

Válaszok a Dr. Ferenczy György bírálatában feltett kérdésekre

Mindenekelőtt szeretném megköszönni bírálóm alapos munkáját és a dolgozatra adott értékelését. Az alábbiakban igyekszem megválaszolni az általa feltett kérdéseket.

1. A mesterséges intelligencián alapuló szerkezet előrejelzés minőségi ugrást jelent a szekvencia-szerkezet kapcsolat feltárásában. Az áttörést ebben az AlphaFold eljárás kidolgozásához köthetjük, amelynek megjelenését megelőzték a doktori műben ismertetett kutatások. Indokolt ezért a kérdés, hogy az AlphaFold minden korábbinál jobb minőségű szerkezet előrejelző képessége milyen hatással van a vizsgált módszerek jövőbeni fejlesztésére és alkalmazására, így például az alacsony komplexitású szekvenciák detektálására? Milyen erre vonatkozó tapasztalat van, és hogyan befolyásolhatja ez a dolgozatban tárgyalt magányos alfa-helikális, funkcionálisan rendezetlen és fibrilláris fehérjeszakaszok detektálására kidolgozott eljárások fejlesztésének és alkalmazásának irányát?

A terület fejlődésének üteme miatt nem könnyű erre a kérdésre kellően általános választ adni, az alábbiak a 2025 elején általam ismert kutatásokra és trendekre vonatkoznak. Az AlphaFold2 eljárást elsősorban globuláris, monomer fehérjék szerkezetpredikciójára tervezték, és ezen a területen kiemelkedően teljesített. Ugyanakkor igen korán felismerték, hogy más típusú szerkezetek, pl. membránfehérjék predikciójára is sok esetben használható eredményt ad (Hegedűs et al. 2022), valamint kiderült, hogy számos olyan szakasz, amelyet az AlphaFold2 saját becslése alapján alacsony megbízhatósággal modellezett, jól megfeleltethető funkcionálisan rendezetlen szakaszoknak (Zhao et al. 2023). Fontos megjegyezni, hogy már korábban is jelen voltak gépi tanuláson alapuló, kifejezetten rendezetlen szakaszokat felismerő módszerek is (pl. SPOT-Disorder, Hanson et al. 2019). A nem-globuláris szakaszok predikciójának pontosabbá tétele, illetve az erre vonatkozó, gépi tanulást felhasználó eljárások fejlesztése kiemelt prioritás a szerkezetkutatók körében, és ennek előmozdítására jött létre pl. az ML4NGP COST együttműködés (<https://ml4ngp.eu>). Itt már nem csupán az a cél, hogy ezeket a szakaszokat felismerjük, hanem az is, hogy konkrét szerkezetbecslést kapjunk ezekre, aminek nehézsége, hogy ezekben az esetekben sokaságalapú leírásra lenne szükség. A gyakorlati megvalósítás szűk keresztmetszete a kellő mennyiségű kísérleti adat, illetve az ezekből meghatározott szerkezeti sokaságok elérhetősége, az elmúlt két évtizedben elért áttörések ellenére az adatok mennyisége és diverzitása még mindig messze alatta marad pl. a globuláris fehérjékre vonatkozó szerkezeti adatokénak. A főleg, nem kizárólag rendezetlen fehérjék sokaságait gyűjtő Protein Ensemble Database jelenleg mintegy 500 rekordot tartalmaz, sokszor ugyanazon fehérje különböző állapotait (<https://proteinensemble.org>, Ghafouri et al. 2024). Ezzel együtt már léteznek olyan gépi tanulás alapú módszerek, melyek képesek szerkezeti sokaságokat létrehozni (Janson et al. 2023)

A magányos alfa-hélixek esete különösen érdekes, mert az AlphaFold2 sok esetben helyesen ismeri fel ezen szakaszok határozott helikális szerkezeti preferenciáit, nem ritkán hosszú, "semmibe lógó", egyébként magas megbízhatóságúnak becsült helikális szegmenseket modellezve. Magukra a modellekre tekintve azonban nem egyszerű a SAH és a coiled-coil szakaszok elkülönítése, illetve az esetleges predikciós hibák kiszűrése. Az AlphaFold2 modellekben megtalálható ilyen helikális szakaszokra egyébként felfigyeltek már kutatók, és ezeket elemezve mintegy „újra felfedezték” a SAH szerkezeti elemet (Triandafillou et al. 2023).

