

Doktori (Ph.D.) értekezés tézisei

Biológiai hálózatok átfedő modularizálását végző számítógépes programok és azok alkalmazási területei

Szalay-Bekő Máté

Témavezetők:

Dr. Csermely Péter egyetemi tanár

Semmelweis Egyetem, Orvosi Vegytani, Molekuláris Biológiai és
Patobiokémiai Intézet

Dr. Papp Balázs tudományos főmunkatárs

Magyar Tudományos Akadémia, Szegedi Biológiai Központ,
Biokémiai Intézet

Biológus Doktori Iskola

Szegedi Tudományegyetem, Természettudományi és Informatikai Kar



Szeged
2013

Bevezetés

Az elmúlt másfél évtizedben a biológiai kutatások területén ugrásszerűen megnőtt a mérési lehetőségek és adatok száma, utat nyitva ezzel a biológia informatikai módszerekkel támogatott rendszerszintű elemzése számára. A hálózat kutatás érdekes és új nézőpontok megjelenését hozta be a biológiába, azáltal hogy a biológiai rendszereket (például a sejt jelátviteli vagy regulációs folyamatait) gráfokkal modellezi, ahol a gráf pontjai biológiai entitásokat (például fehérjéket vagy géneket), míg a gráf élei az entitások közötti fizikai vagy logikai kölcsönhatásokat fejezik ki.

A disszertációmban bemutatott tudományos munkám célja, hogy a bonyolult biológiai rendszerek megértését segítsem elő olyan informatikai programok fejlesztésével, amelyek segítenek az élő rendszereket leíró hálózatos modellek szerkezetének feltérképezésében.

A természetes rendszerek túl fejlettek és ellenállóak ahhoz, hogy szerkezeti felépítésük véletlenszerű legyen. Ezért a komplex biológiai, társadalmi vagy akár üzleti folyamatokat modellező hálózatok vizsgálata során az egyik legfontosabb feladat a hálózat szerkezetének feltérképezése. Ezen hálózatokra általában jellemző az egymással szorosan együttműködő pontok modulokba, sűrű hálózati régiókba való csoportosulása. Például egy sejt fehérje-fehérje kölcsönhatási hálózatában ilyen csoport lehet a riboszómát alakító, vagy az RNS poszttranszkripció módosulásáért felelős fehérjék funkcionális csoportja. A nagy, akár több ezer vagy több tízezer pontot tartalmazó komplex hálózatokban a csoportok megtalálása bonyolult feladat, amelyre több mint száz eltérő megoldást adnak a szakirodalomban (Fortunato, 2010). Ezen megoldások túlnyomó többsége a hálózat egy adott pontját csak egyetlen modulba sorolja be. Ennél több információt nyújtanak az átfedő modularizálási algoritmusok (ahol a pontok több modul tagjai lehetnek egyszerre), illetve a fuzzy eljárások (ahol rendelkezésre áll az az információ, hogy az adott pont melyik modulhoz mennyire tartozik). **A dolgozatomban a ModuLand nevű fuzzy modularizálási eljárás informatikai megvalósítását és az eljárás gyakorlati példákon való alkalmazását ismertetem.**

Célkitűzések

A tudományos munkám során az alábbi célkitűzéseket foglalmaztam meg.

1. Egy olyan felhasználó barát modularizációs számítógépes program megvalósítása, amely konkrét biológiai kérdések megválaszolását teszi lehetővé komplex biológiai hálózatos modellek modulszerkezetének elemzésével.
2. A fenti modularizációs program alkalmazhatóságának vizsgálata az *Escherichia coli* baktérium Met-tRNS szintetáz fehérjéje aminosav hálózat modulszerkezetének elemzésével.
3. A fenti modularizációs program alkalmazhatóságának vizsgálata a *Buchnera aphidicola* és az *Escherichia coli* baktériumok metabolikus hálózatainak összehasonlításával.
4. A fenti modularizációs program alkalmazhatóságának vizsgálata az élesztő modulszerkezetének elemzésével, illetve az expressziós dinamika szempontjából érdekes tulajdonságot mutató date és party hub fehérjék modulszerkezetben betöltött szerepének elemzésével.

