



Bírálat

Dosztányi Zsuzsanna

“A rendezetlen fehérjék bioinformatikai vizsgálatai”

c. MTA doktori értekezéséről

Dosztányi Zsuzsanna MTA doktori értekezését rövid értekezés formájában adta be magyar nyelven. Az MTA doktori eljárás szabályzata ugyan lehetővé teszi, hogy a részletes, gyakran 150 oldalt is meghaladó terjedelmű értekezés helyett ezt a formát válassza a jelölt, de megjegyzi, hogy „a rövid értekezés kivételes forma ... [csak] általánosan elismert, kiemelkedő, nagy hatású tudományos teljesítmény” esetén engedélyezheti a tudományos osztály. Ezt az értékelést a Biológiai Tudományok Osztálya elvégezte, és a döntést csak megerősíteni tudom. Az *MTA Biológiai Tudományok Osztályának követelményei az MTA doktora cím elnyeréséhez* 300 független hivatkozást határoz meg minimum követelménynek a hagyományos, és 400-at a tézises formához. Dosztányi Zsuzsanna idézettsége ezt több mint 20-szorosan meghaladja, az MTMT 8705 független, a Google Scholar pedig 13200 hivatkozást sorol fel. Ez utóbbin Dosztányi Zsuzsannának 10 olyan cikkét találtam, ami önmagában meghaladja a 300-as minimum hivatkozási számot. Mindezek alapján megállapítható, hogy **a kérelmező tudományos teljesítménye kellően kiemelkedő és nagy hatású, hogy a doktori eljárás rövid értekezés alapján is lefolytatható legyen.**

A mintegy 30 oldalas, jól követhető magyar nyelvű összefoglaló egy rövid bevezetéssel indul, majd azt egy mintegy 10 oldalas irodalmi áttekintés követi. Végig vezet a szakterület nagyjából 20 éves történetén, ami egybevág Dosztányi Zsuzsanna eddigi tudományos karrierjének időszakával. Vázolja a folyamatot ahogy a klasszikus szerkezet-funkció paradigma fokozatosan átadta a helyét egy összetettebb képnek, melyben kiemelt szerepet játszanak a doktori mű fókuszában álló rendezetlen fehérjék. Elmagyarazza, hogy a rendezetlenség nem rendetlenséget jelent, hanem egyfajta flexibilitást, aminek egyre több funkcióját ismerjük fel, a rendezetlen szegmensek által kialakított kölcsönhatások kinetikai és termodinamikai tulajdonságai széles skálán mozognak. Mivel a szerző eredményei bioinformatikai jellegűek, pár oldalban vázolja az általános szerkezet becslő módszereket, a fehérje szerkezeti adatbázisokat, majd kitér a kimondottan rendezetlen fehérjékre vonatkozó módszerekre és a rendezetlen kötőhelyek előrejelzésére.

A disszertáció célkitűzéseit 4 pontban sorolja fel, nevezetesen:

1. Rendezetlenség predikciós eljárás kifejlesztése egy újfajta energia becslő eljárás alapján
2. A rendezetlen fehérjék kölcsönhatásainak jellemzése az interaktóm és a komplexek szintjén

3. Predikációs módszer kifejlesztése a rendezetlen kötőhelyek felismerésére
4. A rák és a rendezetlenség kapcsolatának vizsgálata mutációs adatok alapján

Ezt követően a felhasznált módszerek, programozási környezet és kulcs-adatbázisok rövid leírása következik. Az eredményeket 4 fő- és összesen 10 alpontban foglalja össze mintegy 13 oldal terjedelemben. Az alpontok megfeleltethetők a tézispontoknak. Majd egy tömör Konklúzió, végül a 10 tézis felsorolása következik.

