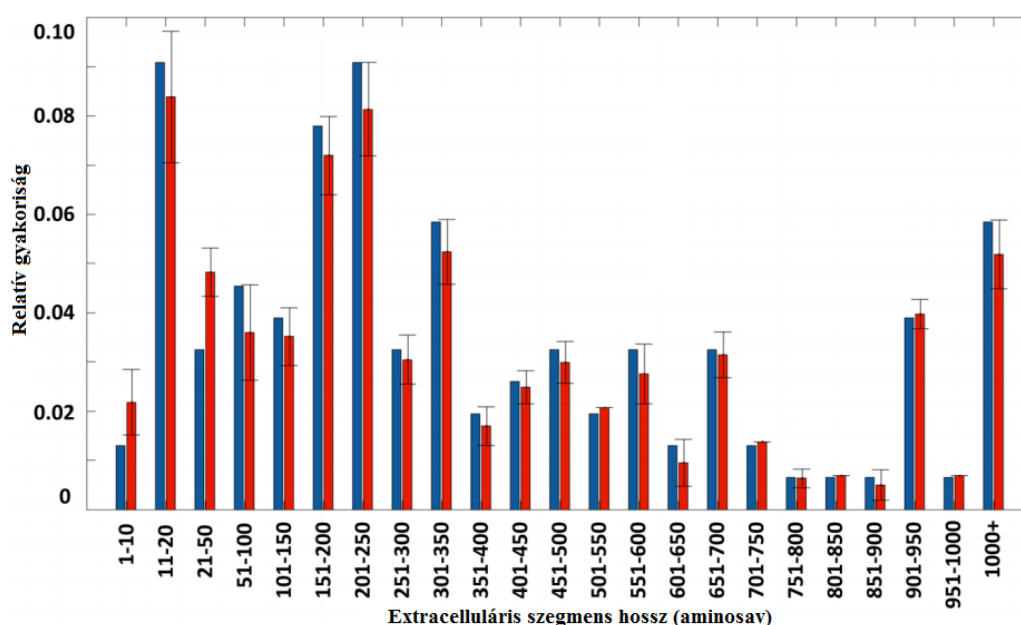


25. ábra: A várt és detektált, valamint a várt és nem detektált lizinek környezetének szekvencia logói (előtte-utána 10 aminosav).

A szekvencialogók egyértelműen azt mutatják, hogy a kidolgozott módszer preferáltan azokat a lizineket azonosítja, amelyeket követ egy pozitívan töltött aminosav (25. ábra: Várt&Detektált/Utána), amely mivel tripszin enzim hasítóhely, így lehetőséget biztosít megfelelő méretű peptidek keletkezésére, valamint a peptidek pozitív ionizáltsági állapotát is segítheti, amelyek fontosak az MS analízis szempontjából. Továbbá a legnagyobb érték a jelölt

aminosav előtti és utáni (± 1) pozícióban volt, amely befolyásolhatja a kémiai reakciót a jelölési folyamat alatt.

A jelölési módszer eredményességét, a jelölés környezetének aminosav szintű összetételén túl, befolyásolhatja a jelölendő lizint tartalmazó extra-citoszólikus hurok régió hossza. Így azoknak az extra-citoszólikus szegmenseknek a hosszeloszlását is kiszámoltuk, amelyekből azonosítottunk kovalensen módosított lizin aminosavakat. Ezenkívül kiszámoltuk, hogy ha random választanánk ki hasonló számú jelölhető lizint a fehérjék extra-citoszólikus régióiban, mint amennyit ténylegesen azonosítottunk az azonosított fehérje halmazon, akkor annak a hosszeloszlása hogyan viszonyulna a megfigyelthez.



26. ábra: Kovalensen módosított lizint tartalmazó extra-citoszólikus szegmensek hosszúság eloszlása az összes lizint tartalmazó szegmenshez képest (kék oszlopok). A jelölt lizin pozíciók számának megfelelő mennyiségű lizint véletlenszerűen kiválasztottunk az extra-citoszólikus loopokban, majd a random kiválogatott lizineket tartalmazó szegmensek hosszeloszlását számoltuk, a randomizációt megismételtük ezerszer és a szórást is feltüntettük (piros oszlopok).

A 26. ábra alapján megállapítottuk, hogy a jelölt lizineket tartalmazó extra-citoszólikus hurkok hosszeloszlása és random kiválogatott loopok hosszeloszlása között nincs szignifikáns különbség, ami arra utal, hogy a Szulfo-NHS-SS-biotin reaktivitására nincs hatással az extra-citoszólikus szegmensek vagy domének hossza.

5 Diszkusszió

A TM fehérjék a biológiai rendszerekben, gyógyszerhatóanyag-tervezésekben betöltött jelentős szerepük miatt különböző módszerekkel próbálnak a lehető legtöbb szerkezeti és funkcionális információt megszerezni. A hagyományos 3D szerkezet megoldó eljárásokkal a TM fehérjék nehezen vizsgálhatóak a kedvezőtlen fiziko-kémiai tulajdonságaik miatt, így a megoldott TM fehérje térszerkezetek száma a PDB adatbázisban rendkívül alacsony. Az ismert TM fehérjék szerkezeti terének kitöltésére ezért alternatív és alacsonyabb felbontású szerkezeti információt szolgáltató technikák is elterjedtek, például a topológia meghatározó kísérleti és bioinformatikai módszerek. Ezeket a módszereket kombinálva is szokták használni az eredmények megerősítésére, úgymint kísérleti cikkek eredményeinek alátámasztása topológia becslő módszer eredményével és megfordítva. Egyes predikciós módszerek pontossága tovább növelhető a kísérleti eredmények kényszerfeltételként való megadásával.

Munkám kiindulási fázisában az irodalom és adatbázisok szisztematikusan átvizsgálásával topológiaadatokat gyűjtöttünk, amelyet a TOPDB adatbázisban rögzítettünk. Az adatbázis jelentős felhasználási köre a következőkből adódik: kiindulópont lehet a referenciaszettek létrehozásához becslő algoritmusokhoz¹⁶⁰, a TM fehérjékhez kapcsolható adatbázisok fejlesztéséhez¹⁶¹⁻¹⁶³ és az egyes TM fehérjék kísérletes jellemzéséhez^{164,165}. A TOPDB adatbázis frissítésénél az adatok minőségi és mennyiségi válogatására is figyelmet fordítottunk. Adatbázisokból, illetve az ott hivatkozott cikkekéből, valamint további szakirodalmi cikkekéből gyűjtöttünk topológiaadatokat. Ezek közé tartozik:

- A TMDET algoritmus által definiált topográfia, amelyet a PDB adatbázisban megtalálható egyes 3D szerkezetekből ad meg, ezeket az eredeti 3D struktúrát leíró cikkben lévő topológiaadatokkal kiegészítettünk.
- TM fehérjék szolubilis doménjeinek megoldott szerkezete.
- Az utóbbi évek kísérleti módszereiből szerzett topológiaadatok.
- *Saccharomyces cerevisiae* és *Escherichia coli* teljes TM proteomjának topológia analízise^{100,101}.
- A HT technikákkal generált topológiaadatok, úgymint N- vagy O-glikoziláció pozíciói.

A TOPDB adatbázis a frissítés által 75211 kísérleti módszerrel meghatározott topológiaadatot tartalmaz (közel 3-szor annyi, mint az előző verzióban lévő adatmennyiség).

A kísérleti adatok a csoport által fejlesztett CCTOP algoritmusban használhatóak megszorításként, amelyek jelentősen növelik a becslés megbízhatóságát és pontosságát. A topológia predikció pontosságát tovább növeltük bioinformatikai eszközök felhasználásával, pl a membrán egy adott oldalán előforduló domének felhasználásával (TOPDOM), az eddigi információk homológ fehérjékre való kiterjesztésével, valamint topológia becslő algoritmusok felhasználásával. Az ExTopoDB adatbázis a TOPDB adatbázishoz hasonló, szintén kísérleti módszerekkel meghatározott topológiaadatokat tartalmaz¹⁶⁶. Az adatbázis azonban 2013. július elseje óta nincs frissítve, a teljes TM fehérje szám csak 2143 (158 organizmusból), ezzel szemben a TOPDB adatbázisban jelenleg 282 organizmus 4190 fehérjéjéhez van topológiai adat összegyűjtve. A TOPDB, így a topológiaadatok jelenleg elérhető legfrissebb, rendszerezett, nyilvánosan hozzáférhető legnagyobb tárhelye.

