

# Proteomika – az új kihívás

Hudecz Ferenc

A proteom fogalma (proteome, *protein complement to a genome*) az utóbbi években vált a tudományos köztudat részévé. A proteomika a proteom, vagyis az élő szervezetben előforduló összes, szerkezetében akár a legkisebb mértékben eltérő fehérje megismerésével foglalkozó tudományterület, amely a genommal kapcsolatos kutatás mintájára, annak kiegészítőjeként jött létre, de ma már a genomikától független, önálló diszciplína.

A proteomika meg kívánja ismerni a fehérjék szerkezetét, biológiai funkcióját és ezek térbeli és időbeli változását. Nemcsak úgy tekinti a fehérjét, mint izolált molekulát, hanem figyelembe veszi a fehérje és környezete közötti kölcsönhatást. Vizsgálódási körébe tartozik a fehérje eredetének meghatározása, rendszertani besorolása, a különböző forrásból származó, de azonos biológiai funkciót ellátó proteinek szerkezetének összehasonlítása, a fehérje jelenlétének vagy hiányának igazolása egészséges, illetve patológiás körülmények között. Célja – a korszerű kémiai analitika eszközeit felhasználva – az igen kis mennyiségben, illetve koncentrációban jelen lévő fehérjék kimutatása, azonosítása, szerkezetük meghatározása; a fehérjékkel kapcsolatos ismeretek taxonómiai, szerkezeti vagy funkcionális rendszerezése, a kísérleti adatok megerősítése (validálás), valamint az így képződő adathalmazok (adatbázisok) információtartalmának elemzése. Sokak nézete szerint ide sorolhatóak a proteomika módszertanával foglalkozó témakörök (például nagy érzékenységű tömegspektrometriás módszerek, fluoreszcencián alapuló mikroszkópos technikák, DNS-csipek fejlesztése stb.). Már most jól látszik, hogy a proteomika jelentősen segíti a korszerű gyógyszerkutatást és az orvostársadalmat a mai nál hatékonyabb diagnosztikai eljárások és terápiás szerek kifejlesztésében.

**proteomika, proteom, fehérjeszerkezet, fehérjeelválasztás, protein-bioinformatika**

## PROTEOMICS – THE NEW CHALLENGE

The term "proteome" (proteome, *PROTEin complement to a genOME*) by now a generally accepted expression in biomedical science meaning a complete complements of proteins. The discipline "proteomics", coined after proteome, deals with the analysis of the complete set of proteins occurring in the living organism. This includes the identification and quantification of proteins, the determination of protein localisation, modifications, interactions, activities and function. Performing comparative studies is an important part of proteomics for the analysis of proteins in health and disease. The knowledge generated is already used for improved diagnostic procedures and development of new drugs and therapies. During the proteome analysis, as outlined in this paper, even very small quantities (concentrations) of proteins are measured, then the protein is identified and its structure is elucidated. This procedure is followed by functional studies. An important part of proteomics is the collection and validation of numerical databases suitable for data mining. There is a general understanding that methodology driven research (e.g. NMR, mass spectrometry, DNA chips) is also an integrated part of this discipline. It is already sensed that the analysis of proteome can lead to the discovery of new proteins proving targets for drug research and to the establishment of new procedures with a perspective of improved diagnosis and therapy.

**proteomics, proteome, protein structure, separation and isolation of proteins, protein bioinformatics**

dr. Hudecz Ferenc (levelezés/correspondence): MTA-ELTE Peptidkémiai Kutatócsoport/Research Group of Peptide Chemistry at Eötvös Loránd University, Hungarian Academy of Sciences; H-1117 Budapest, Pázmány Péter sétány 1/A. E-mail: hudecz@szerves.chem.elte.hu

Érkezett: 2002. november 18. Elfogadva: 2003. január 10.

*Szerkesztőségi megjegyzés:* E témához szorosan kapcsolódó közleményeket olvashatnak korábbi számainkban: Falus A, et al. Funkcionális genomika. LAM 2002;12(6-7):365-70., valamint Sasvári-Székely M. A Humán genom projekt. LAM 2003;13(2):102-9.

Az álom lényegében megvalósult. 2001 februárja óta gyakorlatilag ismerjük a teljes humán genom nukleotidszekvenciáját (1). A genom felépítésének ismerete azonban még nem jelenti a szervezetben előforduló, a genom által kódolt fehérjék mindegyikének ismeretét. Mára már az is kiderült, hogy a genomnak csak egy része kódol fehérjéket, ezek azonosítása azonban nem egyszerű feladat. Ma még nem tudják teljes bizonyossággal azonosítani az exon régiókat, az eltérő hasításból származó variánsokat („splicing variants”) és a kisméretű proteineket kódoló szegmenseket. A fehérjék számának becslését tovább nehezíti a proteinek szintjén tapasztalt polimorfizmus. Sok esetben már egy-egy aminosav-rész cseréje egy másikkal patológias állapotot hozhat összefüggésbe (például a sarlósejtes anaemia). Mégis az ismert polimorfizmusok többsége, az aminosavszekvenciában előforduló – akár jelentősebb – eltérés, nincs hatással a fehérjék biológiai viselkedésére. A polimorf fehérjék azonosítása, „felismerése” tehát nem nyilvánvaló.

## A humán proteom, a fehérjék száma

Becslések és éles viták tárgyát képezi a proteineket kódoló gének számának meghatározása. Ma ez a szám 20 000 és 35 000 között mozog. Csupán polipeptidláncok („naked” proteins) genetikai kódolásáról lehet beszélni, hiszen a génben rejlő információból ma még nem tudjuk megjósolni a proteinoldalláncokon, illetve a fehérje N- és/vagy C-terminálisán bekövetkező – transzláció utáni („post-translational”) – kémiai változásokat. A riboszómákon történt fehérjeszintézist követően ugyanis a molekula sokféle változáson mehet keresztül. Mai tudásunk szerint több mint százféle kémiai módosítás következhet be. A genom szintjén bekövetkező alternatív hasítás („splicing”) és a poszt-transzlációs módosítások miatt a humán genom a korábban genetikusok által valószínűsített 100 000-féle fehérjével szemben közel egymillió, kémiai szerkezetét tekintve különböző protein jelenlétét teszi lehetővé a szervezetben.

A számok már önmagukban is világosan jelzik, hogy mind a gének, mind pedig a fehérjék működésének megismerése komplex feladat, jókora kihívást jelent a modern természet- és orvostudomány számára. A probléma felismerését követte annak nevének nevezése. Így vált a tudományos köztudat részévé a proteom fogalma (proteome: *protein complement to a genome*). (Lásd a szójegyzéket!) A proteomika a proteom, vagyis az élő szervezetben előforduló összes, szerkezetében akár a legkisebb mértékben eltérő fehérje megismerését tűzte ki célul. A genommal kapcsolatos kutatás (genomika) analógiájaként, annak kiegészítőjeként jött létre, de ma már a genomikától független, önálló diszciplína.

