

Fehérjemolekulák biológiailag releváns belső mozgásainak modellezése sokaság alapú megközelítésekkel

DISSZERTÁCIÓ



Pázmány Péter Katolikus Egyetem
Információs Technológiai és Bionikai Kar

Készítette
Harmat Zita

Témavezető
Dr. Gáspári Zoltán

Budapest, 2022

Ha nem az Úr építi a házat,
építői hiába dolgoznak rajta.
Ha nem az Úr őrzi a várost,
őre hasztalan virraszt felette.

Hiábavaló hajnal előtt kelnetek,
és késnetek a lefekvéssel,
akik a fáradság kenyerét eszitek,
hiszen ő álmában is megad mindent
annak, akit kedvel.

Zsolt 127, 1-2 Káldi Neovulgáta

Tartalomjegyzék

Kivonat	4
Bevezető	8
1. Szakirodalmi áttekintés	11
1.1. Fehérjeszerkezetek predikciója és modellezése	12
1.2. Fehérjék belső dinamikája	14
1.2.1. Mágneses magrezonancia (Nuclear Magnetic Resonance, NMR) spektroszkópiai mérések	15
1.2.2. NMR paraméterek	15
1.2.3. A ligandumkötés modelljei	17
1.2.4. Rendezetlen fehérjék és sokaságaik	18
1.3. A munkám során vizsgált fehérjék áttekintése	19
1.3.1. A gasztrotropin fehérje	19
1.3.2. A miozin VI fehérje SAH doménje	21
1.3.3. A Cd3ε fehérje	21
2. Célkitűzések	22
3. Módszerek	24
3.1. A GROMACS program	24
3.2. A CoNSEnsX+ webszerver	24
3.3. A főlánc torziós szögek hierarchikus Dirichlet modellje	25
3.4. A DIPEND program alkotóelemei	25
3.4.1. A Python3 környezet	25
3.4.2. A ChimeraX program	25
3.4.3. Az SCWRL4 program	26
3.5. A MUMO (Minimal Under-restraining Minimal Over-restraining) módszer	26
3.6. Molekuladinamikai szimulációk	26
3.7. Főkomponens-analízis	28
3.8. A gasztrotropin S^2 adatainak és hidrogénkötéseinek elemzése	29
3.9. A gasztrotropin szerkezetek amid kémiai eltolódás különbségeinek vizsgálata	30
3.10. A gasztrotropin fehérje ligandumkötésének vizsgálata dokkolással	30
4. Eredmények	31

4.1.	A gasztrotropin fehérje vizsgálata	31
4.1.1.	A molekuladinamikából származó sokaságok megfelelnek a kísérleti paramétereknek	31
4.1.2.	A gasztrotropin sokaságok a hordó kinyílásának két külön módját mutatják meg	31
4.1.3.	A PCA elemzésben azonosított két mozgásban résztvevő aminosavak nagymértékben átfednek az NMR módszerrel mért különböző kicserélődési gyakoriságokat mutató mozgásokban érintett aminosav csoportokkal	37
4.1.4.	Az <i>apo</i> formában található rejtett " <i>holo</i> -féle" konformáció részben kitekeredett	38
4.1.5.	A dokkolás eredménye	39
4.2.	A DIPEND eljárás	42
4.2.1.	A program felépítése	42
4.2.2.	Alsokaságok kiválasztása	45
4.2.3.	A programmal előállított szerkezeti sokaságok elemzése	46
5.	Diszkusszió és tézisek	52
5.1.	A gasztrotropin fehérje vizsgálata	52
5.2.	Első téziscsoport	53
5.3.	Fehérjeszerkezet modellezés a DIPEND program segítségével	54
5.4.	Második téziscsoport	55
5.5.	Kitekintés	56
6.	Köszönetnyilvánítás	57
	Irodalomjegyzék	66
	Függelékek	67

Kivonat

A fehérjék nagyon fontos biomolekulák, hiszen az életfolyamatok nagyrésze a fehérjék létén, mozgásán és egymással való kölcsönhatásán alapszik. Nem statikus objektumok, hanem állandó mozgásban vannak, belső dinamikájuk van. Szerkezetüket ezért leginkább atomi szerkezeti sokaságokkal lehet jellemezni. A munka során három fajta fehérjét modelleztem mások által közzétett kísérleti adatok alapján: a globuláris gasztrotropin fehérje, a fibrilláris miozin VI SAH doménje és a rendezetlen CD3 ϵ fehérje citoplazmikus doménje.

A gasztrotropin fehérje epesavakat köt és azok szállításában vesz részt. Ligandumkötésének atomi szintű részletei a sokrétű NMR mérések ellenére nem teljesen tisztázottak. A mérések egy az *apo* állapottal lassú kicserélődésben lévő nagy energiájú állapotot is mutattak, mely hasonlít a *holo* állapothoz. A mérés során $|\Delta\omega(^{15}N)|$ paramétereket is meghatároztak.

Kutatásom során olyan molekuladinamikai szimulációkat végeztem az *apo* és a *holo* NMR-rel Tőke Orsolya kutatócsoportja által meghatározott atomi szerkezeteken, melyekben az általuk végzett NMR mérésekből származó NOE és S^2 paramétereket építettem bele a Minimal Under-restraining Minimal Over-restraining protokollt használva. Az ebből származó szerkezeti sokaságok megfelelően visszaadják az általuk mért kémiai eltolódási paramétereket. Megkötés nélküli molekuladinamikai szimulációkból egy nagyméretű sokaságot nyertem, melyet alkotó minden egyes szerkezetnél egyenként az *apo* állapot kémiai eltolódásaihoz képesti ^{15}N kémiai eltolódások különbségeit összehasonlítottam a Tőke Orsolya csoportja által mért $|\Delta\omega(^{15}N)|$ értékekkel. Sikertült néhány olyan konformert azonosítani, melyek nagy korrelációt és kis RMSD-t eredményeztek ezekkel a paraméterekkel összehasonlítva. Ezek egy részénél a helikális sapka és a C/D hurok részben kitekerekedik. Az eredmények alapján a ligandumkötésben a "portál hipotézissel" összhangban ennek a részleges kitekerekedésnek szerepe van.

Kutatásom során egy olyan programot fejlesztettem, mely úgy prediktálja 100 aminosavnál rövidebb rendezetlen fehérjeszakaszok atomi szintű szerkezeti modelljeit, hogy a Dunbrack labor által meghatározott valószínűségek alapján véletlenszerűen választ torziós szöveget a szekvenciális szomszédokat figyelembe véve. Emellett a valószínűségi eloszlás kombinálható egy a felhasználó által megadott eloszlással aminosavanként, így már ismert *a priori* szerkezeti preferenciák is modellezhetőek. A humán Cd3 ϵ rendezetlen szakaszán és a miozin fehérje SAH doménjén mutattam be az eljárás használatát. A generált szerkezetek jó megfelelést mutatnak az e szakaszokra meghatározott NMR kísérleti adatokkal.

Fehérjeszerkezetek szekvencia és *a priori* szerkezeti preferenciák alapján végbemenő atomi szintű modellezése remélhetőleg közelebb viszi a tudományos közösséget a fehérjék

funkciójának és működésének és ezen keresztül sok biológiai folyamat jobb megértéséhez.

Abstract

Proteins are important biomolecules, because a lot of biological processes are based on the existence and dynamics of proteins and their interaction. Proteins are not static molecules, they are constantly in motion, they have internal dynamics. Therefore their structure can best be modelled with an ensemble of atomic level structures. In the course of this work, three kinds of proteins were modelled based on experimental data published by others: the globular protein gastrotropin, the fibrillar myosin VI SAH domain and the intrinsically disordered Cd3ε cytosolic domain.

The protein gastrotropin binds bile acids and plays a role in their circulation. The atomic level details of its ligand binding is still not entirely elucidated in spite of several NMR measurements. These experiments revealed a high energy state of the protein, which is in slow exchange with the *apo* state and which is similar to the *holo* state. In these measurements $|\Delta\omega(^{15}N)|$ parameters were also determined.

On the course of my research I performed molecular dynamics simulations on the *apo* and on the *holo* state on atomic structure models determined by others with NMR measurements. These simulations incorporate NOE and S^2 parameters published by others. The simulations use the a Minimal Under-restraining Minimal Over-restraining protocol. The ensembles derived from these simulations have a good correspondence with the chemical shift parameters determined by the group of Orsolya Tóke. I also generated a large pool of conformers with unrestrained simulations. I compared the ^{15}N chemical shift differences of the conformers compared to the *apo* structures. These differences were compared with the $|\Delta\omega(^{15}N)|$ parameters. I successfully identified a few conformers, which have a good correlation and small RMSD with these parameters. In some of these conformers the helical cap and the C/D loop is partially unfolded. The results point to the role of this unfolding in the ligand-binding of the protein in accordance with the "portal hypothesis".

In the course of my research I developed a program which predicts the atomic level structure of protein segments shorter than 100 amino acids using probabilities derived by the Dunbrack lab and chooses dihedral angles randomly based on these probabilities. The program can combine this distribution with a user defined distribution residue by residue, therefore being able to model already known *a priori* structural preferences. The working of the program has been demonstrated with an intrinsically disordered segment of the human Cd3ε and the helical part of the porcine myosin VI protein. NMR experimental parameters are available in the scientific literature for these segments. The generated structures show good correlation with these parameters.

The atomic level modelling of protein segments based on sequence and *a priori* structural preferences hopefully brings the scientific community closer to the understanding of the function and working and hence the understanding of many biological processes.

Közlemények

Jelen disszertáció a következő közlemények alapján készült:

- Harmat Z, Szabó AL, Tőke 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:

- Konferencia absztraktok:
 - 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

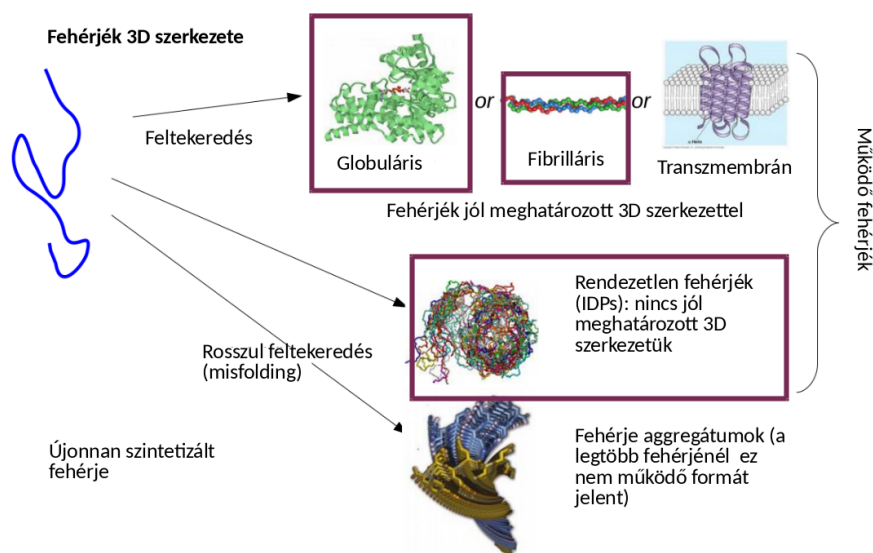
Rövidítések listája

Rövidítés	Angol teljes kifejezés	Megfeleltetés magyarul
BMRB	Biological Magnetic Resonance Data Bank	biológiai mágneses rezonancia adatbázis
CPMG	Carr-Purcell Meiboom-Gill Relaxation Dispersion	Carr-Purcell Meiboom-Gill relaxációs diszperzió
CA	α carbon atom	α szénatom
ED	Essential Dynamics	eszenciális dinamika
I-BABP	ileal Bile Acid-Binding Protein	vékonybélbeli epesavkötő fehérje
IDP	Intrinsically Disordered Protein	rendezetlen fehérje
iLBP	intracellular Lipid-Binding Protein	intracelluláris lipiddkötő fehérje
ITAM	Immunoreceptor Tyrosine-based Activation Motif	immunoreceptor tirozin-alapú aktivációs motívum
FABP	Fatty Acid-Binding Protein	zsírsavkötő fehérje
GCA	Glycocholic Acid	glikokólsav
GCDA	Glycochenodeoxycholic Acid	glikokenodezoxikólsav
GUI	Graphical User Interface	grafikus felhasználói felület
NMR	Nuclear Magnetic Resonance	mágneses magrezonancia
NOE	Nuclear Overhauser Effect	nukleáris Overhauser effektus
MD	Molecular Dynamics	molekuladinamika
MHC	Major Histocompatibility Complex	nagy sejtkompatibilitási komplex
MUMO	Minimal Under-restraining Minimal Over-restraining	minimális alulhatározás minimális túlhatározás módszer
PCA	Principal Component Analysis	főkomponens analízis
PDB	Protein Databank	fehérje adatbázis
RDC	Residual Dipolar Coupling	reziduális dipoláris csatolás
RMSD	Root Mean Square Deviation	legkisebb négyzetes eltérés
SAH	Single α -Helix	magányos α -hélix
SAXS	Small-angle X-ray Scattering	kisszögű röntgen szórás
TCR	T cell antigen Receptor	T-sejt antigén receptor

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 (1. ábra).



1. ábra. A fehérjefeltokeredés különböző fázisai és lehetőségei. Lila téglalappal vannak jelölve azok a fajta fehérjeszerkezetek, melyek tanulmányozása e munka része. Dobson és munkatársai [Dobson, 2003] cikke alapján rajzolva.

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 -szálas régiót tartalmaznak, és alakjuk megközelítően gömb alakú. A fibrilláris típusba a jellemzően hosszú fonalszerű alakokkal 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 mozgékonyosságuk miatt számos konformáció közötti folyamatos átrendeződésben vannak.

A fehérjék alakjának változatossága döntő fontosságú szerteágazó 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 röntgenkristallográfia 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 kísérleti adatokkal való kombinálása. A modellek segítenek megérteni, pontosítani a kísérleti adatokat, a kísérleti adatok pedig realiztikusabbá 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 a dologzatban később bemutatandó 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 pontosságát 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 [Jenny Gu (Editor), 2009].

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 CD3 ϵ citoplazmikus doménje pedig a funkcionálisan rendezetlen fehérjék közé tartozik. A gasztrotropin esetében megkötéseket tartalmazó molekuladinamikai szimulációkat futtattam, míg a másik két fehérje esetében egy előzetes konformerkönyvtárat hoztam létre, melyből szelekcióval állítottam elő a kísérleti paramétereknek megfeleltethető sokaságokat. A konformerkönyvtárak előállítását az általam fejlesztett DIPEND (DIordered Protein Ensembles from Neighbor-dependent Distributions) program 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 CD3 ϵ fehérje esetében alkalmazunk, két lépése van. Az elsőben egy nagyszámú szerkezetből álló, diverz konformációs teret lefedő sokaságot hozunk létre a DIPEND programmal. Ennek egyenletesen le kell fednie a fehérje konformációs terének minél nagyobb részét. A második lépésben ebből választunk a CoNSEnsX⁺ program segítségével egy alsokaságot, mely megfelel a kísérleti adatoknak. A két lépés különválasztása azért előnyös, mert így mindkét lépés tetszés szerint alakítható az adott fehérje sajátosságait figyelembe véve.

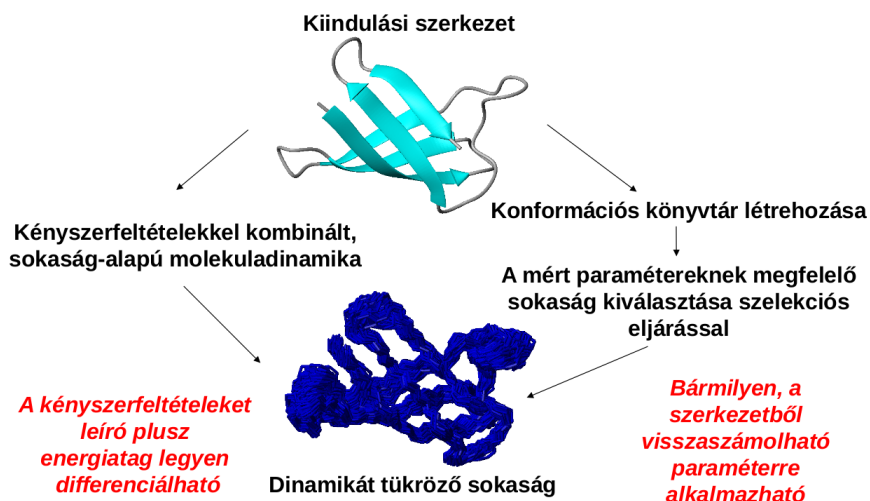
A fentebb leírtakkal összhangban a munka során a szerkezeti modellek validálásában szakirodalmi adatokból származó kísérleti adatokat használunk fel.

1. fejezet

Szakirodalmi áttekintés

Minden fehérjeszerkezet modellnek tekinthető, bármennyire kísérleti adatokon is alapul a szerkezet létrehozása, hiszen ez a teljes fehérje működésének csak néhány aspektusát tudja leírni. Ezeket a modelleket különböző megközelítéssel, módszerrel lehet létrehozni. Vannak olyan módszerek, melyek az adott fehérjére nem használnak kísérleti adatokat, hanem rokon fehérjékre vagy fehérjékre általánosságban vonatkozó általános adatokat használnak fel, ilyenek a szerkezet predikcióból származó modellek.

Vannak ún. kísérleti szerkezetmeghatározásból származó modellek, ahol az adott fehérjéből származó kísérleti adatokat használnak fel a modellépítéshez, de itt is szükséges általános adatok felhasználása, ez történik például egy molekuladinamikai szimulációnál az erőter használataánál, hiszen annak adatai nem az adott fehérjére vonatkoznak. Van olyan szerkezeti modellezés is, mely e kettő között helyezkedik el, azaz kísérleti paraméterekkel megkötött molekuladinamikai szimulációkból származó szerkezeti sokaság mint szerkezeti modell jön létre vagy egy generált szerkezeti sokaságból kísérleti adatok segítségével választott alsokaság lesz a szerkezeti modell.



1.1. ábra. Egy szerkezetsokaság generálásának kétféle módja

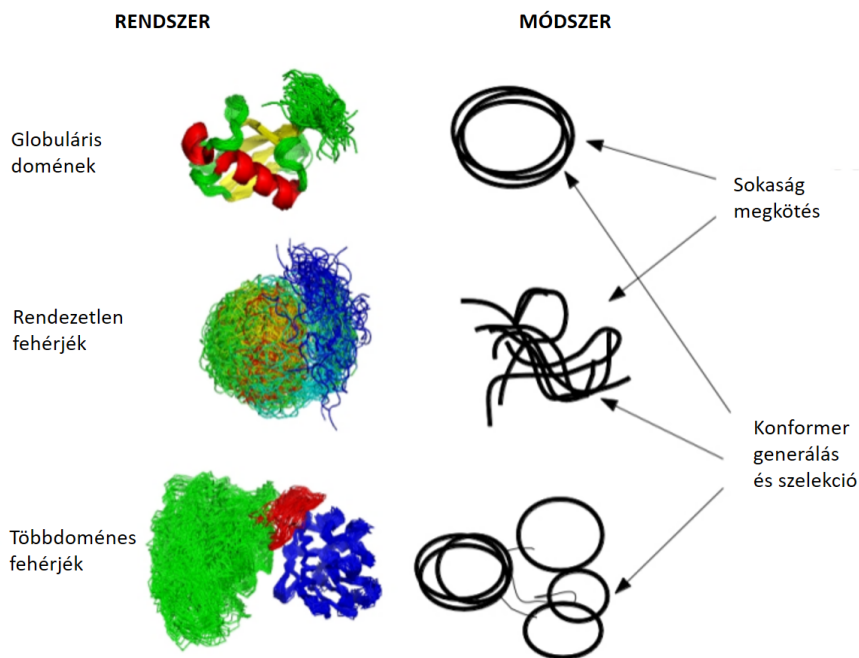
Az 1.1. ábrán összefoglaltam a kétféle szerkezeti modell generálását: a kísérleti adatokkal megkötött molekuladinamikai szimulációk és a nagyszámú konformerből kísérleti adatok alapján történő szerkezeti sokaság választását.

