

## 1.5.4. Fehérje hálózatok vizsgálata

### 1.5.4.1. Bevezetés

Az elmúlt másfél évtized kutatásai rámutattak, hogy a fehérjék nem önállóan, kisebb komplexekben vagy útvonalakban, hanem nagyobb, összetettebb rendszerben, hálózatokba rendeződve működnek a sejtekben. Ahhoz, hogy ehhez a megfigyeléshez eljussunk, több tízezer publikáció és több százezer kutató aprólékos és precíz munkájára volt szükség. Valamint arra, hogy a számítógépes kapacitás és az ezzel együtt fejlődő bioinformatikai adatbázisok, alkalmazások lehetővé tegyék a létrejött adatmennyiség megfelelő feldolgozását, ábrázolását és elemzését. A hálózatkutatás és a rendszerbiológia (*systems biology*) jelentősége nem abban áll, hogy egy helyen sok információt tudunk elérni és elemezni. Ezeknek a modern megközelítéseknek az ereje az adott rendszer ún. emergens tulajdonságainak vizsgálatában áll. Emergens tulajdonságnak nevezzük azokat a jellemzőket, amely az adott rendszer komponenseire külön-külön nem jellemzőek vagy nem értelmezhetőek, de a rendszer egészére igen. Ilyen emergens tulajdonság például a robusztusság, a redundancia és a modularitás. A robusztusság a rendszer ellenállóképességét mutatja valamilyen zavar esetén. A redundancia párhuzamos megoldásokat mutat, amelyek biztosítják egy komplex rendszer stabil működését erősebb behatások ellen. A modularitás pedig arra a jelenségre utal, hogy a stabil és sikeres rendszerek csoportokba rendeződnek, amelyek probléma esetén leválaszthatóak. Érdekessége a rendszer-szintű kutatásoknak, hogy gyakran párhuzamot lehet vonni teljesen eltérő rendszerek felépítése és emergens tulajdonságai között, függetlenül attól, hogy a sejtek természetes fehérjehálózatáról van szó, vagy mesterségesen létrehozott hálózatokról, mint amilyen az internet vagy egy áramellátási hálózat (Barabási 2002, Csermely 2006).

A biológia tudományterületén belül a fehérjék hálózatának kutatása az egyik leginkább elterjedt, legtöbb adattal és bioinformatikai alkalmazással rendelkező irányvonal. Ugyanakkor fontos megjegyezni, hogy egyre több adat áll rendelkezésre a molekuláris szabályozási hálózatokról (fehérje-DNS, mRNS-miRNS és miRNS-lncRNS kapcsolatok), valamint a fehérjehálózatok kutatását évtizedekkel megelőzte akár a metabolikus, akár az ökológiai hálózatok elemzése (pl.: táplálékhálózatok).

A fehérje hálózatok elemzése számtalan előnnyel járhat a biológiában. Különböző fajok fehérje hálózatainak összehasonlító vizsgálata az elméleti és evolúciós biológiai kutatások során precízebb törzsfajlódási modelleket és új fehérje funkciókat eredményeztek. Ennek fő oka, hogy a fehérjék közötti kapcsolatok sokkal gyorsabban változnak az evolúció során, mint maguk a fehérjéket kódoló gének, vagy azok szabályozási régiói. Egy-egy aminosav csere nem szükségszerűen változtatja meg egy fehérje térszerkezetét vagy funkcióját, de egy kapcsolat létrejöttét alaposan befolyásolhatja. Következésképp, annak érdekében, hogy megértsük több öröklődő betegség és a rák pathomechanizmusát, és közelebb kerüljünk a gyógyításukhoz, fontos feltérképezni a mutációval rendelkező fehérjék kapcsolódó partnereit. Gyakran a kísérletek kiértékeléséhez nélkülözhetetlen, hogy ne csak a vizsgált fehérje tulajdonságait, hanem ezen partnerek tulajdonságait is megvizsgáljuk. Végül fontos megemlíteni még, hogy bizonyos fehérje kapcsolatok gyógyszeres célzása hatékonyabb terápiás megközelítés lehet, mint magának a teljes fehérjének a gyógyászati befolyásolása.

### 1.5.4.2. Fehérje-fehérje kapcsolatok

Kezdjük a fehérje hálózatok világának megismerését a hálózatot alkotó fehérje-fehérje kapcsolatok definíciójával és leggyakoribb kísérletes kimutatásával. Fehérje-fehérje kapcsolat (FFK, angolul PPI) alatt a fizikai kapcsolódást értjük, amikor két fehérje

doméneken vagy motívumokon keresztül kapcsolódik egymással. Ez nem feltétlenül jár poszt-transzlációs módosulással. A FFK-k lehetnek irányítottak, azaz meghatározható a jel terjedés iránya, vagy irányítatlanok. Affinitásbeli különbségek is lehetnek a kapcsolatok között, amelyek meghatározhatják a kapcsolat erejét és állandóságát. Az FFK-k lehetnek irreverzibilisek (például proteáz enzimek által végzett vágások a szubsztrátfehérjén), vagy reverzibilisek (például egy kináz enzim foszforilációja a szubsztrátfehérjén). Nem minden FFK fordul elő mindenhol, a fehérjék expressziószabályozásának és sejten belüli precíz transzportjának köszönhetően beszélhetünk kompartment- és szövetfüggő kapcsolatokról, valamint állandóan jelenlévőkről (*house-keeping*), illetve adott külső vagy belső behatás következtében létrejövő kapcsolatokról. Utóbbiak alkotják a sejtek jelátviteli hálózatát. Azaz a jelátviteli FFK-k általában irányítottak, és egy külső jel hatását közvetítik általában a sejtmag felé. Megjegyezzük, hogy léteznek egyéb fehérje kapcsolatok, ahol a partner nem fehérje, hanem más természetű molekula (pl.: DNS, RNS, lipid vagy szénhidrát). Ezen kapcsolatok közül a fehérje-DNS kapcsolatok rendszere már egyre jobban ismert az epigenetika és a transzkripcionális kapcsolatok kutatása révén. A közeljövőben a lipidomika térnyerésének köszönhetően valószínűleg egyre többet fogunk megtudni a lipidek segítségével létrejövő FFK-ról. A továbbiakban kizárólag FFK-ról lesz szó, de több kísérletes és számítógépes módszer alkalmas egyéb kapcsolattípusok kimutatására és vizsgálatára is.