A nagy mennyiségű rendelkezésre álló kísérleti adat ellenére, még számos humán TM fehérjéhez egyáltalán nem találtunk semmilyen topológiaadatot, ezért egy saját kísérleti módszert fejlesztettünk ki, az MSTOP eljárást. A módszert mind elméleti módon, mind különböző kísérleti rendszerekben, három sejttípuson, RBC-n, HL60 és K562 sejtvonalakon teszteltük. A módszer alapkomponense egy jól jellemzett kémiai anyag (Szulfo-NHS-SS-biotin), amelyet korábban sejtfelszíni fehérjék izolálására használtak. Az általunk fejlesztett eljárásban a legfontosabb változtatás a hasonló módszerekhez képest, hogy az extra-citoszólikus hozzáférhető primer aminosav tartalmazó fehérjeszegmensek jelölése után membránt preparáltunk (ezzel is elősegítve a membránkötött fehérjék relatív gyakoriságának növekedését a citoszólikus fehérjékkel szemben), illetve a preparátum fehérjéit először proteáz kezelésnek vetettük alá és csak a jelölt peptideket dúsítottuk (csökkentve a céljaink szempontjából irreleváns jelöletlen peptidek számát). A módosítás pontos szekvenciális helyét tandem MS készülékkel vizsgáltuk és azonosítottuk. Ezek az azonosított extra-citoszólikus pozíciók a CCTOP algoritmus topológia becslésében megszorításként használhatóak, amely így a megfelelő TM fehérje topológia predikciójának pontosságát növelik. A módszer által generálható megszorítások hatását a CCTOP becslés pontosságára megmutattuk egy TM fehérje teszthalmazon (10. ábra). A 25, 50, 75 és 100% extracelluláris lizin megszorítás hozzáadásával a pontosság növekedése nem lineárisan változott. A predikció pontosságának maximális lehetséges növekedésének közel fele elérhető mindössze 20% extra-citoszólikus lizin megszorítással (ami megegyezik az MSTOP topológiai adatszerző módszer által meghatározott lizinek százalékos arányával). Továbbá igazoltuk, hogy a kísérletileg meghatározott extra-citoszólikus megszorítások csak akkor növelik a pontosságot, ha az

adatok hamis pozitív (azaz citoszólikus) találatoktól mentesek vagy szinte mentesek (már 4% hamis pozitív arány is jelentősen csökkenti, a becslés pontosságát). Ezért a módszerünket úgy optimalizáltuk, hogy minimalizáljuk annak a lehetőségét, hogy a TM fehérjéket intracellulárisan jelöljük. Ennek érdekében az alkalmazott membrán impermeábilis jelölővel maximalizáltuk a sejtfelszíni jelölést, úgy, hogy az a sejtek integritását ne befolyásolja, valamint meggyőződünk arról is, hogy citoszólikus jelölődés konfokális mikroszkópon permeabilizált körülmények esetén sem mutatkozott a feljelölt sejteken. Meghatároztuk a biotinált minták teljes biotintartalmának kikötéséhez elegendő avidin agaróz gyöngy mennyiségét is. Az aspecifikus háttérpeptidek jelentős százalékát sikerült eliminálni a biotinált peptideket is hordozó avidin agaróz gyöngy mosási lépéseinek számának és puffer összetételének megfelelő beállításával. Az utóbbi két módosítással, a teljes fehérjéket izoláló protokollokhoz képest jelentősen csökkentettük a minták komplexitását, amely ugyan kevesebb fehérjét eredményezett, de csökkentette a hamis pozitív találatok lehetőségét.

Az eddig elvégzett kísérletek által összesen 730 szekvenciális pozíciót azonosítottunk 198 TM fehérjéből, a három különböző sejtípus vizsgálataival. Ezekre a sejtípusokra más cikkekből ismert proteomikai elemzésének eredménye alapján megállapítható, hogy az azokban azonosított TM fehérjékhez képest kevesebbet detektáltunk, amely következő okok miatt lehetséges:

- Nem hozzáférhető primer aminok lehetnek a TM fehérjék extra-citoszólikus szegmenseiben.
- Tripszin proteolitikus enzim használata esetén a jelölt lizin mellett a tripszin már nem tud hasítani, emiatt hosszabb peptidek keletkeznek, amelyek már nem azonosíthatók LC-MS/MS-el.
- Az egyes analizált peptideken posztranszlációs módosulás is előfordulhat, amely szintén nehezítheti az azonosítást.
- A fehérjék különböző mennyisége a biológiai membránokban szintén befolyásolja a detektálhatóságot (a relatív intenzitások miatt), a dinamikus tartománynak megfelelő jelölt peptid-mennyiségeket tudunk azonosítani¹³⁰.

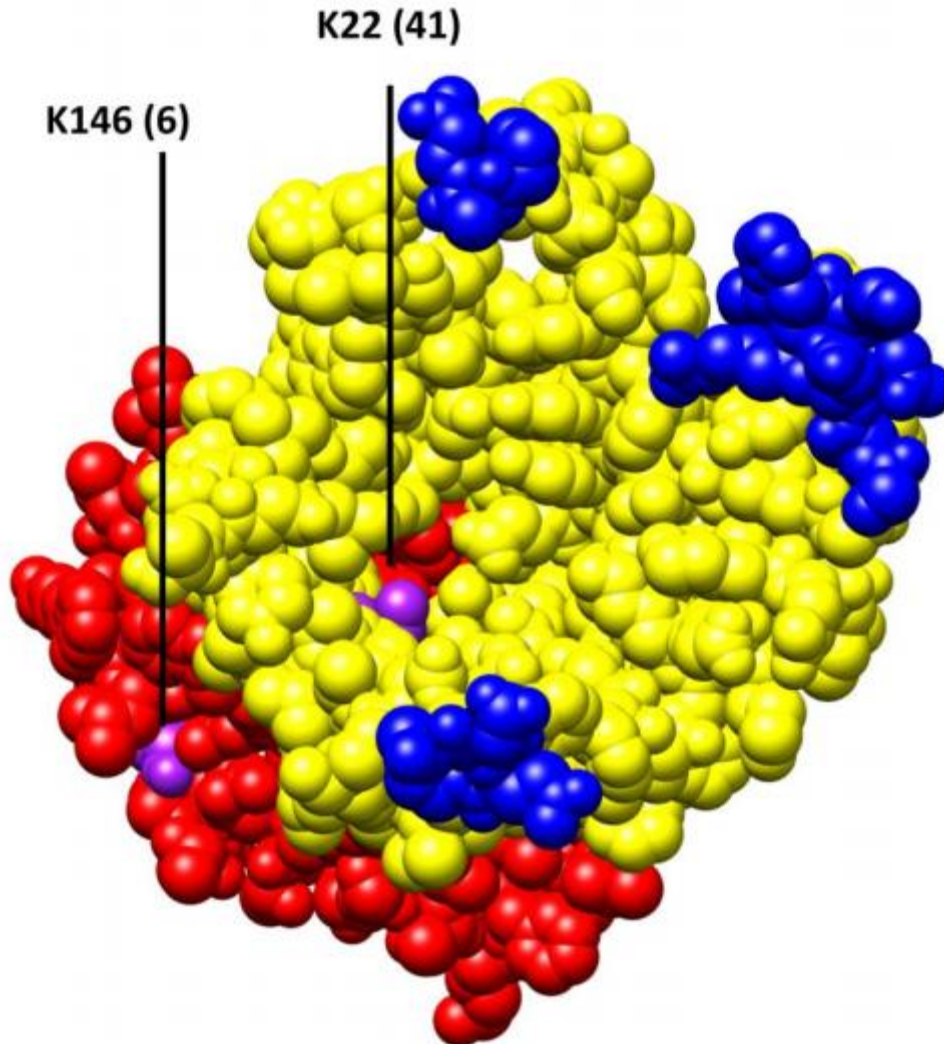
A jelölt pozíciók validálására a publikusan elérhető, TOPDB adatbázisban összegyűjtött adatokat használtuk. A 730 jelölt pozícióból 450 pozícióhoz találtunk független kísérleti módszerrel szerzett adatokat, amelyek 98,7% -ban megerősítették, hogy a kapott pozíciók extra-citoszólikusak. Az MSTOP módszer a jelölt TM fehérjében azonosított extra-

citoszólíkus lizinek százalékos aránya (az elméleti extra-citoszólíkus lizinekhez képest), valamint független kísérletekkel jellemzett pontossága a 4.3 fejezetben ismertett bioinformatikai számítások tükrében az elvárásoknak megfelelő. Növeli az olyan módszerek pontosságát, amelyek a megszerzett adatokat megszorításként használják (pl.: CCTOP). A módszer ezért segíti a humán TM fehérjék topológia becslés pontosságának jelentős növekedését.

A TM fehérjék extra-citoszólíkus része mellett, számos gyakori citoszólíkus és extra-citoszólíkus fehérje peptidjein is jelölést azonosítottunk. Az extra-citoszólíkus oldalon horgonyzott fehérjékhez tartozó jelölt peptidek a módszerből adódóan várhatóak voltak, ellenben a citoszólíkus jelölések nem. Az eddigi tapasztalatok alapján egyértelműen kijelenthető, hogy ezek a jelölések nem, vagy nagyon kevéssé történtek a citoszólíkus TM fehérjeszegmenseken, csak a gyakori citoszólíkus gyakori fehérjéket érintették. A citoszólíkus fehérje jelölések eredhetnek sérült sejtekből (4.4.2 fejezet 15. ábra 2-es részén egy sérült sejt látható, az egész sejt zöld fluoreszcenciával rendelkezik, míg a többi esetén csak a membrán határvonalán detektálható a zöld szín), melyek akár vissza is tapadhatnak a sejtek felszínére, mint a normál vérkeringés során¹⁶⁷. Detektáltunk jelölt hiszton fehérjéket, amelyek DNS-t kötnek, de kimutatták, hogy a hisztonok, sejt-specifikus granulafehérjéket tartalmazó neutrofil extra-citoszólíkus csapdákhöz is kötnek¹⁶⁸. Más, a módszerünk által is jelölt citoszólíkus fehérjékről is publikálták, hogy előfordulhatnak az a sejtfelszínen (pl.: tubulin¹⁶⁹, aktin¹⁷⁰). Továbbá Hofmann és munkatársai, akik a „sejtfelszín elfogás” (CSC) technikát használták, hasonló szintű intracelluláris jelölést tapasztaltak, amikor négy Hodgkin és négy non-Hodgkin limfóma sejtvonalból tanulmányozták a sejtfelszíni fehérjéket¹³⁴. A fenti tények, valamint TM fehérjékhez tartozó jelölések extra-citoszólíkus validálása megerősítette, hogy a módszerünk hosszútávon alkalmas lehet topológiaadat meghatározására, tartva az alacsony hamis pozitív találati arányt.

