

**Fehérjemolekulák biológiailag releváns belső
mozgásainak modellezése sokaság alapú
megközelítésekkel**



Harmat Zita
A PhD disszertáció tézisei

Pázmány Péter Katolikus Egyetem
Információs Technológiai és Bionikai Kar

Témavezető:
Dr. Gáspári Zoltán

Budapest, 2022

1. Bevezető

A fehérjék a legfontosabb biomolekulák közé tartoznak. Meghatározzák a sejtek alakját és mozgását, katalizálják az élethez legfontosabb reakciókat, lehetővé teszik az anyagtranszportot a sejtek és környezetük között, és még számos más molekuláris folyamat meghatározó szereplői. Működésük pontos leírása elengedhetetlen az életfolyamatok megértéséhez és az azokba való – pl. orvosi célú – beavatkozások tervezéséhez.

A fehérjék szerkezetileg négy nagy csoportra oszthatók: globuláris, fibrilláris, transzmembrán és rendezetlen fehérjékre. Az első három csoportba jól meghatározott, időben állandó háromdimenziós szerkezettel rendelkező fehérjék tartoznak. A globuláris fehérjék lokális, ún. másodlagos szerkezete változatos, több α -helikális és/vagy β -szálal régiót tartalmaznak, és alakjuk megközelítően gömb alakú. A fibrilláris típusba a jellemzően hosszú fonalszerű alakkal rendelkező fehérjék tartoznak. A transzmembrán fehérjék sok szempontból többnyire a globuláris fehérjékhez hasonlítanak, de jellemzőjük, hogy átnyúlnak a biológiai membránokon. A funkcionálisan rendezetlen fehérjék az előző három csoporttal ellentétben nem rendelkeznek egyetlen jellemző szerkezettel, hanem nagyfokú belső mozgékonyaságuk miatt számos konformáció közötti folyamatos átrendeződésben vannak.

A fehérjék alakjának változatossága döntő fontosságú szerkeázó biológiai szerepük meghatározásában. Ma már jól ismert, hogy minden fehérje, nem csupán a rendezetlenek, dinamikus objektum, és belső mozgásai a szerkezettel együtt szervesen hozzájárulnak a molekuláris funkció betöltéséhez. Szerkezetükről atomi képet először a röntgenkrisztallográfia adott, mely egy statikus, kristályos állapotban lévő szerkezetet írt le. Később azonban egyéb kísérleti módszerekkel is lehetett vizsgálni a fehérjéket, például oldott fázisú NMR mérésekkel. E kísérletek során sokkal több és részletesebb információt kaphatunk a fehérjék belső dinamikájáról, mozgásairól, több időskálát is lefedve. Ezt a dinamikát szerkezeti sokaságmodellel is le tudjuk írni.

Mivel a fehérjék szerkezetének és belső dinamikájának kísérletes felfedezése idő- és pénzigényes, egyre inkább teret hódítanak az *in silico* fehérjeszerkezeti és dinamikai számítások. Ezeknél azonban nem szerencsés, ha egyáltalán nem kapcsolódnak a kísérleti adatokhoz, mivel nehezzé válik a modellek pontosságának, azaz a valósággal való megfelelésüknek a biztosítása és annak ellenőrzése is. Jelenleg igen hangsúlyos irányzat emiatt az *in silico* modellezés a kísérleti adatokkal való kombinálása. A modellek segítenek megérteni, pontosítani a kísérleti adatokat, a kísérleti adatok pedig realisztikusabbá teszik a modelleket. Ezt a kísérleti modelleket felhasználó számítógépes atomi fehérje szerkezet modellezést, mint korszerű módszert

elterjedten használják már évek óta. Megjegyezhető, hogy valójában a koncepció nem új abban az értelemben, hogy a hagyományos fehérjeszerkezet-meghatározó módszerek adatszegények, azaz a meghatározandó térbeli koordináták száma lényegesen meghaladja a kísérletileg mért paramétereket, emiatt nagy mértékben támaszkodnak *a priori* ismertnek vett, elsősorban geometriai (kötéshosszak, kötésszögek stb.) adatokra, melyeket a modellezési eljárásokkal rokon módon építenek be a protokollokba. Ez leginkább az NMR-spektroszkópiával nyert adatokból történő szerkezetszámolási lépés során érhető tetten. A jelenleg egyre inkább terjedőben lévő "hibrid" eljárások sajátossága, hogy a kísérleti paraméterek és a modellezésben használt erőter súlya utóbbi felé jobban eltolódik, hangsúlyosabbá téve annak szerepét, a használt kísérleti paraméterek pedig leginkább már rendelkezésre álló szerkezetek esetében értelmezhetőek. Erre példa az S^2 rendparaméter, vagy fehérjekomplexek felépítéséhez használt dokkolási eljárásokban alkalmazott kísérleti kényszerfeltételek.

