

Az elmúlt öt évben végzett munka egy több mint harminc éve az OTKA/NKFI által finanszírozott, és egy már elnyert NKFI projekt támogatásával továbbiakban is folytatódó hosszú távú program része. Sajnos sok nehézséggel kellett megküzdeni ebben az öt évben. Projektünk 2012-es indulásától 2014 nyaráig a négy főmunkatársam közül három távozott Lendület-csoport vezetőnek a csoportból. A három húsz éves gyakorlattal rendelkező, világ színvonalú kutatót kezdőkkel kellett „pótolni”. A nagy összegű NK típusú projektből a létszámot részben tudtuk pótolni. A húszéves tapasztalatot viszont nem. Az MTA TTK működésével, főleg a pénzügyi és gazdasági helyzettel is rengeteg, a napi sajtóból is ismert, probléma volt. Mindezek ellenére megpróbáltuk teljesíteni vállalásainkat. A transzmembrán fehérjék, a stabil térszerkezetű, vizes közegben oldható fehérjék és a rendezetlen fehérjék témakörében elért eredményeket külön fejezetben, a fejezeteken belül közelítőleg időrendi sorrendben mutatom be. Ez egy csoporttámogató NK típusú projekt volt. Tehát nem minden a pályázatban résztvevő kutatók által, a pályázat terhére, végzett munkában szerepeltem olyan mértékben, ami indokolta volna a társszerzőségemet. Ezekben természetesen nem szerepelek.

Az eredmények:

### 1. Transzmembrán fehérjék

A korábbi évtizedek legnagyobb visszhangot kiváltott cikkeink, köztük a három, egyenként több, mint ezer hivatkozású eredeti közlemény, a transzmembrán fehérjék vizsgálata témakörből kerültek ki. Mostanra ezt a témát egykori diákom majd munkatársam Tusnád Gábor saját Lendület csoportjában műveli. Csoportomból való kiválása után is bent maradt ebben a projektben. Eleinte készültek közösen is cikkek, de később inkább saját diákjaival és a Lendület projekt terhére dolgozott. Ugyanakkor sok hasznos szakmai konzultációt folytatunk.

Az első ilyen cikk egy fehérje szerkezet gépi analizálását jelentő számos kereső motorral ellátott web szervert mutat be. A szerver bármilyen fehérje analizálására alkalmas, mégis itt a transzmembrán fehérjékről szóló résznél mutatom be. Nemcsak azért mert létrehozását ilyen fehérjék analizálásának igényének köszönheti. Hanem azért is mert ez az a meglehetősen bonyolult, általában sok alegységet tartalmazó fehérje típus, ami az irodalomban fellelhető korábbi nyilvános web szerverekkel nem tanulmányozható. A szerver, kutatók számára szabadon elérhető a világhálón. Mint ahogy a NAR folyóiratban 2012-ben megjelent cikkünk is. Kozma D, Simon I and Tusnád GE: *CMWeb: an interactive on-line tool for analysing residue-residue contacts and contact prediction methods*, Nucleic Acids Res 40, W329-W333, 2012.

Ugyanebben a nyolcas impakt faktorú folyóiratban jelent meg a 2013. évben az előzőhöz hasonlóan Kozma Dániellel és Tusnád Gáborral írt cikkünk egy új PDBTM adatbázisról. Ez a munka nem a nyolc évvel korábban, Kozma Dániel helyett, Dosztányi Zsuzsannával készített adatbázis korszerűsítése, hanem sok szempontból egy teljesen új adatbázis szerver. A két szerver megjelenése közötti időszakban új szerkezeti elemeket fedeztek fel. Ilyenek, mindenekelőtt a membránba csak benyúló, de azon át nem menő hurkok, vagy a membrán felszínén haladó, félig a vizes fázisba érő felületi hélixek. Ezek az új elemek egészen új leírást tettek szükségessé. Kozma D, Simon I and Tusnád GE: *PDBTM: Protein Data Bank of transmembrane proteins after 8 years*, Nucleic Acids Res 41, D524-D549, 2013.

Magyar Csaba kollégám az ugyancsak TTK-s Kardos Julianna csoportjával együttműködve készített munkájában egy transzmembrán neurotransmitter szubsztrát kötését vizsgálta molekula dinamikai szimulációval. Molekula dinamikai számításokban itt került felhasználásra első ízben a stabilizációs centrum mint az átmeneti állapotok stabilitásának jellemzője. Ennek eredménye a CURRENT DRUG DISCOVERY TECHNOLOGIES folyóiratban jelent meg 2014-ben. Simon Á, Bencsura Á, Héja L, Magyar C, Kardos J: *Sodium-Assisted Formation of Binding and Traverse Conformations of the Substrate in a Neurotransmitter Sodium Symporter Model*, CURRENT DRUG DISCOVERY TECHNOLOGIES 11:(3) pp. 227-233, 2014.