A fejlesztett módszer által szolgáltatott extra-citoszólíkus lizin aminosav pozíció megszorítások között az irodalomban lévő adatokkal összevetve összesen 6 db ellentmondó pozíciót találtunk (azaz a pozíció lokalizációja vagy citoszólíkus vagy membránrégió volt korábban). A 6-ból 4 pozícióhoz tartozik két homológ fehérjeszerkezet a PDB adatbázisban. Az ADP/ATP transzlokáz 2 (ADT2) fehérjéhez társítható megoldott szerkezet (PDB: 2C3E, 89% szekvencia azonosság¹⁷¹), amely alapján a 147-es lizin citoszólíkus és a 23-as membrán régióban lokalizálódik. A 23-as pozícióban lévő lizin, amelyet 41-szer detektáltunk a kísérletünkben, a kifelé nyitott szerkezet üregében helyezkedik el (27. ábra), amely így

hozzáférhető a Szulfo-NHS-SS biotin reagens számára. Hasonlóan a 147-es lizin pozíció esetében, amelyet csak 6-szor detektáltunk jelölve, a jelölődés azzal magyarázható, hogy a fehérje nyitott csatornáján a jelölő ágens keresztül tud jutni és jelöli a fehérje ezen részét.



27. ábra: Az emberi ADT2 fehérje homológjának a szerkezete (a szarvasmarha mitokondriális ADP-ATP transzlokáz fehérjéjének szerkezete, PDB: 2C3E), a színezése a topológia alapján történt (kék: extra-citoszólíkus, sárga: membrán régió, piros: citoszólíkus). A kovalensen módosított lizineket lila színnel jeleztük, a lizinek pozícióját is jelöltük, és zárójelben feltüntettük, hogy az adott pozícióban levő lizint hány kíséretben azonosítottuk jelölve.

További két ellentmondó pozíciót azonosítottunk az emberi facilitált glükóz transzporter 1 (GTR1) fehérjén, amely pozíciók elhelyezkedését a fehérje megoldott szerkezetén vizsgáltuk (PDB:4PYP, 99%-os szekvencia azonosság⁴⁹). A fehérje ebben a szerkezetben befelé nyitott állapotban található, amelyben a kötött ligand kordinátáit is

meghatározták. A 245 és 256-os pozíciókat 3-3-szor találtuk meg jelölve, amelyek a térszerkezetben egy olyan α -hélixben találhatóak, amely a citoszólikus belépési oldallal szemben vannak. Feltételezhetően a jelölő ágens eléri ezeket a pozíciókat nyitott konformációban. Az extra-citoszólikus lizineket lényegesen többször találtuk meg ebben a fehérjében, például a 117-es lizint több mint 50-szer detektáltuk. Alacsonyabb szintű jelölési hatékonyságot figyeltünk meg az eddigi kísérleteink alapján a 38, 183 és 300 pozíciókban, amely összefüggésben lehet a fehérje nyitott és zárt állapotai közötti folyamatos átmenettel, ami közben az áthaladó ligand részben lefedheti ezeket a pozíciókat.

Fejlesztett módszerünk mellett egy jelentős konkurens stratégia létezik, amellyel szintén extra-citoszólikus szegmenseket tud meghatározni: ez a sejt felszíni N-glikopeptidek azonosítására épül, amellyel számos sejt vonal sejt felszíni proteomját jellemezték¹²⁴. A glikoziláció a jelenlegi ismereteink szerint az extra-citoszólikus szegmensekben bekövetkező sejten belüli természetes módosulás, így a glikocsoportokra specifikus gliko-peptid dúsítási stratégia specifikusabban képes extacelluláris peptideket azonosítani. Azonban az extra-citoszólikusan azonosítható lizinek száma négyszer nagyobb, mint az extra-citoszólikus N-X-S/T motívum, ahol az N-glikoziláció történhet. Ebből következik, hogy az extracellulár lizin azonosítások több adatot szolgáltatathatnának a topológia predikciókhoz. Egyetlen tudományos közleményben találtunk az általunk kidolgozott stratégiához hasonlóan Szulfo-NHS-SS-biotin reagenssel módosított lizineket tartalmazó peptid listát (Hoffman és mtsi, Lys-CSC technika)¹³⁴, amelyben jelentős mennyiségű citoszólikus TM fehérjerégióban azonosítottak jelölést. A jelölt lizinek 17%-a citoszólikus, míg 83%-a extra-citoszólikus TM fehérjeszegmensekből származik a CCTOP algoritmus által becsült topológia alapján. A munkánkban tárgyalt paraméter optimalizációkkal a hamis pozitív találati arányt jelentősen csökkentettük, a jelölésünk pontosságát 98,7%-ra emeltük, amely szignifikánsan javítja a megszorításokkal végzett topológia predikciót.

A TM fehérje topológiai adatok gyűjtésével a TOPDB adatbázishoz, valamint az MSTOP módszer által szolgáltatott kísérleti eredmények felhasználásával pontosabb topológia becslést értünk el jelentős mennyiségű humán TM fehérje esetében. A későbbiekben tervezzük a további adatok szerzését a topológia predikciók támogatása érdekében, amelyhez elsődlegesen további sejt vonalakat és organelumokat szeretnénk felhasználni. A szakirodalom szerint ezek más-más TM fehérjéket expresszálhatnak, a tandem tömegspektrometriával kimutatható mennyiségben¹²⁴, valamint más proteázok bevonásával növelnénk az adott sejt típusból izolált fehérjék extra-citoszólikus lefedettségét. A jelenlegi és

a későbbi kísérletekből származó adatok segíthetik az egyedi TM fehérjék topológia becslését és szerkezeti modellezését, ezáltal segítheti a szerkezet-funkció kapcsolat jobb megértését.

6 Irodalomjegyzék

1. Goluszko, P. & Nowicki, B. Membrane cholesterol: a crucial molecule affecting interactions of microbial pathogens with mammalian cells. *Infect. Immun.* **73**, 7791–6 (2005).
2. Harayama, T. & Riezman, H. Understanding the diversity of membrane lipid composition. *Nat. Rev. Mol. Cell Biol.* **19**, 281–296 (2018).
3. Cooper GM. *Structure of the Plasma Membrane. The Cell: A Molecular Approach. 2nd edition. Sunderland (MA): Sinauer Associates* (2000).
4. Saher, G., Brügger, B., Lappe-Siefke, C., Möbius, W., Tozawa, R., Wehr, M. C., Wieland, F., Ishibashi, S. & Nave, K. High cholesterol level is essential for myelin membrane growth. *Nat. Neurosci.* **8**, 468–75 (2005).
5. Becker, T., Gebert, M., Pfanner, N. & van der Laan, M. Biogenesis of mitochondrial membrane proteins. *Curr. Opin. Cell Biol.* **21**, 484–93 (2009).
6. Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. & Walter, P. *Membrane Proteins. Molecular Biology of the Cell. 4th edition. New York: Garland Science* (2002).
7. Ikezawa, H. Glycosylphosphatidylinositol (GPI)-anchored proteins. *Biol. Pharm. Bull.* **25**, 409–17 (2002).
8. Long, S. B., Hancock, P. J., Kral, A. M., Hellinga, H. W. & Beese, L. S. The crystal structure of human protein farnesyltransferase reveals the basis for inhibition by CaaX tetrapeptides and their mimetics. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 12948–53 (2001).
9. Hayashi, N. & Titani, K. N-myristoylated proteins, key components in intracellular signal transduction systems enabling rapid and flexible cell responses. *Proc. Jpn. Acad. Ser. B. Phys. Biol. Sci.* **86**, 494–508 (2010).
10. Del Angel-Mosqueda, C., Gutiérrez-Puente, Y., López-Lozano A. P., Romero-Zavaleta R. E., Mendiola-Jiménez A., Medina-De la Garza C., M.-M. M. & D. la G.-R. M. A. Epidermal growth factor enhances osteogenic differentiation of dental pulp stem cells in vitro. *Head Face Med.* **11**, 29 (2015).

11. Salomao, M., Zhang, X., Yang, Y., Lee, S., Hartwig, J. H., Chasis, J.A., Mohandas, N. & An, X. Protein 4.1R-dependent multiprotein complex: new insights into the structural organization of the red blood cell membrane. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 8026–31 (2008).
12. Wallin, E. & von Heijne, G. Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Sci.* **7**, 1029–38 (1998).
13. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–80 (2001).
14. Dobson, L., Reményi, I. & Tusnády, G. E. The human transmembrane proteome. *Biol. Direct* **10**, 31 (2015).
15. Chothia, C., Levitt, M. & Richardson, D. Structure of proteins: packing of alpha-helices and pleated sheets. *Proc. Natl. Acad. Sci. U. S. A.* **74**, 4130–4 (1977).
16. White, S. H. & Wimley, W. C. Membrane protein folding and stability: physical principles. *Annu. Rev. Biophys. Biomol. Struct.* **28**, 319–65 (1999).
17. Zhang, X. C. & Han, L. How does a β -barrel integral membrane protein insert into the membrane? *Protein Cell* **7**, 471–7 (2016).
18. Kozma, D., Simon, I. & Tusnády, G. E. PDBTM: Protein Data Bank of transmembrane proteins after 8 years. *Nucleic Acids Res.* **41**, D524–9 (2013).
19. Lomize, M. A., Pogozheva, I. D., Joo, H., Mosberg, H. I. & Lomize, A. L. OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Res.* **40**, D370–6 (2012).
20. Chothia, C. Proteins. One thousand families for the molecular biologist. *Nature* **357**, 543–4 (1992).
21. Koonin, E. V, Wolf, Y. I. & Karev, G. P. The structure of the protein universe and genome evolution. *Nature* **420**, 218–23 (2002).
22. Oberai, A., Ihm, Y., Kim, S. & Bowie, J. U. A limited universe of membrane protein families and folds. *Protein Sci.* **15**, 1723–34 (2006).
23. Goder, V. & Spiess, M. Topogenesis of membrane proteins: determinants and dynamics. *FEBS Lett.* **504**, 87–93 (2001).