Egy modellnek két fő ismérve van: a pontossága (accuracy) és a precizitása (precision). A pontossága fejezi ki azt, hogy mennyire felel meg a valóságnak, a precizitása pedig azt tükrözi, hogy a felhasznált kísérleti adatok mennyire pontosak (kísérleti hiba). Egy pontos, de nem precíz modellnél nagy a szórása az adatoknak, de azok átlaga megfelel a valóságnak. Egy precíz, de nem pontos modellnél kicsi a szórás, de az átlag messze jár a valós értéktől. A két eset közül a pontos, de precíz modell limitáltan még használható, de a nem pontos, ámde precíz modell sajnos használhatatlan. Egy modell szabotosságát úgy tudjuk ellenőrizni, ha például többféle kísérleti módszer adatait használjuk fel, hiszen annak kicsi az esélye, hogy ugyanazt a hibát több kísérletnél is elkövetik. Emellett felmerülhet az is, hogy a kísérleti adatokat a modellünk felépítésekor tévesen értelmezzük, vagy rosszul dolgozzuk fel. Ez a hiba észrevehető például akkor, ha az egyik kísérletből származó adatoknak megfelel a modellünk, de egy másik kísérletből származó adatoknak nem felel meg, ez esetben gond van a modellünk pontosságával [1].

A dolgozatban három különböző fehérjetípus szerkezeti sokagáinak előállítását és elemzését mutatom be: a globuláris fehérjék csoportját a gasztrotropin fehérje képviseli, a fibrilláris csoportot a miozin VI molekula magányos α -helikális (SAH) doménje, a CD3e citoplazmikus doménje pedig a funkcionálisan rendezetlen fehérjék közé tartozik. A gastrotropin esetében megkötéseket tartalmazó molekuladinamikai szimulációkat futtattam, míg a másik két fehérje esetében egy előzetes konformerkönyvtárat hoztam létre, melyből szelekcióval állítottam elő a kísérleti paramétereknek megfelelően a sokaságokat. A konformerkönyvtárak előállítását az általam fejlesztett

DIPEND (DIordered Protein Ensembles from Neighbor-dependent Distributions) programsor segítségével valósítottam meg. Mindegyik fehérjénél a fehérjesokaság kísérleti paramétereknek való megfelelésének vizsgálatához felhasználtam a kutatócsoport más tagjai által fejlesztett CoNSEnsX⁺ programot.

A fehérjeszerkezet-modellezés azon módjának, melyet a miozin és a CD3e fehérje esetében alkalmazunk két lépése van. Az elsőben egy nagyszámú szerkezetből álló, diverz konformációs teret lefedő sokaságot hozunk létre a DIPEND programsorral. Ennek egyenletesen le kell fednie a fehérje konformációs terének minél nagyobb részét. A második lépésben ebből választunk a CoNSEnsX⁺ program segítségével egy sokaságot, mely megfelel a kísérleti adatoknak. A két lépés különválasztása azért előnyös, mert így mindkét lépés tetszés szerint alakítható az adott fehérje sajátosságait figyelembe véve.

A fentebb leírtakkal összhangban a munka során a szerkezeti modellek validálásában szakirodalmi adatokból származó kísérleti adatokat használunk fel.

2. Módszerek

2.1. A DIPEND program alkotóelemei

A programot túlnyomórészt Python 3-ban írtam. A Guido van Rossum által a nyolcvanas évek végén megalkotott programnyelv (<https://www.python.org/doc/>) interpretált, magasszintű és nagyon népszerű. Objektumorientált, de más programozási stílusokat is támogat. Jól olvasható, átlátható és fejleszthető kódot eredményez. Különleges ismeretjelje, hogy a programblokkokat zárójelek helyett a behúzás mértékével választjuk el.

Az UCSF egyetemen fejlesztett Chimera molekulanézegető és molekulamodellező program [2] új változata, a ChimeraX [3] (<https://www.rbvi.ucsf.edu/chimerax/download.html>) egy jól karbantartott, folyamatosan fejlesztett, regisztráció fejében szabadon elérhető, sokrétű program, mely kitűnően alkalmas molekula modellezésre. Nagy előnye a programom szempontjából, hogy meghívható parancssorból is paraméterekkel felhasználói közbeavatkozás nélkül. A DIPEND program a ChimeraX-nek a szekvencia alapján megadott torziós szögekkel történő fehérjemodell-építő, az adott torziós szöveget beállító illetve a sztérikus ütközés ellenőrző funkcióját hívja meg.

A Dunbrack labor által fejlesztett SCWRL4 program [4] (<http://dunbrack.fccc.edu/SCWRL3.php/>) licenz ellenében szabadon felhasználható. Egy fix főláncra fehérje oldalláncok atomi modelljeinek sztérikus ütközés nélküli optimális felépítését teszi lehetővé. Parancssorban meghívható, gyors, és egy pdb fájlt használ bemenetnek, és ugyanúgy egy pdb fájlt ad ki kimenetnek, megtartva a lánc- és aminosav számozást. Főláncfüggő rotamer könyvtárat használ, melyekre energiát számol ki gyakoriságuk és a többi oldalánccal való közelség alapján. A sztérikus ütközéseket k-diszkrét politópokkal, azaz 3-dimenziós formák alkalmazásával küszöböli ki. A sok lehetőség közül egy saját gráf fa-dekompozíciós módszerrel választ és a választást heurisztikával gyorsítja meg, mely mindig hamar ad eredményt, noha nem biztos, hogy megtalálja a minimális energiájú konformert.