1.1. Fehérjeszerkezetek predikciója és modellezése

Az elmúlt évtizedekben nagyon sok fehérje szekvenciáját határozták meg, azonban atomi szerkezete csak jóval kevesebb fehérjének ismert. A PDB adatbázisban [Berman et al., 2000] fellelhető szerkezetek száma legalább három nagyságrenddel kisebb, mint a UniProt adatbázisban [UniProt, 2020] fellelhető szekvenciák száma [Pakhrin et al., 2021]. A fehérjék szerkezetének kísérleti meghatározása sok időt és pénzt emészt fel, ezért egyre népszerűbbek az *in silico* predikciós eljárások. A fehérjeszerkezetek predikciója során sok fehérje szerkezetéből származó összesített adatokat építenek be a predikciós módszerbe, mely ezen adatok alapján határoz meg egy lehetséges fehérjeszerkezetet vagy szerkezeti sokaságot, azonban az adott fehérjéről, mint rendszerről nem épít be semmilyen adatot a modelljébe, vagyis nem rendszerspecifikus. Még mindig sok a megválaszolatlan kérdés a megdöbbentően jó eredményeket produkáló módszerek, mint pl. az AlphaFold2 [Pakhrin et al., 2021] megjelenése ellenére. Az AlphaFold2 a fehérje szekvenciájához megkeresi a rokon szekvenciákat, majd ezeket egymáshoz illesztve megkeresi a konzervált és kevésbé konzervált aminosavakat. Ezt az információt és az esetleg elérhető templát szekvenciákból az aminosavak közötti távolságinformációkat felhasználva aminosavanként meghatározza, hogy melyek vannak közel egymáshoz, és melyek vannak távolabb egymástól. Az AlphaFold2 ezt és a lokális szerkezetre vonatkozó információt egységben kezeli, és a mély neurális hálózaton többször is átfuttatja ezeket az adatokat, végül egy rövid optimalizáció után a predikált szerkezetet ad ki, mely egydoménes fehérjék esetében igen pontosnak bizonyul [Pakhrin et al., 2021]. Lényeges aspektusa azonban a módszernek, hogy bár a becsült jósági paraméter az egyes régiók esetében összefüggésben van a belső dinamikával, így pl. az AlphaFold2 a rendezetlenség becsülésére is használható, az egyetlen szerkezet, amelyet kiad, önmagában nem képes a fehérje belső mozgásának, belső dinamikájának modellezésére, ami például rendezetlen fehérjék esetében különösen is fontos.

A fehérjeszerkezetek kísérleti adatok alapján való modellezése és fehérjék szerkezetének meghatározása hasonlóak abban, hogy mindkettő esetben kísérleti adatokat, valamint a szerkezet predikciókban is használt *a priori* adatokat is használunk. Ezek például a kötéshosszak, kötőszögek és torziós szögek, vagyis olyan adatok, melyeket egy molekuláris dinamikai erőter is definiál. A kísérleti adatok pedig például NMR mérésekből származó paraméterek is lehetnek. A különbség a szerkezet meghatározás és modellezés között, hogy a modellezésnél kevesebb kísérleti adatot használunk, mint a szerkezetmeghatározásnál, valamint a modellezésnél előfeltételezésekkel is élünk, melyek általában szerkezet meghatározásnál nem állnak rendelkezésre. A fehérjeszerkezetek predikciója pedig egyáltalán nem használ fel kísérleti adatokat, melyek az adott fehérjére vonatkoznak. A tudományos közösség nagyarányú erőfeszítése ellenére sok fehérje esetében a belső dinamika még mindig tisztázatlan. Egy peptidlánc torziós szögei nagyon sok kombinációt alkothatnak, ahogyan ezt Levinthal is már évtizedekkel ezelőtt megírta [Levinthal, Cyrus, 1968]. A rendezetlen fehérjeszakaszok esetében, ahol kevés az α -helikális illetve β -szálas másodlagos szerkezeti szakaszok aránya, még inkább bővül a lehetőségek tárháza, hiszen még az α -helikális és β -szálhoz tartozó szögtartomány-megkötések sem érvényesek. Amikor egy

fehérje kezdő szerkezeti sokaságát építjük fel, mindenképpen le kell szűkíteni egy heurisztikával a lehetőségeket. Az egyik lehetőség erre a modell alulról felfelé megépítése, ahol a lokális kölcsönhatásokra építünk, vagyis a szekvenciális szomszédok egymással való kölcsönhatására.



1.2. ábra. A különböző rendszerek, és sokaság-modellezésükhöz illő módszerek, adaptálva kutatócsoportunk 2013-ban megjelent cikkéből [F. Ángyán and Gáspári, 2013]

Az egyik legjobb módszer rendezetlen fehérjék szerkezeti modellezésére egy diverz, nagyszámú szerkezeti sokaság generálása, majd abból NMR és SAXS (Small-angle X-ray Scattering) mérésekből számított kísérleti paraméterek alapján egy alsokaság kiválasztása (1.2. ábra) [F. Ángyán and Gáspári, 2013]. E paraméterek a fehérje mozgásainak sokaság- és időátlagát írják le, emiatt e paramétereknek csak egy szerkezeti modellekből álló sokaság tud megfelelni, nem várható el, hogy egyetlen szerkezet teljesítse ezt a megfelelést, ezért beszélünk alsokaságról, a röntgenkristallográfiában megszokott egyetlen szerkezeti modell helyett. E sokaságok tárolására jött létre a Protein Ensemble Database [Lazar et al., 2021],[Quaglia et al., 2021].

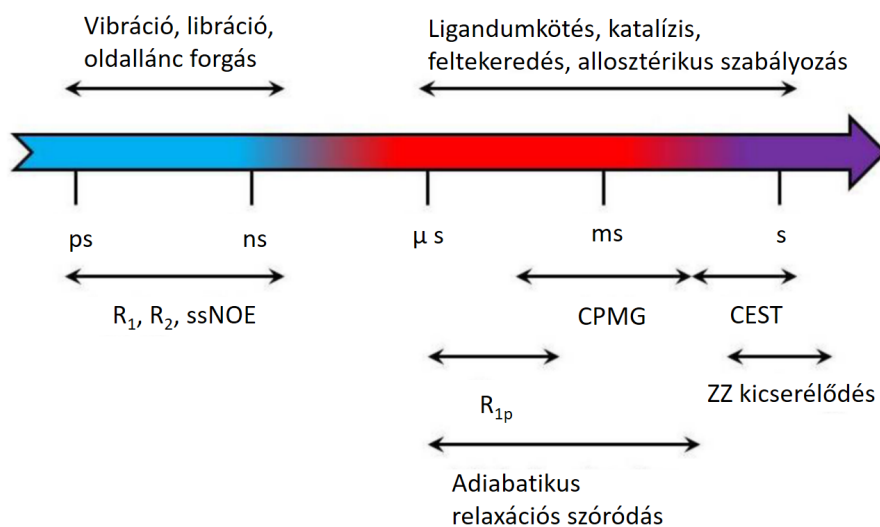
A rendezetlen fehérjeszakaszok modellezésére több program is rendezkedésre áll. Az egyik az ENSEMBLE [Krzeminski et al., 2013], amely kísérleti adatokkal dolgozik, például kémiai eltolódásokkal, és olyan modellsokaságot épít a Trades program segítségével [Feldman and Hogue, 2002], mely megfelel ezen adatoknak. Sajnos a programot nem tudtam használni, mert a mai Linux rendszerekkel már nem kompatibilis a nyilvánosan elérhető verzió.

Egy másik ilyen program a Flexible Meccano [Ozenne et al., 2012], amely fehérje kristályszerkezetek rendezetlen (nem α -helikális és nem β -szálas) szakaszaiból vett fragmentsekből számol valószínűségi eloszlást a torziós szögekre, és így épít fel szerkezeti sokaságokat. Ezt a programot sem sikerült működésre bírnom a program által használt elavult Java verzió miatt.

Egy már meglévő szerkezetsokaságból kísérleti adatok alapján való választáshoz több szoftver áll rendelkezésre. A kutatócsoportunk által fejlesztett CoNSEnsX⁺ webserver [Ángyán et al., 2010] új verziója [Dudola et al., 2017] módosított "mohó szelekció" eljárással képes alsokaságot szelektálni, mely alsokaság a bemeneti konformerek segítségével elérhető legjobb megfelelés – a felhasználó által beállított súlyozás melletti – alsó becslésének tekinthető. Emellett a webszerver támogatja a BME (Bayesian Maximum Entropy) eljárást oly módon, hogy az annak használatához szükséges bemeneti állományokat előállítja, illetve a BME kimeneteként kapott súlyozást felhasználja a megfelelés számításához és megjelenítéséhez [Bottaro et al., 2020], [Dániel Dudola, 2020]. Sanchez-Martinez és Crehuet által írt maximum entrópia alapján RDC adatokat felhasználva választ alsokaságot [Sanchez-Martinez and Crehuet, 2014].

Ez a fajta modellezés hasonlít az NMR adatok alapján lévő szerkezetmeghatározás elvére, mivel mindkettőben alulhatározott a probléma, azaz jóval kevesebb kísérleti paraméter áll rendelkezésre, mint a rendszer szabadsági fokai.

1.2. Fehérjék belső dinamikája



1.3. ábra. Időskálák, melyeken fehérje belsődinamikai mozgások végbemennek, és az őket vizsgáló kísérleti módszerek, adaptálva Chao és Byrd cikkéből [Chao and Byrd, 2020].

A fehérjék belső mozgásai különböző időskálákat ölelnek fel, e mozgások jellemzően más-más típusú biológiai funkciókkal hozhatóak összefüggésbe (1.3. ábra). A ps-ns skálán a fehérjék kötéseinek vibrációja, oldalláncainak mozgása történik, a μ s-tól másodperces időskáláig lejátszódó mozgások pedig a ligandum kötés, katalízis, feltekeredés, allostérikus szabályozás [Chao and Byrd, 2020]. A gyors mozgások a fehérjék konformációs entrópiájának is lényeges meghatározói lehetnek.

1.2.1. Mágneses magrezonancia (Nuclear Magnetic Resonance, NMR) spektroszkópiai mérések

A mágneses magrezonancia a mágnesesen aktív atommagok gerjesztésén alapuló módszer, oldat- és szilárd fázisban is alkalmazható. Munkám során oldatfázisú vizsgálatok eredményeit használtam fel. Ennek során a (citoplazmikus vagy extracelluláris) fehérjéket természetes környezetükhöz nagyon hasonló körülmények között vizsgáljuk egy mintacsőben. Az erős mágneses térbe helyezett atommagokat ún. rádiófrekvenciás sugárzással gerjesztve és az ebből származó jeleket detektálva és feldolgozva – itt tovább nem részletezett módon – a fehérje térszerkezetének és belső dinamikájának különböző aspektusaibe nyerhetünk betekintést, ahogy ez az 1.3. ábrán is látszik. Az NMR-spektroszkópia két nagyon fontos aspektusa az atomi szintű felbontás, azaz hogy az egyes aminosavak adott atomjairól egyedileg kaphatunk információt, illetve a különféle mérések által felölelt, igen tág időskála.

1.2.2. NMR paraméterek

A kémiai eltolódás paraméterek fontos információt adnak a fehérje szerkezetéről és belső dinamikájáról. A kémiai eltolódás – némileg leegyszerűsítve – az adott atomra egyedileg jellemző rezonanciafrekvencia, mely függ az atom kémiai típusától és a körülötte lévő elektronsűrűségtől, melyet a kémiai kötésekön kívül a térben közel lévő csoportok is befolyásolnak. Ezáltal a fehérjemolekulák esetében a kémiai eltolódások szorosan összefüggnek a másodlagos és harmadlagos szerkezettel.

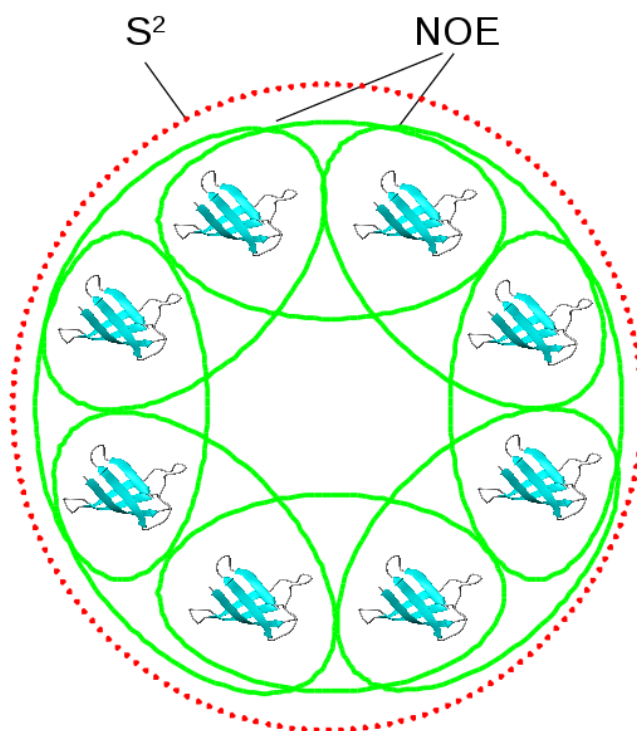
A CA és a HA kémiai eltolódások arról tudósítanak, hogy a peptidláncnak milyen a lokális konformációja. Titrálási kísérletekben a kémiai eltolódások változásainak követése információt adhat a kötőhelyről, a kölcsönhatás során bekövetkező szerkezeti-dinamikai változásokról és a kötési állandóról is [Gáspári and Perczel, 2010].

Az S^2 általános rendparaméter az amid N-H kötés mozgékonyaságáról ad információt¹. Értéke 0 és 1 között van. A 0 értéknél a kötés teljes konformációs szabadságot élvez, 1-nél teljesen megkötött a mozgása, vagyis csak a fehérjemolekulával együtt mozog a kötés [Gáspári and Perczel, 2010]. Definíciója mindig egy sokaságra vonatkozik, hiszen egy szerkezetre értéke mindig 1. Kiszámolása előtt a sokaság szerkezeteinek illesztése szükséges. [F. Ángyán and Gáspári, 2013]. E paraméter NMR relaxációs adatokból való származtatásának módját Lipari és Szabó dolgozta ki [Lipari and Szabo, 1982].

Az NOE (Nuclear Overhauser Effect) paraméterek atomok távolságáról adnak információt, így a fehérjék atomi szintű szerkezetmeghatározásához használhatóak. Legtöbbször a hidrogénatomok közötti távolságát használjuk. Ezek a távolságok a fehérjék folytonos mozgásai miatt nem egzaktak, hanem távolságsávok határozhatók meg. Két közel, lévő, de egymással át nem lapoló hidrogén atom távolsága 1.8 Å. A sávok tipikusan 1.8-2.5, 1.8-3.5 illetve 1.8-5.0 Å távolság értékekbe esnek. Amennyiben nagyon flexibilis a fehérje, akkor csak a szekvenciában egymás melletti aminosavak hidrogénjei vagy egy aminosavon belüli hidrogének távolságáról kapunk információt [Gáspári and Perczel, 2010]. Sokaság-

¹Létezik metilcsoportra vonatkoztatott változata is, az jelen dolgozat szempontjából nem releváns

átlagolásuk úgy történik, hogy egy azonos atomcsoport távolságainak a $-6.$ hatványát vesszük (r^{-6}), majd összegzés után ennek megfelelő fokú gyököt vonunk. Kényszerfeltételként való alkalmazásuk multireplika szimulációk esetében ideálisan szerkezetpárok között történik a túlillesztés elkerülése végett, ahogyan ezt a MUMO (minimal under-restraining minimal over-restraining) megközelítés is teszi [Richter et al., 2007], ugyanis ez az átlagolási módszer kettőnél több molekula esetére alkalmazva akár nagyon nagy egyedi eltéréseket is megengedhet.



1.4. ábra. A MUMO (Minimal Under-restraining Minimal Over-restraining) módszer ábrája, Richter és munkatársai cikke [Richter et al., 2007] alapján rajzolva.

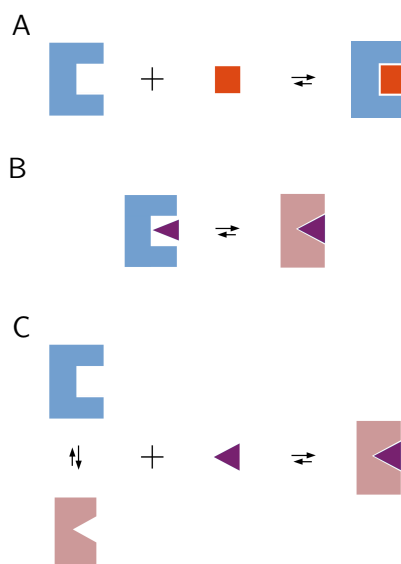
Az RDC (residual dipolar coupling, maradvány diopoláris csatolás) paraméterek sokat elárulnak a fehérjék szerkezetéről és belső dinamikájáról, egészen a μ s-os mozgástartományig [F. Ángyán and Gáspári, 2013]. Az atommagok közötti vektorok relatív orientációját és dinamikáját írják le. Alapvetően az adott atomok közötti vektor és a külső mágneses tér kölcsönös irányultságától függenek, ezért a hagyományos oldatfázisú mérésekben, vagyis izotróp esetben kiátlagolódnak a fehérje forgó mozgása miatt, mert a két atom által a mágneses térhez képest meghatározott vektor egyenletes valószínűséggel veszi fel a lehetséges térbeli irányokat. Azokban az esetekben tudják őket mérni, ahol a vektor iránya nem véletlenszerű. Ennek eléréséhez a fehérjéket olyan ún. rendezett közegbe (oriented medium) teszik, mely felhígított kristályszerű szerkezettel rendelkezik, ilyenek a bicellák és a bakteriofágok. Jó esetben ez a kísérleti elrendezés nem változtatja meg a fehérjék dinamikáját [Gáspári and Perczel, 2010]. Kiértékeléséhez szükség van a molekula mágneses térhez képest való orientációjának megbecslésére, amely orientáció egy sokaság tagjainál különböző lehet, szerkezetről szerkezetre változhat szterikus és elektrosztatikus okok miatt is [F. Ángyán and Gáspári, 2013]. Megjegyzendő, hogy iga-

zán részletes dinamikai információ kinyeréséhez többféle közegben mért, egymástól lineárisan független RDC paraméterkészletek szükségesek, és ezek kinyerése messze nem triviális [Higman et al., 2011].