24. van Klompenburg, W., Nilsson, I., von Heijne, G. & de Kruijff, B. Anionic phospholipids are determinants of membrane protein topology. *EMBO J.* **16**, 4261–6 (1997).
25. Goder, V., Junne, T. & Spiess, M. Sec61p contributes to signal sequence orientation according to the positive-inside rule. *Mol. Biol. Cell* **15**, 1470–8 (2004).
26. Rapoport, T. A., Goder, V., Heinrich, S. U. & Matlack, K. E. S. Membrane-protein integration and the role of the translocation channel. *Trends Cell Biol.* **14**, 568–75 (2004).
27. Beltzer, J. P., Fiedler, K., Fuhrer, C., Geffen, I., Handschin, C., Wessels, H. P. & Spiess, M. Charged residues are major determinants of the transmembrane orientation of a signal-anchor sequence. *J. Biol. Chem.* **266**, 973–8 (1991).
28. Denzer, A. J., Nabholz, C. E. & Spiess, M. Transmembrane orientation of signal-anchor proteins is affected by the folding state but not the size of the N-terminal domain. *EMBO J.* **14**, 6311–7 (1995).
29. Goder, V. & Spiess, M. Topogenesis of membrane proteins: determinants and dynamics. *FEBS Lett.* **504**, 87–93 (2001).
30. Pfeffer, S., Burbaum, L., Unverdorben, P., Pech, M., Chen, Y., Zimmermann, R., Beckmann, R. & Förster, F. Structure of the native Sec61 protein-conducting channel. *Nat. Commun.* **6**, 8403 (2015).
31. Rapoport, T. A. Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. *Nature* **450**, 663–9 (2007).
32. Watson, H. Biological membranes. *Essays Biochem.* **59**, 43–69 (2015).
33. von Heijne, G. Net N-C charge imbalance may be important for signal sequence function in bacteria. *J. Mol. Biol.* **192**, 287–90 (1986).
34. Elofsson, A. & von Heijne, G. Membrane protein structure: prediction versus reality. *Annu. Rev. Biochem.* **76**, 125–40 (2007).
35. Alberts B, Johnson A, Lewis J, et al. *The Endoplasmic Reticulum. Molecular Biology of the Cell. 4th edition. New York: Garland Science* (2002).

36. Ismail, N., Hedman, R., Schiller, N. & von Heijne, G. A biphasic pulling force acts on transmembrane helices during translocon-mediated membrane integration. *Nat. Struct. Mol. Biol.* **19**, 1018–22 (2012).
37. Lodish H, Berk A, Zipursky SL, et al. *Section 17.7, Protein Glycosylation in the ER and Golgi Complex. Molecular Cell Biology. 4th edition. New York: W. H. Freeman* (2000).
38. Souda, P., Ryan, C. M., Cramer, W. A. & Whitelegge, J. Profiling of integral membrane proteins and their post translational modifications using high-resolution mass spectrometry. *Methods* **55**, 330–6 (2011).
39. Hubbard, S. R. & Miller, W. T. Receptor tyrosine kinases: mechanisms of activation and signaling. *Curr. Opin. Cell Biol.* **19**, 117–23 (2007).
40. Chamberlain, L. H. & Shipston, M. J. The physiology of protein S-acylation. *Physiol. Rev.* **95**, 341–76 (2015).
41. Kavsak, P., Rasmussen, R. K., Causing, C. G., Bonni, S., Zhu, H., Thomsen, G. H. & Wrana, J. L. Smad7 binds to Smurf2 to form an E3 ubiquitin ligase that targets the TGF beta receptor for degradation. *Mol. Cell* **6**, 1365–75 (2000).
42. Gruenberg, J. The endocytic pathway: a mosaic of domains. *Nat. Rev. Mol. Cell Biol.* **2**, 721–30 (2001).
43. Di Guglielmo, G. M., Le Roy, C., Goodfellow, A. F. & Wrana, J. L. Distinct endocytic pathways regulate TGF-beta receptor signalling and turnover. *Nat. Cell Biol.* **5**, 410–21 (2003).
44. Kobilka, B. K. G protein coupled receptor structure and activation. *Biochim. Biophys. Acta* **1768**, 794–807 (2007).
45. Huang, Y. & Thathiah, A. Regulation of neuronal communication by G protein-coupled receptors. *FEBS Lett.* **589**, 1607–19 (2015).
46. Graaf, Cd., Donnelly, D., Wootten, D., Lau, J., Sexton, P. M., Miller, L. J., Ahn, J. M., Liao, J., Fletcher, M. M., Yang, D., Brown, A. J., Zhou, C., Deng, J. & Wang, M.-W. Glucagon-Like Peptide-1 and Its Class B G Protein-Coupled Receptors: A Long March to Therapeutic Successes. *Pharmacol. Rev.* **68**, 954–1013 (2016).

47. Foskett, J. K., White, C., Cheung, K.-H. & Mak, D.-O. D. Inositol trisphosphate receptor Ca²⁺ release channels. *Physiol. Rev.* **87**, 593–658 (2007).
48. Geppetti, P., Veldhuis, N. A., Lieu, T. & Bunnett, N. W. G Protein-Coupled Receptors: Dynamic Machines for Signaling Pain and Itch. *Neuron* **88**, 635–49 (2015).
49. Deng, D., Xu, C., Sun, P., Wu, J., Yan, C., Hu, M. & Yan, N. Crystal structure of the human glucose transporter GLUT1. *Nature* **510**, 121–5 (2014).
50. Sands, J. M. Regulation of renal urea transporters. *J. Am. Soc. Nephrol.* **10**, 635–46 (1999).
51. Levin, E. J., Quick, M. & Zhou, M. Crystal structure of a bacterial homologue of the kidney urea transporter. *Nature* **462**, 757–61 (2009).
52. Higgins, C. F. ABC transporters: from microorganisms to man. *Annu. Rev. Cell Biol.* **8**, 67–113 (1992).
53. Dean, M., Rzhetsky, A. & Allikmets, R. The human ATP-binding cassette (ABC) transporter superfamily. *Genome Res.* **11**, 1156–66 (2001).
54. Kim, M., Turnquist, H., Jackson, J., Sgagias, M., Yan, Y., Gong, M., Dean, M., Sharp, J. G. & Cowan, K. The multidrug resistance transporter ABCG2 (breast cancer resistance protein 1) effluxes Hoechst 33342 and is overexpressed in hematopoietic stem cells. *Clin. Cancer Res.* **8**, 22–8 (2002).
55. Riordan, J. R., Rommens, J. M., Kerem, B., Alon, N., Rozmahel, R., Grzelczak, Z., Zielenski, J., Lok, S., Plavsic, N. & Chou, J. L. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science* **245**, 1066–73 (1989).
56. Zhu, Q., Lee, D. W. K. & Casey, J. R. Novel topology in C-terminal region of the human plasma membrane anion exchanger, AE1. *J. Biol. Chem.* **278**, 3112–20 (2003).
57. Vince, J. W. & Reithmeier, R. A. Carbonic anhydrase II binds to the carboxyl terminus of human band 3, the erythrocyte Cl⁻/HCO₃⁻ exchanger. *J. Biol. Chem.* **273**, 28430–7 (1998).
58. Reeves, J. P., Condrescu, M., Chernaya, G. & Gardner, J. P. Na⁺/Ca²⁺ antiport in the mammalian heart. *J. Exp. Biol.* **196**, 375–88 (1994).

59. Jung, H. The sodium/substrate symporter family: structural and functional features. *FEBS Lett.* **529**, 73–7 (2002).
60. Saier, M. H., Reddy, V. S., Tsu, B. V., Ahmed, M. S., Li, C. & Moreno-Hagelsieb, G. The Transporter Classification Database (TCDB): recent advances. *Nucleic Acids Res.* **44**, D372–9 (2016).
61. Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., Olsson, I., Edlund, K., Lundberg, E., Navani, S., Szigyaró, C. A., Odeberg, J., Djureinovic, D., Takanen, J. O., N. J. & P. F. Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
62. Smyth, M. S. & Martin, J. H. x ray crystallography. *Mol. Pathol.* **53**, 8–14 (2000).
63. Newstead, S., Ferrandon, S. & Iwata, S. Rationalizing alpha-helical membrane protein crystallization. *Protein Sci.* **17**, 466–72 (2008).
64. Parker, J. L. & Newstead, S. Membrane Protein Crystallisation: Current Trends and Future Perspectives. *Adv. Exp. Med. Biol.* **922**, 61–72 (2016).
65. Caffrey, M. & Cherezov, V. Crystallizing membrane proteins using lipidic mesophases. *Nat. Protoc.* **4**, 706–31 (2009).
66. Li, D. & Caffrey, M. Renaturing membrane proteins in the lipid cubic phase, a nanoporous membrane mimetic. *Sci. Rep.* **4**, 5806 (2014).
67. Varga, J. K. & Tusnády, G. E. TMCrys: predict propensity of success for transmembrane protein crystallization. *Bioinformatics* **34**, 3126–3130 (2018).
68. Pan, B. B., Yang, F., Ye, Y., Wu, Q., Li, C., Huber, T. & Su, X. C. 3D structure determination of a protein in living cells using paramagnetic NMR spectroscopy. *Chem. Commun. (Camb).* **52**, 10237–40 (2016).
69. Wüthrich, K. Protein structure determination in solution by NMR spectroscopy. *J. Biol. Chem.* **265**, 22059–62 (1990).
70. Frueh, D. P., Goodrich, A. C., Mishra, S. H. & Nichols, S. R. NMR methods for structural studies of large monomeric and multimeric proteins. *Curr. Opin. Struct. Biol.* **23**, 734–9 (2013).