Ehhez a témakörhöz tartozik, hogy amerikai kollégákkal együttműködve vizsgáltuk a transzmembrán fehérjék természetes környezetének a membránoknak a fázis átalakulását is. Az erről Sugár IP-vel és Chong PL-lel. irt cikkünk 2013-ban a Biophysical Journalban jelent meg. Sugár IP, Simon I and Chong PL: *Series of concentration-induced phase transitions in cholesterol/phosphatidylcholine mixtures*, Biophys J 104, 2448-2455, 2013 .

Ugyancsak Sugár Istvánnal jelent meg a következő évben 2014-ben egy a többi cikkemhez nehezen köthető rendszer biológiai cikkem a Self-regulating genes-ről. Egy analitikus leírást adtunk Poisson reprezentációval a probléma „steady state” megoldásáról. Cikkünk a CEJP-ben jelent meg. Sugár IP, Simon I: *Self-regulating genes. Exact steady state solution by using Poisson representation*, CENT EUR J PHYS 12: (9) 615-627, 2014

## 2. Stabil térszerkezetű, vizes közegben oldható fehérjék

2014-ben megjelent egy a stabil térszerkezetű vizes közegben oldható fehérjék ciszteinjei kovalens állapotáról szóló cikkünk. A cikk lényege, hogy különböző korábban eltérő elméleti háttérrel kidolgozott becslő eljárások összevetésével kimutattuk, hogy ha egy adott hipotézis alapján kidolgozunk egy becslő módszert és az jól működik, az még nem bizonyítja az elvi alapul szolgáló hipotézis helyességét. Ugyanis a biológiai információk között annyi a kölcsönös függés, hogy ugyanhoz az eredményhez nagyon különböző kiindulási alapról is el lehet jutni hibátlan logikával. Cikkünk Amit , Fiser Andrással, Mészáros Bálinttal és Tüdős Évával írtunk a FEBS OPEN BIO folyóiratban jelent meg. Tüdős T, Mészáros B, Fiser A, Simon I: *A word of caution about biological inference - Revisiting cysteine covalent state predictions*, FEBS OPEN BIO 4: 310-314, 2014.

Szintén 2014-ban jelent meg egy másik stabil térszerkezetű vizes közegben oldható fehérjékről szóló cikkünk Magyar Csabával valamint a Targetex Ltd és a Chemaxon Ltd cégek munkatársaival. Tüdős és kardiovaszkuláris betegségek gyógyításában fontosnak tartott foszfodiészteráz inhibitorok azonosítására különböző 2D és 3D módszerek kombinációját használtuk fel. Irodalomból ismert 44 szerkezetileg különböző PDE inhibitor molekulából indultunk ki. A kétdimenziós és a Screen3D szoftver segítségével flexibilis illesztésből számított háromdimenziós hasonlóságok Tanimotó koefficiensei (T2D és T3D) alapján vizsgáltuk a kiindulási molekulákat. A háromdimenziós hasonlóság használható paraméternek bizonyult és a  $T3D > 0.3$  feltétel hasznosnak bizonyult első körös illetve  $T3D > 0.6$  feltétel második körös szűrések elvégzésére. Egy 2D hasonlóságon alapuló első körös szűrés által szolgáltatott

molekulák kísérletileg validált találatok alapján egy második szűrési ciklust is elvégeztünk. A második szűrés során egy kombinált „fúziós” rangsorolást alkalmaztunk, mely során a  $T2D+T3D>1.5$  feltételt alkalmaztuk. Ez a feltétel alkalmasnak bizonyult nagy 2D és kisebb 3D, illetve nagy 3D és kisebb 2D hasonlóságú molekulák azonosítására. A kísérletileg megvizsgált 35 molekulából 10 hatásosnak bizonyult. A kombinált 2D/3D módszerrel azonosított molekulákat különböző foszfodiészterázakkal szemben mutatott specificitását is vizsgáltuk. A Schrödinger Suite programcsomag Glide XP programjának a segítségével bedokkoltuk a hatásos molekulákat a következő foszfodiészteráz PDB szerkezetekbe: 3G45 (PDE4B), 3G4G (PDE4D), 2H42 (PDE5A), 3HR1 (PDE10A). Megkíséreltük a különböző PDE alosztály specificitására magyarázatot adni a fehérje-ligand komplexek szerkezeti analízise alapján. Az egyszerű dokkolási számolásokon túl a kötőhely flexibilitását, illetve a ligandhoz történő konformációs adaptációját modellezendő ún. Induced Fit Docking számításokat is végeztünk. Az így kapott ligand komplex szerkezeteken MM-GBSA kötési szabadentalpia számításokat végeztünk, melyek a docking score-nál sokkal megbízhatóbb egyezést mutatnak a kísérleti eredményekkel általában. Végül a PDE4B esetében az UCR2 domént tartalmazó és nélküli fehérje szerkezetek felhasználásával arra a következtetésre jutottunk, hogy az UCR2 domén szerepet játszik a ligandkötésben. A cikk a MOLECULES című folyóiratban. Dobi K, Hajdu I, Flachner B, Fabo G, Szaszko M, Bognar M, Magyar C, Simon I, Szisz D, Lorincz Z, Cseh S, Dorman G: Combination of 2D/3D Ligand-Based Similarity Search in Rapid Virtual Screening from Multimillion Compound Repositories. Selection and Biological Evaluation of Potential PDE4 and PDE5 Inhibitors., MOLECULES 19: (6) 7008-7039, 2014.