Az első téziscsoport a következő három pontot tartalmazza:

1. *Kifejlesztettem egy újszerű energiabecslő eljárást, ami képes megbecsülni egy adott fehérjeszekvencia kompakt állapotához tartozó párkölcsönhatási energiát. Az energiabecslés alapján megmutattam, hogy a rendezetlen fehérjék becsült energiája kedvezőtlenebb, mint a rendezett fehérjéké, ami rámutatott a rendezetlenség fizikai alapjára is.*
2. *Az energiabecslő eljárás felhasználásával kidolgoztam az IUPred módszert, amely pozíció specifikusan jósl rendezetlenséget az aminosav szekvenciából.*
3. *Az IUPred módszert elérhetővé tettem webszerveren keresztül*

Az IUPred módszer széleskörű használata és a mintegy 3500 hivatkozás önmagáért beszél, nem igényel további részletes értékelést. Sőt hatása még nagyobb, hiszen a módszer beépült számos egyéb web-szerverbe, például PFAM, ELM és PDB adatbázisokba és részét képezi több kombinált rendezetlenség jósló módszernek is. A sikerben fontos szerepe lehet, hogy a módszer egy frappáns ötleten, - a konkrét szerkezettől független statisztikai kvázi-energia mátrixon - alapul, amely szinte minden esetben robusztus eredményeket szolgáltat.

Kérdés: A 10. oldalon megjegyzi, hogy „a statisztikai potenciálok ... általában felülmúlják a fizika alapú energiafüggvényeket”. Mi ennek a magyarázata? Mi hiányzik a fizikai képből? A statisztikai potenciáloknak lehetőséget adnak-e valamilyen fizikai interpretációra?

A következő két tézis az IUPred által feltárt rendezetlenséget próbálja általánosabb kontextusba helyezni:

4. *Rámutattam, hogy a fehérje kölcsönhatási hálózatok központi csomópontjaiban nagyobb arányban fordulnak elő rendezetlen fehérjék.*
5. *Összeállítottam egy, a rendezetlen fehérjék komplexeinek szerkezetét tartalmazó adatszettet és összevettem ezen komplexek tulajdonságait globuláris fehérjék komplexeinek jellemzőivel. Ez alapján feltártam a rendezetlen fehérje komplexek alapvető szerkezeti sajátosságait.*

A 4. tézis új. plauzibilis magyarázatot ad arra, hogy a fehérje kölcsönhatási hálózatokban mely fehérjék jelennek meg sok kapcsolattal rendelkező

csomópontként. Hagyományosan ezt gén-duplikáció és a preferenciális kapcsolódás eredményeként értelmezik. A tézisponthoz kapcsolódó publikációban azonban megmutatták, hogy a nagy fokszámú csomópontok fehérjéinek esetében magasabb volt a rendezetlen aminosavak száma és aránya is. A megállapítás összhangban van az 5. tézispont állításával, miszerint a rendezetlen fehérje jobban tud idomulni számos partner fehérjéhez.

Kérdés: A fehérje hálózatokban gyakran valóban sok homológ fehérje jelenik meg egy-egy csomópont körül, mely a „hagyományos” gén-duplikációs modellt támasztja alá. Mennyiben módosíthatja az eredményeket, ha erre korrigálunk?

Két tézis egy újabb bioinformatikai módszert mutat be:

- 6. Az energiabecslő eljárás alapján kifejlesztettem egy új módszert, amellyel az aminosav szekvenciából jósolhatóak azok a rendezetlen szegmensek, amelyek más fehérjékhez kötődnek és eközben rendezett szerkezetet alakítanak ki.*
- 7. Az energiabecslésen alapuló szekvencia predikciós módszereimet elérhetővé tettem modernizált formában webszerverként és letölthető programcsomagként is.*

Az ANCHOR módszer valamivel később született, mint az IUPred, de széleskörű használata minden kétséget kizáróan alátámasztja ezeknek a téziseknek is a létjogosultságát. Sőt sok szempontból akár fontosabbnak is tekinthető, hogy nem csak szegmentálni lehet rendezett és rendezetlen részekre az aminosav láncokat, hanem becsülni lehet a funkciót is, meghatározni, hogy hol jelennek meg kötőhelyek.