2.2. Molekuladinamikai szimulációk

A molekuladinamikai szimulációkat a GROMACS program 4.5.5-ös verziójának [5],[6] Gáspári Zoltán által módosított változatával futtattuk. Ez a változat képes kezelni az S^2 rendparamétereket és átlagolja az NOE paramétereket minden replikán [7], vagyis a MUMO (Minimal Under-restraining Minimal Over-restraining) megközelítést alkalmazza [8]. Az OPLS-AA erőteret [9] és a TIP3P vízmodellt [10] alkalmaztam ezekben a szimulációkban.

Az *apo* szerkezet modellezéséhez a Protein Data Bank (PDB) adatbázis 1O1U azonosítójú szerkezetének 7. modelljét [11] választottam, mert annak volt az összes modell közül a legnagyobb PRIDE-NMR pontszáma [12]. A PRIDE NMR egy a kutatócsoportunk által még 2007-ben kifejlesztett módszer [12], mely egy sokaságból kiválasztja a legjobb NOE megfelelésűt úgy, hogy az atomtávolságok eloszlását összehasonlíttja kontingencia analízissel a NOE adatokkal. Így képez egy pontszámot, mely a pride score, ami egy 0 és 1 között, hasonlósági valószínűség. A kísérleti paraméterek eloszlását figyelembe veszi és tapasztalataink alapján megbízhatóan kiválasztja azt a kezdő modellt, mely a legjobban tükrözi a kísérleti adatokat. A *holo* szerkezet modellezéséhez a 2MM3 azonosítójú PDB adatbázisbeli szerkezet első modelljét választottuk. A ligandumok topológiáját a TopolGen szkript segítségével építettük fel (http://www.mdtutorials.com/gmx/complex/02_topology.html). A ligandumok glikokólsav (GCA, PDB ligand ID: GCH), valamint glikokenodezoxikólsav (GCDA, PDB ligand ID: CHO) elnevezést kaptak. A hidrogén atomok töltéseinek nehézatomokhoz való igazításához egy Perl scriptet használtunk, illetve egy-egy atom típusát manunálisan is átírtuk, ha az szükségesnek lát-

szott.

NOE kényszerfeltételek csak a *holo* 2MM3 azonosítójú szerkezethez voltak elérhetőek, így az *apo* formához azokat használtuk, melyek a CoNSEnsX⁺ webszerver szerint megfeleltek ennek a szerkezetnek. A kényszerfeltétellistát módosítottuk úgy, hogy 1 Å széles tartományokat alakítottunk ki 4-től 10 Å távolságig, emellett a sztereospecificitást is eltávolítottuk, vagyis szimmetrizáltuk a távolságokat úgy, hogy mindegyik ugyanazon sztereocentrumhoz tartozó atomnál a legnagyobb távolságot vettük figyelembe.

Az *apo* szerkezethez a kémiai eltolódások a BMRB [13] adatbázisból az 19843 azonosítójú adatsorból származnak, a *holo* szerkezethez pedig maguktól a szerzőktől. Az S² rendparamétereket négy hőmérséklet (283, 291, 298, és 313 K) értékre Horváth Gergely és munkatársai [14] munkájából vettük át.

Miután az OPLS-AA erőtérrrel és a TIP3P vízmodellel legyártottuk a topológiát, egy kocka alakú dobozba tettük a molekulamodellt, majd energiainimalizálást végeztünk a konjugát gradiens módszerrel 5000 iterációig. A maximális erőt 200 $\text{kJmol}^{-1}\text{nm}^{-1}$ értékre állottuk. A következő lépésben vízzel vettük körül a molekulamodellt és az egyik vízmolekulát kicseréltük egy nátrium ionra, hogy a rendszer összetöltése semleges legyen. Ezután egy újabb energiainimalizálást végeztünk ugyanezekkel a paraméterekkel, de immár a vizet és az iont is belevéve. Végül egy rövid molekuladinamikai szimulációt végeztünk úgy, hogy a nehézatomok pozíciója lehetőleg ne változzon, azaz egy 1000 $\text{kJmol}^{-1}\text{nm}^{-2}$ értékű kényszerfeltétellel (position restraint) tartottuk ezeket az atomokat a helyükön. Ez a szimuláció 2500 db 2 fs hosszú lépést tartalmazott a kötéshosszak megkötésére a LINCS algoritmust használva [15].

Az ezután következő főszimulációnál 8 replikát futtattunk párhuzamosan az OpenMPI környezetet használva [16]. Az S² kényszerfeltételeket a 8 replikára átlagként alkalmaztuk, az NOE kényszerfeltételeket pedig szomszéd replikánként, azaz páronként vettük figyelembe a MUMO (Minimal Under-restraining, Minimal Over-restraining) protokoll szerint [8]. A szimulációkat 283, 291, 298, és 313 K hőmérsékleteken az adott hőmérsékleteken mért S² rendparaméterekkel végeztük. A kötéshosszakat a LINCS algoritmussal kötöttük meg, 2 fs lépésközzel 2 ns és 6 ns időtartamig futottak, a 8 replikát kombinálva ezek összesen 16 és 48 ns időt tesznek ki. Ugyanezen paraméterekkel, de kényszerfeltételek nélküli kontroll szimulációkat is végeztünk.

A konformációs tér nagyobb mintavételezése érdekében további szimulációkat végeztünk: olyanokat, melyekben csak S² vagy csak NOE kényszerfeltételek szerepeltek. Gyorsított molekuladinamikai szimulációkat (AMD)

és rövid (500 ps) targetált molekuladinamikai szimulációkat is futtattunk az *apo* szerkezeten a *holo* szerkezet kémiai eltolódásait használva, és fordítva, hogy a kettő közti átmenetet el tudjuk érni.

