

MTA doktori értekezés tézisei

A rendezetlen fehérjék bioinformatikai vizsgálatai

Dosztányi Zsuzsanna



ELTE TTK Biológiai Intézet Biokémia Tanszék

Budapest, 2023

1 BEVEZETÉS

A fehérjék mind szerkezeti mind funkcionális tulajdonságait tekintve hihetetlenül változatosak. Ugyanakkor ennek az elképeztő változatosságnak az alapját egy viszonylag egyszerű kémiai felépítés adja: a fehérjék 20 különböző aminosav egymáshoz kapcsolódásával kialakuló láncmolekulák, melyek egyediségét az aminosavak sorrendje, az aminosav szekvencia adja. A fehérjekutatás egyik legalapvetőbb kérdése, hogy hogyan határozza meg az aminosav sorrend a fehérjék térszerkezetét és funkcióját. Ennek a kérdéskörnek az alapjait több mint ötven éve fektették le, többek között Christian Anfinsen, a ribonukleáz refolding vizsgálatával, Cyrus Levinthal, a fehérje feltekeredés, a folding, alapvető paradoxonának megfogalmazásával, illetve John Kendrew az első fehérje, a mioglobinszerkezetének meghatározásával [1]. Bár az azóta eltelt időszakban jelentős előrelépések történtek a szekvencia - szerkezet - és funkció összefüggésének megértésében, mind a mai napig újabb és újabb dolgokat tanulunk a fehérjék sokszínűségéről.

Az egyik alapvetően új irány a rendezetlen fehérjékhez kapcsolódik. A sokáig általánosan elfogadott nézet az volt, hogy a fehérjék megfelelő működéséhez elengedhetetlen, hogy egy jól definiált szerkezettel rendelkezzenek. Ezt a paradigmát írta át a rendezetlen fehérjék funkcionális fontosságának felismerése. Bár ez a jelenség a kezdetekben jelentős vitákat váltott ki [2], a 2000-es évek kezdetétől megindult a rendezetlen fehérjék szisztematikus vizsgálata. Ezekben - a rendezetlen fehérjék kísérletes vizsgálatának nehézségei miatt - döntő szerep jutott a bioinformatikai megközelítéseknek.

Doktori értekezésemben a rendezetlen fehérjék bioinformatikai vizsgálata során elért eredményeimet foglaltam össze. Közel két évtizedes tevékenységem során többek között új bioinformatikai eszközöket fejlesztettem ki, melyekkel felismerhetők a rendezetlen régiók illetve azok kötőhelyei az aminosav szekvenciából. A saját és mások által kifejlesztett módszerek alkalmazásával elemeztem a rendezetlen fehérjék kölcsönhatási tulajdonságait és betegségben betöltött szerepüket. Összességében, vizsgálataim alapvetően járultak hozzá ezen új fehérje osztály jobb megismeréséhez.

2 IRODALMI ÁTTEKINTÉS

2.1 A rendezetlen fehérjék és funkcióik

Az eredendően rendezetlen fehérjék (intrinsically disordered protein/IDP) nem képesek önmagukban egy jól-definiált szerkezet kialakítására, ugyanakkor számos biológiai folyamatban kulcsfontosságú szerepet játszanak [1–3]. A rendezetlen fehérjéket egy konformációs sokasággal lehet jellemezni, aminek jellegzetes molekuláris tulajdonságai, mint például a megnövekedett molekuláris méret, a denaturált állapotba való átmenet hiánya, a periódikus másodlagos szerkezetek hiánya vagy a konformációs heterogenitás számos kísérleti módszerrel megragadható [4]. A szerkezeti sokaság részletes tulajdonságainak leírásához azonban többféle módszer kombinációjára van szükség [5]. A kísérletesen igazolt rendezetlen fehérjék összegyűjtésére jött létre a DisProt adatbázis [6,7]. Az adatbázis folyamatosan bővül, jelenleg (2022 végén) 2469 fehérje 5568 régiójára tartalmaz információt. Ezek annotációjában több kutatócsoport, többek között a mi csoportunk is részt vett [8].

Az eredendően rendezetlen fehérjék alapvető fontosságának felismerése drasztikusan alakította át a fehérjék szerkezete és működése közötti összefüggésről alkotott elképzelésünket [4]. Az ismert példák alapján a rendezetlen fehérjék funkcionális szempontból több kategóriába sorolhatók [9,10]. Funkciójuk eredhet közvetlenül a rendezetlen állapotból, például a globuláris doméneket összekötő linker régiók esetén. Másik tipikus funkciójuk a molekuláris felismerésben való részvétel, melynek során specifikus kötőpartnerekhez, például egy másik fehérjéhez, RNS-hez vagy DNS-molekulához kötődnek. A kötődés révén befolyásolhatják a partner molekulák működését, elősegíthetik fehérjék vagy RNS molekulák feltekeredését, részt vehetnek kismolekulák szállításában és raktározásában vagy irányíthatják nagyobb komplexek összeszerelődését. Egy viszonylag újonnan felismert funkciójuk, hogy koncentráció függő módon képesek fázisszeparációt indukálni, ami kulcsszerepet játszik membrán nélküli organellek kialakulásában [7,11].

A rendezetlenség számos funkcionális előnyt biztosít a kölcsönhatások szempontjából [12]. Partner molekulával való kölcsönhatás indukálhatja egy jól-definiált szerkezet kialakítását, az úgynevezett csatolt feltekeredés és kötődés (coupled folding and binding) során [12]. A nem-kötött és kötött állapot szerkezeti tulajdonságainak finomhangolása lehetőséget biztosít a kötődés specifitásának és affinitásának precíz kontrolljára [12]. A rendezetlen régiók nagyfokú plaszticitásuk révén különböző molekula felszínéhez tudnak alkalmazkodni, ami lehetővé teszi számukra a több partnerhez való kötődést [10,13]. A rendezetlen fehérjék gyakran tartalmaznak kompakt, néhány aminosavból álló rövid lineáris motívumokat, melyek specifikus doménnel való kölcsönhatást közvetítenek [14]. Gyakoriak bennük a különböző poszttranszlációs módosítások is, melyek megkönnyítik működésük szabályozását a sejtekben [15]. A lineáris motívumok és PTM-ek kombinációja révén molekuláris kapcsolók összetett hálózata alakulhat ki [16]. Ezeken keresztül, a rendezetlen fehérjék képesek a sejten belüli jeleket komplex módon integrálni és ezáltal központi szerepet játszanak jelátviteli és szabályozó folyamatokban [10,13].