A rendkívül nagyszámú fehérjével kapcsolatos információk feldolgozásának igénye hozta létre szinte egy időben, 2000-2001-ben a „Human Proteome Organisation” (HUPO) szervezetet, valamint a Svájci Bio-

## SZÓJEGYZÉK

**Genom:** A kifejezés Winkler (1920) nevéhez fűződik, aki a „genes and chromosomes” szavakból alkotta meg. Az IUPAC által elfogadott meghatározás szerint egy szervezet, egy sejt, egy organellum vagy egy vírus teljes kromoszomális és kromoszómán kívüli génjeinek összességét, vagyis a szervezet teljes DNS-állományát értik e kifejezés alatt.

**Genomika:** E kifejezés eredetileg egy újonnan elindított tudományos folyóirat neve volt (Genomics). Ma a genom szekvenciájára vonatkozó adatállományt, új gének felfedezését, a géntérképezést, különböző fajokból, fajtákból származó genomok összehasonlítását, sőt, az új, gén alapú eljárásokat is a genomika fogalmához sorolják (genomics – Roderick, 1986).

**Proteom:** A kifejezés Wilkins (1994) nevéhez fűződik, aki egy olaszországi konferencián (Conference on Genome and Protein Maps, Siena) használta először. Szerkesztése a „protein complement expressed by a genome” szövegből történt. Egy szervezet, egy sejt, egy organellum vagy egy vírus összes proteinjét, a szervezet teljes fehérjeállományát értik e kifejezés alatt.

**Proteomika:** A proteomot alkotó fehérjék mennyiségi és minőségi analízise, a különböző fajokból, fajtákból származó proteomok összehasonlítása; a proteom összetételének vizsgálata különböző körülmények között, új, a fenti problémák megoldását elősegítő eljárások és azok kutatása. Jelenti fehérjék azonosítását (szekvencia, 3D szerkezet) és mennyiségi meghatározását; a kísérleti adatok alapján adatbázisok összeállítását; a fehérje lokalizációját, szerkezeti módosítását (például transzláció utáni), funkciójának, aktivitásának megismerését kísérletileg vagy predikcióval; a proteom tanulmányozását a betegségek diagnosztizálása és terápiája céljából.

Készült a [www.genomicglossaries.com/content/proteomics.asp](http://www.genomicglossaries.com/content/proteomics.asp) honlap felhasználásával.

informatikai Intézet (Swiss Institute of Bioinformatics, SIB) és az Európai Bioinformatikai Intézet (European Bioinformatics Institute, EBI) közös programját. A „Human Proteomics Initiative” (HPI) célja az emberi proteinnel kapcsolatos információk gyűjtése, rendszerezése és közreadása az élettudományokkal foglalkozó közösség számára (2).

A program kezdeményezői azoktól az angol lexikonkészítőktől és filológusoktól vették az ötletet, akik 1857-ben elhatározták, hogy összegyűjtik az összes információt az angol nyelvben használt valamennyi szó jelentésével és használatával kapcsolatban. A monumentális munka több évtizedig tartott és máig páratlan szótárhoz vezetett. Így készült az Oxford English Dictionary (OED). Az OED létrehozásakor – és azóta is – nyelvészek egy csoportja azzal a kéréssel fordult a világ angol nyelven beszélőjéhez, írjanak arról, hogy az egyes szavakat milyen értelemben, összefüggésben használják, használták nyelvi környezetükben. A HPI kezdeményezői szó szerint ezt kérték a világban dolgo-

zó kutatóktól, csak az „angol nyelv/angol szó” kifejezések helyére a „humán proteom/humán protein” kifejezést illesztették.

## A proteomika

A proteomika a fehérjék kutatásával komplex módon foglalkozó tudományterület, amely meg kívánja ismerni a fehérjék szerkezetét, biológiai funkcióját és ezek térbeli és időbeli változását. A proteomika nemcsak úgy tekinti a fehérjét, mint izolált molekulát, hanem figyelembe veszi a fehérje és környezete közötti kölcsönhatást. Ugyancsak a proteomika vizsgálódási körébe tartozik a fehérje eredetének meghatározása, rendszertani besorolása, a különböző forrásból származó, de azonos biológiai funkciót ellátó proteinek szerkezetének összehasonlítása, a fehérje jelenlétének vagy hiányának igazolása egészséges, illetve patológiás körülmények között.

A proteomika célja – a korszerű kémiai analitika eszközeit felhasználva – az igen kis mennyiségben, illetve koncentrációban jelen lévő fehérjék kimutatása, azonosítása, szerkezetük meghatározása. A proteomika fontos területe a fehérjékkel kapcsolatos taxonómiai, szerkezeti vagy funkcionális jellegű ismeretek rendszerezése, a kísérleti adatok megerősítése (validálás), valamint az így képződő adathalmazok (adatbázisok) információtartalmának elemzése. Sokak nézete szerint a proteomika nemcsak a fenti problémákkal kapcsolatos kutatásokat fedi le, hanem idesorolhatók a proteomika módszertanával foglalkozó témakörök (például a nagy érzékenységű tömegspektrometriás módszerek, a fluoreszcencián alapuló mikroszkópos technikák, a DNS-csipek fejlesztése stb.).

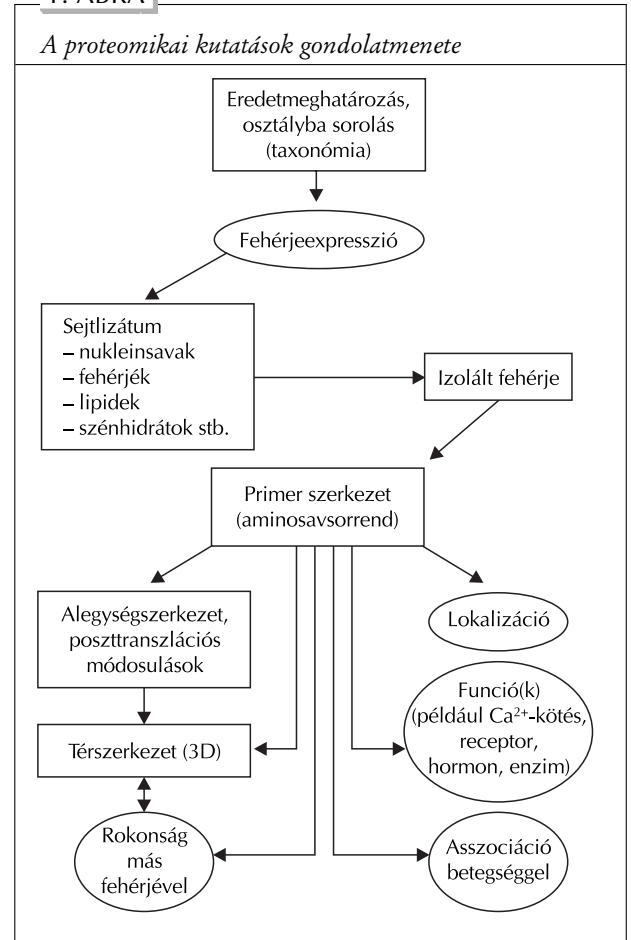
A proteomika kifejezés „divatos” is. Jól körvonalazható törekvés a kémia, a biokémia, a biofizika, valamint más természettudományok és az orvosbiológiai kutatások területén az a felfogás, amely szerint a proteomika inkább egy „komplex” megközelítési mód, mintsem egy kiválasztott fehérjével kapcsolatos kutatás.