Magyar Csabával és a Targetex Ltd munkatársaival a 2015 évben is irtunk egy cikket. Több központi idegrendszeri betegségek gyógyításának egy ígéretes célpontja a 5-hidroxitriptamin receptor 6. Az 5HT<sub>6</sub> receptor ismert, 25 különböző kemotípusba tartozó gátlószereiből kiindulva 2D hasonlósági szűrést ( $T2D>0.65$ ) végezve létrehoztunk egy kiindulási molekulakönyvtárat. A 49 kiindulási molekulát és a 2D szűrésből kapott 11 kísérletesen igazolt molekulát felhasználva farmakofór modelleket építettünk. Az irodalomban ismert volt már egy 4 elemből álló farmakofór modell, amelynél specifikusabb modelleket akartunk építeni. Farmakofór modellek felhasználásával több millió molekulát tartalmazó adatbázisok átvizsgálásával sikeresen azonosítottunk több kísérletileg is igazolt hatékonyságú molekulát, melyek között találtunk egy eddig nem ismert kemotípusba tartozó molekulát is. A cikk a Chemical Biology & Drug Design folyóiratban jelent meg. Krisztina Dobi, Beáta Flachner, Mária Pukáncsik, Enikő Máthé, Melinda Bognár, Mária Szaszko, Csaba Magyar, István Hajdú, Zsolt Lőrincz, István Simon, Sándor Cseh, György Dormán: *Combination of pharmacophore matching, 2D similarity search and in vitro biological assays in the selection of potential 5-HT<sub>6</sub> antagonists from large commercial repository*, Chemical Biology & Drug Design 86 (4) 864-880, 2015.

Magyar Csaba a 2016-ban Szöllősi Dániellel, Erdei Áronnal, Gyimesi Gergellyel és Hegedős Tamással irt cikkében a multidrog rezisztenciát vizsgálta. Az aryl hydrocarbon receptoron (AhR), mint a multidrog rezisztencia háttérében álló széles ligand specificitással rendelkező receptorok PAS-B modénjének modelljén kollaborációban in silico vizsgálatokat végeztek. Alacsony szekvenciális hasonlóságot mutató PAS domének, egy HIF-2a és egy CLOCK szerkezet felhasználásával homológia modelleket építettek. A modelleken DMD és MD molekuladinamikai szimulációkat végeztek azok dinamikus tulajdonságait vizsgálva. Replica exchange szimulációk segítségével meghatározták a különböző templát szerkezeten alapuló

modellek olvadásgörbéjét. A multidrog kötő képességet az MD szimulációk eredményeként kapott szerkezetekbe az Autodock Vina program segítségével bedokkolva több mint egy tucat molekula segítségével vizsgálták. A különböző templatok alapján készül homológia modellek nem mutattak lényegesen eltérő viselkedést vizsgálataik során. A konformációs sokaságba történő dokkolás nem volt képes megkülönböztetni az alacsony és magas affinitással rendelkező ligandokat. Ellenőrzésképpen a fehérje flexibilitását másképpen figyelembe vevő Induced Fit Docking számolásokat is végeztek. Az eredmények ugyanazt mutatták. Ezek alapján valószínűsítették, hogy a ligand specificitásért nem csak a PAS-B domén tehető felelőssé. A ligand kötő zseb mellett az oda vezető csatorna is fontos szerepet játszhat a testidegen molekulák felismerésében. A cikk a PLOS ONE folyóiratban jelent meg. Szöllősi D, Erdei Á, Gyimesi G, Magyar C, Hegedűs T: *Access Path to the Ligand Binding Pocket May Play a Role in Xenobiotics Selection by AhR*, PLoS One 11(1):e0146066, 2016.