Az FFK hálózatos kontextusban történő vizsgálatát az elmúlt évtized nagy áteresztőképességű (*high-throughput*) kísérletes módszereinek fejlesztései tették lehetővé. Az alábbiakban röviden bemutatjuk a leginkább elterjedt technikákat (1. ábra):

Az **élesztő-kettős-hibrid** páronkénti (bináris) kapcsolatok élesztőn belüli kimutatását teszik

lehetővé. A vizsgálat során az élesztő kísérletes rendszerként szolgál, amelyben bármilyen faj fehérjéinek a kapcsolódását lehet páronként vizsgálni. Az eljárás lényege, hogy két plazmiddal transzfektáljuk az élesztősejteket. Az egyik plazmid a minket érdeklő, ún. csali (*bait*) fehérjét kódolja egy élesztő transzkripció faktor (pl.: Gal4) DNS-kötő doménjével együtt. A másik plazmid egy másik, a zsákmány (*prey*) fehérjét kódolja, az élesztő transzkripció faktor aktivációs doménjével. Egyik fuzionált fehérje (ha jól van tervezve) sem képes önmagában transzkripciót beindítani. Ha egyszerre jelen van a két plazmid, akkor sem biztos, hogy bármilyen transzkripció létre tud jönni. Csak abban az esetben lesz funkcionális az élesztő transzkripció faktor, ha a csali és zsákmány fehérjénk kapcsolódik egymással, ezáltal a transzkripció faktor két doménje is olyan közel lesz egymáshoz, hogy funkcionális egységet tudnak alkotni. Az így bekövetkező transzkripció eseményt a célgén expressziójával, leggyakrabban a célgénbe épített színes riportter rendszer segítségével tudják kimutatni. Azaz amennyiben van kapcsolat, a riportergén a génbe épített fluoreszcens jellel együtt expresszálódik, és ez kimutatható. A módszer előnye, hogy egy csali fehérje kapcsolódását akár az adott faj összes ismert fehérjéjével együtt, egyesével le lehet tesztelni. A módszer több hátránnyal is rendelkezik:

1. Élesztőben zajlik a kapcsolódás, így olyan kapcsolatok, amelyeknek más, például emlős enzimek előzetes aktiválására is szükségük van, nem fognak létrejönni.
2. Átmeneti, rövid ideig tartó vagy gyenge affinitású kapcsolatokat nehéz így kimutatni.
3. A klasszikus élesztő-kettős-hibrid rendszerben a kapcsolódás az élesztő sejtmagjában történik, így mindkét fehérjének be kell oda jutnia.

4. Olyan fehérjék, amelyek maguk is rendelkeznek transzkripcionális aktivitással vagy sok savas oldalláncot tartalmaznak, nem vizsgálhatóak ezzel a módszerrel mivel megzavarják a riporter rendszert.

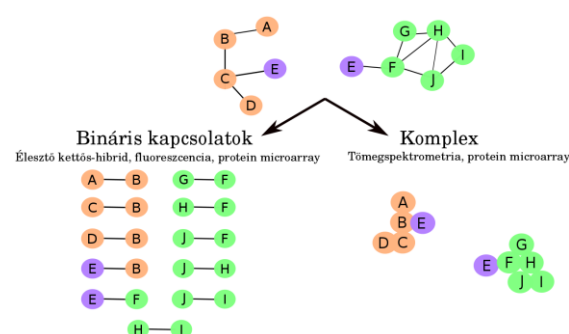
Az elmúlt években kifejlesztésre került memberánkött és ún. split-ubiquitin alapú élesztő-kettős-hibrid eljárás is, amelyek segítségével membrán fehérjék és akár transzkripciós faktorok kapcsolódásai is kimutathatók.

Az **affinitástisztítással összekötött tömegspektrometria** bármilyen *in vivo* rendszerből képes a stabil kapcsolatok kimutatására. A módszer első lépése egy kiválasztott fehérje tisztítása, kinyerése egy sejtből, ahol normál mennyiségben expresszálódik. A tisztítási eljárás során azok a fehérjék, amelyek a kiválasztott fehérjével közvetlenül, vagy egy nagyobb komplexek keresztül kapcsolódnak szintén kinyerésre kerülnek. Ezt követően a fehérje komplexek vagy Western-blot technikával vagy leginkább különböző tömegspektrometriai módszerek segítségével kerülnek azonosításra (ld. 1.2.8 fejezet). A módszer előnye, hogy a fehérje kapcsolatok vizsgálata az eredeti sejtípusból, vagy ahhoz közeli rendszerből történhet (szemben az élesztő-kettős-hibrid módszerrel, ahol élesztőben zajlik a vizsgálat). A módszer hátránya, hogy kapcsolódó fehérjék listáját kapjuk eredményül, a kapcsolódási pontokat/sorrendet nem. Utóbbi azért okozhat gondot, mert a kapott fehérjelistán nem minden fehérje kapcsolódik egymással.

Vannak továbbá **fluoreszcens módszerek (FRET és BiFC)**, amelyekkel két előre megjelölt fehérje sejtben belüli kapcsolódását tudjuk mikroszkóppal megfigyelni. Előnyük, hogy teljesen *in vivo* a kimutatás, ráadásul sejtben belüli lokalizációs információt is kapunk a kapcsolat helyéről. Hátrányuk, hogy nagyobb vizsgálatokat (*screeneket*) nem tesznek lehetővé, és sok kontroll konstruktra van szükség a helyes kiértékeléshez. Leginkább

akkor szokták alkalmazni ezeket a fluoreszcens módszereket, ha néhány konkrét kapcsolatot szeretnének megerősíteni, amelyeket élesztő-kettős-hibriddel vagy tömegspektrometriával találtak.

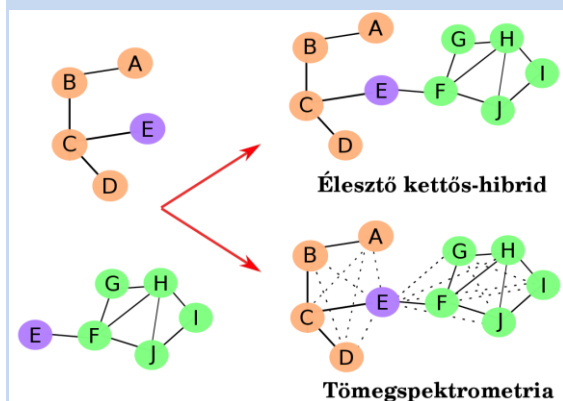
Nagy lehetőségeket rejt magában a **fehérje microarray (fehérje chip)** megközelítés, amely nagy mennyiségű fehérje egyidejű kapcsolódását tudja kimutatni. Az a módszer lényege, hogy egy felülethez, (ami lehet üveg, nitrocellulóz membrán vagy valamilyen specifikus gyöngy) hozzákötünk különböző fehérjéket. A tesztelendő fehérjetartalom általában fluoreszcensen jelölve van. Amikor egy tesztelendő fehérje kapcsolódik a chip-hez rögzített valamelyik fehérjével, a fluoreszcens jel az adott fix helyen látszik, és így lehet kimutatni. Mivel tudjuk, hogy melyik pozícióban milyen fehérjét helyeztünk el, lehet tudni, hogy mely fehérjék között történt kapcsolódás. Ezzel a módszerrel egyszerre sok fehérje kapcsolódását lehet kimutatni olcsón és gyorsan. Hátránya, hogy a felületen úgy kell elhelyezni a fehérjéket, hogy megfelelő térszerkezettel rendelkezzenek, ráadásul hosszú ideig, amíg a vizsgálat le nem zajlik. A megközelítés mind a kiértékelésében, mind a specifikus és szelektív fehérje kapcsolatok kialakításában több kihívást tartogat.