A CD3e fehérjeszerkezeten egy explicit vizes közegben futtatott hagyományos szimulációt is elvégeztünk, 1 μ s hosszan AMBER99SB-ILDN erőterrel SPC/E explicit vízmodellel a GROMACS 2020-as verziójával. Neutralizáció és egy rövid energiaminimalizálás után 1 ns NVT és NPT ekvilibrálást végeztünk. A szimuláció 300 K hőmérsékleten folyt, GPU gyorsítással. A sokaságot úgy generáltuk, hogy minden 50 ns-nál belevettük az adott szerkezetet, így a sokaság 20001 konformerből áll. Minden szerkezetet belevettünk, a kezdeti nyújtott szerkezetek is reprezentálva legyenek.

2.3. Főkomponens-analízis

A főkomponens-analízist a ProDy [17] nevű Python modullal végeztük, a molekulák mozgásait pedig a VMD molekulamodellező program [18] NM-Wiz moduljával jelenítettük meg.

2.4. A gastrotropin S2 adatainak és kémiai eltolás adatainak elemzése

Az S^2 rendparaméterek esetében összehasonlítottuk az eredeti PDB-beli sokaságokét a molekuladinamikai szimulációval MUMO módszerrel nyert sokaságokéval. Azokat a pontokat, ahol 0.2-nél nagyobb eltérés mutatkozott, a további elemzésnél elhagytuk. A MUMO szimulációk esetében ez maximum 5 pontot jelent.

A szerkezetekből a Shiftx2 programmal [19] számoltuk vissza a kémiai eltolódás értékeket.

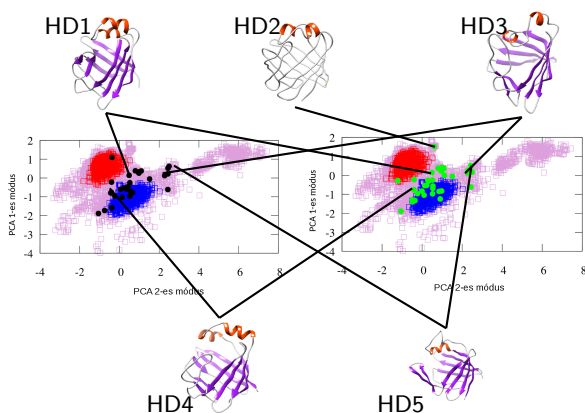
2.5. A gastrotropin szerkezetek amid kémiai eltolódás különbségeinek vizsgálata

A ^{15}N kémiai eltolódásokat minden szerkezetből a Shiftx2 programmal [19] számoltuk vissza. Az összes konformációnál kiszámoltuk az abszolútértékét a visszaszámolt kémiai eltolódásoknak és az összes *apo* szerkezet visszaszámolt kémiai eltolódásainak. Ezeket a különbségeket összehasonlítottuk a kísérletesen meghatározott $|\Delta\omega(^{15}\text{N})|$ értékekkel, melyeket CPMG relaxációs diszperziós NMR mérésel határoztak meg minden aminosavra, ahol ezt meg lehetett tenni [14]. Miután a különbségeket 0 és 1 közé normaltunk, korrelációs és RMSD értékeket számoltunk ki belőlük. A kísérleti $|\Delta\omega|$

értékek három hőmérsékletre vonatkozóan állnak rendelkezésre, a szimulációk viszont többféle hőmérsékleten, így a három hőmérséklet különbségeinek az átlagát vettük. Azokat a szerkezeteket választottuk ki elemzésre, melyek a legnagyobb korrelációt és a legkisebb RMSD-t adták.

3. Új tudományos eredmények

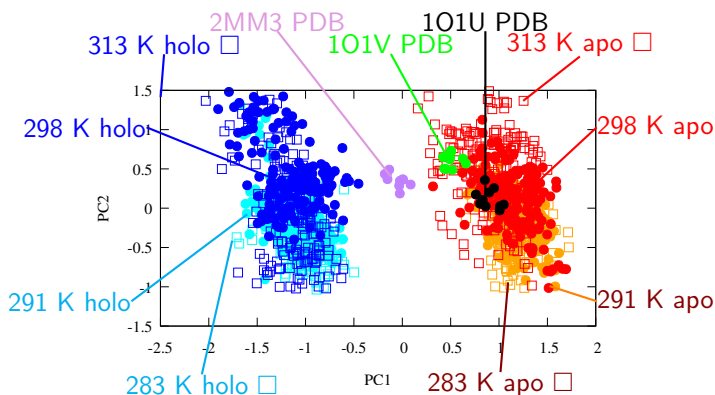
3.1. Első tézis csoport



1. ábra. Az *apo* MUMO (piros) és a *holo* MUMO (kék) sokaságok PCA scatter plotja, valamint a többi generált sokaság (lila négyzetek), melyeket a rejtett állapotkonformációk kiválasztására használtam. A ^{15}N kémiai eltolódások és a kísérletes $|\Delta\omega(^{15}\text{N})|$ értékek közötti különbségek közötti 0.35 fölötti korrelációt elérő konformereket fekete ponttal jelöltem baloldalt. Ugyanezen különbségből számolt RMSD-knél a 0.00603 alatti konformereket zöld ponttal jelöltem a jobboldalon. A kiválasztott konformereket HD1-HD5 megnevezéssel kis szerkezetekkel rajzoltam fel és a hozzájuk tartozó pontokat is jelöltem.