71. Arora, A., Abildgaard, F., Bushweller, J. H. & Tamm, L. K. Structure of outer membrane protein A transmembrane domain by NMR spectroscopy. *Nat. Struct. Biol.* **8**, 334–8 (2001).
72. Gautier, A., Mott, H. R., Bostock, M. J., Kirkpatrick, J. P. & Nietlispach, D. Structure determination of the seven-helix transmembrane receptor sensory rhodopsin II by solution NMR spectroscopy. *Nat. Struct. Mol. Biol.* **17**, 768–74 (2010).
73. Cheng, Y. Single-Particle Cryo-EM at Crystallographic Resolution. *Cell* **161**, 450–457 (2015).
74. Nogales, E. The development of cryo-EM into a mainstream structural biology technique. *Nat. Methods* **13**, 24–7 (2016).
75. Hite, R. K. & MacKinnon, R. Structural Titration of Slo2.2, a Na⁺-Dependent K⁺ Channel. *Cell* **168**, 390–399.e11 (2017).
76. Tusnády, G. E., Dosztányi, Z. & Simon, I. TMDET: web server for detecting transmembrane regions of proteins by using their 3D coordinates. *Bioinformatics* **21**, 1276–7 (2005).
77. Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–32 (1982).
78. Hoffmann, W. Molecular characterization of the CAN1 locus in *Saccharomyces cerevisiae*. A transmembrane protein without N-terminal hydrophobic signal sequence. *J. Biol. Chem.* **260**, 11831–7 (1985).
79. Heijne, G. The distribution of positively charged residues in bacterial inner membrane proteins correlates with the trans-membrane topology. *EMBO J.* **5**, 3021–7 (1986).
80. von Heijne, G. Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rule. *J. Mol. Biol.* **225**, 487–94 (1992).
81. Cserző, M., Wallin, E., Simon, I., von Heijne, G. & Elofsson, A. Prediction of transmembrane alpha-helices in prokaryotic membrane proteins: the dense alignment surface method. *Protein Eng.* **10**, 673–6 (1997).
82. Rost, B., Casadio, R., Fariselli, P. & Sander, C. Transmembrane helices predicted at 95% accuracy. *Protein Sci.* **4**, 521–33 (1995).

83. Tusnády, G. E. & Simon, I. The HMMTOP transmembrane topology prediction server. *Bioinformatics* **17**, 849–50 (2001).
84. Käll, L., Krogh, A. & Sonnhammer, E. L. L. Advantages of combined transmembrane topology and signal peptide prediction--the Phobius web server. *Nucleic Acids Res.* **35**, W429–32 (2007).
85. Viklund, H. & Elofsson, A. Best alpha-helical transmembrane protein topology predictions are achieved using hidden Markov models and evolutionary information. *Protein Sci.* **13**, 1908–17 (2004).
86. Reddy, A., Cho, J., Ling, S., Reddy, V., Shlykov, M. & Saier, M. H. Reliability of nine programs of topological predictions and their application to integral membrane channel and carrier proteins. *J. Mol. Microbiol. Biotechnol.* **24**, 161–90 (2014).
87. Dobson, L., Reményi, I. & Tusnády, G. E. CCTOP: a Consensus Constrained TOPology prediction web server. *Nucleic Acids Res.* **43**, W408–12 (2015).
88. Tsirigos, K. D., Peters, C., Shu, N., Käll, L. & Elofsson, A. The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. *Nucleic Acids Res.* **43**, W401–7 (2015).
89. Tusnády, G. E., Kalmár, L. & Simon, I. TOPDB: topology data bank of transmembrane proteins. *Nucleic Acids Res.* **36**, D234–9 (2008).
90. Dobson, L., Langó, T., Reményi, I. & Tusnády, G. E. Expediting topology data gathering for the TOPDB database. *Nucleic Acids Res.* **43**, D283–9 (2015).
91. Varga, J., Dobson, L. & Tusnády, G. E. TOPDOM: database of conservatively located domains and motifs in proteins. *Bioinformatics* **32**, 2725–6 (2016).
92. Boyd, D., Traxler, B. & Beckwith, J. Analysis of the topology of a membrane protein by using a minimum number of alkaline phosphatase fusions. *J. Bacteriol.* **175**, 553–6 (1993).
93. Duffy, E. B. & Barquera, B. Membrane topology mapping of the Na⁺-pumping NADH: quinone oxidoreductase from *Vibrio cholerae* by PhoA-green fluorescent protein fusion analysis. *J. Bacteriol.* **188**, 8343–51 (2006).
94. Hanson, B. R., Lowe, B. A. & Neely, M. N. Membrane topology and DNA-binding ability of the Streptococcal CpsA protein. *J. Bacteriol.* **193**, 411–20 (2011).

95. Broome-Smith, J. K., Tadayyon, M. & Zhang, Y. Beta-lactamase as a probe of membrane protein assembly and protein export. *Mol. Microbiol.* **4**, 1637–44 (1990).
96. Wang, J., Chu, B. B., Ge, L., Li, B. L., Yan, Y. & Song, B. L. Membrane topology of human NPC1L1, a key protein in enterohepatic cholesterol absorption. *J. Lipid Res.* **50**, 1653–62 (2009).
97. Lorenz, H., Hailey, D. W. & Lippincott-Schwartz, J. Fluorescence protease protection of GFP chimeras to reveal protein topology and subcellular localization. *Nat. Methods* **3**, 205–10 (2006).
98. Lorenz, H., Hailey, D. W., Wunder, C. & Lippincott-Schwartz, J. The fluorescence protease protection (FPP) assay to determine protein localization and membrane topology. *Nat. Protoc.* **1**, 276–9 (2006).
99. Brach, T., Soyk, S., Müller, C., Hinz, G., Hell, R., Brandizzi, F. & Meyer, A. J. Non-invasive topology analysis of membrane proteins in the secretory pathway. *Plant J.* **57**, 534–41 (2009).
100. Kim, H., Melén, K., Osterberg, M. & von Heijne, G. A global topology map of the *Saccharomyces cerevisiae* membrane proteome. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 11142–7 (2006).
101. Daley, D. O., Rapp, M., Granseth, E., Melén, K., Drew, D. & von Heijne, G. Global topology analysis of the *Escherichia coli* inner membrane proteome. *Science* **308**, 1321–3 (2005).
102. Haider A. J., Briggs D., Self T. J., Chilvers H. L., H. N. D. & K. I. D. Dimerization of ABCG2 analysed by bimolecular fluorescence complementation. *PLoS One* **6**, e25818 (2011).
103. Newton, S. M., Klebba, P. E., Michel, V., Hofnung, M. & Charbit, A. Topology of the membrane protein LamB by epitope tagging and a comparison with the X-ray model. *J. Bacteriol.* **178**, 3447–56 (1996).
104. Liu, X. Y. & Matherly, L. H. Analysis of membrane topology of the human reduced folate carrier protein by hemagglutinin epitope insertion and scanning glycosylation insertion mutagenesis. *Biochim. Biophys. Acta* **1564**, 333–42 (2002).

105. Banerjee, A. & Swaan, P. W. Membrane topology of human ASBT (SLC10A2) determined by dual label epitope insertion scanning mutagenesis. New evidence for seven transmembrane domains. *Biochemistry* **45**, 943–53 (2006).
106. Hong, M., Tanaka, K., Pan, Z., Ma, J. & You, G. Determination of the external loops and the cellular orientation of the N- and the C-termini of the human organic anion transporter hOAT1. *Biochem. J.* **401**, 515–20 (2007).
107. Feramisco, J. D., Goldstein, J. L. & Brown, M. S. Membrane topology of human insig-1, a protein regulator of lipid synthesis. *J. Biol. Chem.* **279**, 8487–96 (2004).
108. Renard, S., Lingueglia, E., Voilley, N., Lazdunski, M. & Barbry, P. Biochemical analysis of the membrane topology of the amiloride-sensitive Na⁺ channel. *J. Biol. Chem.* **269**, 12981–6 (1994).
109. Salaün, C., Rodrigues, P. & Heard, J. M. Transmembrane topology of PiT-2, a phosphate transporter-retrovirus receptor. *J. Virol.* **75**, 5584–92 (2001).
110. Wong, S. C., Zhang, L., Proefke, S. A. & Matherly, L. H. Effects of the loss of capacity for N-glycosylation on the transport activity and cellular localization of the human reduced folate carrier. *Biochim. Biophys. Acta* **1375**, 6–12 (1998).
111. Unal, E. S., Zhao, R., Qiu, A. & Goldman, I. D. N-linked glycosylation and its impact on the electrophoretic mobility and function of the human proton-coupled folate transporter (HsPCFT). *Biochim. Biophys. Acta* **1778**, 1407–14 (2008).
112. Cao, W. & Matherly, L. H. Analysis of the membrane topology for transmembrane domains 7-12 of the human reduced folate carrier by scanning cysteine accessibility methods. *Biochem. J.* **378**, 201–6 (2004).
113. Motoda, H., Sasaki, T., Kano, Y., Ryan, P. R., Delhaize, E., Matsumoto, H. & Yamamoto, Y. . The Membrane Topology of ALMT1, an Aluminum-Activated Malate Transport Protein in Wheat (*Triticum aestivum*). *Plant Signal. Behav.* **2**, 467–72 (2007).
114. Eckert, B. & Beck, C. F. Topology of the transposon Tn10-encoded tetracycline resistance protein within the inner membrane of *Escherichia coli*. *J. Biol. Chem.* **264**, 11663–70 (1989).
115. Levy, S., Nguyen, V. Q., Andria, M. L. & Takahashi, S. Structure and membrane topology of TAPA-1. *J. Biol. Chem.* **266**, 14597–602 (1991).