2.2 A rendezetlen fehérjék predikciója

A rendezett és rendezetlen fehérjék közötti alapvető szerkezeti különbségek tükröződnek eltérő aminosav összetételükben is. Általánosságban, a rendezetlen fehérjékben alacsonyabb a hidrofób aminosavak aránya és gyakoribbak bennük a töltött és poláros aminosavak [2,17]. Ezen specifikus szekvenciális tulajdonságok arra utalnak, hogy nemcsak a globuláris fehérjék szerkezete, hanem a rendezetlenség is kódolva van a szekvenciában, ami alapján predikciós módszerek készíthetők. Az elmúlt körülbelül 20 év alatt több mint 100 rendezetlenség becselő módszert fejlesztettek ki, amelyekre sokban hatottak a globuláris fehérjék szerkezet predikciójára korábban kifejlesztett módszerek. Az alkalmazott megközelítések széles skálát fednek le, mögöttes elv, komplexitás, szükséges futási idő és pontosság tekintetében is [18–20]. A legegyszerűbb megközelítések egyszerű aminosav tulajdonsági skálákon alapulnak [21–23]. A módszerek másik jelentős csoportja gépi tanuláson alapul, egyre inkább támaszkodva mélytanulós eljárásokra [4,24]. Ezek a módszerek azonban kevésbé átlátható módon, alapvetően fekete dobozként működnek, és érzékenyebbek a tanító halmazban esetlegesen előforduló hibákra.

A rendezetlenség predikciós módszerek hatékonyságának összehasonlító értékelésére és a terület fejlődésének nyomon követésére több vizsgálatot is végeztek [19,25,26]. A módszerek teljesítménye függ a kiértékeléshez használt adatbázistól és a használt metrikától is. A 2021-ben közzétett Critical Assessment of Intrinsic Protein Disorder (CAID) kísérlet volt az eddigi legnagyobb kezdeményezés a rendezetlenség predikció értékelésére [28]. Ebben 32 módszer előrejelzési pontosságát és futási idejét értékelték, amihez a DisProt adatbázisban újonnan annotált fehérjéit használták.

A DisProt adatbázisban rendelkezésre álló adatok azt mutatják, hogy a funkcionálisan annotált rendezetlen régiók leggyakoribb funkciója más makromolekulákhoz való kötődés, ezen belül is elsősorban fehérje-fehérje kölcsönhatás kialakítása [12]. Azonban az ismert rendezetlen kötőrégiók száma csak lassan növekszik, ezért nagy szükség van olyan módszerekre, melyek a szekvenciából képesek felismerni a rendezetlen kötőhelyeket. A rendezetlen régiók kötőhelyeinek predikciójára szolgáló eljárások kiindulópontját azok a térszerkezeti komplexek adják, amelyek kísérletesen igazolt rendezetlen fehérje szegmenseket tartalmaznak, melyek csak a kötődés határára vesznek fel rendezett, tradicionális módszerekkel is tanulmányozható szerkezetet. Ezeket a régiókat szokás MoRF-nak (molecular recognition feature) is hívni [29]. A rendezetlen kötőrégiók szekvenciából történő felismerésére kifejlesztett módszerek vagy a már létező rendezetlenség jósoló módszerek valamilyen speciális jellemzőjét aknázzák ki, vagy gépi tanulós eljárást használnak a kötőrégiók felismerésére és ezek megkülönböztetésére mind a globuláris fehérjéktől, mind az ismert kötődésben nem résztvevő rendezetlen szegmensektől [30–32]. A CAID kiértékelte a rendezetlen kötőrégiók jöslására szolgáló módszerek teljesítményét is, a DisProt kötőrégiókra vonatkozó annotációit használva tesztalmazként [28]. Ez az adathalmaz kevesebb mint 250 fehérje régiót tartalmazott, azonban még ezek az annotációk is gyakran hiányosak illetve pontatlanok voltak. A kategóriában mindössze tíz módszer szerepelt, és a probléma nehézsége a módszerek viszonylag alacsony teljesítményében is

tükröződött. Bár egyre több módszer születik a rendezetlen régiók funkcionális szerepének jóslására is [33], ez a terület kevésbé kiforrott a rendezetlenség predikciókhoz képest is.

2.3 A rendezetlen fehérjék általános jellemzése

A rendezetlen fehérjék jellemzésében, az egyedi példák vizsgálata mellett, nagy szerepet játszottak a bioinformatikai módszerek, köztük az IUPred és az ANCHOR, melyek használata révén feltárhatjuk a rendezetlen fehérjék általános jellemzőit.

A rendezetlenség koncepciójának általánosan elfogadottá válásához nagyban hozzájárult, hogy a különböző genomszekvenciák által kódolt fehérjék vizsgálata alapján széles körben elterjedt jelenségről van szó. Eszerint a rendezetlen fehérjék az életfa minden ágán jelen vannak és előfordulásuk erősen korrelál az organizmusok komplexitásával [34,35]. A DISOPRED2 módszer alapján a hosszú (>30 aminosav hosszú) rendezetlen szegmensek az archea fehérjék 2,0%-ában, a bakteriális fehérjék 4,2%-ában és az eukarióta fehérjék 33,0%-ában fordul elő [36,37]. Ezt a trendet több más vizsgálat is megerősítette [38–40]. A rendezetlenséghez kapcsolódó specifikus funkciók vizsgálta gén ontológiai annotációk és a SwissProt adatbázis kulcsszavain keresztül megerősítette a rendezetlenség fontos szerepét jelátviteli és szabályozó folyamatokban, és megmutatták a rendezetlen fehérjékhez kapcsolódó fő funkcionális kategóriákat [36,37,41]. A rendezetlen fehérjék funkcionális sokszínűsége evolúciós tulajdonságaikban is tükröződik. Szerkezeti megkötések hiányában a rendezetlen fehérjék általában nagyobb evolúciós variabilitást mutatnak [42], ugyanakkor a funkcionálisan fontos pozíciók - ettől jelentősen eltérő módon - erősen konzerváltak is lehetnek [43].