A proteomikát gyakran tekintik azonosnak a funkcionális genomikával vagy éppen fordítva. Tartalmát tekintve a „metabolic engineering”, az „in silico & molecular modeling” típusú fehérjeszerkezet-meghatározás, sőt, a génexpresszió is a proteomika részét képezi (in silico: komputeres). A módszertani kutatás oldaláról pedig idesorolják a fehérjekromatográfiát, az elektroforézist, a fehérje-tömegspektrometriát, az NMR-t és a röntgenkristallográfiát.

## A proteom megismerésének gondolatmenete

A proteomika tárgykörébe tartozó kutatások egyre világosabban kibontakozó gondolatmenetét az 1. ábra foglalja össze. A folyamat a fehérje eredetének meghatározásával, rendszertani besorolásával kezdődik. A vizsgálat bizonyítja, hogy a fehérje szerkezete és funk-

1. ÁBRA



ciója az adott sejt, szövet, szervezet sajátja, vagyis a genom által kódolt. Ez azonban még nem jelenti azt, hogy a fehérje jelen van a biológiai mintában, vagy expresszálódik az adott kísérleti körülmények között. Ismeretes, hogy a génexpresszió programozott történéssorozat, függ a sejtciklustól, de jelentős mértékben befolyásolható külső körülmények által (például hőmérséklet, sugárzás, kémiai ágensek, a médium összetétele, pH-értéke).

A protein expresszióját és kimutatását a vegyület izolálása követi. A tiszta fehérje szerkezetének azonosítása az aminosavsorrend megállapításával (primer szerkezet) veszi kezdetét, majd az alegységstruktúra és a transláció utáni módosítások felderítése kerül sorra. A szerkezetről szerzett információkat akkor tekinthetjük teljesnek, ha a térbeli elrendeződésre (konformáció, 3D szerkezet) vonatkozó adatokat is – kísérletek segítségével – meghatároztuk.

Rendszerint már a struktúrával kapcsolatos részleges adatok birtokában elkezdődik a fehérje sejten belüli, illetve szöveti lokalizációjára vonatkozó adatgyűjtés. Ezzel párhuzamosan zajlik a protein biológiai hatásának, funkciójának (például transzport, enzim, immunreakció) tisztázása. Érdeemes megemlíteni, hogy a proteomika – jelen pillanatban – nem tekinti tárgyának a proteinszintézis, illetve -lebontás folyamatait. (Ezzel foglalkozik a *metabolomkutatás*.) Ugyanakkor nagy jelentőséget tulajdonít a különböző forrásból származó

fehérjék szerkezeti és funkcionális rokonsága vizsgálatának.

A proteomikának az orvosi diagnosztika és terápia számára talán legfontosabb területe azt próbálja meg felderíteni, hogy egyes betegségek miként hozhatók összefüggésbe bizonyos protein(ek) jelenlétével vagy éppen hiányával, esetleg mutációjával. A következőkben a teljesség igénye nélkül számba vesszük a folyamat egyes lépéseivel kapcsolatos problémákat. Az áttekintés izgalmas intellektuális és egyben kihívást jelentő olvasmány lesz, amely megvilágítja e napról napra fejlődő tudományterület sokszor egészen gyakorlati nehézségeit.

## A fehérjék expressziója

A fehérjék azonosításának előfeltétele, hogy jelen legyenek a vizsgálat tárgyát képező sejtben, szövetben, szervezetben. Ezt azt jelenti, hogy „normális” körülmények között vannak olyan fehérjék, amelyek lehetnének, de nincsenek jelen. Ebből a paradoxonból látszik, hogy a természetes körülmények között ki nem fejeződő proteinek megismeréséhez először olyan körülményeket kell előállítani, amelyek előidéznek a fehérje szintézisét. Egy másik probléma forrása az, hogy bizonyos fehérjék természetes körülmények között igen kis mennyiségben expresszálódnak. E kétféle nehézség megoldását segíti a rekombinántechnika. Rekombináns fehérjék gyors és nagy mennyiségű termeltetését eredményező expressziós rendszerek kutatása ugyancsak fontos kapcsot jelenthet új fehérjék, illetve új gének között is. Napjaink egyik legkedveltebb eljárása a célfehérjét kódoló gén beépítése és expressziója bakulovírusban. Az expressziós rendszerek kutatásával foglalkozó *proteomika* célja olyan eljárások kidolgozása, amelyek nagy mennyiségben képesek komplex proteinek gyors termeltetésére.

A nehézségek másik forrása az a megfigyelés, hogy a gének számánál sokszorosan több fehérje fordulhat elő a sejtekben. A humán genom génjeinek számát, mint említettük a bevezetésben, 20 000–35 000-re becsülik, míg a fehérjék „sokfélék”, számuk egyes vélekedések szerint 500 000 és egymillió közé esik. A sejtben megjelenő fehérjék minősége és mennyisége az „expressziós profillal” jellemezhető és jelentősen eltérhet egymástól. A külső körülmények (mint a hőmérséklet, a fény- vagy részecskesugárzás) nemcsak elősegíthetik vagy gátolhatják egyes proteinek expresszióját, hanem megváltoztathatják a fehérjék mennyiségét (koncentrációját) is. O'Farrell és munkatársai (3) két különböző összetételű táptalajon *E. coli* baktériumsejteket tenyésztettek. A fehérjéket kétdimenziós gélelektroforézissel – a proteinek mérete, illetve töltése alapján – választották el egymástól. A „normális”, teljes médiumban és az argininmentes táptalajon tenyésztett sejtek lizátumairól készült gélek összehasonlítása világosan mutatta, hogy bizonyos fehérjék az argininmentes körülmények között nem jelentek meg, míg más fehérjék kisebb vagy nagyobb mennyiségben

vannak jelen. Hasonló módon különböző „expressziós profil” regisztráltak abban az esetben is, amikor a sejtek táptalajához különböző gyógyszer-molekulákat vagy környezetszennyezőket (például nehézfémionokat, toxinokat) adtak. A toxikogenomika-kutatások ezt a jelenséget használják ki gyógyszerjelölt vegyületek által okozott károsodások, például a májkárosodás előjelzésére.