Az általam kívánatosnak tartott predikciós munkafolyamat úgy nézne ki, hogy egy kezdeti modellből azonosítja a globuláris, fibrilláris, transzmembrán és funkcionálisan rendezetlen szakaszokat, és ezek felismerése után ezek mibenlétét és határait iteratív módon finomítja, adott esetben akár külön megbecsülve az egyes régiók mobilitását, vagy explicit módon szerkezeti sokaságot adva kimenetként. Természetesen felmerül a multimerizáció lehetősége is, erre a 4. kérdésre adott válaszban még visszatérek.

2. A fehérjeszerkezeti sokaságok előállítása néhány replika néhány 10 nanoszekundumos szimulációjával történt. Mint azt említi is a 3.5.1 fejezetben, ennél ma lényegesen hosszabb szimulációk is rutinszerűen végezhetőek, bár – amennyiben jól értem – ennek korlátja a MUMO eljárás CPU-hoz kötött implementálása. Kérdésem, hogy látja-e értelmét és lehetőségét hosszabb szimulációk, és más mintavételezést javító eljárások bevezetésének és alkalmazásának a sokaságok előállításban?

Valóban, az általam implementált eljárás a CPU-alapú párhuzamosítás miatt nem profitál az utóbbi években történt, GPU-alapú gyorsításra irányuló fejlesztésekből, és több replika esetében a GPU-alapú megoldás implementálása túlmutat a jelenlegi lehetőségeinken. A megkötéseket használó, sokaság-alapú megoldások esetében nagyrészt igaz az, hogy a bejárt konformációs teret elsősorban a megkötések jellege határozza meg, bár természetesen itt is tapasztalható, hogy a hosszabb szimulációs idő javíthatja a mintavételezést. A kisméretű, globuláris domének esetében ezzel együtt a CPU-t használó sokaság-alapú számítások véleményem szerint az esetek többségében megfelelőek, hiszen itt igaz az a közelítés, hogy egy jól definiált átlagos szerkezet körüli, a szerkezet nagymértékű megváltozását nem okozó fluktuációkat vizsgálunk. A helyzet azonban megváltozik, amikor olyan szakaszokat modellezünk – hosszabb funkcionálisan rendezetlen szakaszok vagy többdoménes, flexibilis linkerekkel összekötött rendszerek –, amelyekben a konformációs tér ennél jóval nagyobb. Ezek esetében a jelenlegi szimulációs eszközök mellett elsősorban nem a szimulációs idő minél hosszabbá tételében, hanem inkább a kezdeti szerkezetek konformációs változatosságának növelésében és több szimuláció futtatásában látom a megoldást. Ennek oka, hogy a tipikus erőterek sokszor a kompakt szerkezeteket részesítik előnyben, így – saját tapasztalatunk alapján – kb. 100 ns nagyságrend után a kialakult kölcsönhatások miatt már nem feltétlenül várható egy, az addigra felvett térszerkezettől drasztikusan eltérő állapot mintavételezése egy flexibilis rendszer esetében. A ma már elérhető, kifejezetten rendezetlen fehérjékre fejlesztett erőterek (Mu et al. 2021) javíthatnak ezen a helyzeten, de véleményem szerint ezzel együtt kérdéses, hogy egy moduláris, rendezett és flexibilis szakaszokat egyaránt tartalmazó rendszer modellezése esetében elkerülhető-e egy összetett munkafolyamat kialakítása a hatékonyabb mintavételezéshez. Összességében tehát igen, fontos és kívánatos a szimulációs idő növelése, azonban ezt véleményem szerint leginkább több rövidebb, 100 ns nagyságrendet felölelő, különböző kezdeti feltételekből indított szimulációval érdemes megvalósítani.