Két évtizede vezettük be a stabilizációs centrumok fogalmát. Azokat a nem kovalens kölcsönhatásokkal összetartott speciális klasztereket melyek elsődlegesen felelősek a natív fehérje szerkezet fenntartásáért. Ebből a témából több cikket írtunk melyekre több száz hivatkozást kaptunk, és számos más szerző szerző is foglalkozott a témával. Meglepő módon, eddig senki sem vizsgálta, hogy bár a klaszterek neve a stabilitásra utal, van-e közülük a fehérjék hő stabilitásához. Egy nagy pont mutációs információkat és termodinamikai paramétereket összegyűjtő adatbázis statisztikus elemzésével a kérdésre pozitív válasz adtunk. Cikkünk amit Magyar Csabával, Sávoly Zoltánnal és indiai közreműködőnk M Michael Gromiha-val írtunk a BBRC-ben jelent meg 2016-ban. Csaba Magyar, M. Michail Gromiha, Zoltán Sávoly, István Simon: *The role of stabilization centers in protein thermal stabilization and human diseases*, Biochem Biophys Res Com 471, 57-62, 2016.

Legutóbbi, 2017-ben az első cikkünk amit Magyar Csabával és a Targetex Ltd munkatársaival írtunk, új potenciális hatóanyagok glutaminil cikláz gátlására fragmens alapú drog tervezéssel történt keresését mutatjuk be. A glutaminil cikláz egy ígéretes célfehérje neurodegeneratív betegségek gyógyításában. Irodalomból ismert gátlószereinek felhasználásával 2D hasonlósági szűrés alapján összeállított fragmens-könyvtár kísérletileg leghatékonynak ítélt 15 elemét vizsgáltuk a Schrödinger Suite 2015 programcsomaggal. A Glide program segítségével bedokoltuk a fragmenseket a glutaminil cikláz 3si0 PDB szerkezetébe. Fragmensek esetében nem várhatjuk, hogy a pontozófüggvények gyógyszermolekulákra jellemző pontértékeket szolgáltatassanak, mivel azok nem fragmensekre, hanem drog jellegű molekulákra lettek optimalizálva. A lehető legjobb pontértékek elérése érdekében a fehérje szerkezetét is flexibilisen kezelő Induced Fit Docking számításokat végeztünk a fragmenseken. A legjobb pontszámokat kapott fehérje-fragmens komplexek szerkezeti összehasonlítása azt mutatta, hogy egy fragmens a fehérje más részéhez kötődik, mint a többi. Ezt a fragmenst egy vele komplementer felszínhez kötődő másik fragmenssel manuálisan összekötöttük és az így kapott molekulán újabb IFD számolást végeztünk. A számolás kedvezőbb kötődést jósolt, mint egyes irodalmi gátlószerek esetében. Azonban ez a molekula nem elérhető kereskedelmi forgalomban és a kémiai szintézis is nagy kihívás lett volna. Ezt a problémát megkerülendő 2D hasonlósági keresést végeztünk kereskedelmi forgalomban kapható molekula adatbázisokban. A találatokon újból IFD számításokat végeztünk. Mivel a sztenderd pontozó függvények az aktív és inaktív molekulák szeparációjára lettek optimalizálva, ezek nem használhatóak kis kötési szabadentalpia különbségek becslésére. Ezért az így kapott fehérje-ligand komplexeken MMGBSA kötési szabadentalpia számolásokat végeztünk. Ez a módszer tapasztalataink szerint sokkal jobb

egyezőést mutat a kísérleti eredményekkel. Mivel azonban nem akartuk az IFD számítások eredményét sem teljesen figyelmen kívül hagyni a molekulákat az IFD docking score és az MMGBSA számítások eredményeiből kapott konszenzus Z-score alapján rangsoroltuk. A listában legelől szereplő molekulák kísérleti validálása folyamatban van. A cikk a MOLECULAR DIVERSITY folyóiratban jelent meg. Szaszko M, Hajdú I, Flachner B, Dobi K, Magyar C, Simon I, Lőrincz Z, Kapui Z, Pázmány T, Cseh S, Dormán G: *Identification of potential glutaminyl cyclase inhibitors from lead-like libraries by in silico and in vitro fragment-based screening*, MOLECULAR DIVERSITY 21 (1) 175-186, 2017.