## A fehérjekomponensek izolálása

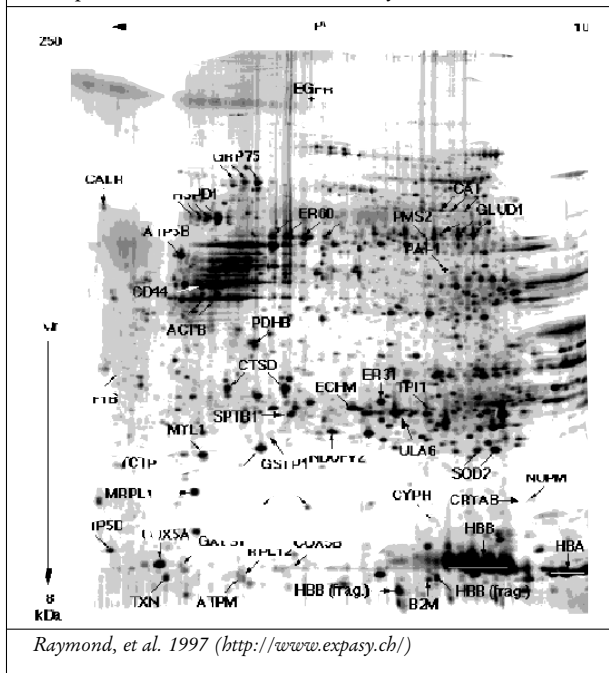
A sejtekben jelen levő fehérjék jelentős része a citoszolban oldott állapotban található, de nem elhanyagolható a sejtorganellumokban, membránban „rögzített” protein mennyisége. (A sejtek átlagos fehérjekoncentrációját 300 g/l-re becsülik.) Ahhoz, hogy tiszta, szerkezet- és funkcióvizsgálatokra alkalmas fehérjéhez jussunk, először a sejtek struktúrátságát kell megszüntetni. Ezt el lehet érni a sejtek „sokkolásával”. Különböző detergenssek, a hőmérséklet drasztikus emelése vagy csökkentése egyaránt alkalmas sejtlyázatok előállítására. A sejtlyázatok készítésével kapcsolatban azonban többféle nehézséggel számolhatunk. Például esetenként a sejtek dezintegrációja nem megy végbe teljesen, és a lizátum nemcsak molekulákat, hanem el nem roncsolt sejtorganellumokat, molekulakomplexeket is tartalmaz. A hidrofób (például membránból származó) fehérjék rossz oldékonyságuk miatt aggregátumokat hozhatnak létre. Másfajta probléma, hogy a lízis következményeként az addig „ellenőrzés alatt álló” enzimek (proteázok, foszfatázok stb.) aktivitása kontroll nélkülivé válik, és elindul a fehérjék degradációja. Például a lizoszómákba zárt enzimek térbeli elkülönülése a membránstruktúra felbomlásával megszűnik. A fehérjék enzimkatalizálta hidrolízise, részleges lebomlása azt is jelentheti, hogy olyan vegyületek is megjelennek a lizátumban, amelyek az élő sejtben nem fordulnak elő.

A sejtlyázum olyan oldat, amely komplex keverék, számos nem protein jellegű komponenssel. Érdemes megjegyezni, hogy nemcsak a lizátumok, hanem a *test-folyadék*ok (például vérplazma, cerebrospinalis folyadék) is fontos forrásai a proteomikai kutatásoknak. Ezek szintén komplex oldatok jelentős fehérjetartalommal. Becslések szerint például a humán plazmában 72 g/l a fehérjekoncentráció.

A fenti keverékek alkotórészei közül a fehérjék elválasztása elsősorban a proteinek eltérő mérete, töltése és esetenként specifikus, nagy affinitású kötődése alapján valósítható meg. Az alkalmazott módszerek túlnyomó többsége kromatográfia és kétdimenziós (2D) gélelektroforézis. A 2. ábra egészséges humán vastagbélből származó epithelsejtek lizátumáról (4), a 3. ábra pedig humán plazmáról (5) készült fehérjekeverékek komponenseit mutatja be. Az alkotórészeket először méret, majd pedig izoelektromos pont alapján választották el egymástól 2D gélelektroforézissel. Az ábrán látható pöttyök száma jelzi a fehérjekomponensek sokféleségét, a foltok mérete pedig a mennyiségi kü-

2. ÁBRA

Egészséges humán vastagbél-epithelium fehérjetérképe a sejtizlátum 2D gélelektroforézise alapján. A gélen a komponensek láthatóvá tétele ezüsfestéssel történt



Raymond, et al. 1997 (<http://www.expasy.ch/>)

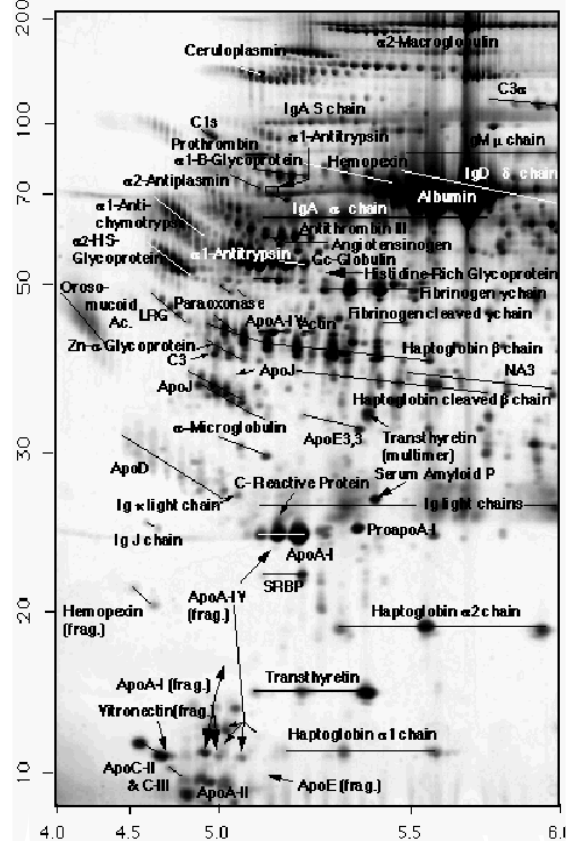
lönbségekre utal. Egy folt azonban nem biztosan csupán egyféle vegyületet tartalmaz, hiszen a sejten belül több hasonló méretű és töltésű protein is lehet. Ezek viselkedése a gélelektroforézis körülményei között egyforma.

A célfehérje izolálását a komponensek sokfélesége, nagy száma mellett nehezé teszi az egyes molekulaféleségek eltérő mennyisége, koncentrációja is (6). A nagy mennyiségben jelen lévő protein („abundant protein”) mellett egy nyomokban előforduló, hasonló méretű és töltésű komponens kimutatása csak nagyon érzékeny analitikai eljárással lehetséges. Egy másik „gyakorlati problémát” jelent az, hogy egy kis koncentrációjú fehérje megfelelő mennyiségben történő izolálásához igen nagy mennyiségű testfolyadékra vagy sejtizlátumra lehet szükség. Ez pedig nagyszámú sejtől álló mintát vagy olyan nagy mennyiségű testfolyadék felhasználását jelenti, amennyi egy személyben nincs is.