A teljesség kedvéért megemlítem a coarse-grained szimulációk lehetőségét, melyek természetesen nagymértékben megnövelik a mintavételezés hatékonyságát és számos esetben, mint pl. fázisszeparációval létrejött kondenzátumok vizsgálata (Dignon et al. 2019), megfelelő eszközt jelenthetnek. Az általam leírt kutatásokban ezek használatának korlátja, hogy tapasztalataink szerint

nem triviális egy ilyen szimulációból a megfelelő lokális geometriával rendelkező atomi szintű modellek előállítását, amelyek pl. az NMR-paraméterekkel való kvantitatív egyezések megállapításához feltétlenül szükségesek lennének.

3. A 71. oldalon tárgyalja, hogy azokat a predikciós eljárásokat tartja értékesnek, amelyek a fiziológiásan releváns szerkezeteket képesek előrejelezni, és ezzel kapcsolatban megjegyzi, hogy az AlphaFold fiziológiásan irreleváns monomer szerkezeteket javasol az emberi vázizomban található miozin II-re. Ezzel kapcsolatban általánosan is szeretném kérdezni, hogy véleménye szerint mi okozza, hogy az AlphaFold fiziológiásan nem releváns szerkezeteket javasol, és hogyan tudjuk azonosítani ezeket, illetve mennyire lehetünk biztosak abban, hogy egy javasolt szerkezet fiziológiás körülmények között nem valósul meg?

Az AlphaFold2 esetében a dolgozatban bemutatott predikció elsősorban azért nem megfelelő, mert az AlphaFold2 alap verziója nem tud multimer szerkezeteket kezelni. Ez természetesen nem probléma egy hozzáértő kutató számára, az én fenntartásom, hogy amennyiben az ilyen, a fiziológiás állapotot nem tükröző modelleket nagy számban tesszük elérhetővé, mint ahogyan az most pl. az UniProt adatbázisban is megfigyelhető, akkor az torzíthatja az adatbázist használókban ezen fehérjék nagyobb léptékű szerkezeti szerveződéséről kialakult képet. Az AlphaFold-Multimer, és különösen a 2024-ben bemutatott AlphaFold3 (Abramson et al. 2024) kifejezetten képes multimerek, fehérjekomplexek modellezésére, és megfelelő beállítások mellett hasznos és releváns eredményeket adnak.

A fiziológiásan releváns szerkezet megállapítása természetesen nem egyszerű feladat, a bemutatott predikció esetében a kritika a miozin II molekulák kísérletesen meghatározott szerkezetei, a coiled coil szerkezetek jellemzői, illetve a motorfehérje működési mechanizmusának ismeretében fogalmazódott meg. Ebben még a specializáltabb predikciók sem mindig nyújtanak támpontot, pl. a drebrin esetében a kutatók számára meglepő volt, hogy a fehérje monomerként viselkedett a predikciók alapján benne megtalálható coiled coil motívum ellenére – a feloldás, hogy ez a coiled coil valójában egy SAH (Peckham & Knight 2009).

A kérdés másik oldala, hogy egy becsült szerkezet ténylegesen megvalósul-e fiziológiás körülmények között. Egy coiled coil motívum segítségével dimerizálódó fehérje esetében természetesen elvben előfordulhat, hogy a riboszómán történő keletkezése után egy ideig monomer formában van jelen a sejtben. Az általam ismert új eredmények azonban arra utalnak, hogy számos stabil komplex összeszerelődése már a transzláció során megkezdődik (Badonyi & Marsh, 2023), így a monomer állapotok életideje a sejtben elhanyagolhatóan tekinthető ezekben az esetekben. Összefoglalva, ha egy fehérje az in vitro kísérletek és a rendelkezésre álló in vivo funkcionális adatok alapján is multimerként viselkedik, akkor egy monomer állapotot tükröző modell relevanciája még akkor is megkérdőjelezhető, ha a feltekeredése során esetlegesen felvesz ehhez hasonló állapotot.