Mivel e paraméterek mindegyike más-más aspektusát emeli ki a szerkezeti modelleknek, több paraméter egyidejű beépítése a modellalkotásba növeli annak pontosságát. Az egyik ilyen módszer a MUMO (Minimal Under-restraining Minimal Over-restraining) [Richter et al., 2007] (1.4. ábra), melynek során több (általában 8) párhuzamos molekuladinamikai szimulációt (replikát) futtatunk, ahol egy-egy időpillanatban a 8 replika S^2 paraméter átlagának meg kell felelnie egy adott értéknek minden aminosavra, a NOE megfeleléseknek pedig szerkezetpáronként kell teljesülnie. Ehhez némileg hasonló módszert használtak többek között az ubikvitin fehérje gyors időskálájú belső dinamikájának feltérképezésére [Lindorff-Larsen et al., 2005].

1.2.3. A ligandumkötés modelljei

Fehérjék ligandumkötésének magyarázatára az évtizedek során több elmélet is született. A Fischer-féle kulcs-zár modell a legegyszerűbb közülük: itt a fehérje és a ligandum is statikus, és megtalálva egymást pontosan egymásba illenek (1.5. A ábra). Ez az elmélet igen fontos mérföldköve volt a molekuláris komplementaritás felismerésének, és megtermékenyítőleg hatott a fehérje-ligandum kölcsönhatást vizsgáló kutatásokra, azonban mivel a partner molekulákat inherens módon statikusnak kezeli, a biomolekulák dinamikus mivoltának felismerésével pontosításra szorult. Jelenleg két másik fajta elmélet létezik: a Koshland-féle indukált illeszkedés modell, mely szerint a ligandumkötés indukál a fehérjén egy olyan konformációváltozást, melynek következtében a fehérje felveszi a ligandumkött konformációt, mely már jobban illeszkedik a ligandum szerkezetéhez (1.5. B ábra).



1.5. ábra. A ligandumkötés háromféle elmélete (A) Fischer-féle kulcs zár modell (B) Indukált illeszkedés (C) Konformációs szelekció. Boehr és munkatársai cikke alapján rajzolta [Boehr et al., 2009]

A másik elmélet a konformációs szelekció elmélete. Ez szerves kapcsolatban van a fehérje konformációs tájkép elméletével, mely szerint egy fehérje nagyon sok különböző,

hasonló energiaszintű konformációt felvehet, melyek között ott vannak azok a valamivel magasabb energiájú konformációk is, melyek a kötött állapothoz nagyon hasonlóak, csak ligandum nélküliek (1.5. C ábra). E konformációk összessége, ha az energiaszinteket ábrázoljuk, egy hegyes-völgyes tájképhez hasonlít. Az alacsonyabb energiaszintű konformációk, azaz a völgyek jobban populáltak, vagyis több van belőlük, mint a többiből. A konformációs szelekció elmélete szerint a ligandum megjelenése ezt a tájképet alakítja át úgy, hogy a magasabb energiájú, a kötött állapothoz hasonló konformációkból több legyen, melyekbe a ligandum be is köt. Boehr és munkatársai mindezekről leírják [Boehr et al., 2009], hogy a fehérjék ligandumkötésében, vagy akár fehérje-fehérje kölcsönhatásokban, sőt RNS-fehérje kölcsönhatásokban is a konformációs szelekció és az indukált illeszkedés is szerephez jut. A ligandum a konformációs szelekció szerint választja ki a neki megfelelő konformációt, azonban az oldalláncok és a főlánc kölcsönhatásának optimalizációjában az indukált illeszkedés játszik szerepet, vagyis a ligandum kötése alakítja ki az oldalláncok és a főlánc viszonyának kötött állapotra jellemző formáját [Boehr et al., 2009]. E szerzők szerint az, hogy a két mechanizmus közül melyik jut érvényre, attól függ, hogy a fehérje és a ligandum között eleve milyen erős az interakció. Ha erős, akkor az indukált illeszkedés jut nagyobb szerephez, ha gyengébb, akkor a konformációs szelekció lesz domináns [Boehr et al., 2009].

Annak kiderítésére, hogy a fehérje végtelenül sok konformációs lehetősége közül melyek azok, amelyek gyakran megjelennek, és melyek azok, amelyek hasonlóak a ligandumkötöthöz, a kutatók kísérleti módszereket alkalmaznak. Ilyenek a röntgenkristallográfia, a kisszögű röntgenszórás (SAXS), a mágneses magrezonancia (NMR) mérések és egyéb fehérjeszerkezet meghatározó módszerek. Jelen munka során NMR mérések adataival foglalkozunk, ezek közül is különösképpen az S^2 , NOE, kémiai eltolódás és RDC paraméterekkel.

1.2.4. Rendezetlen fehérjék és sokaságaik

Vannak olyan fehérjék, amelyekben nincs vagy nagyon kevés az olyan szakasz, amelynek stabil térszerkezete lenne. Ezek a fehérjék különösen sok konformációt felvehetnek, így nagyon diverz dinamikával és szerkezettel rendelkeznek. Sok biológiai folyamatban részt vesznek, például a sejtfolyamatok transzkripció és transzláció általi szabályozásában, a fehérjék feltekeredésében mint chaperon fehérjék, a DNS kondenzálásában, az mRNS érésében és splicingjában, a fehérjék szállításában, stb. A rendezetlen fehérjék meghibásodása a homeosztázis felbomlásához vezet, például neurodegeneratív betegségekhez, rákhoz, cukorbetegséghez, szív- és érrendszeri betegségekhez. Aggregálódásuk ún. prion betegségekhez is vezet. Emellett kórokozók, például vírusok rendezetlen fehérjéket használnak arra, hogy a gazdasejt jelátvitelét megzavarják. Ezért is övezi érdeklődés ezen fehérjéket, mint gyógyszer-célpontokat [Dunker et al., 2001], [Tompá, 2012].

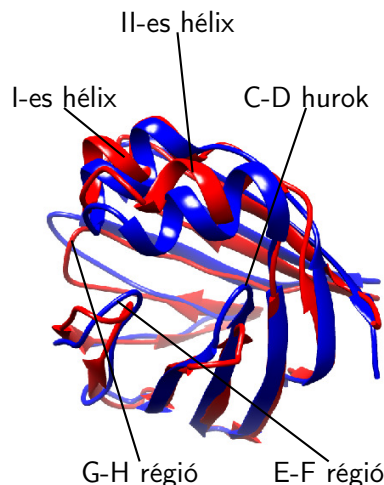
A már említett jelenség, hogy a fehérjék állandóan mozgásban vannak, azaz sok konformer állandó kicserélődésében, még hangsúlyosabb a rendezetlen fehérjék esetében. A legtöbb kísérleti paraméter, mint például az NMR kísérletekből származóak e konformer átlagát írják le [F. Ángyán and Gáspári, 2013]. Ezért eredményezheti a dinamika és

a szerkezet jobb jellemzését a rendezetlen fehérjék sokaság modellezése egy-egy statikus szerkezeti modellhez képest.

1.3. A munkám során vizsgált fehérjék áttekintése

1.3.1. A gasztrotropin fehérje

A gasztrotropin, más néven I-BABP vagy FABP6 az epesavak májból való transzportjában játszik szerepet [Smathers and Petersen, 2011]. Az epesavak a májban szintetizálódnak koleszterinből, majd a proximális vékonybélbe szekretálódnak az epéből, majd újrahasznosulnak a vérkeringésen keresztül [Zwicker and Agellon, 2013], [Trauner and Boyer, 2003]. Ez az újrahasznosítás biztosítja azt, hogy csak egy kis részét kell az epesavaknak szintetizálni. A gasztrotropin fehérje ebben az újrahasznosítási folyamatban játszik szerepet a disztális vékonybél felszívódásban szerepet játszó nyálkahártyasejtjeiben lévő kötési kölcsönhatások útján [Horváth et al., 2012] [Lin et al., 1990], emellett a koleszterin háztartásban is részt vesz [Smathers and Petersen, 2011], [Alrefai and Gill, 2007].



1.6. ábra. Az apo humán gasztrotropin (piros, PDB azonosító 1O1U 7-es modell) és a holo gasztrotropin (kék, PDB azonosító 2MM3 1-es modell). UCSF Chimera-val készült a kép [Pettersen et al., 2004].

A gasztrotropin az intracelluláris lipidkötő fehérjék (iLBP) családjának tagja, mely 15-kDa méretű fehérjékből áll. Ezek zsírsavakat, retinoidokat, koleszterint, és epesavakat kötnek [Veerkamp and Maatman, 1995]. Emellett kimutatták, hogy ezek a fehérjék szerepet játszanak a sejtmagi hormonreceptorok transzkripciós aktivitásának stimulálásában is [Sessler and Noy, 2005] [Ayers et al., 2007] [Nakahara et al., 2005]. Ebben a fehérjecsaládban a gasztrotropin alcsaládja különleges abban a tekintetben, hogy ezek a fehérjék két ligandumot [Tochtrop et al., 2002] [Eliseo et al., 2007] vagy esetleg hármat is tudnak egyszerre kötni [Turpin et al., 2013] [Kouvatsos et al., 2007]. A PDB adatbázisban 1O1U azonosítóval található a gasztrotropin fehérje ligandum nélküli (*apo*) szerkezeti modellje, melyet Kurz és munkatársai határoztak meg. A kolilaurin kötött formát NMR (Nuclear Magnetic Resonance) spektroszkópiával szintén Kurz és munkatársai határozták meg, és 1O1V azonosítóval található meg [Kurz et al., 2003]. Azóta a humán gasztrotropin fehér-

jének egy olyan formáját is meghatározták NMR spektroszkópiával, ahol ligandumként a glikokólsav és a glikokenodezoxikólsav is jelen van [Horváth et al., 2016]. Rokonaihoz hasonlóan a gasztrotropin fehérje is β -hordó formájú, 10 antiparallel β -szálból áll, melyeket A-tól J-ig jelölünk, és két α -hélixből, melyeket I és II számokkal jelölünk. A β -hordón belül egy körülbelül 1000 \AA^3 méretű kötőzseb foglal helyet [Banaszak et al., 1994]. A fehérje ligandumkötése pozitív kooperativitást mutat [Tochtrop et al., 2002], mely a megkötött epesavak hidroxiláltsági mintázata által vezérelt [Tochtrop et al., 2003]. Ennek megfelelően a hidrogénkötések hálózata fontos közvetítő szerepet játszik a ligandumkötés kooperativitásában [Toke et al., 2006]. Amellett, hogy kommunikáció van a két kötőhely között, a di- és trihidroxil epesavak preferenciát mutatnak a kötés helyét tekintve egymás jelenlétében [Tochtrop et al., 2004]. A ligandum nélküli (*apo*) és a ligandumkötött (*holo*) forma összehasonlításából nyilvánvaló, hogy az epesav kötés úgy jön létre, hogy nagy konformációs változás történik az E-F és a G-H régiókban, valamint a C/D hurokban, és a közelebbi helikális sapkában [Horváth et al., 2012] [Turpin et al., 2013] (1.6. ábra). NMR mérések azt mutatják, hogy az *apo* forma lassú kicserélődésben van egy rejtett, kis populációban jelen levő konformációval, amely bizonyos szerkezeti elemekben a *holo* formához hasonló [Horváth et al., 2014]. Érdekes módon azok az aminosavak, amelyek konformációs fluktuációt mutatnak a μs -ms időskálán, két csoportba oszthatók, egy "lassú" és egy "gyors" csoportba, melyek térben úgy tűnik, elkülönülnek. Míg a "lassú" csoport a helikális régió egy részét, a C/D hurkot és a B és D β -szálakat foglalja magában a fehérje N terminális részén, vagyis szekvenciálisan a fehérje első részén, a "gyorsabb" csoport az EFGH fehérjerészben és szekvenciálisan a fehérje C terminális felében található [Horváth et al., 2014]. Fontos annak a tanulmányozása, hogy a ligandumok hogyan jutnak be a fehérje belsejébe, hiszen ott vannak a ligandum kötőhelyek. A leginkább elfogadott magyarázat erre a "portál hipotézis", mely szerint a C/D és az E/F hurkok és a II-es hélix együttese teszi lehetővé a ligandumok bejutását [Hodsdon and Cistola, 1997] [Jenkins et al., 2002] [Ragona et al., 2014]. NMR szerkezeti és dinamikai kísérletek arra mutatnak, hogy egy konformációs szelekciós mechanizmus érvényesül a ligandumkötésben, mely magában foglal egy zárt és egy nyitott konformáció közötti kicserélődést a humán [Horváth et al., 2016] és a csirke májbeli BABP fehérjerokonokban [Cogliati et al., 2010]. A fehérje belső mozgását feltérképező NMR mérések mellett molekuladinamikai szimulációk is alátámasztják korrelált mozgások jelenlétét a humán gasztrotropin fehérjében és az *apo* formában az E-F régió részleges kitekeredésére is rámutatnak [Horváth et al., 2017].

Egy tanulmányban [Horváth et al., 2019] a fehérjében előforduló hisztidin aminosavak hatását vizsgálták a fehérje belső dinamikájára. Termodinamikai és kinetikai mérésekkel kimutatták, hogy a hisztidinek protonált formája kedvez a ligandumok kötésének. Molekuladinamikai szimulációkat is végeztek, és ezek kimutatták, hogy protonált hisztidinek esetén a fehérje szerkezete stabilabb. Tanulmányukban azt a következtetést vonták le, hogy a ligandum belépése és kilépése allosztérikusan van szabályozva, mely pH függő. Ezzel az e munkában felvázolt eredményeket kiegészítették a pH változásra vonatkozó információkkal.

Fehérje neve	Domén	BMRB azonosító	Szekvencia
CD3 ϵ	citoplazmikus	18889	SLKNRKAKAKPVTRGAGAGGRQGRGQNKERPPVNPDPYEP IRKGQRDLYSGLNQRRRI
Miozin VI	SAH	30591	KQQEEEEAERLRRIQEEMEKERKRREEDQRRRKEEEERRM KLEMEAKRKQEEERKKREDDEKRIQAE

1.1. táblázat. A vizsgált CD3E és Miozin VI fehérje domének BMRB azonosítói és szekvenciái

1.3.2. A miozin VI fehérje SAH doménje

A magányos α -hélix (SAH) domének ritkán fordulnak elő, mert önmagukban az α -hélixek nagyon instabilak, mivel hidrogénkötésekért a víz oldószer, illetve a fehérje főlánc hidrogénkötései versenyeznek. Néhány fehérjénél azonban egy különleges mintázat segítségével mégis stabil tud maradni egy ilyen domén. Ebben a mintázatban általában négy negatív töltésű aminosav, például a glutaminsav, és utána négy pozitív töltésű aminosav, például a lizin és az arginin váltják egymást, ezt hívják ER/K mintázatnak. Ez a domén egyfajta konzolgerenda (cantilever arm) szerepet tölt be a fehérje működésében, hidat képezve a domének között [Sivaramakrishnan et al., 2008]. Ahhoz, hogy olyan fehérjeszegmenst modellezzünk, mely *a priori* ismert szerkezeti preferenciával rendelkezik, a miozin VI fehérje α -helikális SAH doménje tökéletes tesztalany. Ez a helikális rész a miozin motorfehérje emelőkarjának része. Ez a fehérje az izmok összehúzódásában játszik szerepet. A 6OBI PDB azonosítóval rendelkező szerkezetet választottuk referenciának, mely az első olyan NMR szerkezet, ami SAH domént tartalmaz. A hozzátartozó kémiai eltolódás adatok BMRB azonosítója 30591 [Barnes et al., 2019], szekvenciája az 1.1 táblázatban olvasható. Barnes és munkatársai e szerkezet meghatározásához többféle NMR paramétert mértek, például ^{15}N , ^{13}C és 1H kémiai eltolódásokat, $^3J_{HNHA}$ skaláris csatolásokat és $^1N - H$, $^2C' - H$, illetve $^1N - C'$ RDC (reziduális dipoláris csatolásokat).

1.3.3. A Cd3 ϵ fehérje

Az immunrendszer tanulmányozása szempontjából fontos a T sejtek aktiválásának folyamatát megérteni, hiszen így tudjuk a kórokozók elleni védekezést segíteni, illetve az autoimmun betegségeket kezelni. A Cd3 ϵ (T cell antigen Receptor TCR) fehérje a Major Histocompatibility Complex (MHC) része, mely egy T sejt receptor komplex. Egy kórokozóból származó peptidszakasz aktiválja ezt a komplexet, melyhez egy antigén bemutató sejt hozza a peptidet. Ez több esemény megtörténtét okozza, több jelátviteli útvonalon hatva (például a RAS, P53 és az NFAT útvonalakon). Ez végül a T limfociták és végső soron az immunrendszer aktiválásához vezet, ami effektor T sejtek generálását jelenti. Emellett ez a fehérje még a kalcium jelátvitelben is szerepet játszik [Huang and Wange, 2004]. Citoplazmatikus doménje nagyrészt rendezetlen, azonban NMR-rel vizsgálható. Az Cd3 ϵ fehérje 153-207 aminosavainak kémiai eltolódásait tartalmazó adatsor BMRB azonosítója 18889 [Isaksson et al., 2013], szekvenciája az 1.1 táblázatban olvasható.

2. fejezet

Célkitűzések

Munkám során célul tűztem ki egyes kiválasztott fehérjék dinamikus szerkezeti sokaságainak előállítását és jellemzését azok belső dinamikája funkcionális szerepének megértése céljából. A vizsgált rendszerek közé a globuláris, β -hordó szerkezetű gasztrotropin, a miozin VI fibrilláris magányos α -helikális nyaki régiója, valamint a CD3 ϵ funkcionálisan rendezetlen citoplazmikus szakasza tartoztak. A csoportunk korábban adaptálta a globuláris fehérjék ps-ns időskálájú mozgásait jellemző S^2 rendparamatétereket beépíteni képes MUMO protokollt a nyílt forráskódú GROMACS molekuladinamikai programcsomagba. Ezen kívül rendelkezésemre állt a Dudola Dániel által továbbfejlesztett CoNSEnsX⁺ eljárás, mely alkalmas az előállított sokaságok elemzésére és a szelekcióval adott paramétereknek jobban megfelelő alsokaság kiválasztására. Nem állt ugyanakkor rendelkezésünkre olyan, nyílt forráskódú és rugalmasan módosítható eljárás, mely fibrilláris és rendezetlen fehérjék sokaságainak modellezésére alkalmas lehetett volna. Ezért ezen rendszerek vizsgálatához új eljárás kidolgozása is szükségessé vált.