A svájci GeneProt ([www.geneprot.com](http://www.geneprot.com)) vállalkozás, amelynek tudományos tanácsában három Nobel-díjas működik, kutatási programjában emberi szövetben, illetve testfolyadékban található fehérjék azonosítását és terápiás célú felhasználását tűzte ki célul. A vállalkozás 51 nagy teljesítményű tömegspektrométer és 90 nagy teljesítményű folyadékkromatográf (HPLC) mellett 1420 szuperszámítógépet és 32 informatikust foglalkoztat (2002. évi adatok). 2002. szeptember 30-án Genfben jelentették be, hogy ipari méretű proteomikai vizsgálatot indítanak a cerebrospinalis folyadék proteín-összetételének megállapítására különböző degenera-

3. ÁBRA

Humán plazma részleges proteintérképe 2D gélelektroforézis alapján. A minta a 8–200 kDa méret- és a pI=4,0–6,0 izoelektromospon-tartományba eső fehérjekomponenseket ábrázolja



Sanchez, et al. 1995. (<http://www.expasy.ch/>).

tív, illetve vascularis dementia kórképek esetében (például Alzheimer-betegség). A vizsgálatához a cég rendelkezésére áll két liter liquor cerebrospinalis és három liter plazma gondosan jellemzett, valamilyen demenciában szenvedő betegekből. Azért van szükség ilyen nagy mennyiségű folyadékra, mert elsősorban olyan fehérjék azonosítására törekszenek, amelyek igen kis koncentrációban vannak jelen.

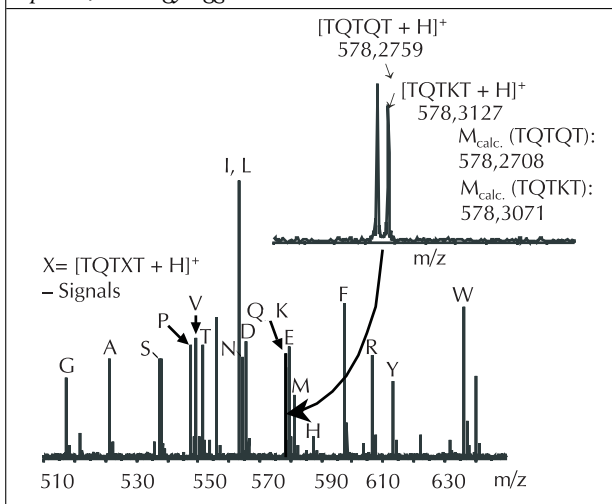
Nemcsak a mennyiségben, illetve koncentrációban mutatózó szélsőségek befolyásolhatják a fehérjekomponensek elválasztását. Más, „extrém” tulajdonságok is nehezítik a kutató munkáját. Például az igen nagy méretű, a túlságosan kicsi vagy a hidrofób, főként membránból származó proteinek izolálása távolról sem tekinthető rutinfeladatnak, ezeket (még) rendszeresen elveszítjük. Gondot okozhatnak az „átlagos” tulajdonságok is. Például a humán szérumból 2D elektroforézissel készült gél 600 foltot tartalmaz. Ebből több mint 200 a pI=5-6 izoelektromospon-tartományban található. Nem csoda, hogy a proteomika szakirodalmának jelentős része fehérjeelválasztás-tudomány, s az új megoldásokat bemutató közlemények bizony gyorsan elérik a „citation classic” fokozatot.

## A fehérjék azonosítása, a primer szerkezet meghatározása

A tiszta, azonos molekulákat tartalmazó proteinpreparátum aminosavsorrendjének (primer szerkezet) meghatározása az első, Edman-lebontás alapján működő szekvenátorok megjelenése óta vált automatizálható, nagy hatékonyságú eljárássá. 1972-ben húsz fehérje primer szerkezetét ismerték, két évtizeddel később 20 000 protein aminosavszekvenciáját őrizték az adatbázisok. Az újabb, ugrásszerű változás a tömegspektrometriás módszer megjelenésével van összefüggésben. Érdemes megjegyezni, hogy a 2002. évi kémiai Nobel-díj „felét” *John B. Fenn* (Virginia Commonwealth University, Richmond, USA) és *Koichi Tanaka* (Shimadzu Corporation, Kiotó, Japán) nyerte el a „biológiai makromolekulák tömegspektrometriás vizsgálatára” kidolgozott ionizációs technika megalkotásáért. A tömegspektrometria jelentősen csökkenti a vizsgálathoz szükséges anyagmennyiséget és a tömegmérés pontossága miatt növeli a kísérleti adatok megbízhatóságát. A 4. ábrán 19 peptid keverékének tömegspektruma látható (7). A világon legkorszerűbbnek tekintett készülékkel felvett spektrumból azt látjuk, hogy a módszer képes különbséget tenni két olyan vegyület között is, amelyek tömege egymástól csupán 0,0368 egységgel tér el. (A hidrogénatom tömege 1,00794 egység.) Ennek alapján érthető, hogy sok esetben elegendő a fehérje néhány fragmensének tömegét megmérnünk ahhoz, hogy – a genom ismerete és megfelelő számítógépes program alapján – nagy biztonsággal meg lehessen mondani, melyik fehérjével van dolgunk. Ezt a korábbi, fehérjelebontáson alapuló szekvenciameghatározás nem tette lehetővé.

### 4. ÁBRA

A TQTX peptidek ( $X=19$  aminosav egyike) komponenseinek kimutatása ESI-FTICR tömegspektrum alapján. Az ábrát dokumentálja, hogy a módszer különbséget tesz két olyan peptid (TQTQT/TQTKT) között is, amelynek tömege egymástól csupán 0,0368 egységgel különbözik



Windberg, et al. 2002., engedéllyel átvéve

Az aminosavsorrend meghatározását nehezíti, ha a fehérje intramolekuláris diszulfidkötéseket tartalmaz, vagy ilyen módon – egyébként egymástól független – polipeptidláncok kapcsolódnak össze. Hasonló problémákat okoznak a nem kovalens kötéseket tartalmazó proteinkomplexek (például enzimkomplexek), valamint a polipeptid oldalláncain – a transláció után – módosult proteinek (például szénhidrát- és/vagy lipid-oldalláncot tartalmazó, bakteriális glükolipoproteinek).

Említettük, hogy esetenként elégséges néhány fehérjefragmens szekvenciáját (Edman-módszer) vagy pontos tömegét ismernünk a fehérje azonosításához, amennyiben a fehérje teljes aminosavsorrendje valamely protein-adatbázisban fellelhető. A fragmens(ek) adatai alapján ehhez a proteinformatika nyújt segítséget azzal, hogy a protein- és genomadatbázisok között teszi lehetővé az „átjárhatóságot”. A fehérjék feldarabolása rendszerint enzimes hasítással történik. A megfelelő méretű és számú fragmenseket eredményező enzim kiválasztása empirikus, kísérleti megfigyeléseken, többszöri próbálkozáson alapul. Újabban, például a prionfehérjék megismerése kapcsán az is világossá vált, hogy léteznek „enzimrezisztens” fehérjék is. Ez esetben nem enzimes, hanem kémiai fragmentálást alkalmaznak, például bróm-ciánnal történő hasítást.

Az előző fejezetben említettük, hogy a humán szérumról készült 2D gélen 600 folt szerepel. Ezek közül az izoelektromos pont alapján 205 esik a  $pI=5-6$  tartományba. Nos, a 205 fehérje közül csupán a felének, 105 proteinek az aminosavsorrendje található meg valamelyik fehérje-adatbázisban. A maradék 100 protein ismeretlen (volt). E fehérjék azonosítását tehát nem segíthette a fragmensek aminosavszekvenciájának, illetve pontos tömegének ismerete és az adatbázis: az összes fragmens aminosavsorrendjét, sőt, még a fragmensek sorrendjét is kísérletesen kellett meghatározni.