1. ábra: Fehérje - fehérje kölcsönhatásokat feltérképező módszerek

**Praktikus megfontolás:** Fontos látni, hogy a különböző módszerek kiegészítik egymást, valamint érdemes mindig tisztában lennünk azzal, hogy egy adott módszer mire érzékeny, és mit nem vesz figyelembe. Erre a legjobb példa, ha összevetjük az élesztő-kettős-hibrid

és az affinitásiztitást követő tömegspektrometria által kapott hálózatokat (2. ábra)



2. ábra: Az élesztő-kettős hibrid és tömegspektrometria által kapott FFK-k. A sima vonalak a valódi kapcsolatokat, a szaggatott vonalak az adott módszerrel kimutatott kapcsolatokat jelzik. Látható, hogy az élesztő-kettős-hibrid módszer nem mutat ki hamis kapcsolatot, de összeköt két komplexet egy közös elem miatt. A tömegspektrometrián alapuló módszer szintén egy komplexnek veszi a fehérjéket, de ráadásul több hamis pozitív kapcsolatot is jelez. A fehérjék részletes vizsgálata, és jól megválasztott kontrollok segítségével lehet ezeket az eredményeket korrigálni, pontosítani.

Három fő forrása van a FFK-nak: 1) adatbázisok, amelyek a korábban bemutatott nagy áteresztőképességű módszerekkel kimutatott kapcsolatokat tartalmazzák, és/vagy kézi vagy gépi módszerrel gyűjtik a cikkekben publikált kyszámú, de részletes kapcsolatokat; 2) bioinformatikai predikciók, amelyek homológia és szerkezeti tulajdonságok alapján jósolnak lehetséges kapcsolatokat.

#### 1.5.4.3. Fehérje-fehérje kapcsolati adatbázisok

Az elmúlt évtizedben több tucat adatbázist készítettek a fehérje-fehérje kapcsolatok tárolására, prediktálására, hogy interneten keresztül elérhetőek legyenek az adatok. Nincs általános szabály, hogy mikor melyiket érdemes használni. A legfontosabb, hogy ismerjük az adatok eredetét, a gyűjtési módszert, és hogy minél kevesebb adatforrásból minél több információhoz

tudjunk jutni. Ugyanakkor a kis méretű adatbázisokat nem szabad lebecsülni, ezek gyakran a legprecízebb, legmegbízhatóbb adatokat tartalmazzák. Az alábbiakban a legerjedtebb adatbázisokat mutatjuk be röviden.

A **BioGRID** (<http://thebiogrid.org>) adatbázisban megtalálható a legtöbb modell szervezetben nagy áteresztőképességű vizsgálattal azonosított FFK. Az adatbázis honlapján konkrét fehérjére is kereshetünk, de az egész adatbázist vagy egy konkrét faj összes FFK-ját (interaktom) is letölthetjük ingyen. A BioGRID géninterakciókat is tartalmaz, és különösen az élesztő-kutatásban tekinthető az egyik legfontosabb FFK adatforrásnak .

Az **IntAct** (<http://www.ebi.ac.uk/intact>) egy olyan FFK adatbázis, amely nagy áteresztőképességű vizsgálattal azonosított FFK-kat is tartalmaz, de főleg olyan publikációkból, amelyek nem egy teljes interaktomot tartalmaznak, hanem néhány fehérje összes azonosított kapcsolatát. Ezen kívül az IntAct mögött egy szakértő kurátor csapat van, akik kézi gyűjtéssel, egyesével is gyűjtik a szakirodalomból a kapcsolati adatokat. Szemben a BioGRIDdel, az IntAct egy-egy FFK-ról nagyon sok további adatot tartalmaz. A pontos publikációs forrás feltüntetése mellett megtalálható itt a kimutatási módszer és minden egyéb körülmény, amelyet a kapcsolatot leíró cikkből ki tudtak nyerni a kurátorok. Az IntAct több egyéb FFK adatbázis adatát is tartalmazza (pl.: MINT, InnateDB, I2D) ugyanilyen részletességgel, így nem szükséges ezeket a forrásokat külön is felkeresnünk. Az adatbázis egyben is letölthető és a weblapon keresztül különböző lekéréseket is tudunk tenni, akár fehérje akár kimutatási módszer szerint, és ezek a találatok is letölthetőek.

A **STRING** (<http://string-db.org>) az egyik leginkább elterjedt FFK adatbázis. Több forrásból tartalmaz FFK-kat: nagy áteresztőképességű vizsgálatokból származó kapcsolatokat, valamint más adatbázisok

adatait is tartalmazza, sőt prediktált kapcsolatokat is. Ez egyben előnye és hátránya is ennek az integrált adatbázisnak. Előny, mert a STRING-nél lehet a legtöbb kapcsolati adatot megtudni egy fehérjéről, de hátrány is, mert a felhasználónak nagyon oda kell figyelnie, hogy milyen eredetű és megbízhatóságú kapcsolatokat használ fel. A STRING tartalmaz olyan FFK-kat is, amelyeket egy algoritmus prediktál az alapján, hogy két fehérje neve együtt kerül említésre egy publikáció összefoglalójában. A STRING készítői megbízhatósági értékeket is rendelnek a kapcsolatokhoz. Ez általában minden adatforrás esetén hasznos. A STRING esetében azonban nem mindig működik megfelelően. Ennek két oka van: az egyik, hogy a megbízhatósági értékeket számoló program felülértékeli a STRING készítőitől származó algoritmusok eredményeit más források (akár kísérletes vizsgálatok) eredményeihez képest. A másik, hogy a megbízhatóságot egy kontroll adatbázis tartalmához képest veszik, ami a STRING esetében a KEGG nevű adatforrás. A KEGG több ezer kutató által idézett, több mint 15 éve működő gén-tulajdonság (annotáció) gyűjtemény. Azonban nem konkrét protokoll alapján készült, gyűjtési torzításokat tartalmazhat, amelyek így a STRING megbízhatósági értékeire is hatással lehetnek. A STRING adatbázist ingyen le lehet tölteni, de ha a kapcsolatok részleteire is kíváncsiak vagyunk, akkor regisztrálni kell, és a regisztrációt követően a publikusnál eggyel régebbi verzióját kapjuk meg a teljes adatbázisnak. Az aktuális teljes verziót csak fizetés ellenében lehet megkapni. Mindezeket nem pusztán a STRING használata kapcsán éreztük fontosnak megjegyezni, hanem példának is szánjuk, hogy mindig oda kell figyelni, bármilyen adatforrást is használunk, hogy honnan származnak az adatok, hogyan lettek feldolgozva, és amit letöltünk az valóban az, amit korábban kiválasztottunk. Ha biztosra akarunk menni, érdemes egy kutatást párhuzamosan két adatbázis adatain elvégezni és értelmezni az esetleges eltéréseket.