A rendezetlen fehérjék funkcionális jelentősége alapján feltételezhető, hogy hibás működésük komoly biológiai következményekkel jár [13]. Ezt a kapcsolatot erősítik egyedi példák is, például az α -szinuklein neurodegeneratív betegségekben, a p53 a ráktípusok széles skáláján, vagy a CFTR fehérje a cisztás fibrózis betegségben betöltött szerepe [44]. Általánosabb szinten is a rendezetlen fehérjék nagyobb arányát figyelték meg a betegségekhez - illetve specifikusan rákhoz - kapcsolódó fehérjék között [45]. A SwissProt adatbázis adatai alapján a rákhoz társítható humán fehérjék 79%-a volt rendezetlen, szemben az összes eukarióta fehérje 47%-ával [46]. A rák és a rendezetlenség közötti kapcsolatot erősítette a különböző ráktípusokban gyakori kromoszóma átrendeződések vizsgálata, melyek gyakran érintettek rendezetlen régiókat [47]. A rendezetlen fehérjék nagyobb arányát figyelték meg a dózisérzékeny fehérjék között [48], összhangban azzal, hogy a rendezetlen fehérjék - kölcsönhatási promiskuitásuk és fáziszeperációban játszott szerepük miatt - sokkal érzékenyebbek lehetnek a fehérjeszint változására [49]. Mindezen vizsgálatok ellenére, néhány kivételtől eltekintve, még most sem ismerjük pontosan a rendezetlen fehérjék szerepét a rák kialakulásában. Ennek a kapcsolatnak a feltérképezését új szintre emelheti a rák genom projektek révén egyre nagyobb mértékben rendelkezésre álló mutációs adatok vizsgálata [50].

3 CÉLKITŰZÉSEK

Kutatásom során alapvető célkitűzésem a rendezetlenség vizsgálatával kapcsolatosan új bioinformatikai eszközök kifejlesztése volt, illetve ezek alkalmazása a rendezetlen fehérjék funkcionális és betegséggel kapcsolatos tulajdonságainak jobb megértése érdekében. Ezeket, a közlemények időrendi megjelenését nagyjából követve, 4 főbb célkitűzésben foglaltam össze:

1. Rendezetlenség predikciós eljárás kifejlesztése egy újfajta energia becslő eljárás alapján
 - a. Az energiabecslő eljárás kifejlesztése és alkalmazása rendezett és rendezetlen fehérje szekvenciák megkülönböztetésére
 - b. Az energiabecslés alapján a rendezetlenség predikcióra szolgáló IUPred módszer kifejlesztése
 - c. Az IUPred módszer elérhetővé tétele webszerveren keresztül
2. A rendezetlen fehérjék kölcsönhatásainak jellemzése az interaktóm és a komplexek szintjén
 - a. A rendezetlenség szerepének vizsgálata a fehérje kölcsönhatási hálózatok központi fehérjeiben
 - b. A rendezetlen fehérje kölcsönhatások molekuláris alapelveinek feltárása a rendezetlen és rendezett fehérjék közötti komplexek szerkezetének elemzése révén
3. Predikciós módszer kifejlesztése a rendezetlen kötőhelyek felismerésére
 - a. A rendezetlen kötőhelyek predikciójára szolgáló ANCHOR módszer kifejlesztése
 - b. Az energiabecslésen alapuló IUPred és ANCHOR módszerek elérésének biztosítása modern webszerver és programcsomag formájában
4. A rák és a rendezetlenség kapcsolatának vizsgálata mutációs adatok alapján
 - a. Különböző mutációk eloszlásának vizsgálata rendezett és rendezetlen fehérje szegmensek között
 - b. A rákban gyakran mutálódó rendezetlen fehérje szegmensek azonosítása és ezek funkcionális és rendszer-szintű tulajdonságainak elemzése
 - c. A rákban mutálódó rendezetlen fehérje szegmensek evolúció eredetének vizsgálata

4 EREDMÉNYEK

Az itt bemutatott eredmények majdnem húsz év munkássága során elért legfontosabb eredményeit foglalják össze a rendezetlen fehérjék bioinformatikai vizsgálatának területén. Ezek alapjául 11 eredeti közlemény szolgált, melyek mindegyikében meghatározó (első vagy utolsó) szerző voltam. Az egyes alfejezetek mögötti számok jelölik a saját közlemények megfelelő hivatkozását.

4.1 Rendezetlenség predikció

4.1.1 Az energia becselő eljárás (1)

Legfontosabb eredményem egy energiabecslő eljárás kifejlesztése volt, ami a szerkezet predikciós eljárások széles körében használatos statisztikai potenciál koncepción alapult. Ezek az ismert térszerkezetekből származtatott energia-jellegű mennyiségek jól le tudják írni az aminosavak közötti kölcsönhatási preferenciák általános jellemzőit. Azonban korábban a számítások mindig egy meghatározott konformációra vonatkoztak. Az energiabecslő eljárás lehetővé teszi, hogy túl tudjunk lépni ezen a megkötésen, és közvetlenül a szekvenciából meghatározzuk az adott szekvenciához tartozó párkölcsönhatási energiát. Az egyes pozíciók energiájának becsléséhez csak az adott aminosav típusát és a fehérje aminosav összetételét vettem figyelembe. A módszer egyszerűsége ellenére az ismert szerkezettel rendelkező fehérjék esetén a számolt és a becsült energiák jól korreláltak egymással. Igazoltam, hogy fehérje szinten a rendezett fehérjék becsült energiája általában kedvező (negatív) volt, ehhez képest a rendezetlen fehérjékre az esetek döntő többségében magasabb energia volt jellemző. Ezek az eredmények alátámasztották a feltételezést, hogy a rendezett és rendezetlen fehérjék megkülönböztethetőek a becsült párkölcsönhatási energiájuk alapján, és rámutatott a fehérje rendezetlenség fizikai alapjára.