A doktori munkám során specifikusan kitűzött céljaim:

1. A gasztrotropin fehérje jellemző belső mozgásainak leírása és azok biológiai jelentőségének elemzése, ezen belül specifikusan:
 - (a) a fehérje *apo* és *holo* állapotát modellező szerkezeti sokaságainak előállításával, amelyek megfelelnek mind a felhasznált, mind független kísérleti adatoknak
 - (b) a gasztrotropin fehérje belső mozgásai szerepének vizsgálata a fehérje ligandumkötésében széles konformációs teret bejáró sokaságok és azok kísérleti adatokkal való összevetésének vizsgálatával
 - (c) a ligandumkötésre vonatkozó portál hipotézis cáfolata vagy megerősítése, illetve a ligandumkötés egyéb jellemzőinek felderítése
2. Rendezetlen és fibrilláris fehérjék szerkezeti sokaságainak előállítására alkalmas eljárás kifejlesztése és ennek segítségével konkrét modellrendszerek vizsgálata:
 - (a) a program alkalmazása a CD3 ϵ fehérje funkcionálisan rendezetlen citoplazmikus doménje belső dinamikáját reprezentáló sokaság előállítására és funkcionális relevanciájának vizsgálatára

- (b) a miozin VI fehérje SAH doménje szerkezeti változatosságának vizsgálata az egyenes, "rúdszerű" szerkezettől való eltérések mértékének modellezése céljából.

3. fejezet

Módszerek

3.1. A GROMACS program

Ezt a hasznos és eszenciális eszközt, melynek neve a Groningen MACHine for Chemical Simulation rövidítése (<https://manual.gromacs.org/documentation/>) a 90-es évek elején kezdték el fejleszteni ANSI C-ben [Van Der Spoel et al., 2005]. Molekuladinamikai számításokat végez, modellezve a fehérjeatomokra ható erőket. Kompatibilis többféle mások által kifejlesztett erőtérrrel, azaz erőparaméterezésekkel. Jól párhuzamosítható és nagyon jól dokumentált. Emellett nyílt forráskódú és szabadon beleépíthetők saját rutinok, modulok. Gasztrotropinnal kapcsolatos számításainkhoz a 4.5-ös verziót [Pronk et al., 2013] használtuk, melyet Gáspári Zoltán évekkel ezelőtt kiegészített egy az S^2 paraméterek sokaságátlagát kényszerfeltételként használható modullal. Más szimulációkhoz pedig a 2020-as verziót használtuk [Abraham et al., 2015].

3.2. A CoNSEnsX+ webszerver

Ezt a programot kutatócsoportunk implementálta. Ebben személy szerint nem vettem részt, csupán felhasználtam a programot a generált sokaságok elemzésére. A webszerver (github.com/PPKEBioinf/consensx.itk.ppke.hu) [Ángyán et al., 2010] [Dudola et al., 2017] egy több modellből álló PDB szerkezetet vár bemenetként és egy paraméterfájlt, mely lehet egy NOE távolságokat tartalmazó NMR-star formátumú fájl, és egy másik NMR-star formátumú fájl, melyben különböző NMR paraméterek vannak megadva. Ezek a PDB vagy a BMRB adatbázisból letölthetőek. A kétféle paraméterfájl egyszerre is alkalmazható. A webszerver különböző mások által fejlesztett programokat (SHIFTX [Neal et al., 2003] és PALES [Zweckstetter and Bax, 2000]) felhasználva visszaszámolja a szerkezeti sokaságból az NMR paramétereket, majd ezeket összehasonlítja a beadott paraméterekkel. Egy új fejlesztés nyomán másodlagos kémiai eltolódásokkal is tud számolni a webszerver szekenciális aminosavsomszédokkal korrigált random coil értékekből [Tamiola et al., 2010]. A skaláris csatolásokat (J couplings) az alapbeállításban szereplő Karplus paraméterekkel számoltuk ki [Wang and Bax, 1996]. A CoNSEnsX⁺ RMSD, vagy Pearson korreláció értékeket számol ki a beadott és sokaságból visszaszámolt értékek között. RDC paraméterek esetén pedig ugyanezt Q faktorialis megteszi. Ezeket a megfelelési paramétereket külön-

féle módon jó minőségben ábrázolja is. Mindemellett egy alsokaságot is ki tud választani a kezdeti sokaságból egy mohó algoritmus használatával a konformációk kombinációinak kísérleti adatokkal való összevetése alapján, vagyis kevés szerkezettel indulva akkor áll meg a kiválasztással, ha újabb szerkezetek hozzáadása már nem javít a sokaság kísérleti paramétereknek való megfelelésén. Az RMSD vagy Pearson korreláció vagy RDC esetén Q faktor közül egyet optimalizál, vagyis a visszaadott alsokaság jobb lesz ebben a paraméterben, mint a kezdeti sokaság. A különböző NMR paraméterek súlyozhatók is a kiválasztás szempontjából, azaz egymáshoz képesti fontosságuk, esetleg megbízhatóságuk, vagy az, hogy mennyire akarjuk az adott paramétert a szerkezeti modellben felhasználni, számszerűleg is megadható. Vagyis például egy elemzésnél RMSD-re optimalizálva a CA kémiai eltolódásokat 10-es súllyal, míg a CB kémiai eltolódásokat 5-ös súllyal, a többi kémiai eltolódást figyelmen kívül hagyva is képes alsokaságot kiválasztani.

Ezt a webszervert használtuk a szerkezeti sokaságok S^2 , RDC és kémiai eltolódás adatoknak való megfelelésének vizsgálatára. Az NOE megfelelések vizsgálatára emellett Perl szkripteket is használtunk. Ez utóbbiakat a sokaságokban r^{-6} átlagolással számoltuk ki intramolekulárisan és a sokaságon belül is.

3.3. A főlánc torziós szögek hierarchikus Dirichlet modellje

Ting és munkatársai [Ting et al., 2010] egy olyan módszert fejlesztettek ki, amellyel a különböző ϕ és ψ szögek valószínűségét tudták kiszámolni egy bizonyos aminosavra, melynek egy bizonyos bal (N terminális felőli) és jobb (C terminális felőli) szomszédja van. Ehhez hierarchikus Dirichlet modellt használtak, mely egy matematikai modell. Csak két egymás utáni aminosavval dolgoztak, mivel három egymás utáni aminosav túl sok esetet, vagyis túl sok alcsoportot eredményezne. Adataikat 21 nagy csoportra bontották, a 20 aminosavra és a prolinra, amelynek *cisz* és *transz* állapotát is megkülönböztették. Ezeket a csoportokat 20-20 alcsoportra bontottuk a szerint, hogy melyik a szomszéd aminosav. A kiszámolt valószínűségek Gauss eloszlások kevert összegei.

3.4. A DIPEND program alkotóelemei

3.4.1. A Python3 környezet

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 ismertetőjele, hogy a programblokkokat zárójelek helyett a behúzás mértékével választjuk el.

3.4.2. A ChimeraX program

Az UCSF egyetemen fejlesztett Chimera molekulanézegető és molekulamodellező program [Pettersen et al., 2004] új változata, a ChimeraX [Pettersen et al., 2021] (<https://www.rbvi.ucsf.edu/chimerax/download.html>) egy jól karbantartott, folyamatosan fej-

lesztett, 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 szögeket beállító illetve a sztérikus ütközés ellenőrző funkcióját hívja meg.

3.4.3. Az SCWRL4 program

A Dunbrack labor által fejlesztett SCWRL4 program [Krivov et al., 2009] (<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-dekompozíciós módszerrel választ és a választást heurisztikával gyorsítja meg, mely mindig hamar ad eredményt, noha nem biztos, hogy megtalálja a minimális energiájú konformert.

3.5. A MUMO (Minimal Under-restraining Minimal Over-restraining) módszer

A túlillesztés és az alulillesztés elkerülése végett egy több replikával rendelkező molekuladinamikai szimulációnál az S^2 rendparamétereket és az NOE paramétereket érdemes más számú replikákon ábrázolni. A MUMO (Minimal Under-restraining Minimal Over-restraining) módszer [Richter et al., 2007], ahol 8 replikát alkalmazva az S^2 rendparaméterek mind a 8 replikán átlagolódnak, azonban az NOE paraméterek csak két replikánként párosával átlagolódnak, ahogy az 1.4. ábrán is látható. Ahogyan ezt Richter és munkatársai is leírták [Richter et al., 2007], az NOE paramétereket túlillesztjük akkor, ha két replikánál többre átlagoljuk őket, az S^2 paramétereknél azonban szükséges a 8 replikára átlagolás az alulillesztés elkerülése végett.

3.6. Molekuladinamikai szimulációk

A molekuladinamikai szimulációkat a GROMACS program 4.5.5-ös verziójának [Pronk et al., 2013] Gáspári Zoltán által módosított változatával futtattuk. Ez a változat képes kezelni az S^2 rendparamétereket és átlagolja az NOE paramétereket minden replikán [Fizil et al., 2015], vagyis a MUMO (Minimal Under-restraining Minimal Over-restraining) megközelítést alkalmazza [Richter et al., 2007]. Az OPLS-AA erőteret [Kony et al., 2002] és a TIP3P vízmodellt [Jorgensen et al., 1983] alkalmaztuk ezekben a szimulációkban.

Az *apo* szerkezet modellezéséhez a Protein Data Bank (PDB) adatbázis 1O1U azonosítójú szerkezetének 7. modelljét [Kurz et al., 2003] választottam, mert annak volt az összes modell közül a legnagyobb PRIDE-NMR pontszáma. A PRIDE NMR egy a kutatócsoportunk által még 2007-ben kifejlesztett módszer [Ángyán et al., 2010], mely egy sokaságból kiválasztja a legjobb NOE megfelelésűt úgy, hogy az atomtávolságok eloszlását összehasonlítja 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 szkriptet használtunk, illetve egy-egy atom típusát maunálisan is átírtuk, ha az szükségesnek látszott.

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 kényszerfeltétel-listát 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 [Ulrich et al., 2007] adatbázisból az 19843 azonosítójú adatsorból származnak, a *holo* szerkezethez pedig maguktól a szerzőktől. Az S^2 rendparamétereket négy hőmérséklet értékre: 283, 291, 298, és 313 K fokra Horváth Gergely és munkatársai [Horváth et al., 2014] munkájából vettük át.

Miután az OPLS-AA erőterrel és a TIP3P vízmodellel legyártottuk a topológiát, egy kocka alakú dobozba tettük a molekulamodellt, majd energiaminimalizálást végeztünk a konjugát gradiens módszerrel 5000 iterációig. A maximális erőt $200 \text{ kJmol}^{-1}\text{nm}^{-1}$ értékre állítottuk. A következő lépésben vízzel vettük körül a molekulamodellt és az egyik vízmolekulát kicseréltük egy nátrium ionra, hogy a rendszer össztöltése semleges legyen. Ezután egy újabb energiaminimalizálást 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 [Hess et al., 1997].

Az ezután következő főszimulációnál 8 replikát futtattunk párhuzamosan az OpenMPI környezetet használva [Gabriel et al., 2004]. Az S^2 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 [Richter et al., 2007]. A szimulációkat 283, 291, 298, és 313 K hőmérsék-

leteken az adott hőmérsékleteken mért S^2 rendparaméterekkel végeztük. A kötéhosszakat 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 S^2 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-a verziójával. Neutralizáció és egy rövid energiainimalizálás után 1 ns NVT és NPT ekvilibrálást végeztünk. A szimuláció 300 K hőmérsékleten folyt, GPU gyorsítással. A sokaságot úgy generáltuk, hogy minden 50 ps-nál belevettük az adott szerkezetet, így a sokaság 20001 konformerből áll. Minden szerkezetet belevettünk, hogy a kezdeti nyújtott szerkezetek is reprezentálva legyenek.

3.7. Főkomponens-analízis

A főkomponens-analízis egy kiváló módszer arra, hogy egy adatsor dimenzióit lecsökkentjük, így nyerve ki a legfontosabb tendenciákat belőlük. Fehérjeszerkezet-konformációkra alkalmazva a kiindulási dimenzióink Descartes-koordinátarendszerbeli pontok nagy halma, amik bizonyos mozgások irányában lévő dimenziók mentén vetítődnek le e módszert használva. Ez egy korrelációs mátrix segítségével történik. Ezen új irányok a sajátvektorok. Így a módszer a szerkezet különböző, egymással lineárisan nem összefüggő mozgásait rangsorolja amplitúdó, vagyis a variancia adott módusa által lefedett hányada szerint [David and Jacobs, 2014].

A módszer során egy C mátrix készül, mely a CA atomok koordinátáit tartalmazza. A sajátérték-dekompozícióból sajátvektorok és sajátértékek jönnek ki. Ezeket a sajátértékeket a módus száma függvényében ábrázolva egy könyökplotot kapunk, ahol rendszerint az első néhány módus értéke nagyon magas a többihez képest, és ezek az értékek szigorúan monoton csökkennek. Annál a pontnál, amettől nagyrészt csökken a meredekség, vagyis a könyökpontnál van az utolsó olyan módus, amit figyelembe veszünk. Ezen módok varianciaösszegének el kell érnie legalább a 80%-át a teljes varianciának. A vizsgálatnál összesen 20 mód bőven elég nagy fehérjékre is [David and Jacobs, 2014]. A fehérjék szerkezetére számolt PCA-ra épülő módszer az eszenciális dinamika (ED), mely az eszenciális mozgásokra derít fényt. A kis és nagy mozgások közül leginkább csak a nagy mozgások elemezhetők ezzel a módszerrel, de biológiai jelentőségük is nagyobb, mint a kis mozgásoké. Kis mozgásokhoz a PCA egy-egy fehérjeszakaszra való alkalmazása javasolt. Mivel a PCA egy lineáris módszer, a nemlineáris komponenseket nem biztos, hogy jól meg tudja jeleníteni. A PCA-nak hátránya, hogy a módus irányok nem teljesen pontosak mivel véges a mintavételezése. Emellett néhány kiugró adat is félre tudja vinni az eredményt, főleg az első néhány módus irányát, ezért fokozottan kell figyelnünk erre a biológiailag

releváns mozgások elemzésénél. Ez azzal is biztosítható, ha nem egy-egy módusra koncentrálnak az eredmények értékelésénél [David and Jacobs, 2014]. Fehérjék esetében még az is fontos szempont, hogy a szerkezet funkcionálisan releváns részeinek elmozdulásait vizsgáljuk, mert a szokásosan mozgékony N- és C-terminális régiók könnyen dominálhatják a szerkezetek varianciáját, miközben számos esetben ezek biológiai relevanciája korlátozott.

A PCA analízis előtt a szerkezeteket illeszteni kell egymáshoz. A PCA módusok alkalmazásai lehetnek arra, hogy egy fehérje két állapota közötti különbséget és a kettő közötti utat feltérképezzék. Az elmozdulás vektorokat a módusok (például az első két módus) által kifeszített altérbe levetítve meg lehet vizsgálni, hogy a konformációs teret hogyan mintavételezik a különböző konformációk [David and Jacobs, 2014].

A főkomponens-analízist a ProDy [Bakan et al., 2011] nevű Python modulval végeztük, a molekulák mozgásait pedig a VMD molekulamodellező program [Humphrey et al., 1996] NMWiz moduljával jelenítettük meg.

3.8. A gastrotropin S^2 adatainak és hidrogénkötéseinek 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 ProDy programmal [Bakan et al., 2011] főkomponens-analízist is végeztünk, a vizualizációhoz pedig a VMD program NMWiz modulját [Humphrey et al., 1996] használtuk. A szerkezetekből a Shiftx2 programmal [Han et al., 2011] számoltuk vissza a kémiai eltolódás értékeket.

Paraméter	Érték
A donor és akceptor atomok maximális távolsága	3.9 Å
Akceptor nehézatom és hidrogén atomok maximális távolsága	2.5 Å
A donor, a hidrogén és az akceptor atom közötti minimális szög	90 fok
A donor, az akceptor atom és az akceptorral kovalens kötésben lévő nehézatom közötti minimális szög	90 fok
A hidrogén, a karbonil oxigén és a karbonil szén közötti minimális szög	90 fok

3.1. táblázat. A hidrogénkötések detektálásának kritérium paraméterei Xu és munkatársai [Xu et al., 1997], illetve Baker és munkatársai [Baker and Hubbard, 1984] által kiadott közlemények alapján.

A hidrogénkötések jelenlétét vagy épp hiányát egy Perl szkripttel vizsgáltuk meg, mely a Baker és munkatársai által közzétett [Baker and Hubbard, 1984] és Xu és munkatársai által említett [Xu et al., 1997] távolság és szög alapú paramétereket használ, melyek a 3.1. táblázatban vannak feltüntetve.

3.9. A gastrotropin szerkezetek amid kémiai eltolódás különbségeinek vizsgálata

A ^{15}N kémiai eltolódásokat minden szerkezetből a Shiftx2 programmal [Han et al., 2011] 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 [Horváth et al., 2014]. Miután a különbségeket 0 és 1 közé normáltuk, 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.

3.10. A gastrotropin fehérje ligandumkötésének vizsgálata dokkolással

A GCA és GCDA dokkolását egyes kiválasztott gastrotropin szerkezeteken Szabó András László végezte a Schrödinger Glide programmal [Friesner et al., 2006]. A 2MM3 PDB két ligandumos szerkezet alapján határozta meg a kötőhelyeket. A szerkezetet a beolvasása után több molekulára osztotta. Ezután a GCA vagy a GCDA ligandumokat olvasztotta össze a fehérjével és 'Receptor Grid Generation Tool' eszközzel egy hálót generált a ligandum köré alapbeállításokkal. Az adott ligandumok dokkolását a 'Ligand docking' eszközzel végezte alapbeállításokkal, viszont beállította, hogy aminosavanként írjon ki kölcsönhatási pontszámokat. A második ligandum dokkolásához az első ligandum 2MM3 szerkezetéhez leghasonlóbb pozíciója alapján az első ligandumot összeolvasztotta a fehérjeszerkezettel, és a második ligandumhoz generált négyzethálót. Az összes lehetséges állapothoz (csak GCA, csak GCDA, GCA+GCDA, GCDA+GCA) 32 különböző pozíciót generált és értékelt ki. A MacroModel rutinnal határozta meg a pozíciók teljes energiáját az OPLS3 erőterrel és a vizet mint oldószert használta.

4. fejezet

Eredmények

4.1. A gasztrotropin fehérje vizsgálata

4.1.1. A molekuladinamikából származó sokaságok megfelelnek a kísérleti paramétereknek

A várakozásokkal összhangban a MUMO módszerrel generált szerkezeti sokaságok jobban megfeleltek a kísérleti S^2 paramétereknek, mint a PDB adatbázisból származó szerkezetek, vagy a megkötés nélküli szimulációból származó sokaságok (4.1. táblázat). Érdekes módon az *apo* MUMO az S^2 paraméterekkel számolva 283 K hőmérsékleten, valamint a *holo* az S^2 paraméterekkel számolva 291 K hőmérsékleten csak mérsékelt felel meg ezeknek az adatoknak, míg a többi megkötött sokaság jó megfelelést mutat. Ennek oka nagyon alacsony, 0.3-nál kisebb S^2 értékek megjelenése, melyek leginkább a hurokrégiókban található, és a szimulációkban nem jelennek meg. A visszszámolt és a kísérleti S^2 paramétereket a függelékben elhelyezkedő ábrákon jelenítettük meg.