116. Weston, S., Czieso, S., White, I. J., Smith, S. E., Kellam, P. & Marsh, M. A membrane topology model for human interferon inducible transmembrane protein 1. *PLoS One* **9**, e104341 (2014).
117. Ozohanics, O., Turiák, L., Drahos, L. & Vékey, K. Comparison of glycopeptide/glycoprotein enrichment techniques. *Rapid Commun. Mass Spectrom.* **26**, 215–7 (2012).
118. Ozohanics, O., Turiák, L., Puerta, A., Vékey, K. & Drahos, L. High-performance liquid chromatography coupled to mass spectrometry methodology for analyzing site-specific N-glycosylation patterns. *J. Chromatogr. A* **1259**, 200–12 (2012).
119. Steentoft C., Vakhrushev S. Y., Joshi H. J., Kong Y., Vester-Christensen M. B., Schjoldager K. T., Lavrsen K., Dabelsteen S., Pedersen N. B., Marcos-Silva L., Gupta R., Bennett E. P., Mandel U., Brunak S., Wandall H. H., L. S. B. & C. H. Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *EMBO J.* **32**, 1478–88 (2013).
120. Steentoft C., Vakhrushev S. Y., Vester-Christensen M. B., Schjoldager K. T., Kong Y., Bennett E. P., Mandel U., Wandall H., L. S. B. & C. H. Mining the O-glycoproteome using zinc-finger nuclease-glycoengineered SimpleCell lines. *Nat. Methods* **8**, 977–82 (2011).
121. Vakhrushev S. Y., Steentoft C., Vester-Christensen M. B., Bennett E. P., C. H. & L. S. B. Enhanced mass spectrometric mapping of the human GalNAc-type O-glycoproteome with SimpleCells. *Mol. Cell. Proteomics* **12**, 932–44 (2013).
122. Zhu J., Sun Z., Cheng K., Chen R., Ye M., Xu B., Sun D., Wang L., Liu J., W. F. & Z. H. Comprehensive mapping of protein N-glycosylation in human liver by combining hydrophilic interaction chromatography and hydrazide chemistry. *J. Proteome Res.* **13**, 1713–21 (2014).
123. Chen R., Jiang X., Sun D., Han G., Wang F., Ye M., W. L. & Z. H. Glycoproteomics analysis of human liver tissue by combination of multiple enzyme digestion and hydrazide chemistry. *J. Proteome Res.* **8**, 651–61 (2009).

124. Bausch-Fluck D., Hofmann A., Bock T., Frei A. P., Cerciello F., Jacobs A., Moest H., Omasits U., Gundry R. L., Yoon C., Schiess R., Schmidt A., Mirkowska P., Härtlová A., Van Eyk J. E., Bourquin J. P., Aebersold R., Boheler K. R., Z. P. & W. B. A mass spectrometric-derived cell surface protein atlas. *PLoS One* **10**, e0121314 (2015).
125. Mendoza, V. L. & Vachet, R. W. Probing protein structure by amino acid-specific covalent labeling and mass spectrometry. *Mass Spectrom. Rev.* **28**, 785–815 (2009).
126. Calabrese, A. N. & Pukala, T. L. Chemical Cross-linking and Mass Spectrometry for the Structural Analysis of Protein Assemblies. *Aust. J. Chem.* **66**, 749–759 (2013).
127. Leitner A., Joachimiak L. A., Unverdorben P., Walzthoeni T., Frydman J., Förster, F. & Aebersold, R. Chemical cross-linking/mass spectrometry targeting acidic residues in proteins and protein complexes. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 9455–60 (2014).
128. Stengel, F., Aebersold, R. & Robinson, C. V. Joining forces: integrating proteomics and cross-linking with the mass spectrometry of intact complexes. *Mol. Cell. Proteomics* **11**, R111.014027 (2012).
129. Leitner, A., Walzthoeni, T., Kahraman, A., Herzog, F., Rinner, O., Beck, M. & Aebersold, R. Probing native protein structures by chemical cross-linking, mass spectrometry, and bioinformatics. *Mol. Cell. Proteomics* **9**, 1634–49 (2010).
130. Roesli, C., Mumprecht, V., Neri, D. & Detmar, M. Identification of the surface-accessible, lineage-specific vascular proteome by two-dimensional peptide mapping. *FASEB J.* **22**, 1933–44 (2008).
131. Niehage C., Steenblock C., Pursche T., Bornhäuser M., Corbeil, D. & Hoflack, B. The cell surface proteome of human mesenchymal stromal cells. *PLoS One* **6**, e20399 (2011).
132. Karhemo P. R., Ravela S., Laakso M., Ritamo I., Tatti O., Mäkinen S., Goodison S., Stenman U. H., Hölttä E., Hautaniemi S., Valmu L., Lehti, K. & Laakkonen, P. An optimized isolation of biotinylated cell surface proteins reveals novel players in cancer metastasis. *J. Proteomics* **77**, 87–100 (2012).
133. Bausch-Fluck, D., Hofmann, A. & Wollscheid, B. Cell surface capturing technologies for the surfaceome discovery of hepatocytes. *Methods Mol. Biol.* **909**, 1–16 (2012).

134. Hofmann, A., Thiesler, T., Gerrits, B., Behnke, S., Sobotzki, N., Omasits, U., Bausch-Fluck, D., Bock, T., Aebersold, R., Moch, H., Tinguely, M. & Wollscheid, B. Surfaceome of classical Hodgkin and non-Hodgkin lymphoma. *Proteomics. Clin. Appl.* **9**, 661–70 (2015).
135. Weekes, M. P., Antrobus, R., Lill, J. R., Duncan, L. M., Hör, S. & Lehner, P. J. Comparative analysis of techniques to purify plasma membrane proteins. *J. Biomol. Tech.* **21**, 108–15 (2010).
136. Glish, G. L. & Vachet, R. W. The basics of mass spectrometry in the twenty-first century. *Nat. Rev. Drug Discov.* **2**, 140–50 (2003).
137. Aebersold, R. & Goodlett, D. R. Mass spectrometry in proteomics. *Chem. Rev.* **101**, 269–95 (2001).
138. Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. & Whitehouse, C. M. Electrospray ionization for mass spectrometry of large biomolecules. *Science* **246**, 64–71 (1989).
139. Wilm, M. Principles of electrospray ionization. *Mol. Cell. Proteomics* **10**, M111.009407 (2011).
140. Han, X., Aslanian, A. & Yates, J. R. Mass spectrometry for proteomics. *Curr. Opin. Chem. Biol.* **12**, 483–90 (2008).
141. Zhang, H. & Ge, Y. Comprehensive analysis of protein modifications by top-down mass spectrometry. *Circ. Cardiovasc. Genet.* **4**, 711 (2011).
142. Biemann, K. Appendix 5. Nomenclature for peptide fragment ions (positive ions). *Methods Enzymol.* **193**, 886–7 (1990).
143. Medzihradszky, K. F. & Chalkley, R. J. Lessons in de novo peptide sequencing by tandem mass spectrometry. *Mass Spectrom. Rev.* **34**, 43–63 (2015).
144. Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature* **422**, 198–207 (2003).
145. Perkins, D. N., Pappin, D. J., Creasy, D. M. & Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551–67 (1999).

146. Chalkley, R. J., Baker, P. R., Huang, L., Hansen, K. C., Allen, N. P., Rexach, M. & Burlingame, A. L. Comprehensive analysis of a multidimensional liquid chromatography mass spectrometry dataset acquired on a quadrupole selecting, quadrupole collision cell, time-of-flight mass spectrometer: II. New developments in Protein Prospector allow for reliable and. *Mol. Cell. Proteomics* **4**, 1194–204 (2005).
147. Apweiler, R., Bairoch, A., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., Martin, M. J., Natale, D. A., O'Donovan, C., Redaschi, N. & Yeh, L. S. L. UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res.* **32**, D115–9 (2004).
148. Elias, J. E. & Gygi, S. P. Target-decoy search strategy for mass spectrometry-based proteomics. *Methods Mol. Biol.* **604**, 55–71 (2010).
149. Tusnády, G. E., Kalmár, L., Hegyi, H., Tompa, P. & Simon, I. TOPDOM: database of domains and motifs with conservative location in transmembrane proteins. *Bioinformatics* **24**, 1469–70 (2008).
150. Lowry, O. H., Rosebrough, N. J., Farr, A. L. & Randall, R. J. Protein measurement with the Folin phenol reagent. *J. Biol. Chem.* **193**, 265–75 (1951).
151. Guan, S., Price, J. C., Prusiner, S. B., Ghaemmaghami, S. & Burlingame, A. L. A data processing pipeline for mammalian proteome dynamics studies using stable isotope metabolic labeling. *Mol. Cell. Proteomics* **10**, M111.010728 (2011).
152. Creasy, D. M. & Cottrell, J. S. Unimod: Protein modifications for mass spectrometry. *Proteomics* **4**, 1534–6 (2004).
153. Thomsen, M. C. F. & Nielsen, M. Seq2Logo: a method for construction and visualization of amino acid binding motifs and sequence profiles including sequence weighting, pseudo counts and two-sided representation of amino acid enrichment and depletion. *Nucleic Acids Res.* **40**, W281–7 (2012).
154. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–2 (2012).
155. Langó, T., Róna, G., Hunyadi-Gulyás, É., Turiák, L., Varga, J., Dobson, L., Várady, Gy., Drahos, L., Vértessy, B. G., Medzihradszky, K. F., Szakács, G. & Tusnády, G. E. Identification of Extracellular Segments by Mass Spectrometry Improves Topology Prediction of Transmembrane Proteins. *Sci. Rep.* **7**, 42610 (2017).