4.1.2 Az IUPred módszer (1)

Az energiabecslő eljárás alapján kidolgoztam az IUPred rendezetlenség predikciós módszert. A gyakorlati használhatóságához fontos volt, hogy a predikció pozíció specifikus legyen. Ennek érdekében kismértékben módosítottam az eljárást, és az energiabecslés során nem a teljes szekvenciát, hanem minden pozíció esetén annak meghatározott szekvenciális környezetét vettem figyelembe. Megmutattam, hogy a módszer független teszthalmazon a rendezetlen fehérje pozíciók 76%-át jósolta helyesen rendezetlennek. Az energiabecslés során használt paraméterek pusztán globuláris fehérje szerkezetekből lettek származtatva, rendezetlen fehérjékre vonatkozó információk felhasználása nélkül, ami a megközelítés robusztusságának a kulcsa. Az IUPred módszer mind a mai napig az egyik legnépszerűbb rendezetlenség predikciós módszer. Ennek oka abban kereshető, hogy gyorsan, viszonylag

megbízható predikciókat képes generálni. A módszer beépült számos egyéb adatbázisba és részét képezi több konszenzus rendezetlenség jósló módszernek is [51]. Az elmúlt években számos új módszert fejlesztettek ki a rendezetlenség jóslására, melyek közül sok már megbízhatóbb, mint az IUPred. Azonban a legtöbb esetben az elért kismértékű javulást nem kompenzálja a több nagyságrenddel hosszabb futási időt [28,52].

4.1.3 Az IUPred webservert (2)

Az IUPred módszer sikeréhez az is nagymértékben hozzájárult, hogy könnyen elérhetővé tettem webes felületeken és programcsomag formájában is. Az IUPred webservert eredeti verziója a <http://iupred.enzim.hu> oldalon volt elérhető. Magát a programot C programnyelven írtam, a webservert pedig PHP nyelven készítettem el. A webservert bemenete egy fehérje szekvencia (FASTA formátumban, vagy csak a sima szekvencia). A kimenet minden egyes pozícióban az adott aminosavhoz rendel egy értéket, ami jellemzi annak rendezetlenségre való hajlamát. Ez az érték 0 (teljesen rendezett) és 1 (teljesen rendezetlen) között lehet, általánosságban 0,5-ös érték felett tekintünk egy aminosavat rendezetlennek. Az alapbeállítás egyszerű szöveges (text) kimenet, de a grafikus megjelenítés is választható. Ezt a beadott szekvenciára a szerver azonnal legenerálja a JpGraph software (JpGraph, 2005) program segítségével. Elkészítettem a programcsomag letölthető verzióját is.

4.2 A rendezetlen fehérjék kölcsönhatásainak jellemzése

4.2.1 A rendezetlenség szerepe a fehérje kölcsönhatási hálózatok központi fehérjeiben (3)

Korábbi megfigyelések arra utaltak, hogy a rendezetlen fehérjék egyik fő funkciója fehérje-fehérje kölcsönhatások kialakítása. Felmerült a kérdés, hogy van-e összefüggés egy fehérje kölcsönhatásainak száma és szekvenciális tulajdonságai, mint például fehérje rendezetlenség, vagy alacsony komplexitású régiók megléte között. Munkánkban négy faj, a *Saccharomyces cerevisiae*, *Drosophila melanogaster*, *Caenorhabditis elegans* és *Homo sapiens* publikusan elérhető fehérje interakciós hálózatára vonatkozó adatok alapján vizsgáltuk szisztematikusan ezt a kérdést. A négy faj kölcsönhatási hálózatának elemzésével megmutattuk, hogy a központi (hub) fehérjék esetében magasabb volt a rendezetlen aminosavak száma és aránya is. Valamivel kisebb mértékben, de hasonló tendencia volt megfigyelhető az alacsony komplexitású, illetve ismétlődő szekvencia elemek esetében is, illetve a hub fehérjék általában hosszabbak is voltak. Eredményeim azt mutatták, hogy a kölcsönhatási hálózatokban lévő központi fehérjék egyik meghatározó jellemzője a fehérje rendezetlenség. Ez arra utal, hogy a kölcsönhatási hálózatok evolúciója során előnyt jelentett az, hogy egyes fehérjék növelni tudták képességüket más fehérjékkel való kölcsönhatásra és

ezáltal a hálózat szerveződéséhez kapcsolódó funkciókra specializálódnak. Ennek egyik lehetséges mechanizmusa a fehérje rendezetlenségre épül.

4.2.2 A rendezetlen fehérje kölcsönhatások molekuláris alapelvei (4)

A fehérje kölcsönhatások molekuláris alapelveinek feltárása alapvető fontosságú a fehérjék biológiai funkciójának megértéséhez, melybe betekintést kaphatunk az általuk kialakított komplexek térszerkezetének vizsgálatával. Több korábbi munka is elemezte a különböző komplexek részletes tulajdonságait [53–55], és ezek között már megjelentek rendezetlen fehérjék komplexei is [29,56]. Azonban munkánk volt az első, ami direkt módon kísérletesen igazolt rendezetlen fehérjéken alapult. Ehhez a DisProt adatbázis [57] alapján 39 olyan kísérletesen igazolt rendezetlen fehérjét gyűjtöttünk össze, aminek ismert volt a szerkezete valamilyen másik fehérjéhez kötődve. Ezek interfészét hasonlítottuk össze 72 globuláris fehérje komplex kölcsönhatási felszínével. Bár a két fajta komplex esetén a kölcsönhatási felszín mérete nem tért el jelentősen, rendezetlen fehérjék esetén ez jóval nagyobb részét képezte a felszínnek. Alapvető eltéréseket találtunk komplexek két osztálya között a kölcsönhatást kialakító aminosavak jellegében, a kialakult kontaktusok mértékében és a részt vevő szegmensek számában. Ezek az eltérések arra utalnak, hogy molekuláris felismerés alapelvei eltérőek a rendezetlen fehérjék által kialakított kölcsönhatások során.

4.3 A rendezetlen kötőhelyek predikciója

4.3.1 Az ANCHOR módszer (5, 6)

A rendezetlen fehérjék komplexeinek speciális fiziko-kémiai és szegmentációs tulajdonságai előrevetítették, hogy a rendezetlen fehérjék esetén lehetséges a kölcsönhatásban résztvevő aminosavak predikciója a szekvenciából. Megközelítésünk ezeket a speciális biofizikai tulajdonságokat próbálja megfogni a IUPred metódusban is használt energiabecslő eljárás alapján. A rendezetlen kötőrégiók egyik fő jellemzője, hogy azok egy alapvetően rendezetlen részen belül találhatóak. Ez a tulajdonság közvetlenül jellemezhető az IUPred módszer alapján. A másik fő jellemzőjük, hogy bár saját szekvenciális környezetükkel nem, de speciális globuláris fehérjéhez kötődve kedvező kölcsönhatást tudnak kialakítani. Ennek modellezéséhez meghatározhatjuk azt a becsült energiát, amit az adott aminosav a közvetlen szekvenciális környezetével tudna létrehozni, illetve azt, amit egy átlagosan globuláris fehérjéhez kötődve tudna elérni. Az általunk kifejlesztett módszer, amit ANCHOR-nak neveztünk el, ezen tényezők kombinációja alapján jósol rendezetlen kötőhelyeket a szekvenciából. A módszer továbbfejlesztett változata, az ANCHOR2, kissé eltérő architektúrára épül és paraméterek egy részének optimalizálásához kibővült adatbázist használt. Összességében, módszerünk a globuláris fehérjéken mért 5% fals predikciós ráta mellett, 64%-os hatékonysággal tudta jósolni a rendezetlen kötőhelyek aminosavait, de szegmens szinten a rendezetlen kötőhelyek 72%-át helyesen felismerte. Ugyanakkor a