Fontosnak tartom megjegyezni, hogy mind a 3. mind pedig a 7. tézis egy szoftver közzétételéről szól. Az évek során számos esetben volt részem olyan vitában, ahol megkérdőjelezték, hogy egy szoftver vagy adatbázis létrehozása tekinthető-e önálló tudományos eredménynek. Messzemenően nem értek egyet az ilyen nézetekkel. A tudományos szoftverek részei a tudomány felépítményének, ugyanúgy, mint a kísérletek vagy az elméleti modellek. Ezek is csak szakértő kutatók által hozhatók létre és gyakran jóval nagyobb hatásúak, mint egy-egy klasszikusan tudományosnak elismert eredmény.

Kérdés: Említi a disszertációban, hogy a rendezetlen kötőhelyekhez jóval nehezebb gyógyszermolekulákat tervezni, mint a jól ismert szerkezetekhez. Milyen módszereket lehet egyáltalán alkalmazni? Hozhat-e valamilyen módszer áttörést a közeljövőben?

Végül az utolsó három tézis a rendezetlen fehérjéknek a rákos folyamatokban betöltött szerepéről szól:

- 8. Mutációk vizsgálata alapján megmutattam, hogy a rendezetlen fehérje szegmenseken gyakoribbak a neutrális polimorfizmusok, de ritkébbak bennük a rákos mutációk a globuláris részekhez képest.*

9. Azonosítottam olyan rendezetlen fehérje szegmenseket, amelyek nagyszámú rákos mutációt tartalmaztak, és ez alapján várhatóan aktív szerepet játszanak a rák kialakulásában. Elemeztem ezek jellemzőit funkcionális és rendszer szinten.
10. Evolúciós vizsgálatok alapján rámutattam arra, hogy bár a rendezetlen fehérjék általában kevésbé konzerváltak, a rákban mutálódó példák nagymértékű evolúciós konzerváltságot mutattak.

A dolgozatot szépen megkoronázza, hogy a széleskörűen használt módszerek fejlesztésén túl próbál közvetlen, akár a gyógyászatban is felhasználható összefüggéseket feltárni és evolúciós távlatban is értelmezni a rendezetlen fehérjék elterjedését.

Kérdés: A vizsgálatokban felhasznált COSMIC adatbázis szomatikus mutációkat tartalmaz. Vizsgálták-e, hogy azok az örökletes mutációk melyek növelik a rák kockázatát milyen kapcsolatban vannak a rendezetlen régiókkal?

A 155 elemet tartalmazó, a szakterület teljes történetét és számos aspektusát lefedő referencia lista után került befűzésre a doktori művet megalapozó 11 eredeti angol nyelvű cikk. Dosztányi Zsuzsanna a cikkek mindegyikének első vagy utolsó szerzője. A cikkek 3-5 szerzősek, és egy kivételével csak magyar kutatók szerepelnek a szerzőlistán. Nemzetközi kollaboráció nélkül szinte lehetetlen nivós kutatást folytatni, és Dosztányi Zsuzsanna más tanulmányainál gyakoriak is a széleskörű együttműködések. Ugyanakkor a téziseket megalapozó válogatás még inkább alátámasztja, hogy a bemutatott eredmény saját teljesítmény, nem egy nagy kutatócsoport vagy egy rangsorvezető kiemelt intézmény alapozta meg a nemzetközi sikert.

Általános kérdés: Az utóbbi években az AlphaFold és hasonló nagy gépi tanulás alapú módszerek jelentős átalakulást hoztak a szerkezetbecslő módszerek terén. Milyen változásokat hozhat ez a rendezetlen fehérjék vonatkozásában?

A kérdésekre adott válaszoktól függetlenül, **a felsorolt tézisek mindegyikét új tudományos eredménynek fogadom el.**

A benyújtott doktori művet és az abban foglalt eredményeket **nyilvános vitára alkalmasnak tartom.**

Budapest, 2023. szeptember 12.



Csabai István