156. Boudanova, E., Navaroli, D. M. & Melikian, H. E. Amphetamine-induced decreases in dopamine transporter surface expression are protein kinase C-independent. *Neuropharmacology* **54**, 605–12 (2008).
157. Apweiler, R., Hermjakob, H. & Sharon, N. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim. Biophys. Acta* **1473**, 4–8 (1999).
158. Li, J., Quilty, J., Popov, M. & Reithmeier, R. A. Processing of N-linked oligosaccharide depends on its location in the anion exchanger, AE1, membrane glycoprotein. *Biochem. J.* **349**, 51–7 (2000).
159. Shnitsar, V., Li, J., Li, X., Calmettes, C., Basu, A., Casey, J. R., Moraes, T. F. & Reithmeier, R. A. F. A substrate access tunnel in the cytosolic domain is not an essential feature of the solute carrier 4 (SLC4) family of bicarbonate transporters. *J. Biol. Chem.* **288**, 33848–60 (2013).
160. Klammer, M., Messina, D. N., Schmitt, T. & Sonnhammer, E. L. L. MetaTM - a consensus method for transmembrane protein topology prediction. *BMC Bioinformatics* **10**, 314 (2009).
161. Marsico, A., Scheubert, K., Tuukkanen, A., Henschel, A., Winter, C., Winnenburger, R. & Schroeder, M. MeMotif: a database of linear motifs in alpha-helical transmembrane proteins. *Nucleic Acids Res.* **38**, D181–9 (2010).
162. Goudenège, D., Avner, S., Lucchetti-Miganeh, C. & Barloy-Hubler, F. CoBaltDB: Complete bacterial and archaeal orfeomes subcellular localization database and associated resources. *BMC Microbiol.* **10**, 88 (2010).
163. Lo, A., Cheng, C.-W., Chiu, Y.-Y., Sung, T.-Y. & Hsu, W.-L. TMPad: an integrated structural database for helix-packing folds in transmembrane proteins. *Nucleic Acids Res.* **39**, D347–55 (2011).
164. Farrell, K. B., Tusnady, G. E. & Eiden, M. V. New structural arrangement of the extracellular regions of the phosphate transporter SLC20A1, the receptor for gibbon ape leukemia virus. *J. Biol. Chem.* **284**, 29979–87 (2009).
165. Butler, E. K., Davis, R. M., Bari, V., Nicholson, P. A. & Ruiz, N. Structure-function analysis of MurJ reveals a solvent-exposed cavity containing residues essential for peptidoglycan biogenesis in *Escherichia coli*. *J. Bacteriol.* **195**, 4639–49 (2013).

-
166. Tsaousis, G. N., Tsirigos, K. D., Andrianou, X. D., Liakopoulos, T. D., Bagos, P. G. & Hamodrakas, S. J. ExTopoDB: a database of experimentally derived topological models of transmembrane proteins. *Bioinformatics* **26**, 2490–2 (2010).
167. Skvortsova, T. E., Rykova, E. Y., Tamkovich, S. N., Bryzgunova, O. E., Starikov, A. V., Kuznetsova, N. P., Vlassov, V. V. & Laktionov, P. P. Cell-free and cell-bound circulating DNA in breast tumours: DNA quantification and analysis of tumour-related gene methylation. *Br. J. Cancer* **94**, 1492–5 (2006).
168. Semeraro, F., Ammollo, C. T., Morrissey, J. H., Dale, G. L., Friese, P., Esmon, N. L. & Esmon, C. T. Extracellular histones promote thrombin generation through platelet-dependent mechanisms: involvement of platelet TLR2 and TLR4. *Blood* **118**, 1952–61 (2011).
169. Quillen, M., Castello, C., Krishan, A. & Rubin, R. W. Cell surface tubulin in leukemic cells: molecular structure, surface binding, turnover, cell cycle expression, and origin. *J. Cell Biol.* **101**, 2345–54 (1985).
170. Dudani, A. K. & Ganz, P. R. Endothelial cell surface actin serves as a binding site for plasminogen, tissue plasminogen activator and lipoprotein(a). *Br. J. Haematol.* **95**, 168–78 (1996).
171. Nury, H., Dahout-Gonzalez, C., Trézéguet, V., Lauquin, G., Brandolin, G. & Pebay-Peyroula, E. Structural basis for lipid-mediated interactions between mitochondrial ADP/ATP carrier monomers. *FEBS Lett.* **579**, 6031–6 (2005).

7 Mellékletek

7.1 Felhasznált anyagok melléklete

Sigma-Aldrich: 1,10-Phenanthroline; 2,2'-Thiodiethanol; Bromoethylamine, BSA- Bovine Serum Albumin; CaCl₂; CuSO₄*5H₂O; EDTA- Ethylenediaminetetraacetic acid; F12 HAM médium; FITC (Fluorescein isothiocyanate) konjugált anti-biotin antitest; Folin-Ciocalteu reagens; jóacetamid, KCl; Kimotripszin; MS Grade Tripszin; Na₂CO₃; Na₂-tartarát*2H₂O; NaCl; NaHCO₃; NaOH; NEM- N-ethylmaleimide; NH₄HCO₃; PFA- Paraformaldehyde; PMSF- Phenylmethylsulfonyl fluoride; SDS- Sodium dodecyl sulfate; TCEP- Tris(2-carboxyethyl)phosphine; Termolizin; TLCK- Na-tosyl-L-lysine chloromethyl ketone hydrochloride; TPCK- N-p-Tosyl-L-phenylalanine chloromethyl ketone; Tripszin; TritonX-100

Thermo Scientific: DTT- Ditiotreitól; FBS- Fetal Bovine Serum; NeutrAvidin Agarose; PBS- Phosphate-buffered saline; Pen/Strep- Penicillin Streptomycin Solution; RPMI 1640 médium; Szulfo-NHS-SS-Biotin; TBS- Tris-buffered saline

Bio-Rad: Coomassie Brilliant Blue (CBB) R-250, Tris- Tris(hydroxymethyl)aminomethane; Tween-20

Millipore: ECL Western Chemiluminescent HRP Substrate

New England Biolabs: PNGase F- N-glycosidase F

EuroClone: Prestained Protein Sharpmass VI

Molecular Probes: Hoechst 33342

Merck: HRP (Horseradish peroxidase)-konjugált avidin; PVDF (Polyvinylidene difluoride) membrán; ZnCl₂

Waters: Rapigest

Biomarker Kft.: TexasRed[®] Streptavidin

7.2 Felhasznált eszközök melléklete

Sigma: 2 ml-es potter cső és homogenizáló rúd; 26G-s tű; Corning[®] bottle-top vákuumszűrő rendszer

Thermo Scientific: C18 Spin oszlopok; Linear Trap Quadrupole (LTQ)-Orbitrap Elite; Orion 2-Star pH mérő; Pierce™ Spin Columns - Snap Cap, Dionex Ultimate 3000 NanoLC System (Sunnyvale), Acclaim™ PepMap™ 100 C18 Nano-Trap Columns

Beckman: 10,4ml-es polikarbonát csövek (355603); L7-55 Ultracentrifuga-70.1 Ti fix rotorral

Eppendorf: 5810 R és 5804 R centrifuga; Galaxy 170R inkubátor; T25, T75 sejttenyésztő flaska

Greiner Bio One: 6ml-es EDTA tri-kálium cső

Bio-Rad: Bio-Dot® Microfiltration System; ChemiDoc™ XRS+; TC20™ Automated Cell Counter; PowerPac™ Universal és gélfuttató kádak

Waters: Bruker Maxis II ETD Q-TOF; nanoAcquity UPLC trap oszlop; BEH300C18 nanoAcquity UPLC oszlop

Hirschmann: Bürker-kamra

Biosan: CH-100 Termoblokk; End-to-end forgató

Applied Biosystem: FACS Attune® Acoustic Focusing Cytometer (Blue/Violet)

AnalytikJena: Spektrofotométer

Esco: Streamline® Class II BSC lamináris fülke

Merck: Ultratiszta víz előállító berendezés

Promega: Vac-Man™ Vacuum Manifold

Elma Schmidbauer GmbH: Elmasonic S 30 (H)-Vízfürdős szonikátor

Zeiss: ZEISS LSM 710

7.3 Alkalmazott módszerek melléklete

7.3.1 A felszíni jelölés eredményességének ellenőrzése

Először az RBC-k és CHO emlőssejtek membránfehérjéinek jelölhetőségét vizsgáltuk Szulfo-NHS-SS-biotin reagenssel, a 3.3.2 és 3.3.3. fejezet lépéseit elvégezve. A CHO sejtek letapadó sejtek, amelyre azért esett a választás, mert a konfokális mikroszkópon kontrasztosabb képeket lehet kapni velük. A 24-lyukú plate lemez welljeibe 13 mm-es óraüveget helyeztünk, majd azon növesztettük a CHO sejteket ~80%-os konfluenciáig. A

médiüm eltávolítása után 2-szer PBS pufferrel átmostuk a wellek sejtjeit, majd következett a jelölési reakció (a negatív kontroll minták jelölő ágens nélküli pufferben inkubáltak).

A jelölt és jelöletlen sejteket 6%-os BSA-val PBS pufferben blokkoltuk 4°C-on 10 percig, ezzel csökkentve a következő lépésben az sztreptavidin nem specifikus kötődését.

A sejteket biotinra specifikus, 2 µg/ml-es TexasRed fluoreszcens festékkel konjugált sztreptavidinnel inkubáltuk egy órán keresztül 4°C-on (a TexasRed gerjesztési és emissziós maximuma: 596/615nm). Végül a sejteket egyszer mostuk PBS pufferrel, majd 0,1 µg/ml Hoechst 33342 DNS festékkel (gerjesztési és emissziós maximuma: 350/461 nm) inkubáltuk 2 percig, a festék feleslegét szintén PBS pufferrel távolítottuk el.

A jelöletlen és jelölt mintákban a festékek jelenlétét, lokalizációját Zeiss LSCM 710 konfokális mikroszkóppal vizsgáltuk 63x NA=1.4 Plan Apo objektív felhasználásával. A festékek helyzetét a sejtek adott síkjában a megfelelő hullámhosszúságú gerjesztő lézerek felhasználásával, valamint emissziós spektrumukra jellemző hullámhosszúságú fény kiszűrésére alkalmas szűrőkockákkal rögzítettük/fényképeztük.

A jelölt és jelöletlen RBC-ékből membránpreparátumot is készítettünk (a 3.3.4 fejezet alapján), majd TexasRed konjugált sztreptavidinnel inkubáltuk a keletkezett vezikula membránokat. Szintén konfokális mikroszkóppal vizsgáltuk meg, hogy a jelölés megmaradt-e a membránpreparálás után is (12. ábra E és F).