DisProt adatbázisban annotált rendezetlen fehérje szegmensek kevesebb, mint 50%-át jósolta kötőhelynek, tehát az általános rendezetlen részekről is meg tudta különböztetni a kötőhelyeket. A rendezetlenség predikciós módszerekhez képest szerényebb eredmények ellenére, az ANCHOR és ANCHOR2 módszerek a rendezetlen fehérjék funkcionális helyeinek jellemzésének alapvető eszközeivé váltak [13,58]. Ezt igazolta a CAID értékelés eredménye is, ahol a független adatokon az ANCHOR módszer bizonyult a legjobbnak a rendezetlen kötőrégiók azonosításában [28].

4.3.2 Az energiabecslő eljárásokon alapuló módszerek webservere (6, 7, 8)

2017-ben a szervert átköltöztettük az ELTÉ-re, és elkészítettük a webserverek új verzióját, az IUPred2A-t (6). Ebben egyesítettük az energiabecslésen alapuló módszereink - az IUPred és az ANCHOR - elérését. Az ANCHOR szerver eredeti verziójához, ami az <http://anchor.enzim.hu> oldalon volt elérhető, alapvetően az IUPred szerver esetében használt programokat szabtuk át, mind a bemenet, mind a kimenet kezelését tekintve (7). Az új verzióban az egyik fő változás, hogy a korábbi C, illetve Perl programnyelvek helyett áttértünk a PYTHON program nyelvre. Míg az IUPred programban csak egy kisebb hibajavítást tettünk, az ANCHOR programot teljesen átdolgoztuk, és az új ANCHOR2 verziót tettük elérhetővé. A webservert egy további opcióval bővítettük, ami lehetővé tette redox-érzékeny rendezetlen régiók szekvencia alapú predikcióját is (6). A fejlesztések révén az IUPred2A egy modern megjelenésű, dinamikus webservert lett, ami támogatja az összes jelenleg használatos web-böngészőt. A predikciós kimenet értelmezésének elősegítése érdekében egyéb információkat is feltüntettünk. Ezek körét tovább bővítettük az IUPred szerver legújabb verziójában egy olyan új megjelenítővel, amivel az adott szekvenciához tartozó szekvencia illesztés és a predikciós profil összekapcsoltan tanulmányozható eukarióta modell organizmusokban (8). A módszereket elérhetővé tettük programcsomag formájában is. A szerver legújabb verziója <https://iupred.elte.hu> illetve <https://iupred3.elte.hu> oldalakon érhető el.

4.4 A rendezetlenség és a rák kapcsolata

4.4.1 A rendezetlenség biológiai kockázata (9)

Az egyedi példák és általános elemzések megalapozták a rák és a rendezetlenség közötti kapcsolatot. Ezek alapján felmerült, hogy a rendezetlenség egyfajta biológiai kockázatot jelent [15738986, 18573080]. A rendezetlenség jósló módszerek megjelenése és a különböző szekvencia variációs adatok robbanásszerű növekedése lehetővé tette ennek a kérdésnek a részletesebb vizsgálatát. Munkánkban elsők között elemeztük, hogy hogyan befolyásolják a fehérjék szerkezeti tulajdonságai a mutációk eloszlását. Eredményeink azt mutatták, hogy a neutrális polimorfizmusok gyakoribbak voltak a rendezetlen részekben, ezzel

szemben a rákos mutációk sokkal inkább a rendezett részekre estek. Megfigyeltük azt is, hogy a rákkal kapcsolatba hozott fehérjék átlagosan hosszabbak, több bennük a rendezetlenség, több kölcsönhatásban vesznek részt és speciális funkciók kapcsolódnak hozzájuk. Azonban ezek a tulajdonságok egymással is korrelálnak. A különböző tényezők között kölcsönös információt számoltunk, és megmutattuk, hogy a rendezetlenség és a rákos mutációk közötti kapcsolat indirekt, a funkción keresztül jön létre. A leggyakrabban mutálódó példák vizsgálata alapján, a p53 vagy a PTEN viselkedése összhangban van az általános trenddel. Bár a rendezetlen szegmensek alapvető funkcionális szerepet játszanak ezen fehérjék működésében, a rákos mutációk jellemzően a rendezett részre esnek. Ugyanakkor találhatóak olyan példák is, ahol a mutációk döntő része rendezetlen részre esik. Ilyen például a β -catenin, vagy az APC fehérje. Hasonló példák azonosítása révén további betekintést nyerhetünk a rendezetlenség és a rák kapcsolatáról.