### 3. Rendezetlen fehérjék

A rendezetlen fehérjék működése leggyakrabban egy rendezett fehérjéhez kötődve valósul meg. A kölcsönhatás típusok számon tartására korábban kísérletes térszerkezeti információk alapján létrehoztak egy ELM adatbázist. Egy korábbi OTKA projektünk keretében mi is készítettünk egy ANCHOR nevű algoritmust illetve web szervert a kölcsönhatásban lévő rendezetlen fehérje szegmenseknek az aminosav sorrendből való becslésére. Az ELM adatbázis nem nagyon használható kötőhelyek azonosítására, mert az egyes lineáris motívumok nagyon sokszor fordulnak elő a fehérjékben és ezek közül csak kevés a tényleges kötőhely. Rámutattunk, hogy ez a hiba jelentősen csökkenthető az ELM adatbázis és az ANCHOR algoritmus együttes alkalmazásával. Mészáros Bálinttal és Dosztányi Zsuzsával írt cikkünk 2012-ben a PLOS ONE folyóiratban jelent meg. Mészáros B, Dosztányi Z and Simon I: *Disordered Binding Regions and Linear Motifs-Bridging the Gap between Two Models of Molecular Recognition*, PLoS ONE 7, e46829, 2012

2013-ban egy nagyobb nemzetközi csapat tagjaként a projekt egyik senior közreműködője, Dosztányi Zsuzsanna, részt vett egy rendezetlen fehérjék adatait tartalmazó adatbázis létrehozásában. Az adatbázisba nem csak a kísérletes munkákkal igazolt rendezetlenségű fehérjék, hanem a legjobb teljesítményű becslő módszerek eredményei is belekerültek. Ez természetesen kis mértékben csökkentette az adatbázis pontosságát de egy nagyság rendel növelte a benne található adatok számát. Az erről szóló cikk a NAR folyóiratban jelent meg. Oates ME, Romero P, Ishida T, Ghalwash M, Mizianty MJ, Xue B, Dosztányi Z, Uversky VN, Obradovic Z, Kurgan L, Dunker AK and Gough J: *D2P2: database of disordered protein predictions*, Nucleic Acids Res 41, D508-D516, 2013.

Ugyanebben az évben, szintén nemzetközi együttműködéssel, Dosztányi Zsuzsanna részt vett egy részlegesen rendezetlen mitokondriális fehérjét vizsgálatában is. Erről a munkáról a PLOS ONE folyóiratban számoltak be. Hayward DC, Dosztányi Z, Clark-Walker GD: *The N-terminal intrinsically disordered domain of Mgm101p is localized to the mitochondrial nucleoid*, PLoS One. 2013;8: e56465, 2013.

Intézetünk egy másik csoportjával együttműködve Dosztányi Zsuzsanna és Mészáros Bálint részt vett a humán proteomban található MAPK kötőmotívumok felderítésében és analizálásában. Munkájuk során kísérletes és elméleti csoportokkal együttműködve kifejlesztettek egy bioinformatikai protokollt, mely képes az ismert MAPK motívumok alapján kapott, nagy számú lehetséges MAPK-kötő fehérje közül azonosítani azokat, melyeknél a valódi kötés esélye

nagy. A protokoll számos meglévő predikciós algoritmus a feladathoz alakított, specializált változatát (többek között a már bemutatott lineáris-motívum optimált ANCHOR-t) kombinálva figyelembe veszi a fehérjék szekvenciális, szerkezeti és evolúciós tulajdonságait, majd minden egyes találathoz rendel egy pontszámot mely alapján azok sorba rendezhetőek. A legjobb találatokat kísérletes MAPK-kötési esszékben vizsgálva a protokoll és maguk a MAPK kötőmotívumok is újradefiniálhatóak. Az iteratív eljárás végén nem csak számos új, eddig azonosítatlan humán MAPK szubsztrátot lehetett azonosítani, de az eddigi elnagyolt MAPK motívum definíciókat is sikeresen újradefiniálni és pontosítani számos MAPK alosztályra. A cikk a rangos, tíz feletti impact faktorú, MOLECULAR SYSTEMS BIOLOGY folyóiratban került publikálásra 2015-ben. Zeke A, Bastys T, Alexa A, Garai Á, Mészáros B, Kirsch K, Dosztányi Z, Kalinina OV, Reményi A: *Systematic discovery of linear binding motifs targeting an ancient protein interaction surface on MAP kinases*, Mol Syst Biol 11 (11) 837, 2015.