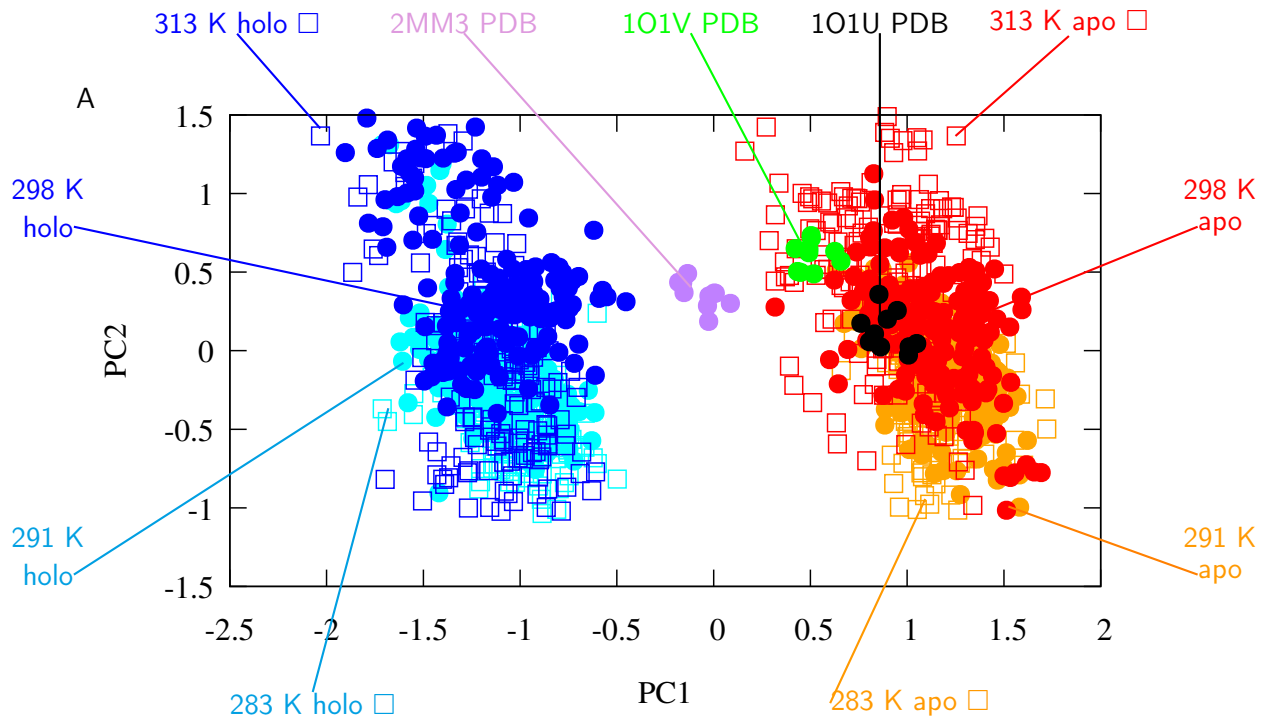
A nem teljesülő NOE kényszerfeltételek száma 0.5 % alatt van az összes MUMO sokaság esetében annak ellenére, hogy nagyobb RMSD értékeik vannak, mint a PDB adatbázisban fellelhető szerkezeteknek. Az amid N és H kémiai eltolódásoknak való megfelelések ugyanabban a tartományban vannak a MUMO sokaság és a PDB adatbázis szerkezeteinél.

4.1.2. A gasztrotropin sokaságok a hordó kinyílásának két külön módját mutatják meg

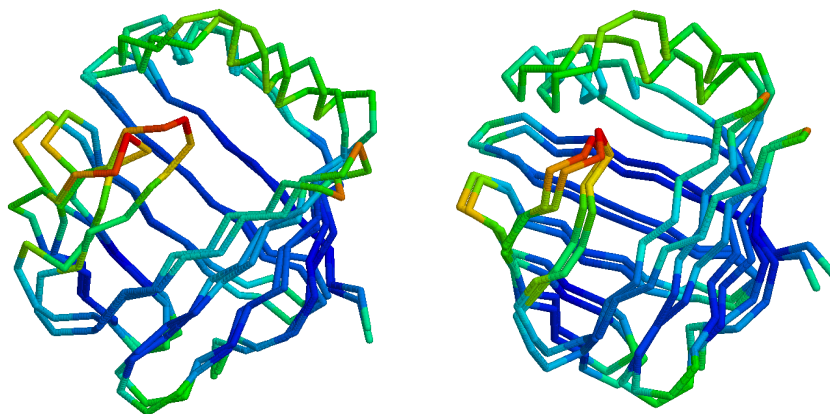
A sokaságok főkomponens-elemzése (PCA) két meghatározó módus jelenlétét mutatja, mindkettő a hordó kinyílásának felel meg (4.1. ábra). A két módus menti mozgás végpontjai a 4.2. ábrán láthatók. Ezeket I. és II. típusú kinyílásoknak neveztük el. Az I. kinyílás jól felismerhetően elkülöníti az *apo* és a *holo* szerkezeteket az 1-es módus mentén, mely a hordó kinyílását jelenti az F és a G β -szálak között. A szerkezetet a hélixek felől nézve, az *apo* és a *holo* szerkezetek közötti szerkezeti változás az E/F és a G/H hurkok óramutató járásával megegyező elfordulása, melyhez a C/D hurok és a II-es hélix óramutató járásával ellenkező irányú kis amplitúdójú mozgása is társul. Ezek az E/F és a G/H hurkok közötti rés megjelenését eredményezik a hordó "tetején". Ezzel ellentétben a II. típusú mozgás a 2-es módus mentén leginkább csak az I. hélixet és a C/D hurkot

	Sokaság méret	Hőmérséklet (K)	Fólianc RMSD (Å)	Fólianc S^2 korreláció	Korrigált fólianc S^2 korreláció	Nem teljesülő NOE feltételek (%) ($r^{-6} > 0.5A$)	Amid N kémiai eltolódás	Amid H kémiai eltolódás
1O1U pdb	10	305	0.66 \pm 0	0.502 ^a	0.857	0.00	0.709	0.596
<i>apo</i> MUMO	168	283	1.01 \pm 0.03	0.675	0.873	0.29	0.832	0.72
<i>apo</i> MUMO	168	291	1.12 \pm 0.04	0.8	0.841	0.29	0.828	0.729
<i>apo</i> MUMO	168	298	1.27 \pm 0.06	0.797	0.883	0.12	0.834	0.738
<i>apo</i> MUMO	168	313	1.45 \pm 0.04	0.963	0.972	-	0.834	0.705
<i>apo</i> nem megkötött	48	283	1.46 \pm 0.09	0.307	-	-	0.827	0.724
<i>apo</i> nem megkötött	48	291	1.45 \pm 0.05	0.351	-	-	0.832	0.703
<i>apo</i> nem megkötött	48	298	1.53 \pm 0.03	0.301	-	-	0.833	0.731
<i>apo</i> nem megkötött	48	313	1.5 \pm 0.07	0.6	-	-	0.829	0.733
2MM3 pdb	10	293	0.52 \pm 0	0.297 ^b	0.315	0.00	0.773	0.593
<i>holo</i> MUMO	168	283	1.28 \pm 0.03	0.726	0.756	0.41	0.784	0.548
<i>holo</i> MUMO	168	291	1.31 \pm 0.07	0.544	0.731	0.46	0.78	0.589
<i>holo</i> MUMO	168	298	1.46 \pm 0.15	0.884	0.930	0.36	0.767	0.552
<i>holo</i> MUMO	168	313	1.5 \pm 0.08	0.724	0.870	0.36	0.758	0.551
<i>holo</i> nem megkötött	48	283	1.92 \pm 0.64	0.073	-	-	0.782	0.531
<i>holo</i> nem megkötött	48	291	2.34 \pm 0.65	0.296	-	-	0.777	0.537
<i>holo</i> nem megkötött	48	298	1.82 \pm 0.07	0.484	-	-	0.772	0.493
<i>holo</i> nem megkötött	48	313	2.59 \pm 2.09	0.272	-	-	0.775	0.544

4.1. táblázat. RMSD értékek, S^2 paraméterek, NOE megfelelés, amid N és H kémiai eltolódás korrelációk^a S^2 adatok 313 K hőmérsékleten (a 4 adatsor közül a legnagyobb korrelációjú adat)^b S^2 adatok 298 K hőmérsékleten (a 4 adatsor közül a legnagyobb korrelációjú adat)

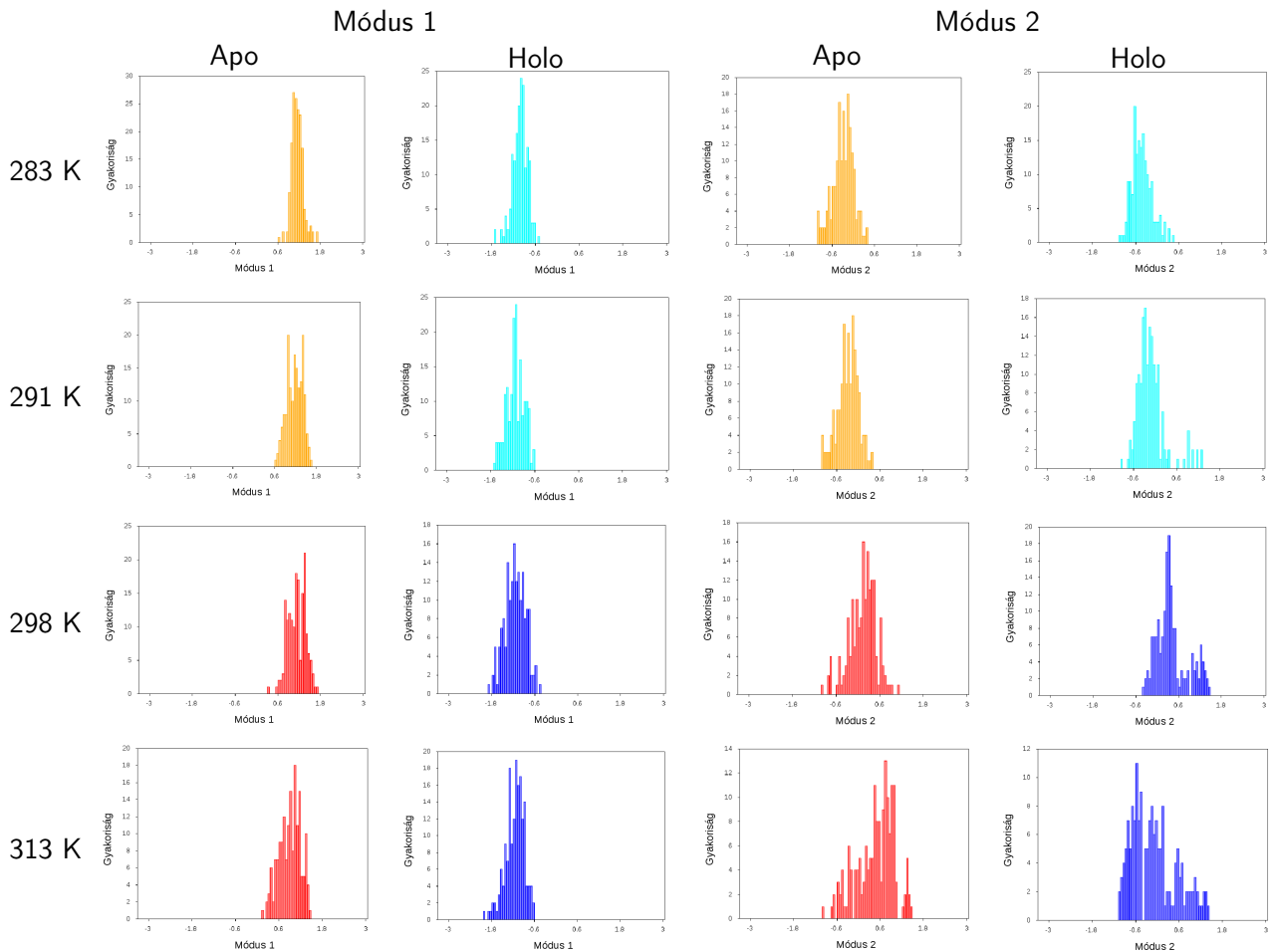


4.1. ábra. A szimulációból származó és a kísérletileg meghatározott szerkezetek főkomponens-analízis (PCA) scatter plotja.



4.2. ábra. A két módus mentén lévő mozgások végállapotai I. módus baloldal, II. módus jobboldal. A színezés "hőmérséklet" szerint történt színátmenetesen, ahol piros a legmozgékonyabb, kék a legkevésbé mozgékony régiókat jelöli.

érinti, leginkább az interhelikális rés szélesedését valamint a D és az E β -szálak közötti kinyílást hozza létre.

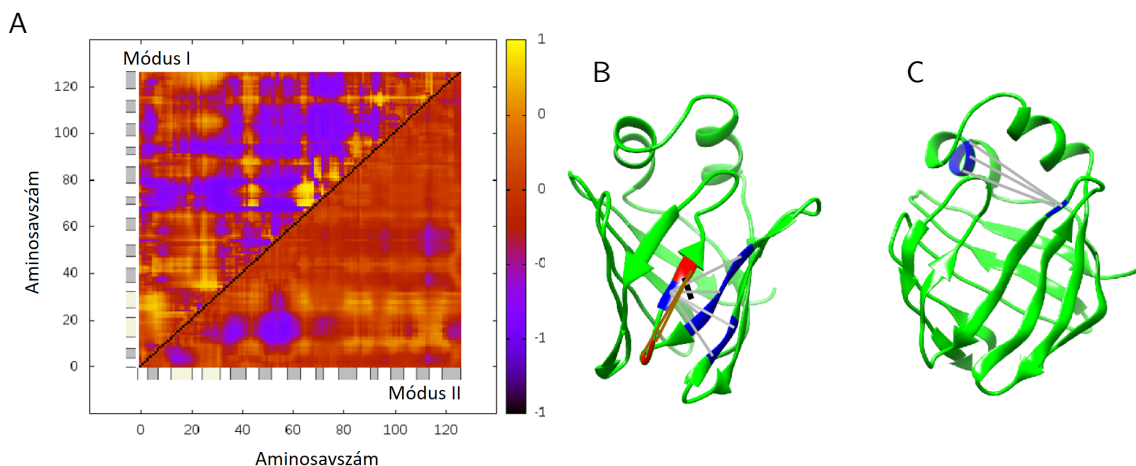


4.3. ábra. A sokaságok főkomponens-analízis scatter plotjaiból származtatott histogramjai a két módus mentén

Figyelemre méltó, hogy a II. típusú mozgás az *apo* és a *holo* szerkezetekben is megjelenik. Magasabb hőmérsékleten a konformációs tér mintavételezése is nagyobb a sokaságoknak, ahogyan ez a 4.3. ábrán is látszik. A portál hipotézis szerint a második hélix N terminális része, a C-D és az E-F hurok egy portál régiót alkot, ennek kinyílása és becsukódása, illetve a helikális rész helicitásának változása szabályozza azt, hogy a fehérje mennyire elérhető a ligandumok számára. A II. típusú mozgás a portál hipotézissel is összeegyeztethető, melyet a humán zsírsavkötő fehérjékben részletesen leírtak [Cai et al., 2012] [Lücke et al., 2000].

Ezek a szerkezeti változások még részletesebben leírhatók a $C\alpha$ (CA) atomok közötti távolságok ábrázolásával. A 4.4. ábrán a $C\alpha$ távolságok és a PCA módusok korrelációja látható minden aminosavpárra mátrixként. Ezen $C\alpha$ - $C\alpha$ távolságok korrelációja alapján a fehérje legmozgékonyabb régiói az I. típusú mozgásnak megfelelő régiók: a terminálisok közelében és a D/E régióban, az E β -szálat is magában foglalva. A II. típusú mozgást figyelembe véve az 50. aminosav körüli régiók (C β -szál) és a 35. aminosav (az összekötő rész a II. hélix és a B β -szál között) tűnnek a legjobban együtt mozgó régióknak.

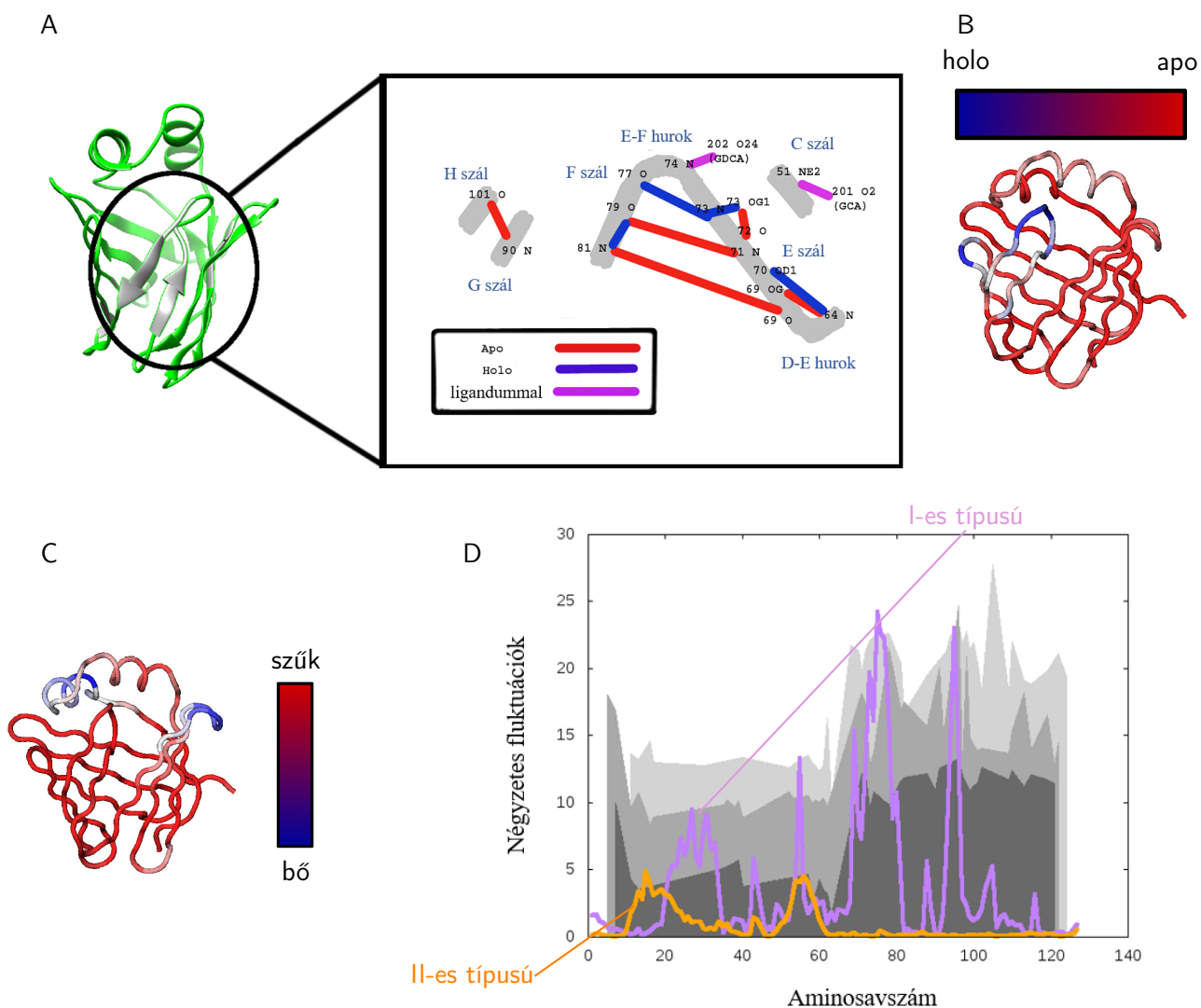
Az I. típusú mozgással a legjobb korrelációt mutató $C\alpha$ - $C\alpha$ távolságok a következő



4.4. ábra. **A:** $C\alpha$ atom távolságok korrelációi az 1-es módushoz képest (átló fölötti rész) és a 2-es módushoz képest (átló alatti rész) minden aminosavra. A helikális részeket bézs színű téglalapokkal, a β -szál részeket pedig szürke téglalapokkal jelöltem. **B:** Az 1-es módussal lévő korrelációk közül legmagasabb (piros, barna vonalak) és a legalacsonyabb (kék, szürke vonalak) távolságok a szerkezeten ábrázolva. **C:** A 2-es módussal lévő korrelációk közül a legalacsonyabb (kék, szürke vonalak) távolságokat a szerkezeten ábrázoltam.

aminosavak között vannak: 47–69, 48–69, 60–69, 61–69, 62–69, 63–69 (korreláció -0.96) és 66–70, 67–70 (korreláció 0.96). A II. típusú kinyílással legjobban korreláló aminosavpárok 16–58, 17–58 és 18–58 (korreláció -0.92). Ezen aminosavpárok szerkezetre vetítése látható a 4.4. ábra B és C részén. Az eredmények azt mutatják, hogy a C és az E illetve a D és E β -szálak közötti növekvő távolságok az I. típusú kinyílással korrelálnak és az I. hélix eltávolodása a hordó tetejétől (a C/D hurok körül) a II. típusú kinyílással korrelál. Azt is érdemes megjegyezni, hogy azok az aminosavak, amelyek az átlagos szerkezettől a legnagyobb eltérést mutatják nem esnek egybe azokkal, amelyek a legnagyobb atomok közötti távolságváltozást mutatják.