4.4.2 Rákban gyakran mutálódó rendezetlen szegmensek (10)

Korábbi elemzésünket kiterjesztve, a célunk olyan példák azonosítása és elemzése volt, ahol a rákos mutációk közvetlenül rendezetlen részeket érintenek. Munkánk során a COSMIC adatbázist használtuk kiindulásként [59] és az ebből összegyűjtött mutációkat rávetítettük a fehérje szekvenciákra. Az iSiMPre módszer alapján [60] olyan régiókat azonosítottuk, ahol több minta összesítése esetén a mutációk feldúsultak, ami rákban betöltött fontos, irányító szerepükre utal. Összesen 145 fehérjében 225 szignifikánsan mutálódott régiót találtunk magas szignifikancia szinttel ($10e-6$). Ezek közül 47 esett rendezetlen régióra, ami kevesebb, mint amit a rendezetlenség előfordulása alapján vártunk (kb. 30%), de így is jelentős halmaz. A rákban gyakran mutálódó rendezetlen szegmensek legnagyobb csoportja rövid lineáris motívumok általi kölcsönhatások kialakításában vett részt, de egyéb típusú funkcionális modulok is érintettek voltak, például autoregulációs helyek, linker régiók vagy DNS és RNS kötő részek. Megfigyeltük, hogy a transzkripció és transzláció szabályozása mellett a degradációs rendszer is jellemzően rendezetlen fehérje részeken keresztül mutálódik, például degron motívumokat érintve [61]. A mutálódott rendezetlen szegmensek fontosságát támasztotta alá a kölcsönhatási és jelátviteli hálózatokban betöltött központi szerepük. Ennek révén perturbációjuk érintheti az ismert rákos ismertetőjegyek (cancer hallmark) [62] mindegyikét. A legtöbb betegminta tartalmazott mind rendezett mind rendezetlen részre eső mutációkat. Azonban azon minták esetében, ahol rendezetlen részre eső mutációk dominálnak, jóval kisebb az esélye hogy létezik a kezelésre jelenleg elérhető gyógyszermolekula. Ez jól mutatja, hogy nagy igény lenne olyan új megközelítésre, melyek révén a rendezetlen fehérjék is targetálhatóak lennének.

4.4.3 A rákban mutálódó rendezetlen régiók evolúciós eredete (11)

A rendezetlen fehérjék sajátos szerkezeti és funkcionális tulajdonságokkal rendelkeznek, és ez evolúciós tulajdonságaikban is tükröződik. A több szerkezeti megkötöttséggel rendelkező rendezett fehérjékhez képest általánosságban a rendezetlen fehérjék viszonylag friss evolúciós találmányok és szekvenciájuk nagyobb variabilitást mutat [42]. Azonban a bennük található funkcionálisan fontos régiók erősen konzerváltak is lehetnek. Ezért felmerült a kérdés, hogy evolúciós eredetüket tekintve, mi a jellemző a rákban szignifikánsan mutálódott rendezetlen fehérje szegmensekre. Vizsgálatunkhoz a filoztratigráfias eljárást használtuk, amely visszavezeti az egyes gének eredetét makorevolúciós átmenetekhez [63]. Meglepő módon, a rákban mutálódott rendezetlen fehérje régiók ősi eredetűek voltak, a régiók többsége a többsejtű állatok megjelenéig visszavezethető volt. Néhány példánál még ennél is ősibb evolúciós eredetet találtunk. A rendezetlen régiók funkcionalitásának fontosságát támasztja alá, hogy megjelenésüket követően gyorsan rögzültek, és a fehérje család tagjainak további duplikációja során megőrződtek és újabb szabályozó elemekkel gazdagodhatnak. Elemzésünk révén világosabb képet kaptunk a fehérjék kulcsfontosságú szabályozó elemeinek kialakulásáról, ráirányítva a figyelmet a moduláris szerveződés figyelembe vételének fontosságára a fehérjék evolúciós eredetének vizsgálata során.

5 KÖSZÖNETNYILVÁNÍTÁS

Mindenekelőtt köszönettel tartozom Simon István akadémikusnak, korábbi csoportvezetőmnek, aki már egyetemi éveim, majd PhD munkám során is témavezetőm volt az Enzimológiai Intézetben és akinek szakmai és emberi támogatására mindig számíthattam.

Szeretném megköszönni a Simon csoport tagjainak, elsősorban Fiser Andrásnak, Tusnady Gábornak, Fuxreiter Mónikának, Magyar Csabának a sok segítséget és az inspiráló légkört. Az Enzimológiai Intézet többi kutatójáról sem szeretnék megfeledkezni. Külön szeretném kiemelni Tompa Pétert, akinek a rendezetlen fehérje kutatás is, és személy szerint én is nagyon sokat köszönhetek.

Hálás vagyok Nyitray Lászlónak, aki támogatta a Lendület pályázatomat, amellyel elindíthattam önálló kutatócsoportomat az ELTE Biokémiai Tanszékén. Köszönet illeti Kovács Mihályt és a tanszék valamennyi munkatársát, hogy befogadtak és mindenben támogattak.

Köszönettel tartozom korábbi és jelenlegi diákjaimnak, elsősorban Mészáros Bálintnak, Pajkos Mátyásnak és Erdős Gábornak, a közös munkálkodásért és együtt gondolkodásért.

Szeretnék köszönetet mondani minden szerzőtársamnak itthon és külföldön, akik iránymutatásukkal és támogatásukkal hozzájárultak kutatóvá válásomhoz, valamint részt vettek a dolgozatban is bemutatott eredmények elérésében.

Végül, de nem utolsósorban megkülönböztetett hála illeti Szüleimet, akik mindenben támogattak, és páromat, Markot, akinek bátorítása és türelme nélkül nem jutottam volna el ideig.