**Praktikus megfontolás:** Az adatbázisokban nagyon gyakori, hogy az FFK-k két olyan fehérje között kerültek kísérletesen leírásra, amelyek különböző fajokból származnak. Például egérbe transzfektálnak egy emberi fehérjét és ott vizsgálják, hogy milyen fehérjékkel kapcsolódik. Épp ezért miután letöltjük egy adott faj interaktomját, érdemes kiszűrni a más fajba tartozó fehérjéket a kapcsolati listáról.

#### 1.5.4.4. Fehérje-fehérje kapcsolatok prediktálása

A kísérletes kimutatás gyakran határokba ütközik. Ilyenkor segít a bioinformatikai módszerekkel történő kapcsolat predikció. Ennek két fő fajtája lehet: a szekvencia és a szerkezet alapú predikció. A szekvencia alapú predikció arra épül, hogy ha egy fajban kísérletesen kimutattak egy FFK-t két fehérje között, és ennek a két fehérjének van rokon fehérjéje (ún. ortológja) egy másik fajban, akkor feltételezhetjük, hogy a kapcsolat is jelen van ebben a másik fajban. Ezt a fajta prediktált FFK-t interológoknak, ortológia alapján prediktált interakciónak hívják. Fontos figyelembe venni, hogy bár a szekvenciák, amelyek alapján két fehérjét ortológoknak minősítünk konzerváltabbak, mint a fehérjék közötti kapcsolatok. Egy FFK gyakran egy-egy aminosav különbség miatt jöhet létre vagy tűnhet el. Így ezeket a predikciókat érdemes mindig fenntartással kezelni, viszont hasznosak lehetnek különösen akkor, ha több fajban is kimutatták már az adott fehérjék közötti kapcsolatot.

A szerkezet alapú predikciókat a fehérjék egymáshoz kapcsolódásának szerkezeti információit felhasználva tehetünk. Az egyik ilyen módszer a domén-domén kapcsolatok felhasználása. Kísérletes módszerekkel sikerült kimutatni, hogy milyen fehérje domének képesek egymással kapcsolódni. Ezeket a folyamatosan bővülő információkat tartalmazza például a DOMINE adatbázis (<http://domine.utdallas.edu>). Van olyan adatbázis is, amely a negatív kapcsolatokat,

azaz azokat a domén-domén kapcsolatokat tartalmazza, amelyek közismerten nem tudnak egymással kapcsolódni. Ilyen adatbázis a Negatome (<http://mips.helmholtz-muenchen.de/proj/ppi/negatome>). Ezeket a domén-domén kapcsolati információkat lehet kombinálni a fehérjék domén összetételével, amelyet az UniProt-ról, vagy a PFAM adatforrásról (<http://pfam.xfam.org>) lehet megszerezni. Ezáltal lehetőség van prediktálni két fehérje kapcsolódását, ha tartalmaznak olyan doméneket, amelyek közismerten képesek egymással kapcsolódni.

A másik szerkezet alapú FFK prediktáló módszer az enzimek célszekvenciáira, illetve regulátor szekvenciáinak specifikusságára épül. Olyan adatbázisok, mint az ELM (<http://elm.eu.org>), azt az információt tartalmazzák, hogy egy enzim milyen aminosav szekvencia (motívum) mentén képes poszt-transzlációs módosítást (pl.: foszforiláció) végezni. Ezt az információt, hasonlóan az előzőekhez, fel lehet használni FFK prediktálásra: ha ismerjük, hogy egy fehérje milyen enzimátikus domént tartalmaz, akkor az ELM információi segítségével akár proteom-szinten is meg tudjuk keresni a lehetséges szubsztrátok listáját. Természetesen egy aminosav szekvencia nem biztos, hogy funkcionális motívumként működik minden fehérje esetében. Épp ezért fontosak az ELM-ben is megtalálható fehérje-térszerkezeti szűrő programok (pl.: StructureFilter), amelyek a fehérjék felszínén, domének között vagy rendezetlen régiókban található motívumokat emelik ki.

**Praktikus megfontolás:** Gyakran két félmegoldás kiadhat egy egészet. Így van ez a FFK prediktálás kapcsán is: ha egy szekvencia alapon prediktált FFK-nak van szerkezet-alapú (domén-domén vagy domén-motívum) bizonyítéka is, akkor nagyobb valószínűséggel lehet valós. Ezeket a metszetben lévő prediktált FFK-kat érdemes fókuszált kísérletekkel bizonyítani.

#### 1.5.4.5. Fehérje hálózatok vizsgálata

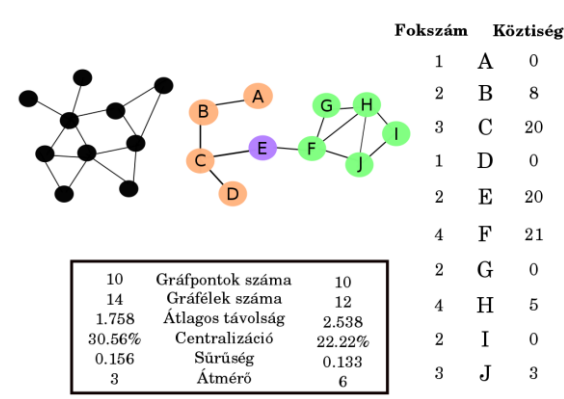
A hálózatokat szokás gráfnak is hívni, így a fehérje hálózatok is egyfajta gráfok, amelyeken alkalmazhatóak a gráfok vizsgálatára kidolgozott módszerek. Fehérjehálózatban, a fehérje-fehérje kapcsolat jellegétől függően, a kapcsolatok (tehát a gráf élei) lehetnek irányítottak vagy szimmetrikusak. Egy A-B fehérjekölcsönhatás például irányított az „A enzimfehérje által katalizált reakció végterméke belép a B enzimfehérje által katalizált reakcióba” típusú esemény, ilyenkor a hálózat élei az enzimfehérjék által katalizált reakciók kapcsolatát reprezentálják. Irányítatlan egy C-D fehérjekölcsönhatás például a „C fehérje és a D fehérje együttesen komplexet alkot” típusú kapcsolatban. Módszertani szempontból elmondható, hogy az irányítottság megváltoztatása az adatbázisok kezelése során is és a hálózatelemzés során is viszonylag egyszerű.