Alkalmazott módszerek

A munkámhoz felhasználtam Dr. Kovács István kollégám által kidolgozott ModuLand fuzzy modularizáló algoritmusokat, amiket egy közös publikációmban tettünk közre (Kovács és mtsai., 2010). Szintén támaszkodtam ezeknek az algoritmusoknak a Linux operációs rendszeren megvalósított változataira, melyek kezdeti verzióit Palotai Robin, Szuromi Gábor és Zalányi Balázs segítségével közösen implementáltunk. A saját munkám során főként az algoritmusok optimalizációjával, a felhasználóbarát kezelőfelület megalkotásával és számos biológiai példa elemzésével foglalkoztam (Szalay-Bekő és mtsai, 2012).

A kutatócsoportom tagjaival közösen definiált módszert egy Cytoscape nevű (Shannon és mtsai, 2003) hálózat ábrázoló és elemző

keretrendszerbe építettem. Az így nyert, biológusok számára kényelmesen használható programot¹ egy nemzetközi bioinformatikai folyóiratban megjelent megosztott első szerzős cikként publikáltam (Szalay-Bekő és mtsai., 2012). Az implementáció során C++ és Java programozási nyelveket használva több optimalizációt is megvizsgáltam és bevezettem a programok futásidejének csökkentése érdekében.

Az általam megvalósított program segítségével vizsgálhatóvá válik tetszőleges irányítatlan hálózat hierarchikus modulszerkezete. A program képes a modulok központi régióinak meghatározására, valamint a hálózat egyes pontjainak a modulszerkezetben betöltött szerepének számszerű jellemzésére. Például egy fehérje-fehérje kölcsönhatási hálózat esetén megállapítható, hogy az adott fehérje mennyire játszik központi szerepet egy funkcionális modulban, vagy épp mennyire alkot hidat különböző funkcionális modulok között. A számos különböző mérőszámról és a modulszerkezetet jellemző adatot a program képes igen egyszerűen (például Excel formátumú) fájlokba kimenteni.

Eredmények

Az általam fejlesztett plug-in-t számos biológiai kutatás során felhasználtam részben én, részben a kutatócsoportom tagjai és részben más nemzetközi kutatócsoportok.

Az *E. coli* baktérium Met-tRNS szintetáz fehérjéjének térszerkezetét modellező aminosav hálózat esetében (Szalay-Bekő és mtsai., 2012) összevettem a hálózat modulszerkezetét a fehérje domain szerkezetével, illetve megvizsgáltam az enzim katalitikus központja és antikodon kötőhelye közötti konformációs változások továbbadásáért felelős aminosavak modulszerkezetben betöltött pozícióját és ezek sajátosságait. Kimutattuk, hogy a fehérjén belüli jelterjedés továbbításáért felelős aminosavak között a statisztikailag vártnál sokkal nagyobb mértékben feldúsulnak a ModuLand Cytoscape plug-in által modulközép pozícióban vagy modulok közötti híd pozícióban megjelölt aminosavak.

1 Az általam megvalósított ModuLand Cytoscape plug-in letölthető a következő honlapról: <http://www.linkgroup.hu/modules.php>

A ModuLand plug-in használatával munkatársaimmal közösen összehasonlítottuk (Szalay-Bekő és mtsai., 2012) a *Buchnera* és *E. coli* baktériumok metabolikus folyamatait leíró hálózatokat, kimutatva az előbbi élőlénynél a szabadon élő, illetve az utóbbi esetében a szimbióta életmódból fakadó különbségeket a metabolikus hálók modulszerkezeteiben. Azt találtuk, hogy a szimbióta *Buchnera* baktérium moduljai átlagosan sokkal nagyobb méretűek és jobban átfednek egymás között, illetve a modulokra jutó metabolikus funkciók átlagos száma is magasabb. Az eredményünk összhangban van a szakirodalom hasonló elemzéseivel (Parter és mtsai., 2007; Mihalik és Csermely, 2011), amelyek általában a nagyobb stressznek kitett élőlények metabolikus vagy fehérje-fehérje kölcsönhatási hálózatai esetén figyeltek meg hasonló modulszerkezeti eltéréseket.