Míg az 1O1U PDB azonosítójú kísérletesen meghatározott *apo* szerkezet az *apo* sokasággal egy csoportba esik, a két különböző ligandummal rendelkező NMR szerkezet, melynek 2MM3 a PDB azonosítója az *apo* és a *holo* MUMO sokaságok között helyezkedik el. Ez azt mutatja, hogy a molekuladinamikából származó *holo* sokaságok az első főkomponens mentén nagyobb mozgást mutatnak, mint az NMR-rel meghatározott szerkezetek. Máshol is felmerült a jelenség, hogy kísérleti adatoknak megfelelő sokaságok "felnagyítják" a különbségeket a különböző állapotok között [Czajlik et al., 2017]. Ez a felnagyítás valószínűleg annak a következménye, hogy az NOE feltételeket a sokaságra átlagoljuk, nem követeljük meg minden szerkezettől azok teljesítését, így nagyobb konformációs teret tudunk lefedni a konvencionális szerkezetmeghatározó módszerekhez képest. Egy másik különbség ezen módszerekhez képest, hogy itt más az egyensúly az erőterparaméterek és a kísérleti paraméterek között.



4.5. ábra. Az I. és II. típusú mozgások leírása. (A) Hidrogénkötések, melyek a MUMO sokaságokban az apo és a holo szerkezetek között a legtöbbször megváltoznak. A fekete számok az aminosav számok, a fekete betűk az atomok jelei, zöld betűkkel a másodlagos szerkezeti elemeket jelöltem. A piros vonalak a főleg az apo formában jellemző hidrogénkötéseket, míg a kék vonalak a főleg a holo formában megjelenő hidrogénkötéseket, a ligandummal létrejött kötéseket pedig lila vonalak jelölik. A 73-as treonin központi szerepe figyelemre méltó. (B) Szerkezeti mozgás az első módus mentén: hordó nyílás (C) szerkezeti mozgás a második módus mentén (D) 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} értékeket, melyek két aminosav klasztert határoznak meg szürke sáttírozással jelöltem a három hőmérsékleten (283 K, 287 K, 291 K). Csak 30-40 aminosavnál van ms időskálán mérhető k_{ex} érték [Horváth et al., 2014], az átláthatóság és a különböző regionális különbségek hangsúlyozására vonalként rajzoltam fel ezeket az értékeket.]

Kötés (aminosav szám és atomnév)	Fehérjeállapot	Donor régió	Akceptor régió
64 N– 69 OG	<i>apo</i>	D-E	E
64 N– 70 OD1	<i>holo</i>	D-E	E
71 N– 79 O	<i>apo</i>	E	F
73 N– 73 OG1	<i>holo</i>	E-F	E-F
73 N– 77 O	<i>holo</i>	E-F	E-F
73 OG1–72 O	<i>apo</i>	E-F	E
81 N– 69 O	<i>apo</i>	F	E
81 N– 79 O	leginkább <i>holo</i>	F	F
90 N– 101 O	leginkább <i>apo</i>	G	H

4.2. táblázat. A MUMO szimulációk során legjelentősebben megváltozó hidrogénkötések listája

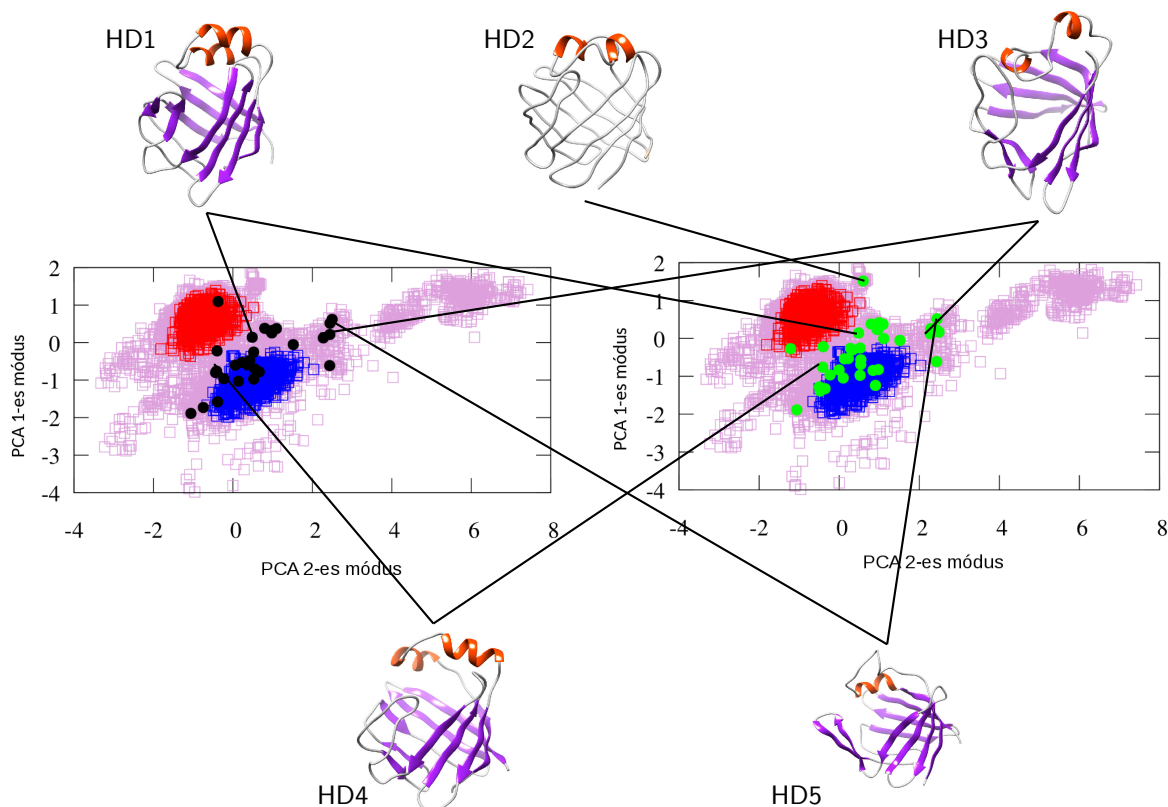
Az *apo* és a *holo* állapotok a hidrogénkötésekben is különböznek (4.5. B ábra, 4.2. táblázat). Ahogyan előző munkákban is megmutatták, az I-BABP fehérjében a hidrogénkötések kiterjedt hálózatot alkotnak [Toke et al., 2006] [Horváth et al., 2014]. Számításaink szerint a hidrogénkötések megjelenésében a legnagyobb különbség egy szálon belüli hidrogénkötés megjelenésében, illetve kettő ilyen kötés elszakadásában jelenik meg az *apo* és a *holo* állapot közötti átmenetben, mely az E és F szálak közötti hordónyírással összhangban történik. Érdekes módon a ligandummal kialakított hidrogénkötések csak néhány szerkezetben vannak jelen (4.5. B ábra), ami egy lazább ligandumkötést sugall, mely dinamikusan változó hidrogénkötések eredménye. A 73-as treonin hidrogénkötései különösen is érdekesek, ahol a $\gamma 1$ OH csoport a *holo* állapotban egy aminosavon belüli hidrogénkötést alkot, mely az *apo* állapotban nem jelenik meg. Ez az aminosav az E/F hurokban valószínűsíthetően kulcsszerepet játszik a ligandumkötés konformáció szelekciós mechanizmusában az EFGH régió közeli aminosavaival együtt a humán I-BABP fehérjében [Horváth et al., 2016].

4.1.3. A PCA elemzésben azonosított két mozgásban résztvevő aminosavak nagymértékben átfednek az NMR módszerrel mért különböző kicserélődési gyakoriságokat mutató mozgásokban érintett aminosav csoportokkal

Összehasonlítva a fehérjerészeket, melyek a konformáció kicserélődési NMR adatok szerinti mozgásokban vesznek részt látható, hogy az aminosav csoport, amely a PCA elemzésben azonosított II. típusú mozgásban leginkább érintett, és az NMR mérésekből származó "lassú" mozgásban résztvevő N terminálison lévő aminosavak csoportja nagymértékben átfed [Horváth et al., 2016], (4.5. D ábra). Noha a CPMG relaxáció diszperziós mérésekből származtatott k_{ex} paraméterek egy jóval lassabb, μs -ms időskálán mozognak a szimulációinkban használt S^2 paraméterek által mutatott időskálához képest, ezek az adatok azt mutatják, hogy a megfigyelt II. típusú hordó kinyílásának köze van az NMR relaxáció diszperziós mérésekből származó lassú konformációs kicserélődésekhez. Úgy gondoljuk, hogy a gyors mozgások előkészítik a nagyobb amplitúdójú lassabb mozgásokat a fehérjé-

ben egy ugyanolyan jellegű kinyílás mentén. A különböző időskálán végbemenő konformációs változást az is alátámasztja, hogy ezek a megfigyelt mozgások hőmérsékletfüggőek, vagyis a II. mozgás mentén a konformerek egyenletesebben oszlanak el magasabb hőmérsékleten. Fontos még az is, hogy a II. típusú mozgás jelen van az *apo* és a *holo* formában is, így a ligandum felvételében és elengedésében is szerepe lehet.

4.1.4. Az *apo* formában található rejtett "*holo*-féle" konformáció részben kitekeredett

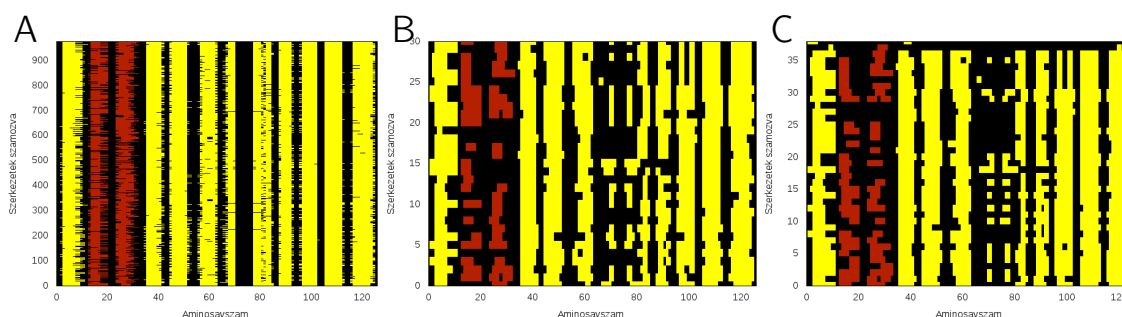


4.6. ábra. Az *apo* MUMO (piros) és a *holo* MUMO (kék) sokaságok PCA scatter plotja, valamint a többi generált sokaság (lila négyzetek), melyeket a rejtett állapotkonformációk kiválasztására használtam. A ^{15}N kémiai eltolódások és a kísérletes $|\Delta\omega(^{15}\text{N})|$ értékek közötti különbségek 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.

A humán gasztrotropin fehérje NMR vizsgálatai megmutatták egy olyan láthatatlan állapot jelenlétét, amely lassú kicserélődésben van a látható *apo* állapottal [Horváth et al., 2012]. Emellett a szerzők arra is következtettek, hogy ez az állapot hasonlít a *holo* állapothoz [Horváth et al., 2016], [Horváth et al., 2014]. Ahhoz, hogy vizsgálni tudjuk ezt az állapotot, egy sok konformációból álló sokaságot gyártottunk, és azokból kiválasztottuk azokat a konformereket, amelyek ezt a nagyobb energiájú állapotot képviselik. Ezt úgy tettük meg, hogy az *apo* állapothoz képest számolt kémiaieltolódás-különbségeket összehasonlítottuk a kísérletesen mért az alapállapot és a rejtett állapot közötti eltérésekre jellemző

$|\Delta\omega(^{15}N)|$ értékekkel és ezekre szelektáltunk. A láthatatlan állapotra vonatkozó szerkezeti információk limitáltak, mivel csak a főlánc ^{15}N értékek állnak rendelkezésre.

Jól látszik, hogy ezek a kiválasztott konformerek, melyek legjobban megfelelnek ezen kísérleti adatoknak, nagy konformációs teret ölelnek fel (4.6. ábra). Közös vonásuk, hogy közelebb vannak a *holo*, mint az *apo* állapothoz, mely összecseng azzal, hogy a korábbi NMR mérésekben "*holo*-féle" jellegű állapotokat meg ennél a nagyobb energiájú állapotnál [Horváth et al., 2016]. Fontos még megjegyezni, hogy e konformerek némelyike megfoghatóbb II. típusú mozgást mutat, mint a MUMO sokaságok. A 4.6. ábra főkomponensei nem ugyanazok, mint a 4.5. ábrán lévők. A HD3 és a HD5 konformerek noha közel vannak a PCA ábrán, a helikális régiójukban a II. típusú mozgás mértéke más. Mindenesetre ezt az aspektust gondolom a leginkább relevánsnak, mivel a MUMO sokaságban világosan megállapítható a II. típusú mozgás jelenléte, míg másfajta mozgások csak a megkötés nélküli sokaságokban jelennek meg.

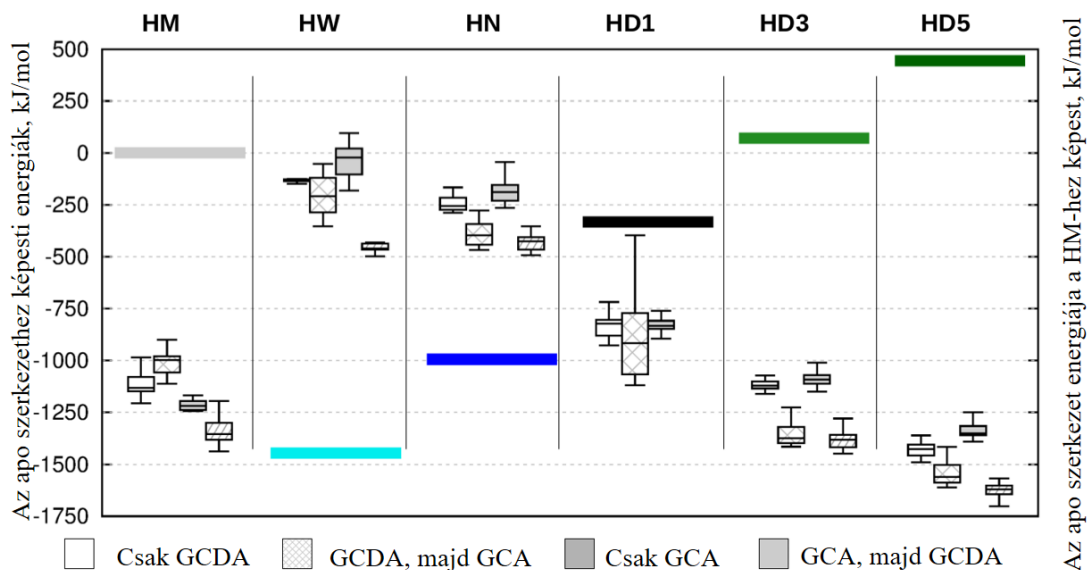


4.7. ábra. *A szimulációkból származó konformációk másodlagos szerkezetei.* Minden oszlop egy aminosavat jelent. A β -szálak sárgával, az α -hélixek barnával, a többi aminosavak feketével jelennek meg. (A) Minden konformáció (B) A nagy korrelációt mutató konformációk (az A konformációk részhalmaza) (C) A legközelebbi RMSD-vel rendelkező konformációk (az A konformációk részhalmaza). Az elemzés a DSSPcont programmal készült [Andersen et al., 2002]. Figyelemre méltó a hélixek rövidülése egyes konformációkban, főleg a B és C paneleken.

Ahogy azt a 4.7. ábra is mutatja, a szimulációs sokaságok másodlagos szerkezete sokszínű az α -hélixek és a β -szálak határainál. Néhány konformációnál szinte az összes α -hélix és β -szál részben vagy egészben kitekeredik. A "*holo*-féle" *apo* konformerek kevés α -hélix és β -szál régióval rendelkeznek. Az E-F régió a leghajlamosabb a kitekeredésre, ahogyan azt a legújabb NMR vizsgálatok is mutatják [Horváth et al., 2017]. Összességében elmondható, hogy az *apo* és a *holo* állapot közötti átmenet nemcsak egy egyszerű kinyílás egy legrövidebb utat bejárva, hanem konformációs átrendeződések komplex egymásutánisága, melyben az intermedierek kitekeredett szerkezetek, mely főleg a helikális sapkát és a C/D hurok régiót érinti és részben hasonlít a II. típusú mozgásra.