Dosztányi Zsuzsával, Mészáros Bálinttal és intézetünk Reményi Attila vezette csoportjával a rákhoz köthető fehérje funkcionális helyek mutációs profilok alapján történő azonosításával foglalkoztunk. A projektünk alapvető célja az volt, hogy kifejlesszünk egy statisztikai analízisen alapuló módszert, amely képes nagy mennyiségű tumor szekvenálási adatból meghatározni, hogy mely humán fehérjék mely részei érintettek egyes ráktípusok kialakulásában. Ehhez a COSMIC adatbázisban lévő pontmutációs, illetve in-frame inzerció és deléciós adatokból indultunk ki. Bár a rákban szerepet játszó fehérjék ezen kívül más módon is tudnak sérülni, ezen adatok adják az ismert szekvenálási eredmények túlnyomó részét. Ezen adatokból kiindulva az általunk kifejlesztett, iSiMPRe nevű módszer képes meghatározni azokat a fehérje-szegmenseket, melyekbe statisztikailag szignifikánsan több mutáció esik, mint amit egy egyenletes mutációs rátából feltételeznénk. Ezt a háttéreloszlást meghatározó mutációs rátát az egyes fehérjékre külön-külön számoljuk, így kompenzálható az egyes genomi régiók közötti mutációs ráta-különbség. Módszerünk újdonsága hogy a pontmutációkat, illetve a lokális inzerciókat és deléciókat képes egységes statisztikai rendszerben kezelni, mellyel lehetőség nyílik olyan régiók azonosítására is, amelyek egyébként a klasszikusan figyelembe vett pontmutációk által csak kevésbé érintettek. A rákban fontos fehérjerégiók azonosítása mellett a projekt egyéb biológiailag szignifikáns eredményekkel is szolgált. A kapott régiók nagyrészt lefedik a már ismert humán rákgéneket, amely azt mutatja hogy az algoritmusunk megközelítése teljesen általános és a tipikus rákgének egy nagyon specifikus csoportját (azon gének halmazát, amely tipikusan genomi átrendeződések során sérül) leszámítva minden tumorigenezisben érintett gén azonosítására alkalmas. Ez azt mutatja, hogy bár a pontmutációk és az inzerciók/deléciók csak egy szűk részét jelentik a lehetséges és előforduló tumorigenikus genomi változások fajtáinak, ezen genetikai változások szignifikáns feldúsulása egy általános kísérőjelensége a rákos elváltozásoknak. Kiemelt fehérje példákon keresztül, melyek funkciójára megfelelő mennyiségű ismeret áll rendelkezésre, megmutattuk, hogy a pontmutációk szigetszerű szignifikáns feldúsulása hogyan tudja biológiai hatásában kiváltani a génvesztést, az amplifikációt, illetve bizonyos korlátozott esetekben akár a génfüziót is. A cikk a Biology Direct folyóiratban jelent meg 2016-ban. Mészáros B, Zeke A, Reményi A, Simon I, Dosztányi Z: *Systematic analysis of somatic mutations driving cancer: Uncovering functional protein regions in disease development*, Biology Direct 11 (1) Paper 23, 2016.

Az egész pályázat legnagyobb volumenű eleme, amire a teljes 26.5 FTE kapacitás ötödét fordítottuk, a rendezetlen fehérjék által kialakított komplexek vizsgálata és a komplexre vonatkozó információk összegyűjtése egy korszerű kereső motorokkal ellátott adatbázisba volt,

olyanba ami a világhálón bárki számára elérhető. A munkát Fichó Erzsébettel és Mészáros Bálinttal végeztük. Mivel ennek az anyag még nem jelent meg és a web szerver sem érhető még el külső felhasználó számára, ezt kissé részletesebben ismertetem.

A rendezetlen fehérjékkel kapcsolatos eddigi kutatásaink középpontjában a rendezetlen fehérjék globuláris partnerekhez való kölcsönhatásai álltak (rendezetlen kötőhelyek és lineáris motívumok kapcsolatának vizsgálata, illetve MAPK lineáris motívum kölcsönhatások). Az ezen a területen kapott eredményeinket és a közben szerzett tapasztalatokat elkezdjük átültetni egy hasonló terület vizsgálatára is. Ennek keretében olyan fehérje komplexek vizsgálatát kezdtük meg, amelyekben nem csak egy, hanem az összes résztvevő partner rendezetlen a monomer állapotában. A lineáris motívum-domén kölcsönhatások esetében leírt kapcsolt rendeződés és kötődés itt is lejátszódik a komplex kialakulásánál, azonban az eddigi rendszerekkel ellentétben ez a jelenség nem csak egy résztvevő fehérjére korlátozódik, hanem a kölcsönhatás összekapcsolódik a komplexet kialakító összes fehérje egyidejűleg lejátszódó foldingjával. Ezen rendszereknek lényegében két állapota létezik: egy, melyben a monomerek rendezetlenek, illetve egy, melyben a kölcsönhatás már kialakult és az összes partner rendezett (ellentétben a 'klasszikus', globuláris fehérjék által kialakított kölcsönhatásoknál, ahol van egy köztes állapot is, ahol a kölcsönhatás még nem alakult ki, de a monomerek már rendeződtek). Ezek alapján az ilyen rendszereket gyakran kétállapotú komplexeknek is nevezik (szemben a 'klasszikus' három állapotú rendszerekkel).