A ModuLand eljáráscsalád alkalmazásával megvizsgáltuk (Kovács és mtsai., 2010) az élesztő fehérje-fehére kölcsönhatási hálózatának modulszerkezetét, ahol a program által meghatározott modulokat sikeresen tudtuk megfeleltetni sejtes funkciókkal. Szintén ebben a hálózatban vizsgáltuk a sok szomszédal rendelkező csomóponti fehérjék kapcsolatainak expressziós mintázatait. Az összes kapcsolatát tipikusan egyszerre expresszáló party hub fehérjéket és a kapcsolataikat a különböző expressziós adatsorokban gyakran váltogató date hub fehérjéket sikeresen elkülönítettük egymástól, miután összefüggést fedeztünk fel ezen fehérjék expressziós mintázatai és a modulszerkezetben betöltött helyzetük között. Ilyen módon képesek voltunk fehérje expressziós dinamikai tulajdonságok jóslására, pusztán a hálózat statikus felépítésének elemzése alapján.

Az általam megvalósított ModuLand plug-in-t 2012-es publikálása óta több mint 150 kutató töltötte le és több nemzetközi publikációban közölt, tölem független kutatásnál felhasználásra került. Például a ModuLand plug-in segítségével vizsgálták (Lehtinen és mtsai., 2013) az oxidatív stressz élesztőre gyakorolt rendszerszintű hatását, míg egy másik tanulmány pedig különböző fajok fehérje-fehérje kölcsönhatási hálózatának illesztését oldotta meg a ModuLand plug-in használatával, olyan módon hogy a feltehetőleg evolúciósan őrződő modulszerkezet minél kevésbé sérüljön a fajok között (Wang és Gao, 2012). Egy harmadik cikkben (Sharma és mtsai., 2013) pedig a Szirtuin enzimcsalád köré épített

fehérje-fehérje kölcsönhatási hálózat 20 funkcionális modulját vizsgálták a ModuLand plug-in segítségével.

Meggyőződésem, hogy a ModuLand Cytoscape plug-in, illetve a jövőben elkészülő hozzá hasonló, átfedő modulok finomszerkezetét feltáró programok képesek sok további érdekes biológiai és gyógyszerkutatói kérdést megválaszolni. A korábban említett konkrét példákon túl a plug-in a jövőben felhasználható lehet például fehérjék funkció jóslásánál, jelátviteli utak keresztcsatlakozásainak (cross-talk) detektálásánál, gyógyszercélpont fehérjék azonosításánál, de akár komplex biológiai rendszerek áttekinthető ábrázolásánál is.

Saját publikációk jegyzéke

2013. november 2.²

Összesített impakt: 50

Független idézetek száma: 187

Házasságkötés előtt használt publikációs név: **Szalay, M.S.**

A disszertációhoz kapcsolódó közlemények

1. Szalay-Bekő, M., Palotai, R., Szappanos, B., Kovács, I.A., Papp, B., Csermely, P. (2012) ModuLand plug-in for Cytoscape: determination of hierarchical layers of overlapping modules and community centrality. *Bioinformatics*, **28**, 2202-2204, IF: 5,3 <http://arxiv.org/abs/1111.3033> – 6 független idézet
2. Kovács, I.A., Palotai, R., Szalay, M.S., Csermely, P. (2010) Community landscapes: a novel, integrative approach for the determination of overlapping network modules. *PLoS ONE* **7**, e12528, IF: 3,7 www.arxiv.org/abs/0912.0161-- 26 független idézet

2 A független idézetek és az *Impact Factor* adatok kigyűjtéséhez a *Web of Science* online eszközt használtam. (<http://www.webofknowledge.com>)

A disszertációtól független saját közlemények

Eredeti tudományos közlemények

1. Fazekas, D., Koltai, M., Türei, D., Módos, D., Pálffy, M., Dúl, Z., Zsákai, L., Szalay-Bekő, M., Lenti, K., Farkas, I.J., Vellai, T., Csermely, P., Korcsmáros, T. (2013) Signalink 2 - a signaling pathway resource with multi-layered regulatory networks. *BMC Systems Biology* **7**(1):7, IF: 3 – 1 független idézet
2. Korcsmáros, T. Szalay, M.S., Rovó, P., Palotai, R., Fazekas, D., Lenti, K., Farkas, I.J. Csermely, P., Vellai, T. (2011) Signalogs: orthology-based identification of novel signaling pathway components in three metazoans. *PLoS ONE* **8**, e19240, IF: 3,7 – 2 független idézet
3. Korcsmáros, T., Farkas, I.J., Szalay, M. S., Rovó, P., Fazekas, D., Spiró, Z., Böde, C., Lenti, K., Vellai, T., Csermely, P. (2010) Uniformly curated signaling pathways reveal tissue-specific cross-talks, novel pathway components, and drug target candidates. *Bioinformatics* **26**, 2042-2050, IF: 5,3 www.signalink.org -- 10 független idézet
4. Wang, S., Szalay, M.S., Zhang, C., Csermely, P. (2008) Learning and innovative elements of strategy update rules expand cooperative network topologies. *PLoS ONE* **3**, e1917, IF: 3,7 www.arxiv.org/0708.2707 -- 21 független idézet