6 HIVATKOZÁSOK

1. Wright PE, Dyson HJ. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol.* 1999;293: 321–331.
2. Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, et al. Intrinsically disordered protein. *J Mol Graph Model.* 2001;19: 26–59.
3. Tompa P. Intrinsically unstructured proteins. *Trends Biochem Sci.* 2002;27: 527–533.
4. Dyson HJ, Wright PE. Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol.* 2005;6: 197–208.
5. Marsh JA, Forman-Kay JD. Ensemble modeling of protein disordered states: experimental restraint contributions and validation. *Proteins.* 2012;80: 556–572.
6. Vucetic S, Obradovic Z, Vacic V, Radivojac P, Peng K, Iakoucheva LM, et al. DisProt: a database of protein disorder. *Bioinformatics.* 2005;21: 137–140.
7. Piovesan D, Tabaro F, Mičetić I, Necci M, Quaglia F, Oldfield CJ, et al. DisProt 7.0: a major update of the database of disordered proteins. *Nucleic Acids Res.* 2017;45: D219–D227.
8. Quaglia F, Mészáros B, Salladini E, Hatos A, Pancsa R, Chemes LB, et al. DisProt in 2022: improved quality and accessibility of protein intrinsic disorder annotation. *Nucleic Acids Res.* 2022;50: D480–D487.
9. Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, Obradović Z. Intrinsic disorder and protein function. *Biochemistry.* 2002;41: 6573–6582.
10. Tompa P. The interplay between structure and function in intrinsically unstructured proteins. *FEBS Lett.* 2005;579: 3346–3354.
11. Alberti S, Gladfelter A, Mittag T. Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell.* 2019;176: 419–434.
12. Berlow RB, Dyson HJ, Wright PE. Functional advantages of dynamic protein disorder. *FEBS Lett.* 2015;589: 2433–2440.
13. Wright PE, Dyson HJ. Intrinsically disordered proteins in cellular signalling and regulation. *Nat Rev Mol Cell Biol.* 2015;16: 18–29.
14. Van Roey K, Uyar B, Weatheritt RJ, Dinkel H, Seiler M, Budd A, et al. Short linear motifs: ubiquitous and functionally diverse protein interaction modules directing cell regulation. *Chem Rev.* 2014;114: 6733–6778.
15. Iakoucheva LM, Radivojac P, Brown CJ, O'Connor TR, Sikes JG, Obradovic Z, et al. The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res.* 2004;32: 1037–1049.
16. Van Roey K, Gibson TJ, Davey NE. Motif switches: decision-making in cell regulation. *Curr Opin Struct Biol.* 2012;22: 378–385.
17. Uversky VN. Natively unfolded proteins: a point where biology waits for physics. *Protein Sci.* 2002;11: 739–756.
18. He B, Wang K, Liu Y, Xue B, Uversky VN, Dunker AK. Predicting intrinsic disorder in proteins: an overview. *Cell Res.* 2009;19: 929–949.
19. Necci M, Piovesan D, Dosztányi Z, Tompa P, Tosatto SCE. A comprehensive assessment of long intrinsic protein disorder from the DisProt database. *Bioinformatics.* 2018;34: 445–452.
20. Zhao B, Kurgan L. Surveying over 100 predictors of intrinsic disorder in proteins. *Expert Rev Proteomics.* 2021;18: 1019–1029.
21. Linding R, Russell RB, Neduva V, Gibson TJ. GlobPlot: Exploring protein sequences for globularity and disorder. *Nucleic Acids Res.* 2003;31: 3701–3708.
22. Galzitskaya OV, Garbuzynskiy SO, Lobanov MY. FoldUnfold: web server for the prediction of disordered regions in protein chain. *Bioinformatics.* 2006;22: 2948–2949.
23. Campen A, Williams RM, Brown CJ, Meng J, Uversky VN, Dunker AK. TOP-IDP-scale: a new amino acid scale measuring propensity for intrinsic disorder. *Protein Pept Lett.* 2008;15: 956–963.
24. Zhao B, Kurgan L. Deep learning in prediction of intrinsic disorder in proteins. *Comput Struct Biotechnol J.* 2022;20: 1286–1294.

25. Liu Y, Wang X, Liu B. A comprehensive review and comparison of existing computational methods for intrinsically disordered protein and region prediction. *Brief Bioinform.* 2019;20: 330–346.
26. Nielsen JT, Mulder FAA. Quality and bias of protein disorder predictors. *Sci Rep.* 2019;9: 5137.
27. Monastyrskyy B, Kryshchak A, Moulton J, Tramontano A, Fidelis K. Assessment of protein disorder region predictions in CASP10. *Proteins.* 2014;82 Suppl 2: 127–137.
28. Necci M, Piovesan D, CAID Predictors, DisProt Curators, Tosatto SCE. Critical assessment of protein intrinsic disorder prediction. *Nat Methods.* 2021;18: 472–481.
29. Mohan A, Oldfield CJ, Radivojac P, Vacic V, Cortese MS, Dunker AK, et al. Analysis of molecular recognition features (MoRFs). *J Mol Biol.* 2006;362: 1043–1059.
30. Oldfield CJ, Cheng Y, Cortese MS, Romero P, Uversky VN, Dunker AK. Coupled folding and binding with alpha-helix-forming molecular recognition elements. *Biochemistry.* 2005;44: 12454–12470.
31. Cheng Y, Oldfield CJ, Meng J, Romero P, Uversky VN, Dunker AK. Mining alpha-helix-forming molecular recognition features with cross species sequence alignments. *Biochemistry.* 2007;46: 13468–13477.
32. Malhis N, Gsponer J. Computational identification of MoRFs in protein sequences. *Bioinformatics.* 2015;31: 1738–1744.
33. Kurgan L. Resources for computational prediction of intrinsic disorder in proteins. *Methods.* 2022;204: 132–141.
34. Schad E, Tompa P, Hegyi H. The relationship between proteome size, structural disorder and organism complexity. *Genome Biol.* 2011;12: R120.
35. Peng Z, Yan J, Fan X, Mizianty MJ, Xue B, Wang K, et al. Exceptionally abundant exceptions: comprehensive characterization of intrinsic disorder in all domains of life. *Cell Mol Life Sci.* 2015;72: 137–151.
36. Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J Mol Biol.* 2004;337: 635–645.
37. Xie H, Vucetic S, Iakoucheva LM, Oldfield CJ, Dunker AK, Uversky VN, et al. Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions. *J Proteome Res.* 2007;6: 1882–1898.
38. Tompa P, Dosztanyi Z, Simon I. Prevalent structural disorder in *E. coli* and *S. cerevisiae* proteomes. *J Proteome Res.* 2006;5: 1996–2000.
39. Mészáros B, Simon I, Dosztányi Z. Prediction of protein binding regions in disordered proteins. *PLoS Comput Biol.* 2009;5: e1000376.
40. Kastano K, Erdős G, Mier P, Alanis-Lobato G, Promponas VJ, Dosztányi Z, et al. Evolutionary Study of Disorder in Protein Sequences. *Biomolecules.* 2020;10. doi:10.3390/biom10101413
41. Bondos SE, Dunker AK, Uversky VN. On the roles of intrinsically disordered proteins and regions in cell communication and signaling. *Cell Commun Signal.* 2021;19: 88.
42. Brown CJ, Johnson AK, Dunker AK, Daughdrill GW. Evolution and disorder. *Curr Opin Struct Biol.* 2011;21: 441–446.
43. Bellay J, Han S, Michaut M, Kim T, Costanzo M, Andrews BJ, et al. Bringing order to protein disorder through comparative genomics and genetic interactions. *Genome Biol.* 2011;12: R14.
44. Tompa P. Intrinsically disordered proteins: a 10-year recap. *Trends Biochem Sci.* 2012;37: 509–516.
45. Uversky VN, Oldfield CJ, Dunker AK. Intrinsically disordered proteins in human diseases: introducing the D2 concept. *Annu Rev Biophys.* 2008;37: 215–246.
46. Iakoucheva LM, Brown CJ, Lawson JD, Obradović Z, Dunker AK. Intrinsic disorder in cell-signaling and cancer-associated proteins. *J Mol Biol.* 2002;323: 573–584.
47. Hegyi H, Buday L, Tompa P. Intrinsic structural disorder confers cellular viability on oncogenic fusion proteins. *PLoS Comput Biol.* 2009;5: e1000552.
48. Vavouri T, Semple JJ, Garcia-Verdugo R, Lehner B. Intrinsic protein disorder and interaction promiscuity are widely associated with dosage sensitivity. *Cell.* 2009;138: 198–208.
49. Babu MM. The contribution of intrinsically disordered regions to protein function, cellular complexity, and human disease. *Biochem Soc Trans.* 2016;44: 1185–1200.
50. Zou H, Pan T, Gao Y, Chen R, Li S, Guo J, et al. Pan-cancer assessment of mutational landscape