1.1.A. Tézis:

Elállítottam S^2 és NOE paraméterekkel végzett megkötött molekuladynamikai szimulációkkal a MUMO módszer segítségével a gasztrotropin



2. ábra. A szimulációból származó és a kísérletileg meghatározott szerkezetek főkomponens-analízis (PCA) scatter plotja.

fehérje szerkezeti sokaságait négy hőmérsékleten, mely sokaságok megfeleltek a modellbe nem beleépített kémiai eltolódás paramétereknek.

A MUMO módszerrel készült sokaságok kémiai eltolódásainak korrelációi a modellbe be nem épített kémiai eltolódásokkal sokkal jobb korrelációt mutattak, mint a PDB adatbázisban fellelhető modellekből álló kis méretű sokaságok (1. táblázat).

1.1.B. Tézis:

Előállítottam különböző molekuladinamikai szimulációkkal egy sok elemből álló szerkezeti sokaságot a gasztrotropin fehérje modellezésére. Főkomponens analízis segítségével megmutattam, hogy a sokaság nagy konformációs teret jár be.

A 1. ábrán látható, hogy ez a szerkezeti sokaság diverz és nagy konformációs teret ölel fel.

1.2.A Tézis:

Az előállított S^2 és NOE kényszerfeltételekkel megkötött sokaságokkal megmutattam, hogy a szabad és a ligandumkötött formában is kétfajta mozgás dominál, mindkettő a hordó kinyílását eredményezi.

E mozgások a sokaságok főkomponens-analízisének (PCA) első két módusa mentén történnek (2. ábra), az elsőben főként az E-F és a G-H hurok vesz részt, míg a másodikban a C-D hurok és a két hélix mozdul el leginkább.

	Sokaság méret	Hőmérséklet (K)	Főlánc RMSD (Å)	Főlánc S ² korréláció	Korrigált főlánc S ² korréláció	Nem teljesítő NOE feltételek (%) ($r^{-6} > 0.5\text{Å}$)	Amid N kémiai eltolódás	Amid H kémiai eltolódás
10IU pdb	10	305	0.66+-0	0.502 ^a	0.857	0.00	0.709	0.596
apo MUMO	168	283	1.01+-0.03	0.675	0.873	0.29	0.832	0.72
apo MUMO	168	291	1.12+-0.04	0.8	0.841	0.29	0.828	0.729
apo MUMO	168	298	1.27+-0.06	0.797	0.883	0.12	0.834	0.738
apo MUMO	168	313	1.45+-0.04	0.963	0.972	-	0.834	0.705
apo nem megkötött	48	283	1.46+-0.09	0.307	-	-	0.827	0.724
apo nem megkötött	48	291	1.45+-0.05	0.351	-	-	0.832	0.703
apo nem megkötött	48	298	1.53+-0.03	0.301	-	-	0.833	0.731
apo nem megkötött	48	313	1.5+-0.07	0.6	-	-	0.829	0.733
2MM3 pdb	10	293	0.52+-0	0.297 ^b	0.315	0.00	0.773	0.593
holo MUMO	168	283	1.28+-0.03	0.726	0.756	0.41	0.784	0.548 ^c
holo MUMO	168	291	1.31+-0.07	0.544	0.731	0.46	0.78	0.589
holo MUMO	168	298	1.46+-0.15	0.884	0.930	0.36	0.767	0.552
holo MUMO	168	313	1.5+-0.08	0.724	0.870	0.36	0.758	0.551
holo nem megkötött	48	283	1.92+-0.64	0.073	-	-	0.782	0.531
holo nem megkötött	48	291	2.34+-0.65	0.296	-	-	0.777	0.537
holo nem megkötött	48	298	1.82+-0.07	0.484	-	-	0.772	0.493
holo nem megkötött	48	313	2.59+-2.09	0.272	-	-	0.775	0.544

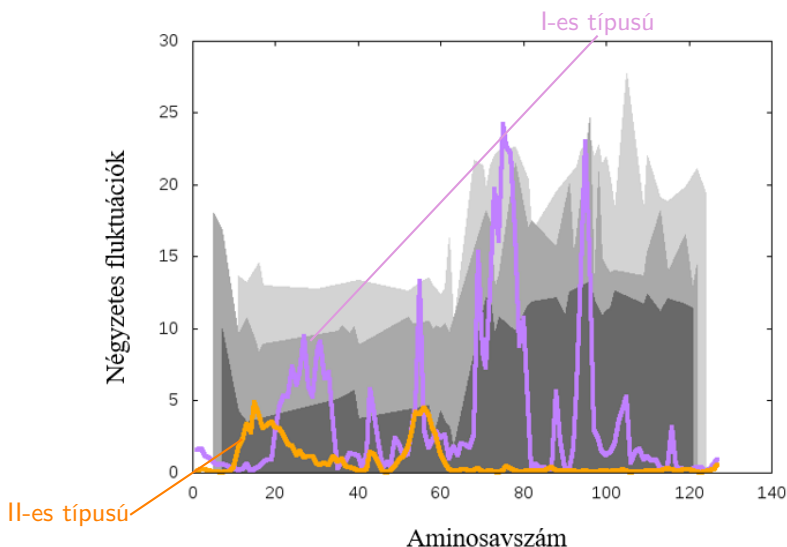
1. táblázat. RMSD értékek, S² paraméterek, NOE megfelelés, amid N és H kémiai eltolódás korrélációk