Szabadalmak

1. Korcsmáros T., Szalay-Bekő M., Palotai R., Szuromi G., Fazekas D., Dunai Zs. (2011) Eljárás és számítógépes rendszer gyógyszerhatóanyagok hatásmechanizmusának szimulációjára. Magyarországi szabadalmi bejelentés, P1100368
2. Szalay, M., Stanojevic, O., Farkas, L. (2010) Automatic use of behavioral information for promotional purposes in communications

system. Nemzetközi Ericsson PCT szabadalmi bejelentés, PCT/SE2010/051312

3. Kovács, I.A., Csermely, P., Szalay, M.S., Korcsmáros, T. (2006) Method for analyzing the fine structure of networks. Nemzetközi PCT szabadalmi bejelentés, PCT/IB2007/05047

Összefoglaló munkák

1. Farkas, I.J., Korcsmáros, T., Kovács, I.A., Mihalik, Á., Palotai, R., Simkó, G.I., Szalay, K.Z., Szalay-Bekó, M., Vellai, T., Wang, S., Csermely, P. (2011) Network-based tools in the identification of novel drug-targets. *Science Signaling* **4**, pt3, IF: 7,6 -- 3 független idézet
2. Palotai, R. Szalay, M.S., Csermely, P. (2008) Chaperones as integrators of cellular networks: changes of cellular integrity in stress and diseases. *IUBMB Life* **60**, 10-18, arxiv.org/0710.1622, IF: 2,8 -- 24 független idézet
3. Korcsmáros, T., Szalay, M.S., Böde. C., Kovács, I.A., Csermely, P. (2007) How to design multi-target drugs: Target-search options in cellular networks. *Expert Op. Drug Discov.* **2**, 799-808, arxiv.org/q-bio.MN/0703010, IF: 2,3 -- 20 független idézet
4. Böde. C., Kovács, I.A., Szalay, M.S., Palotai, R. Korcsmáros, T., Csermely, P. (2007) Network analysis of protein dynamics. *FEBS Lett.* **581**, 2776-2782, arxiv.org/q-bio.BM/0703025, IF: 3,6 -- 51 független idézet
5. Szalay, M.S., Kovács, I.A., Korcsmáros, T., Böde. C., Csermely, P. (2007) Stress-induced rearrangements of cellular networks: consequences for protection and drug design. *FEBS Lett.* **581**, 3675-3680, arxiv.org/q-bio.MN/0702006, IF: 3,6 -- 18 független idézet
6. Korcsmáros, T., Kovács, I.A., Szalay, M.S., Csermely, P. (2007) Molecular chaperones: the modular evolution of cellular networks. *J. Biosci.* **32**, 441-446, arxiv.org/q-bio.MN/0701030, IF: 1,8 -- 14 független idézet

7. Kovacs, I.A., Szalay, M.S., Csermely, P. (2005) Water and molecular chaperones act as weak links of protein folding networks: energy landscape és punctuated equilibrium changes point towards a game theory of proteins. <http://arxiv.org/abs/q-bio.BM/0409030>, *FEBS Lett.* **579**, 2254-2260, IF: 3,6 -- 21 független idézet

A tézisfüzetben hivatkozott további, független közlemények

1. Fortunato, S. (2010) Community detection in graphs. *Physics Reports*, **486**, 75–174.
2. Lehtinen, S., Marsellach, F.X., Codlin, S., Schmidt, A., Clément-Ziza, M., Beyer, A., Bähler, J., Orengo, C., Pancaldi, V. (2013) Stress induces remodelling of yeast interaction and co-expression networks. *Molecular Biosystems*, **9**, 1697–1707.
3. Mihalik, Á., Csermely, P. (2011) Heat shock partially dissociates the overlapping modules of the yeast protein-protein interaction network: a systems level model of adaptation. *PLoS Computational Biology*, **7**, e1002187.
4. Parter, M., Kashtan, N., Alon, U. (2007) Environmental variability and modularity of bacterial metabolic networks. *BMC Evolutionary Biology*, **7**, 169.
5. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, **13**, 2498–2504.
6. Sharma, A., Costantini, S., Colonna, G. (2013) The protein-protein interaction network of the human Sirtuin family. *Biochim. Biophys. Acta*, **1834**, 1998-2009.
7. Wang, B., Gao, L. (2012) Seed selection strategy in global network alignment without destroying the entire structures of functional modules. *Proteome Science*, **10**, S16.