A fentiekből világosan kirajzolódik, hogy a proteomika kialakulásának experimentális hátterét – a korábban tárgyalt elválasztástechnikai módszereken kívül – a szekvenálás modern módszerének, a tömegspektrometriának a robbanásszerű fejlődése biztosítja. Az adatok feldolgozása és elemzése pedig elkerülhetetlenül tette a proteinformatika megjelenését, és mint látni fogjuk, sokoldalú alkalmazását a térszerkezet és a funkcionális tulajdonságok valószínűsítésében is.

### Az annotáció

A fehérjék aminosavsorrendjének ismeretében veszi kezdetét az a folyamat, amelyet a proteomika az „annotáció” fogalmával ír le. E kifejezés nem a szokásos „pár soros ismertetés, tartalmi kivonat”, illetve „feljegyzés, szelvényzet” (Bakos F. Idegen szavak és kifejezések szótára. Budapest: Akadémiai Kiadó; 1983.) értelmében használatos. A proteomikában azt a folyamatot és annak eredményeit írja le, amelynek során megismerjük a protein alegység szerkezetét, poszttranszlációs módosulását, térszerkezetét, lokalizációját, funkcióit,

rokonságát más proteinekkel, valamint szerepét különböző kórképekben. Ez egyben azt is jelenti, hogy az annotáció adatokat rendel egy-egy fehérjéhez. Az újabb kísérleti megfigyelések, mivel gazdagítják az egyes proteinre vonatkozó tudásunkat, „újraírják” az addig ismert adatokat. Példaként az 5. ábra mutatja egy fehérje „személyi igazolványát”, amely az annotáció végeredményeként jött létre és található meg az egyik protein-adatbázisban. A CD44 fehérjéről megtudjuk,

### 5. ÁBRA

Egy annotált fehérje (humán CD44 antigén) „személyi igazolványa” az egyik protein-adatbázisban (SWISS-PROT, <http://www.expasy.ch/>)

NiceProt View of SWISS-PROT: P16070

Entry name	CD44 HUMAN
Primary accession number	P16070
Entered in SWISS-PROT in	Release 14, April 1990
Sequence was last modified in	Release 35, November 1997
Annotations were last modified in	Release 41, March 2002
Protein name	CD44 antigen [Precursor]
Synonyms	Phagocytic glycoprotein I PGP-1 HUTCH-I Extracellular matrix receptor-III ECMR-III GP90 lymphocyte homing/adhesion receptor Hermes antigen Hyaluronate receptor Heparan sulfate proteoglycan Epican CDw44
Gene name	CD44 or LHR
From	Homo sapiens (Human) [TaxID: 9606]
Taxonomy	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

#### References

- [1] SEQUENCE FROM NUCLEIC ACID (VARIOUS ISOFORMS). TISSUE=Lymphoblast; MEDLINE=93101687; PubMed=1465456; [NCBI, ExPASy, EBI, Israel, Japan] Sreaton G.R., Bell M.V., Jackson D.G., Cornelis EB., Gerth U., Bell J.I.; "Genomic structure of DNA encoding the lymphocyte homing receptor CD44 reveals at least 12 alternatively spliced exons." Proc. Natl. Acad. Sci. U.S.A. 89:12160-12164 (1992).  
[2]...

#### Comments

FUNCTION: MAIN CELL SURFACE RECEPTOR FOR HYALURONATE. ADHESION TO MUCOSAL HIGH ENDOTHELIAL VENULE AND TO TYPES I AND VI COLLAGEN. PROBABLY INVOLVED IN MATRIX ADHESION, LYMPHOCYTE ACTIVATION AND LYMPH NODE HOMING.

SUBCELLULAR LOCATION: Type I membrane protein.

hogy 1990-ben került az adatbázisba, honnan származik (emberi), milyen más neveket illették, milyen biológiai funkciói vannak, hol található meg a sejtben.

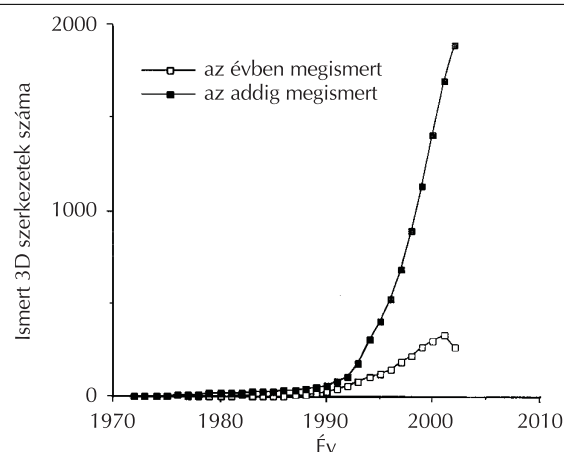
A cikk terjedelme nem teszi lehetővé, hogy az összes, annotációval kapcsolatos lépést bemutassuk. Csúpan néhány probléma megemlítésére szorítkozhatunk a poszttranszlációs módosításokra, a térszerkezet felderítésére és egyes fehérjék patológiás kórképekben betöltött szerepére vonatkozóan.

A sejtben lezajló fehérjeszintézis után a fehérjemolekula sokféle szerkezeti változáson mehet keresztül. A poszttranszlációs módosulás során a protein egy része, például a szignálszekvencia, a transzportért felelős régió vagy a biológiai hatás kifejtését gátló aminosav-részek (Met), propeptidek lehasadhatnak. A molekula szabad aminos csoportjain acetileződhet, de ezekre metilcsoportok is beépülhetnek, például a hisztionfehérjék esetében. A hidroxilcsoportot tartalmazó aminosav-oldalláncon (például Ser, Thr) foszforsav- vagy zsírsavészterek, O-glikozidos kötésben szénhidrát-láncok (oligoszacharidok) kapcsolódhatnak a polipeptidgerinchez. Nem ritka a fehérjén belül vagy két polipeptidlánc között – ciszteinoldallancok igénybevételével – kialakuló diszulfidhíd kötés vagy a glutamin/aszparagin, illetve lizinoldallancok között létrejövő amid típusú kereszt kötés sem. Mai tudásunk szerint több mint 100, kisebb-nagyobb különböző kémiai átalakulás mehet végbe a sejtben, s ezek szoros összefüggésben vannak a fehérje biológiai funkciójával, illetve működésük „ki-be kapcsolásával”.