Bár a kétállapotú komplexek vizsgálata egy logikus lépés a rendezett-globuláris komplexek vizsgálata után, ezen rendszerekből lényegesen kevesebb ismert. Az ezen komplexekre vonatkozó információnak nincs egy jól definiált gyűjtőhelye (mint pl. a fehérje szerkezeteknek a PDB, vagy a motívum-domén kölcsönhatásoknak az ELM adatbázis). Emiatt a kétállapotú komplexekre vonatkozó kutatásaink első lépését egy dedikált adatbázis összeállítása jelentette. A lehető legnagyobb számú kétállapotú komplex azonosításához különböző adatbázisok integrációját valósítottuk meg. Mivel az összegyűjtött komplexeket később szerkezeti vizsgálatokra is fel akarjuk használni, csak olyan kölcsönhatásokat tekintettünk, amelyeknek van megoldott 3D szerkezete. Emiatt a kiindulási adatbázist a PDB jelentette. A PDB-ben lévő nagyjából 100 000 szerkezet közül választottuk ki azokat, amelyek megfelelnek annak a kritériumnak, hogy egyrészt csak fehérjeláncok szerepelnek bennük (a fehérje-DNS és fehérje-RNS kölcsönhatások alapvetően mások, illetve néhány extrém esetet leszámítva a DNS/RNS molekulák nem a kölcsönható fehérje hatására vesznek fel egy meghatározott szerkezetet), másrészt az összes bennük lévő fehérje-régió a kölcsönhatás nélkül, monomer formában rendezetlen. Az első kritérium teljesítése triviális, a második teljesítéséhez pedig egyéb adatbázisokat integráltunk a kiválasztási protokollunkba. A rendezetlenségre vonatkozó kísérletes információk két legnagyobb adatbázisa jelenleg a DisProt és az IDEAL. Az ezekben leírt rendezetlen fehérje-régiókat áttérképeztük UniProt fehérjékre, majd ezekre térképeztük rá a PDB-ben leírt fehérje-régiókat is. Ahol a kísérletesen megállapított rendezetlenség és a PDB-ben szerkezettel rendelkező fehérje-régió jelentős mértékben (legalább 70%-ban) átfedett, ott az adott PDB-ben szereplő fehérje láncot rendezetlennek tekintettük. Az olyan szerkezeteket tartottuk meg, ahol az összes kölcsönható lánc megfelel ugyanilyen kritériumoknak. Mivel a DisProtban és IDEAL-ban csak korlátozott számú fehérjére van rendezetlenséget alátámasztó információ, a keresés kiterjesztéséhez homológián alapuló annotáció-transzfert alkalmaztuk. Ez lényegében azt jelenti, hogy ha A fehérjének van komplex szerkezete, de nincs rá vonatkozó rendezetlenségi annotáció, akkor végignéztük az összes B, C, D... közeli homológiát is. Ha valamelyikre találtunk rendezetlenségi annotációt és a két fehérje közötti hasonlóság elért egy megfelelő mértéket, akkor

úgy tekintettük, hogy A fehérjére is van (áttételes) bizonyítékunk a rendezetlenségre. Nyilvánvaló, hogy a módszer elfogadhatósága nagyon erősen függ attól, hogy mit tekintünk „megfelelő” mértékű hasonlóságnak. Ismert, hogy ha két rendezett fehérje legalább 30%-os szekvenciális egyezést mutat, akkor 90% a valószínűsége, hogy ugyanazt a foldot veszik fel. Ennek mintájára valószínűsíthető, hogy a rendezetlen állapot átviteléhez szükséges hasonlósági mérték sem különbözhet ettől radikálisan, de azért hogy minimalizáljuk az esetleges fals annotációkat, a két fehérje között mi legalább 90%-os szekvencia azonosságot követeltünk meg. Emellett a rendezetlenségi tulajdonság homológok közötti átviteléhez elégséges feltételként fogadtuk el azt is, ha a két homológ fehérje kölcsönható részeikben ugyanazt a konzervált szekvencia-részt találtuk meg a Pfam algoritmussal.

Az adatbázis-integráción alapuló automatizált keresések eredményeit manuális irodalmi keresésekkel bővítettük ki. Nagyszámú olyan komplex ismert, amelynek kétállapotú volta kísérletesen igazolt, azonban az egyes monomer egységeire vonatkozó rendezetlenségi információ – egyelőre – hiányzik a rendezetlenségi adatbázisokból. A kulcsszavakra támaszkodó irodalmi kereséseink eredményei ezt legalább részben képesek ellensúlyozni.