Köszönetnyilvánítás

Köszönöm a Szegedi Tudományegyetemnek, a Biológus Doktori Iskolának és a disszertációm bírálóinak a fokozatszerzési eljárás során nyújtott segítséget.

Ezúton köszönöm témavezetőim segítségét! Különösen hálás vagyok **Csermely Péter professzor úrnak**, aki az elmúlt tíz évben sok különböző (elnöki, mentori, főnöki, témavezetői és különösen baráti) minőségben segített. Szintén köszönöm **Dr. Papp Balázs** segítségét a fokozatszerzéssel és publikációkkal kapcsolatos tudományos és adminisztratív munkákban.

Köszönettel tartozom **Dr. Korcsmáros Tamásnak** a kutatási munkámmal és a fokozatszerzési eljárással kapcsolatos folyamatos és szakszerű tanácsaiért. Szeretném megköszönni a Csermely Péter professzor úr által vezetett **LinkGroup hálózatkutató csoport** tagjainak a sokéves közös munkát és együttgondolkodást. A PhD disszertációmban bemutatott biológiai elemzéseket és informatikai munkák egy részét a csoport számos tagja segítette. Ezúton köszönöm **Dr. Kovács Istvánnak**, a ModuLand eljáráscsalád kitalálójának segítségét és a fuzzy modularizálás területéről folytatott számos érdekes beszélgetést. Szeretném megköszönni **Palotai Robin, Szuromi Gábor** és **Zalányi Balázs** informatikusok segítségét, akik részt vettek a ModuLand programcsomag eredeti, linux operációs rendszeren futó parancssoros verziójának megvalósításában. Szintén köszönöm **Szappanos Balásznak** a *B. aphidicola* és *E. coli* baktériumok metabolikus hálózatainak elemzésében nyújtott segítségét.

Köszönettel tartozom az **Ericsson Magyarország Kft-nek** amiért a fokozatszerzésemet tanulmányi szabadsággal és az eljárási díj egy részének átvállalásával segítette.

Végül és főként köszönöm családomnak, hogy az elmúlt években támogattak a bioinformatikusi tevékenységemet és elviselték az informatikai munkáim mellett kutatásra fordított túlórák sokaságát. Köszönöm szüleimnek, testvéreimnek, főként pedig feleségemnek és fiamnak, **Melindának** és **Márknak** a támogatást!

Társszerzői nyilatkozatok

Alulírott, Kovács István, mint Szalay-Bekő Máté doktori értekezésében felhasznált egyik publikáció (Kovács, I.A., Palotai, R., Szalay, M.S., Csermely, P. (2010) Community landscapes: a novel, integrative approach for the determination of overlapping network modules. *PLoS ONE* 7, e12528.) első szerzője kijelentem, hogy a közleményben közölt eredményeket tudományos fokozat (PhD.) megszerzésére sem én, sem más szerző nem használta fel, és a jövőben sem fogja. A közleményben leírt tudományos eredmények elérésében Szalay-Bekő Máté jelentős mértékben részt vett.

Budapest, 2013. október 30.



Kovács István

Alulírott, Palotai Robin, mint Szalay-Bekő Máté doktori értekezésében felhasznált egyik publikáció (Szalay-Bekő, M., Palotai, R., Szappanos, B., Kovács, I.A., Papp, B., Csermely, P. (2012) ModuLand plug-in for Cytoscape: determination of hierarchical layers of overlapping modules and community centrality. *Bioinformatics*, 28, 2202-2204.) megosztott első szerzője kijelentem, hogy a közleményben közölt eredményeket tudományos fokozat (PhD.) megszerzésére sem én, sem más szerző nem használta fel, és a jövőben sem fogja. A közleményben leírt tudományos eredmények elérésében Szalay-Bekő Máté jelentős mértékben részt vett.

Budapest, 2013. október 30.



Palotai Robin