A kapcsolatok jellemzésére használhatjuk azok előjelét és mérhető fontosságát is. Előjeles hálózatokban minden kölcsönhatás vagy pozitív vagy negatív. Ilyen hálózat lehet egy regulációs kapcsolatrendszer, melyben általában gének közötti kölcsönhatásokat reprezentálunk, de fehérjéket is elemezhetünk így. Súlyozott hálózatokban az A és B fehérje kölcsönhatásának erejét is mérhetjük (például affinitás alapján). Egy súlyozott hálózat általában reálisabb képet ad a vizsgált biológiai rendszerről, de ha a súlyok változékonyak, akkor éppen a súlyozatlan, bináris (van kapcsolat / nincs kapcsolat) hálózat lehet megbízhatóbb. Az egyszerűség kedvéért irányítatlan hálózatokat vizsgálunk a következőkben.

Általánosan elmondható, hogy egy kölcsönhatási hálózat jellemzése két fő csapás mentén történik. Egyrészt mondhatunk valamit magáról az egész hálózatról, ezek a globális vagy makroszkopikus információk. Megadhatjuk, hány pontot (nódust) és hány kapcsolatot tartalmaz, jellemezhetjük a sűrűségét (milyen gazdag a kapcsolatrendszer,

hány kölcsönhatást látunk az elméletileg lehetséges maximumhoz képest), vagy az átmérőjét (milyen messze, azaz hány lépésre van egymástól a két legtávolabbi pont). Ezek a globális mérőszámok lehetővé teszik különböző kölcsönhatási hálózatok összehasonlítását vagy éppen egyazon hálózat időbeli változásainak nyomon követését. Ilyenkor minden információt egyetlen mérőszámba sűrítünk össze: ez nyilván nem alkalmas a részletek bemutatására, de jelezhet valamit a rendszer szintjén (ilyen makroszkopikus indikátor a nemzetek gazdaságát jellemző GDP is).

Másrészt viszont megvizsgálhatjuk a hálózat részletes belső szerkezetét is, ez a lokális nézőpontú elemzés. Ilyenkor egyetlen pont, egyetlen kölcsönhatás vagy pontok egy kisebb halmaza érdekel minket és ezek hálózati pozícióját jellemezzük a hálózat egészén belül. A 3. ábra mutat egy példát a kétféle nézőpontú elemzésre, melyek nyilván nem kizárják, hanem inkább kiegészítik egymást, vagy éppen teljesen más célokat szolgálnak.



3. ábra. Két hipotetikus hálózat globális összehasonlítása (a keretben) és a színes hálózat pontjainak lokális jellemzése itt most csak két mérőszám, a fokszám és a köztiségi centralitás alapján.

A 3. ábra mindkét hálózatában 10 hálózati pontot látunk, de a fekete sűrűbb (több gráfélt, azaz több kapcsolatot tartalmaz). A kisebb átlagos távolság azt jelenti, hogy két véletlenül kiválasztott pont várhatóan közelebb lesz egymáshoz (az átmérője is sokkal kisebb). A

fekete hálózat centralizáltabb, míg a színes jobban szétterül.

A pontok helyzetének részletes elemzése megmutatja, hogy az F és a H pontnak van a legtöbb szomszédja. A H pont kapcsolatai ugyanabból a körből kerülnek ki, míg az F összeköti a hálózati különböző pontcsoportosulásait. Az E pont majdnem ugyanilyen fontos összekötő szerepet játszik, pedig csak feleannyi szomszédja van. Az ilyen „broker” pozícióban lévő fehérjék esetleg pleiotróp hatásúak lehetnek, hiszen különböző funkciójú fehérjék csoportjai között létesíthetnek kapcsolatot (Nguyen és mtsai 2011).

Mіндеzen mérőszámok nyilván sokkal izgalmasabbak és új, váratlan információkkal is szolgálhatnak nagyobb hálózatok esetében.

Nagy, valódi adatbázisok alapján készített hálózatok vizsgálata során azt tapasztalhatjuk (Nguyen és mtsai 2011), hogy a színes hálózat E pontjához hasonló pozícióban gyakran találhatóak chaperon fehérjék (pl. 14-3-3 protein zeta/delta) és a sejten belüli vagy sejtek közötti kommunikációban szerepet játszó fehérjék (pl. Epidermal growth factor receptor).

A lokális vizsgálat során szóba jövő lehetőségek közül olyan hálózatelemző módszereket mutatunk be, melyek egyrészt egyszerűek, másrészt széles körben alkalmazzák őket, harmadrészt pedig csak kevésbé redundáns információt szolgáltatnak.

- Fokszám

Egy tetszőleges hálózatban egy adott pont pozicionális jellemzésének legegyszerűbb módja, ha megadjuk a vele szomszédos (vele összekötött) pontok számát, ezt nevezzük fokszámnak (degree,  $D$ ). A fokszám értéke a leginkább lokális tulajdonsága egy pontnak. Ha a gráf minden pontjára meghatározzuk a fokszámot, jellemezhetjük a hálózat fokszámeloszlását is. A fokszám és a fokszámeloszlás azonban nem mindig ad

precíz jellemzést a pont hálózati pozíciójáról, mert például nem veszi figyelembe az indirekt kölcsönhatásokat: egy pont kapcsolatrendszere úgy is lehet gazdag, ha kevés szomszéd mellett sok második, harmadik, stb. szomszédal rendelkezik.

- Köztség (*betweenness centrality*)

Egy, az indirekt kölcsönhatásokat is figyelembevevő index a köztségi centralitás (*betweenness centrality, BC*). Ez azt határozza meg, hogy egy adott  $i$  pontot milyen valószínűséggel tartalmaznak a további  $j$  és  $k$  pontpárok közötti legrövidebb utak. A standardizált index az  $i$  hálózati pontra ( $BC_i$ ):

$$BC_i = \frac{2 \times \sum_{j < k} g_{jk}(i) / g_{jk}}{(N-1)(N-2)},$$

ahol  $i \neq j$  és  $g_{jk}$  a legrövidebb utak száma  $j$  és  $k$  pontok között,  $g_{jk}(i)$  ezek közül azoknak a száma, melyek tartalmazzák  $i$ -t ( $g_{jk}$  természetesen lehet 1) és  $N$  a gráf csúcspontjainak a száma. A nevező a maximálisan elérhető értékhez viszonyítja az  $i$  pont értékét. A  $BC$  index által centrálisnak tekinthető pontok a felelősek a hálózatban leggyorsabban terjedő hatások fenntartásáért.