^aS² adatok 313 K hőmérsékleten (a 4 adatsor közül a legnagyobb korrélációjú adat)

^bS² adatok 298 K hőmérsékleten (a 4 adatsor közül a legnagyobb korrélációjú adat)

1.2.B. Tézis:

Megmutattam, hogy az aminosavcsoport, amely az általam generált szerkezeti sokaságok PCA elemzése alapján leírt második típusú mozgásban részt vesz, nagymértékben átfed a Tőke Orsolya kutatócsoportja által mért lassú időskálás mozgásban részt vevő aminosavakkal

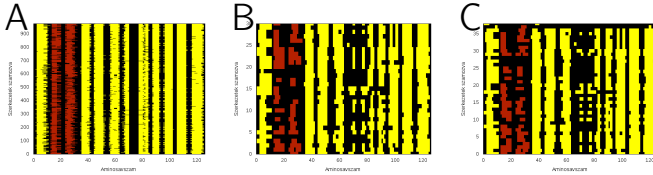


3. ábra. A C α atomok négyzetes fluktuációi az első módus mentén (I. típusú mozgás, lila) és a második módus mentén (II. típusú mozgás, narancssárga). A mért k_{ex} értékek, melyek két aminosav klasztert határoznak meg szürke satírozással vannak jelölve a három hőmérsékleten (283 K, 287 K, 291 K). Csak 30-40 aminosavnál van ms időskálán mérhető k_{ex} érték [14], az átláthatóság és a különböző regionális különbségek hangsúlyozására folyamatosan vannak ezek az értékek felrajzolva.]

Lásd (3. ábra). Tőke Orsolya és munkatársai nem minden aminosavra tudtak mérni az adatokat, így nem tudtuk minden aminosavra megvizsgálni ezt az összefüggést.

1.3. Tézis:

A ligandumok kötőhelyekre való bejutására egy olyan mechanizmust határoztam meg, melynek során a helikális sapka részlegesen kitekeredik.



4. ábra. **A szimulációkból származó konformációk másodlagos szerkezetei.** Minden oszlop egy aminosavat jelent. A β -szálak sárgával, az α -hélixek barnával, a többi aminosavak feketével jelennek meg. (A) Minden konformáció (B) A nagy korrelációt mutató konformációk (az A konformációk részhalmaza) (C) A legkisebb RMSD-vel rendelkező konformációk (az A konformációk részhalmaza). Az elemzés a DSSPcont programmal készült [20]. Figyelemre méltú a hélixek rövidülése egyes konformációkban, főleg a B és C panelen.

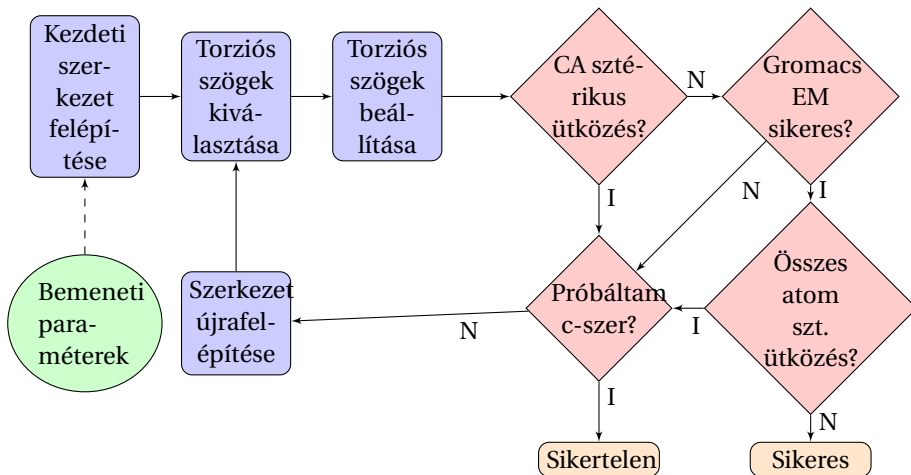
A Tőke Orsolya csoportja által meghatározott rejtett energiaállapotot felvevő kevés konformációval kémiai eltolódás különbség korrelációt tekintve a legmegfelelőbb szerkezetek félig kitekeredettek, főleg a helikális sapka részen (4. ábra). Ebben azonban az erőternek is szerepe lehet.

3.2. Második téziscsoport

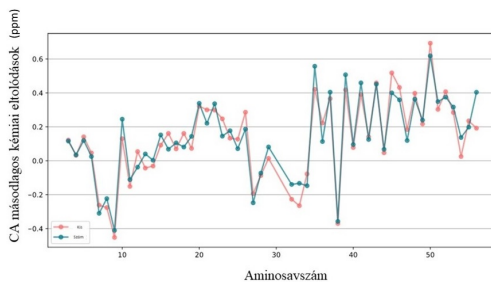
2.1. Tézis:

Létrehoztam a DIPEND nevű programot, ami alkalmas fehérjeszakszok atomi szintű szerkezeti sokaságainak felépítésére lokális szerkezeti preferenciák figyelembe vételével.