7.3.2 Jelölt és jelöletlen sejtek viabilitásának tesztelése

Annak érdekében, hogy igazoljuk, hogy a jelölő ágens nincs hatással a sejtintegritásra, a sejt-maggal rendelkező teszt sejtek (HL60 és K562) három-három sejt csoportján elvégeztük a 3.3.2-3.3.3. fejezetekben ismertetett lépéseket (kezelt/jelölt minták). További 3-3 mintán szintén elvégeztük a megfelelő lépéseket, de a jelölő reagens nélkül (kezeletlen minták). Az összes mintát 1 µg/ml propídium-jodiddal (PI, gerjesztés és emissziós maximuma: 493/636 nm) 5 percen keresztül szobahőmérsékleten inkubáltuk. A sérült sejtekkel ellentétben az ép sejtek a PI-t nem veszik fel, a sérült sejtek DNS-e megfestődik a PI hatására. A DNS magok PI festődését áramlási citometriával (FACS Attune® Acoustic Focusing Cytometer) határoztuk meg, amelyből következtethettünk az egyes minták sérült sejtszám arányára. A kezelt és kezeletlen sejtek PI felvétele közötti különbség szignifikanciáját nem-parametrikus Mann-Whitney teszttel határoztuk meg.

7.3.3 Fehérjemérés

A membránpreparátumok fehérjetartalmának kvantitatív meghatározásához Lowry (Folin) módszert alkalmaztunk. A módszerhez három fő oldat 98:1:1 (A:B:C oldat) arányú elegyét használtuk:

A: 4 mg/ml NaOH és 20 mg/ml Na₂CO₃ ultratiszta vízben

B: 10 mg/ml CuSO₄*5H₂O ultratiszta vízben

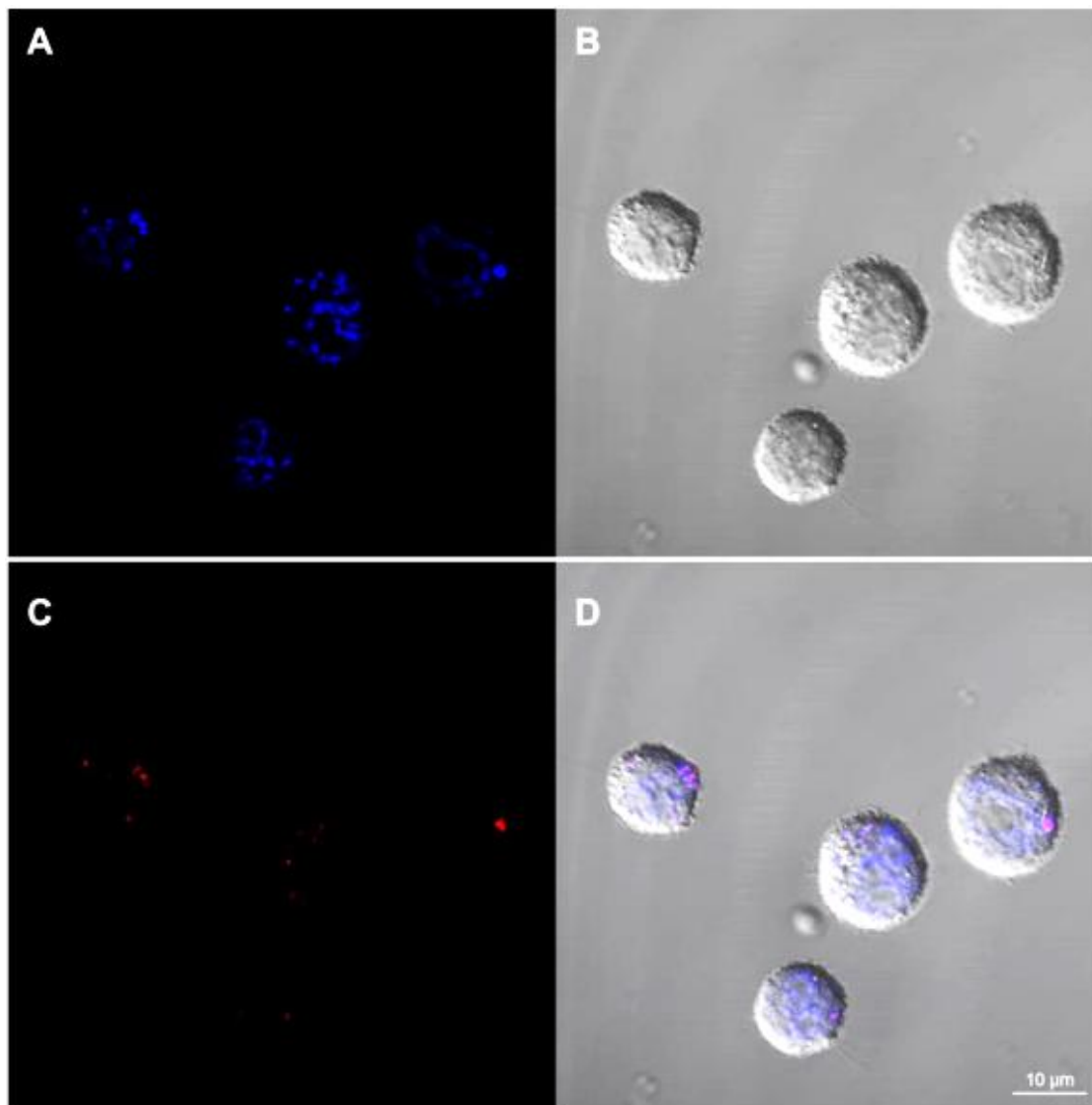
C: 28,5 mg/ml Na₂-tartarát*2H₂O ultratiszta vízben

A fehérje koncentráció meghatározásához szükséges kalibrációhoz a BSA fehérje hígítási sorát, valamint membránpreparátum-minták megfelelő mennyiségét hozzáadtuk a három oldat 1-1ml-es elegyéhez. A vak minták esetében, a membránpreparátumokat kísérő puffer közeget (50 mM NH₄HCO₃, pH=8) adtuk az előzőkhez adott mennyiségben. Végül Folin-Ciocalteu (foszfomolibdát-foszfowolframát) reagenst adtuk a rendszerhez és 45 percig sötétben inkubáltuk az oldatokat.

A kék színváltozást 660nm-en spektrofotométerrel detektáltuk, majd a BSA hígítási sor abszorbancia értékeire kalibrációs egyenest illesztettünk, amely egyenletéből meghatározhatuk a membránpreparátumok fehérjetartalmát.

7.4 Eredmények melléklete

7.4.1 Szulfo-NHS-SS-biotin jelölés nélküli sejtek vizsgálata



28. ábra: Biotinátlan CHO sejtek, TexasRed konjugált avidinnel inkubálva, DNS festése Hoechst 33342 reagenssel. A: Hoechst (DNS marker), B: DIC, C: TexasRed (háttér fluoreszcencia), D: egyesített ábra (az előző három blokk tartalma).

ADATLAP

a doktori értekezés nyilvánosságra hozatalához*

I. A doktori értekezés adatai

A szerző neve: Langó Tamás

MTMT-azonosító: 10054034

A doktori értekezés címe és alcíme: Humán transzmembrán proteom topológiájának meghatározása kombinált kísérleti és elméleti módszerekkel

DOI-azonosító⁴⁶: 10.15476/ELTE.2018.190

A doktori iskola neve: Biológia Doktori Iskola

A doktori iskolán belüli doktori program neve: Szerkezeti Biokémia Program

A témavezető neve és tudományos fokozata: Tusnády Gábor, PhD

A témavezető munkahelye: MTA TTK, Enzimológia Intézet

II. Nyilatkozatok

1. A doktori értekezés szerzőjeként

a) hozzájárulok, hogy a doktori fokozat megszerzését követően a doktori értekezésem és a tézisek nyilvánosságra kerüljenek az ELTE Digitális Intézményi Tudástárban. Felhatalmazom a Természettudományi kar Dékáni Hivatali Doktori, Habilitációs és Nemzetközi Ügyek Csoportjának ügyintézőjét, hogy az értekezést és a téziseket feltöltse az ELTE Digitális Intézményi Tudástárba, és ennek során kitöltse a feltöltéshez szükséges nyilatkozatokat.

b) kérem, hogy a mellékelt kérelemben részletezett szabadalmi, illetőleg oltalmi bejelentés közzétételéig a doktori értekezést ne bocsássák nyilvánosságra az Egyetemi Könyvtárban és az ELTE Digitális Intézményi Tudástárban;

c) kérem, hogy a nemzetbiztonsági okból minősített adatot tartalmazó doktori értekezést a minősítés (*dátum*)-ig tartó időtartama alatt ne bocsássák nyilvánosságra az Egyetemi Könyvtárban és az ELTE Digitális Intézményi Tudástárban;

d) kérem, hogy a mű kiadására vonatkozó mellékelt kiadó szerződésre tekintettel a doktori értekezést a könyv megjelenéséig ne bocsássák nyilvánosságra az Egyetemi Könyvtárban, és az ELTE Digitális Intézményi Tudástárban csak a könyv bibliográfiai adatait tegyék közzé. Ha a könyv a fokozatszerzést követően egy évig nem jelenik meg, hozzájárulok, hogy a doktori értekezésem és a tézisek nyilvánosságra kerüljenek az Egyetemi Könyvtárban és az ELTE Digitális Intézményi Tudástárban.

2. A doktori értekezés szerzőjeként kijelentem, hogy

a) az ELTE Digitális Intézményi Tudástárba feltöltendő doktori értekezés és a tézisek saját eredeti, önálló szellemi munkám és legjobb tudomásom szerint nem sértem vele senki szerzői jogait;

b) a doktori értekezés és a tézisek nyomtatott változatai és az elektronikus adathordozón benyújtott tartalmak (szöveg és ábrák) mindenben megegyeznek.

3. A doktori értekezés szerzőjeként hozzájárulok a doktori értekezés és a tézisek szövegének plágiumkereső adatbázisba helyezéséhez és plágiumellenőrző vizsgálatok lefuttatásához.

Kelt: Budapest, 2018.09.14.....

.....

.....
a doktori értekezés szerzőjének aláírása

*ELTE SZMSZ SZMR 12. sz. melléklet