- Topológiai index

Az indirekt hatások erősségének szerkezeti alapon történő becsléséhez egy általánosított módszer (Jordán és mtsai 2003) lehetővé teszi akármilyen hosszú,  $n$  lépésből álló indirekt kapcsolatok jellemzését is (topological importance,  $TI$ ). A  $j$  fehérje hatása az  $i$  fehérjére  $n$  lépésben  $a_{n,ij}$ . Legegyszerűbb esetben ( $n=1$ )  $a_{1,ij} = 1/D_i$ , ahol  $D_i$  az  $i$  fehérje súlyozatlan (bináris) hálózatban számolt fokszáma. A hatások hosszabb útvonalak mentén összeszorzódnak és párhuzamos, azonos hosszúságú útvonalakra összeadódnak.

$n$  lépésre az  $i$  fehérjét az összes többi  $N-1$  fehérjéből elérő hatások összege:

$$\psi_{n,i} = \sum_{j=1}^N a_{n,ij}$$

és ennek összege 1. Szintén  $n$  lépésre az  $i$  fehérjéből az összes többi fehérjéhez eljutó hatás összege:

$$\sigma_{n,i} = \sum_{j=1}^N a_{n,ji}$$

de ennek összege a különböző fehérjékre már nem állandó. Ez alapján definiáljuk az  $i$  fehérje topológiai fontosságát az  $n$  lépés hosszúságú utak szerint (a bemutatott eredmények mind az  $n = 3$  esetre vonatkoznak):

$$TI_i^n = \frac{\sum_{m=1}^n \sigma_{m,i}}{n} = \frac{\sum_{m=1}^n \sum_{j=1}^N a_{m,ji}}{n}$$

Ezt az indexet bináris és súlyozott hálózatokra is ki lehet számolni (Valentini és Jordán 2010), de a továbbiakban csak a bináris hálózzal foglalkozunk. A súlyozás hatása (a kettő közötti különbség) önmagában is érdekes kutatások alapjául szolgál.

A 3. ábrán látható hálózatban a B hálózati pont szerkezetileg becsült hatása a C pontra  $1/3$ , mert a C három szomszédjának az egyike. A C pont hatása az E pontra  $1/2$ , hasonló megfontolás alapján. Két lépéses, indirekt kölcsönhatásokban gondolkozva megállapítható, hogy a B-E interakció erőssége  $1/6$  ( $1/3 * 1/2$ ). Az E pont két lépéses kapcsolatokban gondolkozva tehát  $1/6$  mértékben függ éppen a B ponttól. Ha fehérje hálózatokban gondolkodunk, akkor a funkcionális egymásrautaltság, a sérülékenység vagy éppen a helyettesíthetőség jellemezhető (és számszerűsíthető) ezzel a fajta hálózatos megközelítéssel.

Nagyobb hálózatok elemzése során izgalmas feltérképezni az indirekt kölcsönhatások



rendszerét, mert ez egyrészt új információk feltárását, másrészt az adatbázis minőségének ellenőrzését segítheti. Példaként, egy 2221 fehérje kölcsönhatási rendszerét magában foglaló hálózat vizsgálata például azt sugallta, hogy a PDZ domént tartalmazó GIPC1 fehérje feltehetően kapcsolatban áll a cukorbetegséget befolyásoló fehérjék csoportjával, bár az adatbázis nem jelzett ilyen közvetlen kölcsönhatást (Ngyuen és mtsai 2011).

- globális indexek (sűrűség, átlagos távolság, átmérő és centralizáltság)

Egy fehérje kölcsönhatási hálózat (és bármilyen más hálózat) alapvető tulajdonsága, hogy hány pontot és élt (kapcsolatot) tartalmaz. A hálózati pontok száma alapján megadható a kapcsolatok számának elméletileg lehetséges maximuma, és ha az aktuálisan jelenlévő kapcsolatok számát ehhez viszonyítjuk, megkapjuk a gráf sűrűségét.

Két adott hálózati pont távolsága egy, ha össze vannak kötve egymással. Ha egy harmadik pont köti össze őket, mint közös szomszéd, akkor távolságuk kettő. Két tetszőleges pont távolsága így megadható az őket összekötő legrövidebb út hosszaként (hány kapcsolaton keresztül jutunk el az egyikből a másikba). Itt megjegyzendő, hogy a gráfélek irányítottasága fontos lehet: irányított kölcsönhatások esetében A és B távolsága esetleg csak 1, de B és A távolsága lehet sokkal nagyobb is. Ezt a hálózatelemzés természetesen kezelni tudja. A hálózat összes pontpárjára kiszámolt távolságértékek közül a legnagyobbat a gráf átmérőjének nevezzük.

A hálózat (topológiai értelemben vett) alakját jól leírja a centralizáltság mérőszáma. Szükséges esetben egy pont az összes többivel össze van kötve, de a többi pont között nincs semmilyen kapcsolat. Ebben a tökéletes csillagalakban a centralizáltság értéke 100%. A másik szélsőség a matekfüzet négyzethálója (ahol a metszéspontok a hálózati pontok): itt minden pont egyformán fontos pozíciójú, négy-négy szomszédal, a

négyzetháló centralizáltsága 0%. 3. ábra fekete és színes hálózata köztes állapotokat mutat, a megfelelő értékek feltüntetésével.

#### 1.5.4.6. Fehérje hálózatok vizsgálata Cytoscape programmal

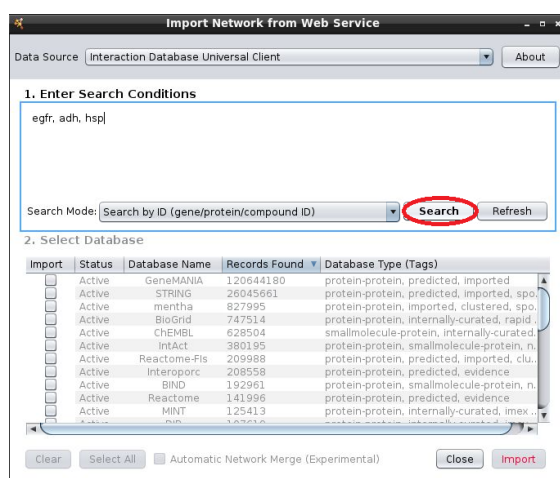
A fehérje hálózatok vizsgálatának legelterjedtebb és legegyszerűbb módja a Cytoscape nevű ingyenes program használata. A Cytoscape-et a honlapjáról lehet letölteni (<http://cytoscape.org>), de érdekessége a programnak, hogy az alapszoftveren kívül több száz kisebb kiegészítő egységet lehet ingyen telepíteni hozzá, amelyek speciális feladatok és lehetőségek elvégzésével növelik a Cytoscape alkalmazhatóságát. Ezeket az ún. plug-in-eket vagy közvetlenül a programból vagy a Cytoscape honlapján böngészve lehet megtalálni és telepíteni.