A fehérjék térszerkezetének megismerésében elért jelentős fejlődést szemléltetik a 6. ábra adatai. A múlt század 50-es éveiben az első lépést a röntgenkristallográfia alkalmazásával John C. Kendrew és Max F. Perutz tették meg, 1959-ben közölték a mioglobint, illetve a hemoglobint térszerkezetére vonatkozó adataikat (kémiai Nobel-díjat kaptak 1962-ben). Később megjelent a magmágneses rezonancia mérésén (NMR) alapuló spektroszkópia, amely az utóbbi évektől lehetővé tette, hogy egyre több és egyre nagyobb méretű

### 6. ÁBRA

Az ismert térszerkezetű fehérjék számának növekedése ([www.rcsb.org/pdb](http://www.rcsb.org/pdb) alapján)



fehérje oldatbeli konformációjára következtethessünk. Az NMR-módszer a nem fehérje kismolekulák szerkezetvizsgálatában rutineljárás, kifejlesztéséért 1991-ben *Richard Ernst* kémiai Nobel-díjat kapott. 2002-ben – a két tömegspektrometriával foglalkozó kutató mellett – *Kurt Wütrich* (ETH, Zürich, Svájc) lett Nobel-díjas. Érdeme az NMR-spektroszkópia alkalmazása fehérjék (például prion) térszerkezetének meghatározásában. Meg kell jegyeznünk azonban, hogy a Protein Adatbankban tárolt 17 222 fehérje-térszerkezet (2002. november 5-i adat) 86%-át krisztallográfiai mérések alapján határozták meg (8).

Az ismert térszerkezetű proteinek száma jóval kisebb, mint az ismert aminosavsorrendű fehérjék száma. Érdemes megjegyezni, hogy az „International Protein Index” (IPI), amely öt adatbázisra épül, 103 590 humán fehérje aminosavszekvenciáját tárolja (2002. november 1.), ugyanakkor e humán proteinek csupán 5%-ának ismert a 3D szerkezete.

A primer, illetve 3D szerkezetű fehérjék száma közötti jelentős eltérés egyik oka a módszerekben, a „technikai” részletekben rejlik. Például a röntgenkrisztallográfiai mérésekhez jó minőségű fehérjekristályokra van szükség, hiszen ez határozza meg a kinyerhető szerkezeti információ mértékét. A fehérjék jelentős része azonban nem vagy csak nagyon nehezen kristályosítható. Elterjedt nézet, hogy a kristályosítás külön szakma, sőt, művészet, amelyre csak „mágikus kezek” képesek. A rendelkezésre álló protein-térszerkezeteket tároló adatbankokban (például a Protein Data Bank) többek között ezért olyan csekély – néhányszor tíz – a membránfehérje-adatok száma. NMR-spektroszkópiával a készülék teljesítményétől függően kisebb és nagyobb méretű fehérjék térszerkezetét is vizsgálni lehet. A nagy teljesítményű készülékek meglehetősen drágák, üzemeltetésük költséges, ezért viszonylag kevés kutatóhelyen állnak rendelkezésre. (Magyarországon is elkélne legalább egy ilyen berendezés.) Az NMR-rel történő 3D szerkezetvizsgálati módszer kétféle mesterségbeli jártasságot igényel: a mérés megtervezését, kivitelezését, valamint a mérési adatokra alapozó számítógépes modellezést. Érdemes megemlíteni azt is, hogy mindkét módszer több milligrammnyi mintát igényel, ami soknak számít, ha nehezen hozzáférhető.

Az ismert primer, illetve 3D szerkezetű fehérjék száma közötti jelentős eltérés motiválta a „térszerkezetjósáslással” foglalkozó informatikai módszerek kialakulását. A predikciós módszereknek két csoportja és ezeken belül számos variánsa jött létre. Az egyik csoportba tartozó eljárások a fehérje primer szerkezetéből indulnak ki és jutnak el egy lehetséges térszerkezeti modellhez. A másik iskola meggondolása szerint a kiindulási pont homológ fehérjék kísérletesen meghatározott térszerkezete. A kétféle megközelítésben közös, hogy már ismert térszerkezetű fehérjékre vonatkozó adatok elemzésére épít és az elemzés alapján felismert összefüggésekből próbál következtetéseket levonni egy ismeretlen térszerkezetű proteinre nézve. Mindkét eljárás lehetőséget ad egy vagy több lehetsé-

ges térszerkezet jóslására és megfelelő modellezőprogramok közbeiktatása után számítógépes megjelenítésükre. Fontos azonban megjegyezni, hogy lényeges különbség van a kísérleti adatokon (krisztallográfia, NMR), illetve predikció révén előállított térszerkezetek valóságtartalma, hitelessége között.

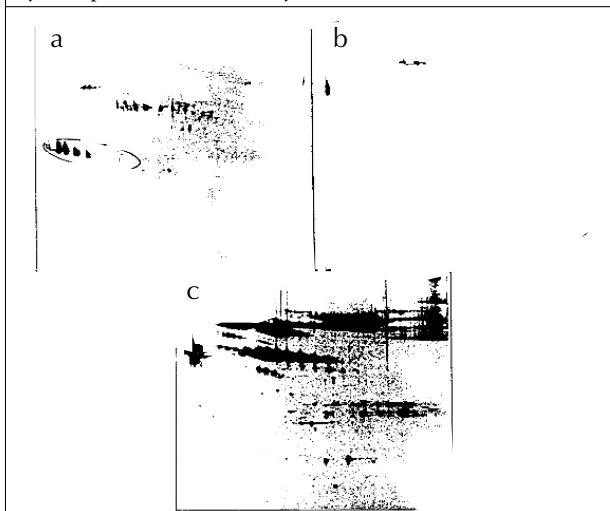
A *proteininformatika* a bioinformatika egyik igen gyorsan fejlődő és Magyarországon fejlesztendő ága. Tárgykörébe tartozik a fehérjékre vonatkozó adatok gyűjtése, rendszerezése, a fehérjék elsődleges és 3D szerkezetét, funkcióját leíró adatbázisok előállítása, fenntartása és folyamatos korszerűsítése (9). A proteininformatika tehát adatbázisokkal és számítógépes programcsomagokkal segíti az annotáció egyes lépéseit, így a molekula térszerkezetének modellezését, a fehérje funkciójának predikcióját, a biológiai hatásért felelős régiók elhelyezkedésének valószínűsítését. Kezeli a proteincsipek adatait, sőt, gyűjti a protein-protein kölcsönhatásokra vonatkozó információkat is. A proteininformatika háttérében álló erőfeszítéseket a következő példa jól szemlélteti. A korábban említett Svájci Bioinformatikai Intézet (Swiss Institute of Bioinformatics) hozta létre a SWISS-PROT adatbázist, amely egyike az IPI által feldolgozott adatbázisoknak. A SWISS-PROT 117 406 fehérje aminosavszekvenciáját tartalmazza (2002. november 1.). E proteinek 7535 fajból származnak, de a legtöbb fehérje humán, illetve egér eredetű. 2001 végén 7600, 2002 novemberében 8665 humán fehérjére vonatkozó adatot találunk az adatbázisban. A kísérletileg igazolt vagy predikció alapján jósolt poszttranszlációs módosítások száma 22 044, míg a polimorfizmusoké – többségük betegséghez is kapcsolódik – 14 132. A bemutatott – és kiragadottnak tekinthető – összegző adatokhoz az intézet dolgozói 27 730 tudományos közleményt dolgoztak fel.