A program a szomszédsági valószínűségi modellt alkalmazva megoldja azt a problematikát, hogy az adott fehérjeszerkezet biológiailag releváns konformációi körül egy jó konformációs tér mintavételezést hajtson végre (5. ábra). A súllyal hozzáadható plusz eloszlással pedig tükrözhetők az *a priori* ismert szerkezeti preferenciák. A program nagyjából 100 aminosav hosszúságig tud megbízhatóan szerkezeteket felépíteni, ennél hosszabb szakaszoknál megnő a szerkezeti ütközések esélye. Ez az egyéni súlyozást használó módokkal és a kicsomózási modullal kitolható, amennyiben egy nyújtott szerkezetet eredményül adó eloszlást ad be neki a felhasználó. Ez azonban megnöveli a futási időt. Összességében a program egy jól használható kezdeti sokaságot tud generálni, melyből más programokkal lehet alsokaságokat szelektálni kísérleti adatok alapján.

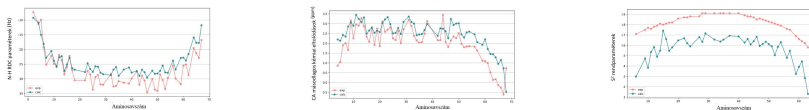


5. ábra. A DIPEND (Disordered Protein Ensembles from Neighbor-dependent Distributions) eljárás folyamatábrája



6. ábra. Megfigyelt és visszszámolt CA másodlagos kémiai eltolódások a CD3 sokaságban.

N-H RDC paraméterek CA másodlagos kémiai eltolódások Főlénc rendparaméterek



7. ábra. A szelektált SAH sokaság Mért (piros) és visszszámolt (kék) NMR paramétere értékei. A számolt értékeket a CoNSensX⁺ szerverrel kaptuk meg.

2.2. Tézis:

A DIPEND programot a humán CD3e fehérje rendezetlen szakaszára alkalmaztam és megmutattam, hogy a DIPEND által generált konformációs sokaság nagyobb konformációs teret jár be és belőle a kísérleti adatoknak jobban megfelelő szerkezeti sokaság szelektálható, mint az egyéb módszerrel generált könyvtárakból.

Ld 6. ábra.

2.3. Tézis:

A DIPEND programot a miozin VI fehérje SAH doménjére alkalmaztam és megmutattam, hogy a DIPEND eljárás egy jól meghatározott előre ismert szerkezeti preferenciával rendelkező fehérje esetében is képes a konformációs tér releváns részét mintavételezni és abból jó lehet egy kísérleti adatoknak megfelelő alsokaságot kiválasztani.

A generált sokaságból szelektált alsokaság NMR adatoknak való megfelelési korrelációja nagyobb, mint a PDB adatbázisból származó szerkezeteké (7. ábra).

4. Közlemények

Jelen disszertáció a következő közlemények alapján készült:

- Harmat Z, Szabó AL, Tóke O, Gáspári Z (2019) Different modes of barrel opening suggest a complex pathway of ligand binding in human gastrotropin. *PLOS ONE* 14(5): e0216142. <https://doi.org/10.1371/journal.pone.0216142>

- Harmat Z, Dudola D., Gáspári Z (2021) DIPEND: An Open-Source Pipeline to Generate Ensembles of Disordered Segments Using Neighbor-Dependent Backbone Preferences. *Biomolecules* 11(10): 1505. <https://doi.org/10.3390/biom11101505>

További közlemények:

- Harmat, Z. (2017) Generation and analysis of dynamic structural ensembles of human gastrotropin PhD PROCEEDINGS ANNUAL ISSUES OF THE DOCTORAL SCHOOL FACULTY OF INFORMATION TECHNOLOGY & BIONICS 2017, HU ISSN 2064-7271
- Harmat, Z. (2018) Characterisation of the Subunit Composition, Stoichiometry and Structure of Calcium Channels Involved in Neurotransmission PhD PROCEEDINGS ANNUAL ISSUES OF THE DOCTORAL SCHOOL FACULTY OF INFORMATION TECHNOLOGY & BIONICS 2018, HU ISSN 2064-7271
- Harmat, Z. (2019) Building the three-dimensional structural model of the postsynaptic protein GKAP PhD PROCEEDINGS ANNUAL ISSUES OF THE DOCTORAL SCHOOL FACULTY OF INFORMATION TECHNOLOGY & BIONICS 2019, HU ISSN 2064-7271
- Harmat, Z. (2020) Development and Optimization of a Pipeline to build Structural Ensemble Models of Intrinsically Disordered Protein Segments PhD PROCEEDINGS ANNUAL ISSUES OF THE DOCTORAL SCHOOL FACULTY OF INFORMATION TECHNOLOGY & BIONICS 2020, HU ISSN 2064-7271

Hivatkozások

- [1] Philip E. Bourne (Editor) Jenny Gu (Editor). *Structural Bioinformatics, 2nd Edition*. J. Wiley & Sons, 2 edition, 2009.
- [2] Eric F Pettersen, Thomas D Goddard, Conrad C Huang, Gregory S Couch, Daniel M Greenblatt, Elaine C Meng, and Thomas E Ferrin. UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem*, 25(13):1605–1612, Oct 2004.
- [3] Eric F. Pettersen, Thomas D. Goddard, Conrad C. Huang, Elaine C. Meng, Gregory S. Couch, Tristan I. Croll, John H. Morris, and Thomas E. Ferrin. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci*, 30(1):70–82, Jan 2021.