A Cytoscape megnyitáskor felkínálja a lehetőséget, hogy új hálózatot építsünk, betöltsük egy korábban elmentett munkánkat, vagy egy publikusan elérhető hálózatot töltsünk le (4. ábra). Az ember és néhány modell organizmus ismert fehérje-fehérje kapcsolati hálózatát is importálhatjuk. Ezek a lehetőségek a program használata közben bármikor választhatóak a File -> Import -> Network -> Public Databases menüpont segítségével. Példaként most töltsük le három fehérje ismert fehérje-fehérje kapcsolatait egy hálózati adatbázisból.



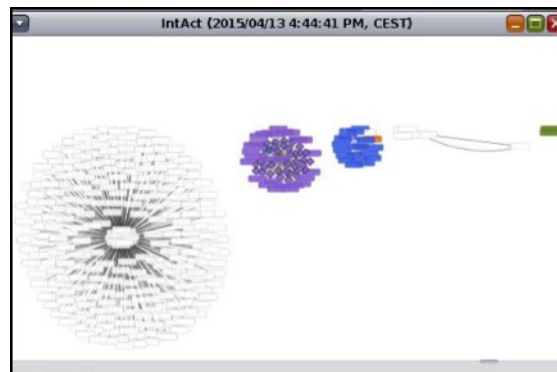
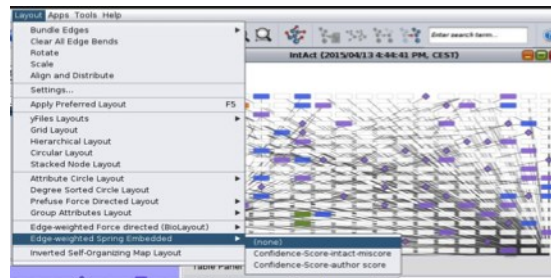
4. ábra: Publikusan elérhető hálózatok betöltése

A keresőmezőbe beírhatunk egy vagy több fehérje nevet, vagy adatbázis azonosítókat (pl.: UniProt). Miután megvan a keresőkifejezésünk, ne felejtünk el a Search gombra kattintani, különben a választható adatbázisok nem lesznek elérhetők (5. ábra). Miután megjelent, hogy melyik adatbázisban pontosan mennyi kapcsolati információ szerepel, bejelölhetjük, hogy melyeket szeretnénk letölteni. Amennyiben több adatbázisból szeretnénk importálni, azok vagy külön hálózatként lesznek megjelenítve, vagy egyből össze is fésülhetjük őket.



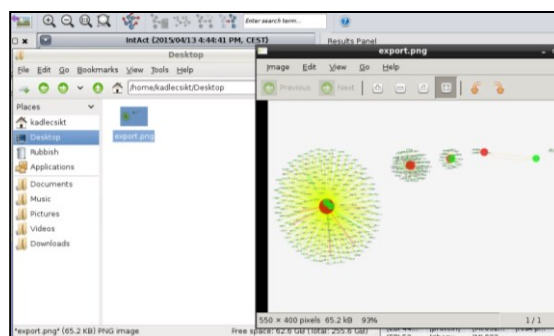
5. ábra: Adott kulcsszóra való keresés

Az importált hálózat alapértelmezésben egy igen fantáziátlan rács elrendezésben jelenik meg. A hálózat pontjait akár kézzel is átrendezhetjük, de a Cytoscape rendelkezésünkre bocsát egy igen könnyen használható layout menüt (6. ábra). A layout a menüsoron keresztül jó pár egyszerűbb és akár matematikailag komplexebb algoritmust is használhatunk, hogy (biológiai) értelmezhető elrendezésben lássuk a hálózatunkat. Példának próbáljuk ki a Layout -> Edge Weighted Spring Embedded elrendezést (6. ábra)!



6. ábra: A hálózati elrendezés kiválasztása (fent), majd ez ebből kapott hálókép (lent)

A hálózat pontjait és kapcsolatait különböző szabályok szerint színezzük. Erre a Style fül segítségével van lehetőség. Ennek köszönhetően exportálás előtt nagyíthatjuk vagy kicsinyíthetjük is a képet megfelelő méretűre. Fontos továbbá megjegyezni, hogy hiába választjuk ki a kívánt formátumot (JPG, PNG, PDF, SVG), a fájlnev nem egészül ki a megfelelő kiterjesztéssel, azt nekünk kell megadni. Érdekes különben például PNG formátumba menteni a hálózatról a képet (7-ábra).



7. ábra: Hálózat elmentése PNG formátumba

**Praktikus megfontolás:** Néhány speciális hálózatelemző index csak a fejlesztők házi használatú programjaiban vagy csak speciálisabb hálózatelemző szoftverekben érhető el. A fent bemutatott centralizációs indexek közül például a fokszámot valószínűleg minden létező szoftver ki tudja számolni, de a TI index kiszámolására csak a CosbiGraph környezet alkalmas (Valentini és Jordán 2010).

#### 1.5.4.7. Fehérje hálózatok kutatásának alkalmazása

A legtöbb esetben egészséges állapotú élőlényekből vagy sejt kultúrából származnak az FFK adatok. Ugyanakkor a legtöbb kutatás valamilyen betegség során fellépő változást, eltérést vizsgál, illetve ezeket az eredményeket felhasználva gyógyászati megközelítésekre tesz javaslatot. Szerencsére az elmúlt években számtalan olyan kísérleti vizsgálat született (például állapot- és betegség-specifikus mRNS expressziós adatok, mutációs adatok, epigenetikai és foszfo proteomikai adatok), amelyek eredményei ráilleszhetőek az emberi fehérje-fehérje hálózatra. A mutációs adatok alapján kiemelhetünk fehérjéket a vizsgálataink során, és összevethetjük a mutált fehérjék mérőszámait a többi fehérje hálózati mérőszámaival. Az expressziós adatok segítségével egy általános fehérje-fehérje hálózatból sejt-, szövet- vagy betegség-specifikus hálózatokat tudunk készíteni, amelyeket akár egymással is összehasonlíthatunk. Egy általános hálózatból úgy lesz specifikus, hogy kiesnek azok a fehérjék, amelyek expresszióját-jelenlétét nem sikerült kimutatni az adott állapotban. Egy fehérje kiesésével természetesen az általa alkotott kapcsolatok is kiesnek. Következésképp, két szövet fehérje-fehérje hálózata jelentős különbségeket is tartalmazhat, mind ránézésre, mind a mérőszámokat statisztikailag elemezve.