Az annotáció folyamatának része a fehérje lokalizációjának meghatározása, amely többek között „riportermolekulával” (radioaktív izotóp, fluorofor, spinjelző) jelzett fehérjék segítségével valósítható meg sejtfrakcionálás, autoradiográfia vagy mikroszkópos technikák kombinálásával. Ezen információk alapján lehetőség nyílik arra, hogy az egyes *fehérjék és bizonyos körképek közötti összefüggés* szisztematikusan vizsgálható legyen az orvosi gyakorlat, a diagnosztika és majdan a terápia számára. Az elgondolás nem előzmény nélküli, hiszen már jóval korábban kimutatták, hogy lehet oki kapcsolat: a sarlósejtes vérszegénység például összefüggésben van a hemoglobin béta-láncában bekövetkezett glutamin-valin aminosavcserével, ami jelentősen befolyásolja a fehérje oxigénmegkötő képességét.

*Sun* és munkatársai 1998-ban publikálták azt a közleményt, amelyben leírják a belső fülben kialakult és klinikailag nehezen diagnosztizálható fekély kimutatására vonatkozó megfigyeléseiket (10). A szerzők a perilympa-folyadék fehérje-összetételét vizsgálták. A kétdimenziós gélelektroforézis és fehérje-előhívás után nyert felvételt mutatja a 7. a. ábra. Összehasonlításként láthatunk egy-egy olyan elektroforetogramot amely a cerebroszpinális folyadékból (7. b. ábra), illetve a szérumból (7. c. ábra) készült. A három kép összevetése

## 7. ÁBRA

Fisztula típusú fekély kimutatása a belső fül labirintusa körüli folyadék 2D gélelektroforézisével nyert fehérjetérkép alapján. Az apolipoprotein D fehérjecsoport (AP30) megjelenik a betegből származó perilymphában (a), de nem mutatható ki sem a cerebrospinalis folyadékban (b), sem a szérumban (c). A minták a  $pI=4,5-8,0$  izoelektromospont-tartományba eső fehérjekomponenseket ábrázolják



Sum, et al. 1998., engedéllyel átvéve

világosan mutatja, hogy a betegből származó perilympa tartalmaz egy molekulacsoportot (bekarikázott rész), amely nincs jelen a másik két testfolyadékban (és nincs jelen az egészséges fülben sem). A részletes vizsgálatok céljára a foltokat kimetszették a gélből, enzimatikusan feldarabolták a bennük levő fehérjéket, és tömegspektrométer segítségével azonosították a fehérjecsaládot. Megállapították, hogy e vegyületek az apolipoproteinek családjába tartoznak, és egymástól csak abban különböznek, hogy a polipeptidgerinchez milyen hosszúságú és összetételű szénhidrátlánc kapcsolódik. E példa szemlélteti azt is, hogy egy klinikai probléma megoldásához, a megfelelő terápia kiválasztásához miként járulhat hozzá a proteomika.

A tudományos ismeretek bővítése mellett jelentős

gyógyszeripari érdeklődés kíséri a proteomika ilyen irányú programjait is. A betegekből és egészségesekből származó cerebrospinalis folyadék protein- és peptidösszetételének összevetésére irányuló, korábban említett kutatás például lehetőséget adhat azoknak a fehérjéknek az azonosítására, amelyek a dementia különböző formáival összefüggésbe hozhatók. Az azonosítás után sor kerülhet a kiválasztott fehérjék/peptidek szintetikus, nagy mennyiségben történő előállítására és pontos szerepük tisztázására. E vegyületek azután diagnosztikumként, illetve gyógyszerként az ipar számára is vonzóak lehetnek. Nem véletlen, hogy a kutatást kezdeményező GeneProt cég partnerei közé tartoznak olyan cégek mint a Novartis, Hewlett Packard, Bruker Daltonics, Waters Corporation és a Swiss Institute of Bioinformatics.

## Kölcsönhatások

A proteomika, úgy tűnik, az új évszázad története lesz. Remélem, e rövid áttekintés meggyőzte arról a tisztelt olvasót, hogy reális lehetőséggé vált az élő szervezetet alkotó összes fehérje szerkezetének belátható időn belüli megismerése. Nehezebb és összetettebb feladatnak tűnik a proteinfunkciók minden részletre kiterő felderítése. A fehérjék biológiai szerepének feltárása sokban múlik azon, hogy a protein-protein, a protein-nukleinsav és a protein-lipid kölcsönhatásokról milyen ismereteink lesznek. Ahogyan a genom megismerése, az elválasztástechnika és a tömegspektrometria, valamint a proteoinformatika új módszereinek létrejötte előfeltétele volt a proteomika kialakulásának, úgy a proteomot alkotó fehérjék megismerése és a kölcsönhatások megfigyeléséhez, jellemzéséhez szükséges módszerek, megközelítések jelentős fejlődése kell a teljes proteinvilág működésének megértéséhez. A másik kihívást a kutatás számára az a közismert tény jelenti, hogy a sejtek, a szervezet és működésük időben folytonosan változik. Ennek feltétele és egyben következménye is a fehérjetérkép időbeli és térbeli mássága.

A problémák ellenére már most jól látszik, hogy a proteomika jelentősen segíti a korszerű orvoslást a ma-inál hatékonyabb diagnosztikai eljárások és terápiás szerek kifejlesztésében.

## IRODALOM

1. Falus A, Szalai Cs, Tóth S. Funkcionális genomika. Lehetőségek, remények, posztgenomiális realizmus. *Lege Artis Medicinae* 2002;12:365-70.
2. O'Donovan C, Apweiler R, Bairoch A. The human proteomics initiative (HPI). *Trends Biotechnol* 2001;19:178-81.
3. O'Farrell PH. The suppression of defective translation by ppGpp and its role in the stringent response. *Cell* 1978;14:545-57.
4. Raymond MA, Sanchez JC, Hughes GJ, Günther K, Riese J, Tortola S, et al. Standardized characterization of gene expression in human colorectal epithelium by two-dimensional electrophoresis. *Electrophoresis* 1997;18:2842-8. (<http://www.expasy.ch/>)
5. Sanchez JC, Appel RD, Golaz O, Pasquali C, Ravier F, Bairoch A, et al. Inside SWISS-2DPAGE database. *Electrophoresis* 1995;16:1131-51. (<http://www.expasy.ch/>)
6. Righetti PG, Stoyanov A, Zhukov MY. The Proteome Revisited. Elsevier, 2001.
7. Windberg E, Hudecz F, Marquardt A, Sebestyén F, Kiss A, Bösze Sz, et al. Analysis of combinatorial libraries of mucin-2 antigen peptides by high resolution mass spectrometry. *Rapid Communications in Mass Spectrometry* 2002;16:34-9.
8. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. *Nucleic Acids Research* 2000;28:235-42.
9. Baxevanis AD, Ouellette BF. *Bioinformatics. A practical guide to the analysis of genes and proteins.* New York: Wiley; 1998. p. 370.
10. Sun Q, Disher MJ, Rustad T, Telian SA, Andrews PC. AP30, a differential protein marker for perilymph and cerebrospinal fluid in middle ear fluid, has been purified and identified as human apolipoprotein D. *Biochem Biophys Acta* 1998;1384(2):405-13.