- [4] Georgii G Krivov, Maxim V Shapovalov, and Roland L Dunbrack. Improved prediction of protein side-chain conformations with SCWRL4. *Proteins*, 77(4):778–795, Dec 2009.
- [5] David Van Der Spoel, Erik Lindahl, Berk Hess, Gerrit Groenhof, Alan E. Mark, and Herman J. C. Berendsen. Gromacs: Fast, flexible, and free. *Journal of Computational Chemistry*, 26(16):1701–1718, 2005.
- [6] Sander Pronk, Szilárd Páll, Roland Schulz, Per Larsson, Pär Bjelkmar, Rossen Apostolov, Michael R. Shirts, Jeremy C. Smith, Peter M. Kasson, David van der Spoel, Berk Hess, and Erik Lindahl. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, 29(7):845–854, 02 2013.
- [7]  Fzl, Zoltn Gspri, Terzia Barna, Florentine Marx, and Gyula Batta. "Invisible" conformers of an antifungal disulfide protein revealed by constrained cold and heat unfolding, CEST-NMR experiments, and molecular dynamics calculations. *Chemistry*, 21(13):5136–5144, Mar 2015.
- [8] Barbara Richter, Joerg Gsponer, Pter Vrnai, Xavier Salvatella, and Michele Vendruscolo. The MUMO (minimal under-restraining minimal over-restraining) method for the determination of native state ensembles of proteins. *J Biomol NMR*, 37(2):117–135, Feb 2007.
- [9] D Kony, W Damm, S Stoll, and WF Van Gunsteren. An improved OPLS-AA force field for carbohydrates. *J Comput Chem*, 23(15):1416–1429, Nov 2002.
- [10] William L Jorgensen, Jayaraman Chandrasekhar, Jeffrey D Madura, Roger W Impey, and Michael L Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2):926–935, 1983.
- [11] Michael Kurz, Volker Brachvogel, Hans Matter, Siegfried Stengelin, Harald Thring, and Werner Kramer. Insights into the bile acid transportation system: the human ileal lipid-binding protein-cholytaurine complex and its comparison with homologous structures. *Proteins*, 50(2):312–328, February 2003.
- [12] Annamria F. ngyn, Balzs Szappanos, Andrs Perczel, and Zoltn Gspri. Consensx: an ensemble view of protein structures and nmr-derived experimental data. *BMC Structural Biology*, 10(1):39, Oct 2010.

- [13] Eldon L. Ulrich, Hideo Akutsu, Jurgen F. Doreleijers, Yoko Harano, Yannis E. Ioannidis, Jundong Lin, Miron Livny, Steve Mading, Dimitri Maziuk, Zachary Miller, Eiichi Nakatani, Christopher F. Schulte, David E. Tolmie, R. Kent Wenger, Hongyang Yao, and John L. Markley. BioMag-ResBank. *Nucleic Acids Research*, 36(suppl_1):D402–D408, 11 2007.
- [14] Gergő Horváth, Orsolya Egyed, and O Toke. Temperature dependence of backbone dynamics in human ileal bile acid-binding protein: implications for the mechanism of ligand binding. *Biochemistry*, 53(31):5186–5198, Aug 2014.
- [15] Berk Hess, Henk Bekker, Herman JC Berendsen, and Johannes GEM Fraaije. Lincs: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, 18(12):1463–1472, 1997.
- [16] Edgar Gabriel, Graham E. Fagg, George Bosilca, Thara Angskun, Jack J. Dongarra, Jeffrey M. Squyres, Vishal Sahay, Prabhanjan Kambadur, Brian Barrett, Andrew Lumsdaine, Ralph H. Castain, David J. Daniel, Richard L. Graham, and Timothy S. Woodall. Open mpi: Goals, concept, and design of a next generation mpi implementation. In Dieter Kranzlmüller, Péter Kacsuk, and Jack Dongarra, editors, *Recent Advances in Parallel Virtual Machine and Message Passing Interface*, pages 97–104, Berlin, Heidelberg, 2004. Springer Berlin Heidelberg.
- [17] Ahmet Bakan, Lidio M Meireles, and Ivet Bahar. ProDy: protein dynamics inferred from theory and experiments. *Bioinformatics*, 27(11):1575–1577, Jun 2011.
- [18] William Humphrey, Andrew Dalke, and Klaus Schulten. VMD: visual molecular dynamics. *J Mol Graph*, 14(1):33–38, Feb 1996.
- [19] Beomsoo Han, Yifeng Liu, Simon W Ginzinger, and David S Wishart. SHIFTX2: significantly improved protein chemical shift prediction. *J Biomol NMR*, 50(1):43–57, May 2011.
- [20] Claus AF Andersen, Arthur G Palmer, Seren Brunak, and Burkhard Rost. Continuum secondary structure captures protein flexibility. *Structure*, 10(2):175–184, Feb 2002.