A legtöbb betegségben érintett gén által kódolt fehérje nem a legmagasabb fokszámmal rendelkezik, és nem a hálózat központjában helyezkedik el. Ez alól egy fontos kivétel a rákban mutációval érintett fehérjék, amelyek jellemzően magas fokszámú, központi fehérjék. A jelenlegi gyógyszer-célpont fehérjéknek is több szomszédja van mint egy átlagos fehérjének. Hasonlóan, a gyógyszer-célpont fehérjék általában központi pozíciót töltenek be a fehérje-fehérje hálózatban. Ugyanakkor minél központibb egy fehérje, minél több kapcsolata van, annál nagyobb az esélye, hogy a gyógyszeres kezelés mellékhatással jár. Épp ezért nagyon fontos, hogy felderítsük egy gyógyszer-célpont fehérje szomszédságát. Ha egy lehetséges gyógyszer-célpont fehérje kapcsolódó partnerei különböző funkciókban vesznek részt, akkor a mellékhatás esélye magasabb, mintha hasonló funkciójú fehérjék lennének a célpont közelében. Ugyanakkor rákellenes terápiaik esetében, amikor a sejt elpusztítása a cél, pont fordítva, azok a legjobb gyógyszer-célpontok, amelyek különböző funkcióira is hatnak. A gyógyszerfejlesztés kihívása, hogy általában ezek a központi célpont fehérjék esszenciálisak az egészséges sejtekben is, így csak célzott terápiával vagy egyéb, komplexebb kezeléssel lehet ezt a megközelítést alkalmazni.

Érdekes módon a FFK-ok maguk is lehetnek gyógyszer-célpontok. Azaz egy hatóanyag molekula nem egy egész fehérje funkciójára van hatással, csak egy adott, specifikus kapcsolatra. Sőt, ez akár olyan kapcsolattípus is lehet, amely több fehérje között is jelen van (például egy domén-domén kapcsolat gátlása). Egy FFK célzása specifikusabb behatást jelent, mint egy egész fehérje funkcióinak teljes gátlása. A fehérje hálózat *in silico* modellezése, hálózati dinamikájának részletes vizsgálata szükséges ahhoz, hogy a több százezer kapcsolat közül kiválasztható legyen a gyógyszeres célzásra leginkább ajánlott FFK (Csermely et al, 2013).

#### 1.5.4.8. Összefoglalás

A hálózatelemzés során át tudjuk tekinteni egy fehérje interakciós partnereit, mely más fehérjékkel áll kapcsolatban a sejtben. A hálózat megmutatja ezen partnerek kapcsolatait is, tehát bepillantást nyújt az indirekt kölcsönhatási rendszerbe is. A teljes hálózatot tekintve segít megérteni egy adott fehérje pozícióját, utalva az abból fakadó funkcionális tulajdonságaira, működésére. Végül pedig magát a rendszert, a teljes kölcsönhatási hálózatot is jellemezhetjük, ily módon jelezhetjük a rendszerszintű változásokat és kereshetjük annak okait. Fontos megjegyezni, hogy a hálózatelemzés eredményei sokszor új kutatások kezdetét jelentik (ötletet adnak, valamit jeleznek), nem feltétlenül a kutatás végső következtetéseit nyújtják.

A hálózatok kutatása többféle segítséget nyújt a fehérjekutatásban. Egyrészt a különböző adatok összegzését és az adatbázisok módszertani egységesítését ösztönzi. Másrészt alkalmas a nagy mennyiségű adat vizualizációjára, ez gondolatébresztő hatású lehet, akár ötleteket is adhat új laboratóriumi kísérletek elvégzéséhez. Harmadrészt az adatbázisok szisztematikus áttekintése módot ad az esetlegesen hiányzó vagy rossz minőségű adatok felülvizsgálatára. Negyedrészt, és ez a legfontosabb, valóban új információkkal, felfedezésekkel szolgálhat.

#### 1.5.4.8. Hivatkozások

Barabási, A.-L., 2002. *Linked: The New Science of Networks*. Perseus Publishing.

Csermely, P. 2006. *Weak Links: Stabilizers of Complex Systems from Proteins to Social Networks*. Springer Verlag, Berlin, pp. 408.

Valentini, R. és Jordán, F. 2010. *CoSBI Lab Graph: the network analysis module of CoSBI Lab*. *Environmental Modelling and Software*, 25:886-888.

Nguyen, T.P., Liu, W.C. és Jordán, F. 2011. *Inferring pleiotropy by network analysis:*

*linked diseases in the human PPI network*. *BMC Systems Biology*, 5:179.

Csermely, P, Korcsmáros, T, Kiss, HJ, London, G, Nussinov, R. 2013. *Structure and dynamics of molecular networks: a novel paradigm of drug discovery: a comprehensive review*. *Pharmacol Ther.* 138(3):333-408.

#### Szójegyzet

**hálózat (gráf):** Hálózati pontokból és gráfélekből álló modell, mely azt mutatja, a vizsgált rendszer mely elemei között van kapcsolat egy kiválasztott kölcsönhatástípus alapján.

**gráfelmélet:** Egymással kölcsönható elemekből álló rendszerek tulajdonságait vizsgáló tudományág. Ilyen rendszer lehet például a sejt, melyben a különböző típusú fehérjék a kölcsönható elemek.

**Nódus (hálózati pont):** Egy hálózati modellben a vizsgált rendszer elemeit mutatja, fehérjehálózatban általában megfelel egy fehérjének (adott esetben egy fehérjekomplexnek).

**Hálózat él:** Egy hálózati modellben a vizsgált rendszer elemei közötti kapcsolatot mutatja. Különböző tulajdonságokkal rendelkezhet: lehet irányított („A” fehérje hatással van „B” fehérjére, de fordítva nem), súlyozott („A” és „B” fehérje kapcsolata erős vagy gyenge, akár számokkal is kifejezve) vagy előjeles („A” fehérje pozitív avagy negatív hatással van „B” fehérjére).

**távolság:** Egy hálózatban két adott pontot összekötő kapcsolatok minimális száma. Szomszédos pontok távolsága egy. Ha két pont között nincs összeköttetés, mert különböző komponensekben vannak, akkor távolságuk végtelen.

**fokszám:** Egy hálózati pont szomszédainak száma, tehát azon nódusok száma, melyeknek a vizsgált ponttól mért távolsága egy.

részgráf: Egy hálózat grápointjainak valamilyen részhalmaza, mely a gráfélek közül is csak az azok között meglévőket tartalmazza.

hálózati modul / motívum: Valamilyen, rendszerint ismétlődő és szerkezeti vagy funkcionális szempontból nevezetesenek tartható kis méretű részgráf.