4.1.5. A dokkolás eredménye

A ligandumkötés további leírásához a számításainkból származó szerkezetek közül néhány kiválasztottal dokkolási szimulációt végzett Szabó András László. A PCA elemzés alapján két szélső szerkezetet választottunk ki, melyek a II. típusú mozgás végpontjai. Három olyan szerkezetet is kiválasztottunk, melyek a legjobban korrelálnak azok közül a rejtett szerkezetek közül, melyek lassú kicserélődésben vannak az *apo* szerkezettel. Összességé-

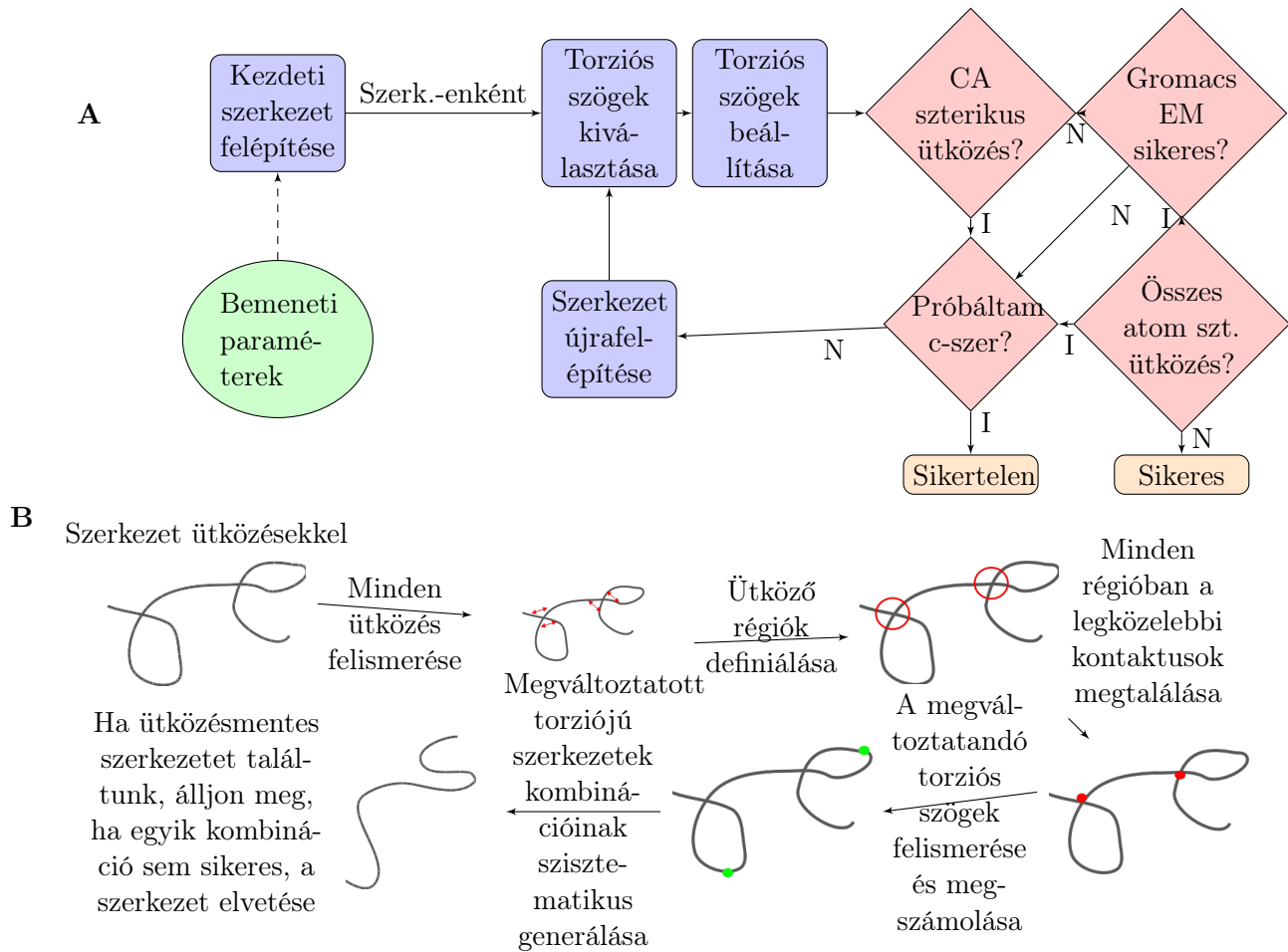


4.8. ábra. A dokkolt gasztrotropin szerkezetek apo szerkezetéhez képesti relatív energiája A kiinduló szerkezetek 2MM3 szerkezetéhez képesti különbségei színes vonalakkal vannak jelölve. Fehér dobozok: csak GCDA, csíkos fehér dobozok: GCDA először, majd GCA, szürke dobozok: csak GCA, csíkos szürke dobozok: GCA először, majd GCDA. HM a holo MUMO szerkezetekből egy reprezentatív szerkezetet jelöl, HN (holo narrow) és HW (holo wide) a holo MUMO szerkezetből PCA alapján két extrém szerkezet, melyek a II. típusú mozgás két végpontjai. A rejtett konformerekből három magas korrelációval rendelkező szerkezetet választottunk ki (HD1, HD3 és HD5).

ben a legjobban azok a szimulációk sikerültek, ahol GCA-t dokkolta először, és a GCDA-t másodsor (4.8. ábra). Ez csak a HD1 szerkezet esetében volt sikertelen, mely az I. típusú mozgás mentén egy középső szerkezetnek számít. A ligandumkötés stabilizálja a félig kitékeredett szerkezeteket, melyek a II. típusú mozgáshoz hasonló kinyílást írnak le. Az apo szerkezetekhez képesti energiák, melyek az 4.8. ábrán színes vonalakként vannak jelölve, egy komplex energiaszint-mintázatot (energy landscape) rajzol ki, ahol a konformációs állapotok és a ligandumkötés egymástól függetlenül járul hozzá a stabilizáláshoz. Ezek az eredmények kompatibilisek azzal a képpel, hogy a félig kitékeredett szerkezetekbe megy be a ligandum, melyet egy szerkezeti kompaktságnövekedés követ, melyet egy I. típusú mozgás mentén lévő haladás ír le, miközben II. típusú kinyílás is történik.

Itt megjegyezzük, hogy a kiszámolt energiák abszolútértékeit nem tarjuk megbízhatónak, hanem egymással összehasonlítva adnak egy képet a ligandumkötés folyamatáról.

A különböző ligandumpozíciók összehasonlítása (4.3. táblázat) és az energiák összehasonlítása (4.8. ábra) ahhoz a megállapításhoz vezet, hogy a nyitott gasztrotropin szerkezetek nagy strukturális sokszínűséggel tudják kötni a ligandumokat nagy affinitás mellett, azonban ennek teljes megalapozásához további vizsgálatok kellenének. A fenti eredmények azt sugallják, hogy a ligandumok dinamikus áthelyeződésen mennek át mind az 1, mind a 2 ligandumot tartalmazó szerkezetekben.



4.9. ábra. **A)** A DIPEND (*DI*sordered *PR*otein *E*nsembles from *N*eighbor-dependent *D*istributions) eljárás folyamatábrája **B)** A DIPEND program kicsomózó részének folyamatábrája

Bemeneti szerkezet	Sikerés GC-DA dokkolások	Sikerés GCA dokkolások GCDA + fehérjéhez	Sikerés GCA dokkolások	Sikerés GC-DA dokkolások GCA + fehérjéhez
HM	32 (22)	32 (24)	32 (13)	32 (25)
HW	32 (9)	30 (25)	32 (14)	32 (20)
HN	32 (12)	32 (29)	32 (32)	32 (23)
HD1	32 (23)	30 (28)	32 (0)	0 (0)
HD3	32 (13)	32 (29)	32 (21)	32 (24)
HD5	32 (5)	32 (30)	32 (28)	32 (7)

4.3. táblázat. *A dokkolási szimulációk összefoglalása a speciális szerkezeteken. A sikeres dokkolások száma, zárójelben pedig a 2MM3 PDB azonosítójú szerkezethez hasonló pózok száma.*

4.2. A DIPEND eljárás

4.2.1. A program felépítése

A program nyílt forráskódú, és elérhető a githubon (<https://github.com/PPKE-Bioinf/DIPEND>). Rendezetlen fehérjeszakaszok atomi szintű szerkezeti modelljeit építi fel úgy, hogy a Dunbrack labor által meghatározott, szomszédságfüggő Ramachandran-térképeket felhasználva [Ting et al., 2010] generál torziós szögeket a szekvencia alapján.

A program egy normál személyi számítógépen, csak egy szálát használva egy kb. 50 aminosav hosszú fehérjeszakasz esetében 1000 szerkezetet kb. 17 óra alatt generál. Egy nagyjából 100 aminosav hosszú szakasznál 1000 szerkezetnél ez már 190 óra (nagyjából egy hét).

A főbb lépések a 4.9. folyamatábrán láthatók. A modell felépítéséhez a ChimeraX [Pettersen et al., 2021] programot hívja meg GUI és grafikus megjelenítés nélkül. A program Linux operációs rendszerben működik. Nagy része Python3-ban íródott, a memória- és sebességoptimalizálás érdekében egy lépést C++-ban oldottam meg. A szekvencia alapján egy hosszú β -szál ($\phi = -65^\circ$ és $\psi = +135^\circ$) épül fel atomi szinten a ChimeraX program [Pettersen et al., 2021] segítségével. Ezután a Dunbrack labor [Ting et al., 2010] kumulatív valószínűségeit felhasználva a program random választ ki egy-egy ϕ és ψ szög-párt minden aminosavra. Például a LEFT módnál az adatsor egy adott aminosav adott balszomszédja esetén 5 fokként minden szögparra tartalmaz egy kumulatív valószínűséget (-180;-180), (-180;-175), (-180;-170), ..., (-75;-35), (-75;-30), (-75;-25), ... , (10;35), (10;40), ... , (175;175) monoton növvő szögparokra. A kumulatív valószínűségek egy nagyon kicsi számtól monoton módon nőnek, (175;175)-nél elérve az 1-et. Minden aminosavparra külön adatsor áll rendelkezésre. Az adatsorokat bináris fájlokban tároljuk a gyorsaság és a helykímélés kedvéért.

A program egy véletlen számot generál 0 és 1 között. Megkeresi azt a szög-párt, aminek kumulatív valószínűsége éppen csak egy kicsit magasabb a dobott számnál, vagyis kiválasztja a megfelelő kumulatív valószínűséggel rendelkező sávot. Például ha 0.77-et dobtunk és van egy 0.76-os és egy 0.78-as szögparunk, akkor a 0.78-ast választjuk. Az

```

1 Given parameters:
2   angle to add: 31
3   base: test_weighted_left_only_
4   cycle: 10
5   dataset: CONLY
6   gromacs optimization: 1
7   mode: WEIGHTED_LEFT
8   number of structures: 5
9   proline phi rotation: 0
10  remaining temporary files: 0
11  random addition: 6
12  unknot max: 7
13  sequence:
14  TTIKIMDPGVGDGATAAKSKRLFKEAPVVVSGPVIIG
15 Proline res 8 phi is not rotated.
16 Proline res 27 phi is not rotated.
17 Proline res 33 phi is not rotated.
18 Proline res 39 phi is not rotated.
19 Proline res 54 phi is not rotated.
20 Proline res 57 phi is not rotated.
21 Proline res 69 phi is not rotated.
22 i: 1, contacts: 1
23 angles to unknot: 2, max tries: 4
24 code: [0, 0] contacts: 0
25 Clash_report_for_the_gmx_optimiz:
    0 -1.0 1000.0
26 # Chosen angles for structure 1:
27 1 T -110 150
28 2 T -65 130
29 3 I -65 140
30 4 K -55 140
31 5 I -55 45
32 6 M 175 -15
33 7 D -65 -155
34 8 P -55 0
35 9 G -60 -5
36 10 V -125 85
37 11 G -50 -10
38 12 D 60 30
39 13 G -165 145
40 14 A 60 45
41 15 T -110 -15
42 16 A -70 150
43 17 A -70 140
44 18 K -75 130
45 19 S -165 170
46 20 K -110 95
647 117 I -120 115
648 118 R -125 -10
649 119 S -85 130
650 120 A -120 140
651 121 A -105 160
652 122 Q -135 5
653 Trying at 0
654 Trying at 1
655 Trying at 2
656 Trying at 3
657 Trying at 4
658 Trying at 5
659 Trying at 6
660 Trying at 7
661 Trying at 8
662 Renaming files based on success...
663 Trying to rename structure 1 to structure 1
664 Trying to rename structure 2 to structure 2
665 Trying to rename structure 3 to structure 3
666 Trying to rename structure 4 to structure 4
667 Trying to rename structure 5 to structure 5
668 Number of successfully optimized structures: 5
669 Cleaning up...
670 Time elapsed: 79.12

```

4.10. ábra. A DIPEND program log fájljának részletei

ehhez tartozó két szög lesz a fehérjeszakasz adott aminosavához tartozó torziós szögpár. Ezt a fajta kiválasztást rulettkerék kiválasztásnak is hívják.

A következő lépésben ChimeraX-szel beállítja a program mindegyik szögpárt a kiválasztott értékekre. Ezután egy C++-ban implementált szerikus ütközés ellenőrzés következik, mely azt nézi meg, hogy vannak-e olyan CA atomok, amik nem szekvenciálisan szomszédok és közelebb vannak, mint 4 Å, ez ugyanis ütközésnek számít. Ha 4 aminosavnál szekvenciálisan közelebb vannak, és mégis 20 Å-nél távolabb térben, akkor láncszakadás van, melyet szintén kiszűr a program.

Ha van néhány ütközés, de nem túl sok, akkor a program kicsomózó modulja lép működésbe. Ez a modul megvizsgálja, hogy mely ütközések esnek egy régióba. Ha e régiók száma egy bizonyos határ alatt van, melyet egy paraméterrel lehet szabályozni, pl. kevesebb, mint 6, akkor a program minden régióban megkeresi az ütközések közül a legkisebb távolsággal rendelkezőt és a két ütköző aminosav között szekvenciálisan félúton lévő aminosavat. Ezen félúton lévő aminosavak torziós szögeihez egy bizonyos szöveget hozzáad vagy kivon belőlük kombinatorikusan, ezzel n ütközésre 2^n szerkezetet hozva létre. Ha valamilyen szerkezetben így már nincs ütközés, befejezi a kicsomózási folyamatot. A kicsomózási modul működése összefoglalva a 4.9. B ábrán látható.

Ha nincs ütközés vagy láncszakadás, a SCWRL4 [Krivov et al., 2009] program segítségével az oldalláncok optimalizálása történik. Ezután a GROMACS [Van Der Spoel et al., 2005] [Abraham et al., 2015] program segítségével a program rövid energiaminimalizálást futtat vákuumban, és figyel, hogy sikerült-e.

Ha sikerült, akkor a ChimeraX szerikus ütközés ellenőrzésével a program megnézi, van-e bármilyen ütközés. Ha nincs, akkor sikeres a modell. A program tárolja, hogy melyik modell sikeres és melyik nem. Minden szerkezetre megismétlődnek a lépések a kezdeti β -szál felépítését kivéve, melyet elég csak egyszer elvégezni. Minden szerkezetet maximum egy bizonyos számszor próbálja meg felépíteni. Például ha 10 modellt szeretnénk felépíteni és 30-szor próbálkozunk, akkor egyszer megpróbálja felépíteni mind a 10-et, majd még 30-szor megnézi, hogy vannak-e sikertelenek, és ha vannak, azokat újrapróbálja. A legjobb

esetben csak 10 felépítés történik, a legrosszabb esetben 310 felépítés, 10 db sikertelen modellel.

A program bemeneti paramétereit a következők:

- A szekvencia (ez az egyetlen kötelező paraméter)
- A generálandó modellek száma
- A felépítés módja
 - LEFT: a torziós szögek kiválasztásánál csak a bal szekvenciális (azaz N terminális felőli) szomszédot veszi figyelembe
 - RIGHT: a torziós szögek kiválasztásánál csak a jobb szekvenciális (azaz C terminális felőli) szomszédot veszi figyelembe
 - TRIPLET: a torziós szögek kiválasztásánál mindkét szekvenciális szomszédot figyelembe veszi
 - WEIGHTED_LEFT: egy a felhasználó által definiált Gauss eloszlást lehet aminosavanként a Dunbrack eloszlással kombinálni
 - WEIGHTED_RIGHT: egy a felhasználó által definiált Gauss eloszlást lehet aminosavanként a Dunbrack eloszlással kombinálni
 - WEIGHTED_TRIPLET: egy a felhasználó által definiált Gauss eloszlást lehet aminosavanként a Dunbrack adatokból származtatott eloszlással kombinálni
- A modellek fájlnev kezdete (például a „bar_” fájlnev kezdetként bar_min_1.pdb, bar_min_2.pdb, stb. fájlneveket ad eredményül)
- A használt Dunbrack-féle adatkészlet (TCBIG vagy Coil only, ld. [Ting et al., 2010])
- A modellépítési próbálkozások maximális száma
- Legyen-e GROMACS optimalizálás (ha esetleg a felhasználónál nem fut a GROMACS)
- Meghagyja-e az ideiglenes fájlokat (diagnosztikai célra)
- A maximum kicsomózandó ütközések száma (0, ha nem akarunk kicsomózódást)
- A kicsomózáshoz hozzáadandó vagy kivonandó szögérték (alapértelmezetten 5 fok)

A program által kimenetként adott logfájl részlete látható a 4.10. ábrán.

A kezdeti szerkezet

Legelső lépésként egy nyújtott konformációt építünk a megadott szekvenciával a ChimeraX program segítségével [Pettersen et al., 2021], ahol a torziós szögek $\phi = -65^\circ$ és $\psi = +135^\circ$.

A modellfelépítési módok

A LEFT és a RIGHT mód a Dunbrack labor adataiból dolgozva a szekvenciális bal (N terminális felőli) illetve jobb (C terminális felőli) szomszédot veszi figyelembe. A TRIPLET módnál származtatott valószínűségekkel számoltam, a honlapjukon lévő utasítások segítségével. Például hogyha cisztein a balszomszéd és glicin a jobb szomszéd, akkor

$$P_{triplet} = \frac{e^{\ln(p_{cl}) + \ln(p_{rg}) - \ln(p_{lall}) - \ln(p_{rall})}}{p_{sum}} \quad (4.1)$$

ahol p_{cl} a cisztein balszomszéd valószínűsége, p_{rg} a glicin jobbszomszéd valószínűsége, p_{lall} az összes balszomszéd valószínűsége, p_{rall} az összes jobbszomszéd valószínűsége és p_{sum} az adott tripletre vonatkozó összes számláló összege. Az ezzel való osztás normalizálja az eredményt.

Ezeket a valószínűségeket minden szögpárra kiszámolva a szögpárokon sorba haladva összeadva és az eredményt normalizálva megkapjuk a kumulatív valószínűségeket az adott hármasra, melyeket a kiválasztásnál használtam.

A WEIGHTED módoknál minden aminosavra egy Gauss eloszlást lehet definiálni. A felhasználó meg tudja határozni a centrális torziós szöveget és egy szórást, emellett egy súlyparamétert, mely 0 és 1 közé esik. Ha 0 a súly, akkor ezt a Gauss eloszlást a program nem veszi figyelembe, vagyis olyan eredményt ad, mintha a WEIGHTED módot nem is használnánk, egységnyi súlynál pedig csak a felhasználó által definiált eloszlást veszi figyelembe az adott aminosavnál. Egy adott szögpárra a kumulatív valószínűség meghatározása a következőképpen történik: $p = w * p_u + (1 - w) * p_d$ ahol w a felhasználó által megadott súly, p_u a felhasználó által megadott Gauss eloszlásból származó kumulatív valószínűség p_d pedig a Dunbrack adatból származó kumulatív valószínűség.