- in intrinsically disordered hotspots reveals potential driver genes. *Nucleic Acids Res.* 2022;50: e49.
51. Dosztányi Z. Prediction of protein disorder based on IUPred. *Protein Sci.* 2018;27: 331–340.
 52. Lang B, Babu MM. A community effort to bring structure to disorder. *Nature methods.* 2021. pp. 454–455.
 53. Jones S, Thornton JM. Principles of protein-protein interactions. *Proc Natl Acad Sci U S A.* 1996;93: 13–20.
 54. Lo Conte L, Chothia C, Janin J. The atomic structure of protein-protein recognition sites. *J Mol Biol.* 1999;285: 2177–2198.
 55. Nooren IMA, Thornton JM. Diversity of protein-protein interactions. *EMBO J.* 2003;22: 3486–3492.
 56. Gunasekaran K, Tsai C-J, Nussinov R. Analysis of ordered and disordered protein complexes reveals structural features discriminating between stable and unstable monomers. *J Mol Biol.* 2004;341: 1327–1341.
 57. Sickmeier M, Hamilton JA, LeGall T, Vacic V, Cortese MS, Tantos A, et al. DisProt: the Database of Disordered Proteins. *Nucleic Acids Res.* 2007;35: D786–93.
 58. Buljan M, Chalancon G, Eustermann S, Wagner GP, Fuxreiter M, Bateman A, et al. Tissue-specific splicing of disordered segments that embed binding motifs rewires protein interaction networks. *Mol Cell.* 2012;46: 871–883.
 59. Forbes SA, Beare D, Bindal N, Bamford S, Ward S, Cole CG, et al. COSMIC: High-Resolution Cancer Genetics Using the Catalogue of Somatic Mutations in Cancer. *Curr Protoc Hum Genet.* 2016;91: 10.11.1–10.11.37.
 60. Mészáros B, Zeke A, Reményi A, Simon I, Dosztányi Z. Systematic analysis of somatic mutations driving cancer: uncovering functional protein regions in disease development. *Biol Direct.* 2016;11: 23.
 61. Mészáros B, Kumar M, Gibson TJ, Uyar B, Dosztányi Z. Degrons in cancer. *Sci Signal.* 2017;10. doi:10.1126/scisignal.aak9982
 62. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell.* 2011;144: 646–674.
 63. Domazet-Loso T, Tautz D. Phylostratigraphic tracking of cancer genes suggests a link to the emergence of multicellularity in metazoa. *BMC Biol.* 2010;8: 66.

7 A TÉZISEK RÖVID ÖSSZEFOGLALÁSA

1. Kifejlesztettem egy újszerű energiabecslő eljárást, ami képes megbecsülni egy adott fehérjeszekvencia kompakt állapotához tartozó párkölcönhatási 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 jósol rendezetlenséget az aminosav szekvenciából.
3. Az IUPred módszert elérhetővé tettem webszerveren keresztül.
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.
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
8. Mutációk vizsgálata alapján megmutattam, hogy a rendezetlen fehérje szegmenseken gyakoribbak a neutrális polimorfizmusok, de ritkábbak 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.

8 AZ ÉRTEKEZÉS ALAPJÁUL SZOLGÁLÓ KÖZLEMÉNYEK

- 1: **Dosztányi Z**, Csizmók V, Tompa P, Simon I. The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. *J Mol Biol.* 2005;347:827-839.
- 2: **Dosztányi Z**, Csizmok V, Tompa P, Simon I. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics.* 2005;21:3433-3434.
- 3: **Dosztányi Z**, Chen J, Dunker AK, Simon I, Tompa P. Disorder and sequence repeats in hub proteins and their implications for network evolution. *J Proteome Res.* 2006;5:2985-2995.
- 4: Mészáros B, Tompa P, Simon I, **Dosztányi Z**. Molecular principles of the interactions of disordered proteins. *J Mol Biol.* 2007;372:549-561.
- 5: Mészáros B, Simon I, **Dosztányi Z**. Prediction of protein binding regions in disordered proteins. *PLoS Comput Biol.* 2009;5:e1000376.
- 6: Mészáros B, Erdős G, **Dosztányi Z**. IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res.* 2018;46:W329-W337.
- 7: **Dosztányi Z**, Mészáros B, Simon I. ANCHOR: web server for predicting protein binding regions in disordered proteins. *Bioinformatics.* 2009;25:2745-2746.
- 8: Erdős G, Pajkos M, **Dosztányi Z**. IUPred3: prediction of protein disorder enhanced with unambiguous experimental annotation and visualization of evolutionary conservation. *Nucleic Acids Res.* 2021;49:W297-W303.
- 9: Pajkos M, Mészáros B, Simon I, **Dosztányi Z**. Is there a biological cost of protein disorder? Analysis of cancer-associated mutations. *Mol Biosyst.* 2012;8:296-307.
- 10: Mészáros B, Hajdu-Soltész B, Zeke A, **Dosztányi Z**. Mutations of Intrinsically Disordered Protein Regions Can Drive Cancer but Lack Therapeutic Strategies. *Biomolecules.* 2021;11:381.
- 11: Pajkos M, Zeke A, **Dosztányi Z**. Ancient Evolutionary Origin of Intrinsically Disordered Cancer Risk Regions. *Biomolecules.* 2020;10:1115.