Az így kapott kétállapotú komplexeket alávetettük egy redundancia-szűrésnek, melyben egy klaszterbe soroltuk az összes olyan komplexet, amelyek azonos számú láncot tartalmaznak és a láncok között páronként legalább 90%-os szekvencia azonosság van. Minden ilyen klaszterből egy reprezentáns elemet választottunk ki, általában a hozzájuk tartozó szerkezetek minősége alapján. Így összesen 158 klasztert kaptunk, melyekre kézzel ellenőriztük a rendezetlenségre vonatkozó annotációkat, hogy a fals pozitív találatok számát minél inkább lecsorítsuk. Végeredményben kaptunk egy nagyon magas minőségű, kontrollált redundanciával rendelkező, annotált adatbázist. A további szerkezeti és funkcionális analízisek megalapozásához az adatbázisban lévő elemeket további osztályokba és alosztályokba soroltuk be. Ez az osztályozás segíthet a kívánt redundancia-szintek beállításához is a későbbi vizsgálatok során. Az adatbázist Mutual Folding Induced by Binding néven egy publikus web-szerver formájában tervezzük közreadni, amely az MTA TTK Enzimológiai Intézet szerverein kerül beüzemelésre.

A kapcsolódóan rendeződő fehérjekomplexek azonosítása után az eredmények publikálásához elengedhetetlen volt egy webszerver, a Mutual Folding Induced by Binding (MFIB) szerver létrehozása. Az elkészült adatbázisnak egyúttal több – általunk támasztott – kritériumnak is meg kellett felelnie: a benne lévő adatoknak többféle szempontrendszer szerint kereshetőnek és áttekinthetőnek kell lenni, egyszersmind letölthetőnek a kutatók és érdeklődők számára.

Az általunk azonosított fehérjeszerkezetek vonatkozó biológiai és szerkezeti annotálása is megtörtént az adatbázis összeállítása során. A létrehozandó szerver egyik legnagyobb kihívása az volt, hogy mindezen információkat egységes formába kellett rendezni, mind az online felület, mind pedig a letölthető fájlok elkészítése során. Minden általunk listázott fehérjeszerkezet egyedi, az adatbázisban tárolt adatokkal összefüggésben álló specifikus azonosítót is kapott.

A webszerver szerveroldala főként PHP scriptnyelven íródott. Ezt egészítik ki a kliensoldalon használt HTML és saját CSS-keretrendszert a különböző open-source JavaScript-könyvtárak. Utóbbiak főként az oldal kezelhetőségének leegyszerűsítését, áttekinthetőségét, és különböző dinamikai illetve animációs elemek használatát tették lehetővé. Az adatbázis adatai egy integrált



MySQL adatbáziskezelő segítségével tárolódnak, majd kérdeződnek le, és jelennek meg az oldalon, a különböző aloldalak funkcióinak megfelelően.

A webszerver elkészült, nyilvánosan elérhető mindenki számára. Az ott közölt adatok többféle formátumban letölthetőek, de egy webes böngésző segítségével is olvashatóak és kereshetőek, kulcsszavas illetve általunk definiált szerkezeti és egyéb csoportosítási szempontok alapján is. A fehérjekomplexek azonosítása során több adatbázis adatait is felhasználtuk (UniProt, UniRef90, DisProt, IDEAL, Pfam), ezekre kereszthivatkozások is találhatóak a webszerver egyes, adott fehérjeszerkezethez tartozó oldalain. Az elkészült webszerver az adatbázis esetleges bővítése esetén automatikusan frissíthető, minimális emberi felügyelettel. További adatok publikálására a webszerver alkalmas, új funkciókkal könnyedén kiegészíthető a jelenlegi konstrukció.

Végül, de nem utolsó sorban, rendezetlen fehérjékről megjelent két nagy terjedelmű összefoglalónk könyv fejezetek formájában. Az elsőt ami 2014-ben a Springer Verlag gondozásában jelent meg. Mészáros B, Dosztányi Z, Magyar C and Simon I: *Bioinformatical Approaches to Unstructured/Disordered Proteins and Their Interactions*, Computational Methods to Study the Structure and Dynamics of Biomolecules and Biomolecular Processes; (Ed. Liwo A) Springer Verlag: pp 525-556, 2014.

A másikat ami 2015-ben szintén a Springer Verlag gondozásában a Methods in Molecular Biology 1261. kötetében jelent meg Dosztányi Zsuzsával és Marco Puntóval írtuk. Punta M, Simon I, Dosztányi Z: *Prediction and analysis of intrinsically disordered proteins*, In: Owens RJ (szerk.) (szerk.) Structural Proteomics (Methods in Molecular Biology, Volume 1261): High-Throughput Methods. New York: Humana Press Inc., Springer, 2015. pp. 35-59., 2015.

Ami a munkáink visszhangját illeti, nagy siker számomra, hogy 2014 után 2016 novemberében újra, ez alkalommal egyedüli magyarként, megkaptam az igen megtisztelő „highly cited researcher” címet az utolsó bő évtizedben megjelent cikkekre kapott hivatkozások alapján., Természetesen figyelembe kell venni, hogy ennek az évtizednek az eleje még nem ennek a projektnek az idejére esett.