4. A dolgozatban jelentős hangsúlyt kapnak az alacsony komplexitású régiók, azok detektálása, szerkezetük előrejelzése és az ilyen szakaszokat tartalmazó fehérjék dinamikus szerkezeti sokaságainak előállítása. Ugyanakkor a fehérjék működésének elemzésében (4.5 fejezet) a hat alkalmazásból csupán egy, a miozin VI (4.5.6 fejezet), amely kifejezetten ehhez kapcsolódik. Az arány eltolódásának módszertani oka van, nevezetesen, a sokaságok hatékonyabb előállítása globuláris fehérjékre, netán a kísérleti adatok eltérő hozzáférhetősége, vagy egyéb szempont?

A dolgozatban bemutatott példák esetében az ok elsősorban módszertani, a globuláris domének vizsgálata a stabil szerkezet és így a korlátozottabb konformációs tér miatt egyszerűbben kivitelezhető. Másodsorban igaz az is, hogy több adat volt ezekhez elérhető: az adott kérdések vizsgálatakor a NOE távolságok és S^2 rendparaméterek meghatározása sokkal inkább számított rutinszerűnek, mint a rendezetlen fehérjék globális szerkezeti jellemzéséhez sok esetben elengedhetetlen RDC, PRE vagy SAXS mérések – jelenleg ezek közül leginkább a SAXS mérések vannak terjedőben az egyszerűbb mintaelőkészítés miatt. Bevallom, hogy saját tapasztalataim is befolyásoltak, az általam sokaságalapú megközelítéssel megoldani próbált első probléma esetében annak komplexitása miatt nem jártam sikerrel egy adott rendezetlen szakasz modellezésével, ezért a későbbiekben egy ideig inkább globuláris fehérjékkel foglalkoztam. A dolgozat leadása óta kutatócsoportunkban meghatároztuk a drebrin fehérje SAH régiójának szerkezeti sokaságát is, és tervben van több további nem-globuláris vagy ilyen szakaszokat is tartalmazó rendszer jellemzése.

Ezúton is még egyszer köszönöm Ferenczy Györgynek a dolgozat bírálatát és a kapcsolódó kérdéseket, bízom benne, hogy sikerült ezek kapcsán érdemi válaszokat adnom.

Kerepes, 2025. 02. 13.

Gáspári Zoltán

Hivatkozások

- Abramson J et al. 2024. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* 630:493-500.
- Badonyi M & Marsh JA 2023. Hallmarks and evolutionary drivers of cotranslational protein complex assembly. *FEBS J* 291:3557-3567
- Dignon GL et al. 2019. Simulation methods for liquid–liquid phase separation of disordered proteins. *Curr Opin Chem Eng* 23:92-98
- Ghafouri H et al. 2024. PED in 2024: improving the community deposition of structural ensembles for intrinsically disordered proteins. *Nucleic Acids Res* 52:D536-D544
- Hanson J et al. 2019. SPOT-Disorder2: Improved protein intrinsic disorder prediction by ensembled deep learning. *Genomics Proteomics Bioinformatics* 17:645-656.
- Hegedűs T. et al. 2022. Ins and outs of AlphaFold2 transmembrane protein structure predictions. *Cell Mol Life Sci* 79:73.
- Janson G et al. 2023. Direct generation of protein conformational ensembles via machine learning. *Nat Commun* 14:774
- Mu J et al. 2021. Recent force field strategies for intrinsically disordered proteins. *J Chem Inf Model* 61:1037-1047
- Peckham M & Knight P 2009. When a predicted coiled coil is really a single α -helix, in myosins and other proteins. *Soft Matter* 5:2493-2503
- Triandafillou CG et al 2023. Pervasive, conserved secondary structure in highly charged protein regions. *PLoS Comp Biol* 19:e1011565
- Zhao B et al. 2023. Comparative evaluation of AlphaFold2 and disorder predictors for prediction of intrinsic disorder, disorder content and fully disordered proteins. *Comp Struct Biotechnol J* 21:3248-3258