A súlyozott módszerek tesztelésénél $\phi = -60$ és $\psi = -40$ torziós szöveget használtunk 10 fok szórással és 0.99-es súllyal.

A szterikus ütközések ellenőrzése

Miután a fentebb leírt módon minden torziós szöveget meghatároztunk, a ChimeraX [Pettersen et al., 2021] programmal egyenként beállítódnak a program GUI és grafikus ablak nélküli meghívásával. A felépített szerkezetnél ellenőrizzük a C++-ban megírt programmal, hogy vannak-e szterikus ütközések, illetve szakadások a peptidláncban. Hogyha a szekvenciálisan nem szomszédos CA atomok közelebb vannak egymáshoz, mint 4 Å, akkor szterikus ütközés van, illetve ha 4 aminosavnyinál közelebb van a szekvenciában két CA atom, mégis távolabb helyezkednek el egymástól, mint 20 Å, akkor láncszakadás van. Ez utóbbinak nem szabadna megtörténnie, szterikus ütközések azonban gyakran előfordulhatnak, főleg ha 100 aminosavnál hosszabb szekvenciával dolgozunk.

4.2.2. Alsokaságok kiválasztása

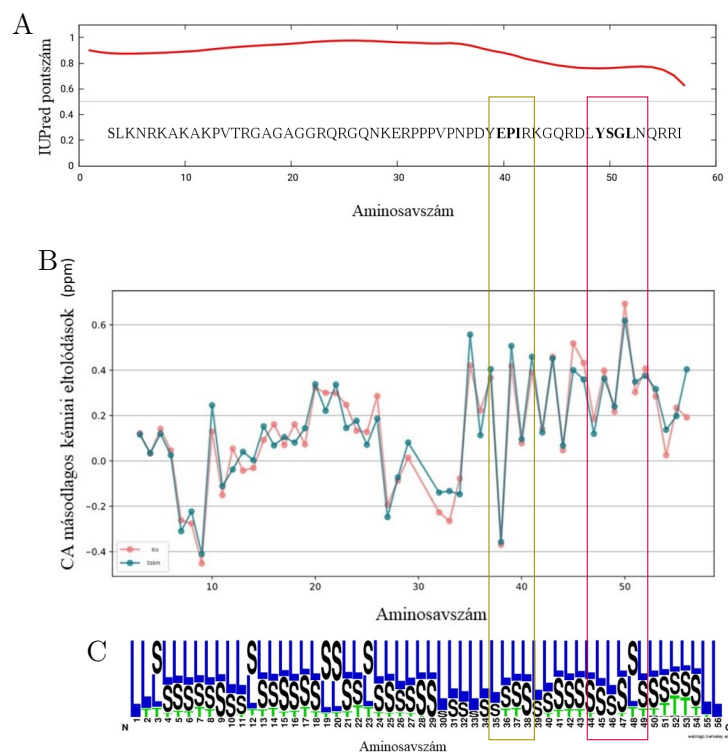
A generált sokaságokból alsokaságok szelekcióját a CoNSEnsX⁺ webservert dockerizált lokális változatával végeztük [Ángyán et al., 2010], [Dudola et al., 2017]. Ez a program

egy mohó algoritmussal kiválasztja azt az alsokaságot, amelyik a legjobban megfelel a kísérleti adatoknak, mely esetünkben legtöbbször kémiai eltolódás paramétereit jelent. Újjonnan másodlagos kémiai eltolódásokkal is tud számolni a program, ezeket is felhasználtuk sokaságaink elemzéséhez.

4.2.3. A programmal előállított szerkezeti sokaságok elemzése

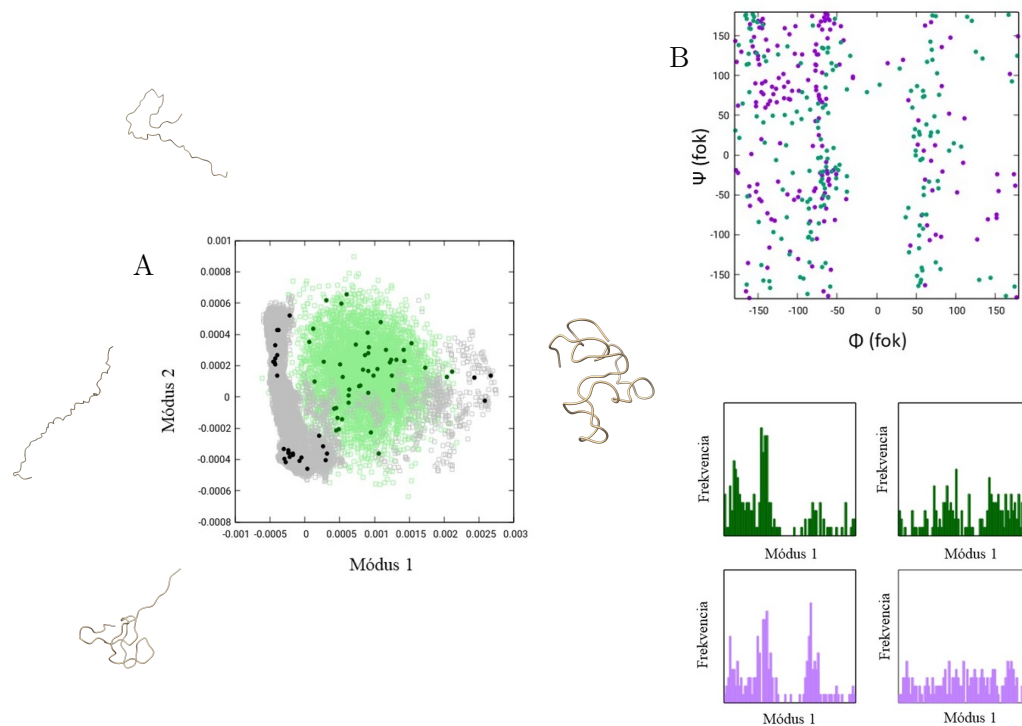
A program tesztelésére két fehérjeszegmens szerkezeti sokaságmodelljét állítottuk elő a programmal, melyeket aztán a CoNSEnsX⁺ webszerverrel elemeztünk. A webszerver immár másodlagos kémiai eltolódásokat is vissza tud számolni szerkezetekből, az elsődlegesek visszaszámolását is megtartva. A másodlagos kémiai eltolódások visszaszámolásához a referenciaértékek Tamiola és munkatársaitól származnak [Tamiola et al., 2010]. A webszerverrel alsokaságokat is választottunk RMSD alapján, mely ugyanaz elsődleges és másodlagos kémiai eltolódásokra is. A másodlagos kémiai eltolódások nagyon érzékenyek bizonyultak a lokális szerkezeti változásokra.

Az előállított Cd3ε fehérjeszegmens sokaság megfelelése a kísérleti adatoknak



4.11. ábra. A Cd3ε sokaság áttekintése **A)** Az IUPred3 szerver által prediktált rendezetlenségi pontszám **B)** Mért és visszaszámolt CA másodlagos kémiai eltolódások **C)** A sokaság másodlagos szerkezeti logója a CA kémiai eltolódás alapján szelektált alsokaságra. A DSSPcont állapotokat átlagoltuk minden modellre. Az ábra a Weblogo programmal készült.

Módszerünk teszteléséhez egy rendezetlen fehérjeszakaszt választottunk, olyan, amelyhez tartozik kémiai eltolódás kísérleti adat, melyet a Biological Magnetic Resonance Data Bank (BMRB) adatbázisból értem el [Ulrich et al., 2007]. A szerkezetet azok közül a mo-



4.12. ábra. *A Cd3ε sokaság áttekintésének folytatása A) A szimulált (szürke), generált (zöld) és szelektált (sötétszürke és sötétzöld) (al)sokaságok PCA scatter plotja. A szélső PCA értékekhez hasonló egy-egy szerkezet szalagábrázolásával. B) Az ITAM1 (sötétzöld) és az ITAM2 (lila) motívumokban lévő összes aminosav Ramachandran ábrái a szelektált alsokaság minden szerkezetére.*

dellek közül választottuk ki, melyek BMRB szekvenciáin az IUPred2A [Mészáros et al., 2018] rendezetlenség predikciós programot futtatva legalább 50 aminosav hosszú, átlagosan legalább 0.5 pontot elérő szakaszok voltak.

Mivel az 57-es izoleucin másodlagos kémiai eltolódásai kétszer olyan nagyok voltak, mint bármely más aminosavé, ezt az aminosavat kihagytuk a számításokból. A kísérleti körülmények, esetleges asszignációs hiba, vagy egyéb még fel nem derített fizikai jelenségek okán lehetnek olyan kísérleti kémiai eltolódás értékek, melyek ezen a legjellemzőbb tartományon kívül esnek. Ennél az 57-es izoleucinnál a kísérleti érték és a program által kiadott értékek nem bizonyulnának összevethetőnek, ezért hagytuk ki ezt az értéket.

	generált 5000		CA + CB RMSD alapján kiválasztott (84 modell)		CA RMSD alapján kiválasztott (45 modell)		MD-ből kiválasztott (CA RMSD, 20 modell)	
	RMSD	corr.	RMSD	corr.	RMSD	corr.	RMSD	corr.
teljes CA kém. elt.	0.458	0.996	0.211	0.999	0.073	1.000	0.400	0.997
másodlagos CA kém. elt.		0.296		0.696		0.953		0.451
teljes CB kém. elt.	0.550	0.999	0.137	1.000	0.601	0.998	0.831	0.997
másodlagos CB kém. elt.		0.233		0.764		0.322		0.301

4.4. táblázat. *A Cd3ε szegmens generált sokaságok megfelelése (RMSD és korreláció értékei) a visszszámolt kémiai eltolódásoknak*

Az előző futtatások alapján a 25-35 aminosavakat magába foglaló szegmensnél hozzáadtunk egy -120, 120 fokos torziós szögekkel és 10 fokos szórással rendelkező eloszlást 0.4-es súllyal. A nagyjából random rendezetlen szerkezeti modellel egyezően a kapott szerkezeti sokaság kémiai eltolódás RMSD-jei és korrelációi jó megfelelést mutatnak, ahogy ez a 4.4. táblázatból és a 4.11. ábrából is látszik. Ami a másodlagos kémiai eltolódásokat illeti, a csak a CA értékekre való szelektálás növelni tudja az azoknak való megfelelést. Ez a CB kémiai eltolódás megfelelést csak kis mértékben javítja, még mindig nem jó megfelelést eredményezve. A CA és CB kémiai eltolódásokat használva egy nagyobb sokaság szelektálódik, alacsonyabb, de még mindig elfogadható korrelációval mindkét adatsorra, mely azt jelenti, hogy a szekvencia mentén lévő lokális konformációk jól reprodukálhatók.

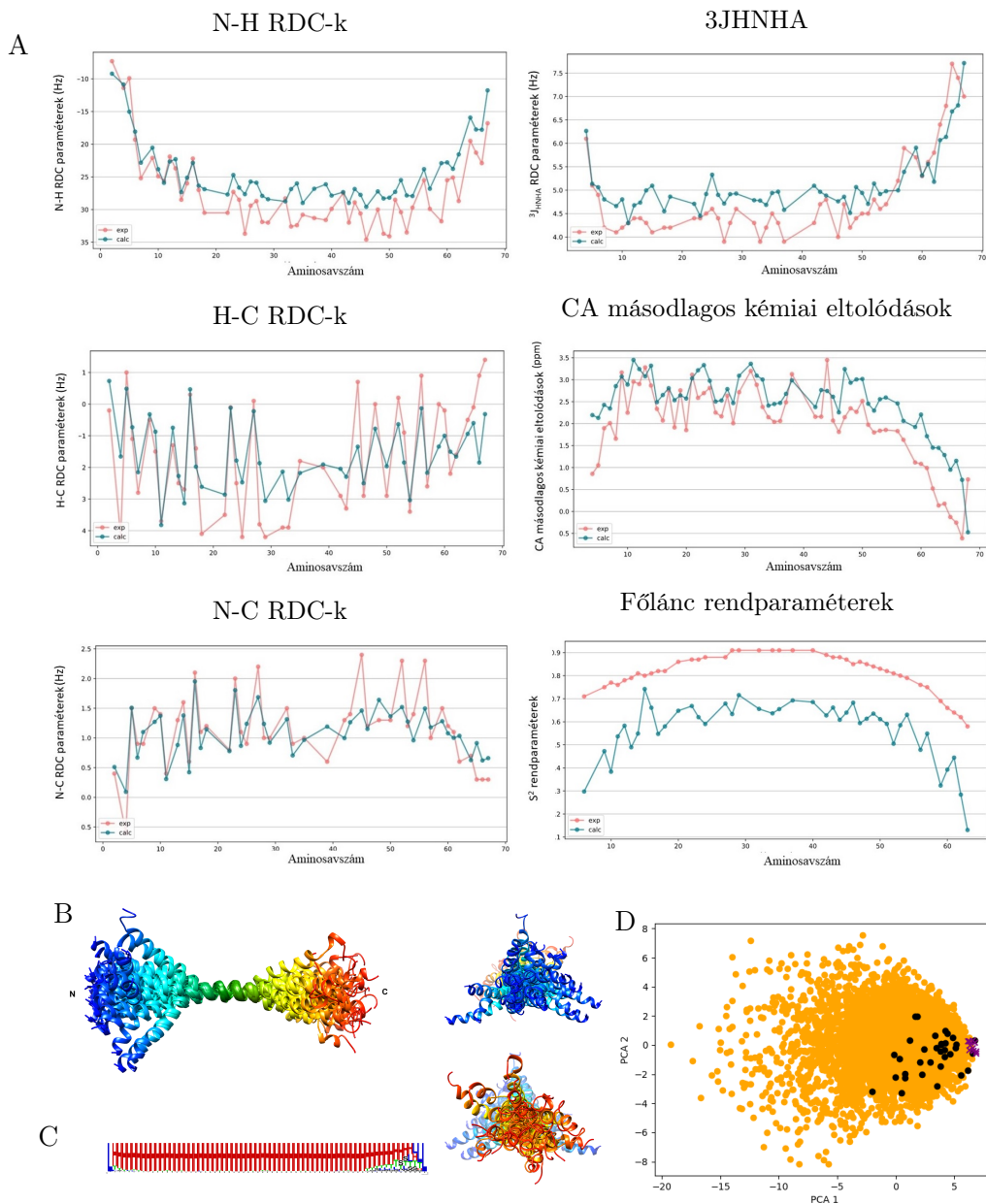
Szelekciót végeztünk az explicit vízben 1 μ s hosszú molekuladinamikai szimulációval kapott sokaságból is. Az ebből választott alsokaság megfelelése rosszabb, mint a DIPEND programmal generálté. Az MD szimuláció során a Cd3 ϵ fehérje gyorsan elér egy kompakt szerkezetet. Főkomponens-analízis alapján az abból kiválasztott alsokaság a kompakt állapotot és a kiindulásihoz közeli nyújtott állapotot is mintavételezi. A DIPEND program által generált szerkezetek ugyanakkor egyenletesebben és jobban mintavételezik a konformációs teret (4.12. ábra). A kísérleti adatoknak való megfelelés teljes vizsgálatához további független adatokra is szükség lenne.

Már korábban leírták, hogy az ITAM (immunoreceptor tyrosine-based activation motif) valamennyire helikális szerkezetet preferál, mely a CA kémiai eltolódásokból is látszik. A sokaságunk esetében ez nem eléggé erős ahhoz, hogy a DSSP bármely aminosavat is helikális szerkezetnek ismerje fel, csak a második motívumban észlelhető jelenléte a "H" állapotnak (4.12. ábra). A "T" állapot sem jelenik meg, hiszen ahhoz négy egymás utáni aminosavnak kellene helikális konformációt felvenni és a "H" állapothoz konzisztens hidrogénkötéseknek kellene létre jönniük, ez viszont elég kicsi valószínűséggel történik meg egy nagyon dinamikus, alacsony másodlagos kémiai eltolódásokkal rendelkező sokaságban. Az ITAM mintázatok 2x4 aminosavainak Ramachandran térképe sem mutat egyértelmű preferenciát az α -helikális torziós szögek iránt, azonban a -54-es ϕ torziós szög közelében lévő enyhe feldúsulás tapasztalható (lásd 4.12. ábra).

A miozin VI magányos alfa-helikális (SAH) szegmense

A 30591 BMRB azonosítójó paraméterfájlt kiegészítettük a hozzá tartozó publikáció [Barnes et al., 2019] mellékletében található ^1N-H , $^2C'-H$ és $^1N-C'$ RDC adatokkal, $^3J_{HNHA}$ skaláris csatolásokkal és S^2 rendparaméterekkel. A szelekcióhoz mindig RMSD-t használtunk. Az RDC adatoknál a Q értékekkel való összehasonlítást nem tesszük meg, mivel a mi számítási módszerünk nem közvetlenül összevethető a Barnes és munkatársai által használt eljárással. Egyrészt ők merev szerkezetekkel számoltak, másrészt mi nem használunk univerzális illesztési tenzort és minden szerkezetet egymástól függetlennek tekintünk, és mindegyiknél külön illesztünk, mielőtt a teljes sokaság RDC-it átlagoljuk.

Mivel egy erősen α -helikális régióról van szó, a Ramachandran térképen szűk területet kellene bejárnia a sokaságnak, ezért indokolt olyan paraméterek választása, melyeknél kis szórással és nagy súllyal dolgozunk. Azonban a másodlagos kémiai eltolódásokat tanul-



4.13. ábra. Mért és visszaszámolt NMR paraméterek és a szelektált SAH sokaság szerkezeti paramétere **A**) A szelektált SAH sokaság Mért (piros) és visszaszámolt (kék) NMR paraméter értékei. A számolt értéket a CoNSEnsX⁺ szerverrel kaptuk meg. **B**) A MYO VI SAH domén kiválasztott 37 konformerének szalag reprezentációja a 28-42 aminosavak mentén illetve. Szivárvány színezés az N terminálistól a C terminálisig. Az ábra az UCSF Chimera programmal készült. **C**) A 37 modellre átlagolt DSSP állapot valószínűségek másodlagos szerkezeti logója. Az ábra a Weblogo programmal készült. **D**) PCA ábra az 1-2 módusok mentén lévő generált 5000 szerkezetnek (narancssárga), a kiválasztott 37 szerkezetnek (fekete) és a 10 PDB adatbázisban található 6OBI azonosítójú szerkezetnek (lila) eloszlásáról. Az 1-es módus a szerkezetek két vége közötti távolságnak felel meg.

mányozva a C terminális szegmensben a helicitás csökkenése várható. Ezért két eloszlást adtunk a program alapbeállításához: egy -58, -47 torziós szög centrumú, 10 fokos szórással és 0.99 súllyal rendelkező eloszlást az 1-55 aminosavakra, és ugyanezt 0.80-as súllyal az 56-68 aminosavakra.

	6OBI (10 modell)		generated 5000		kiválasztott (37 modell)	
	rmsd	corr.	rmsd	corr.	rmsd	corr.
N-H RDC	4.021	0.746	7.304	0.756	3.676	0.917
H-C RDC	1.709	0.456	1.440	0.671	1.058	0.807
N-C RDC	0.822	0.118	0.541	0.576	0.351	0.838
3JHNHA	0.602	0.789	0.765	0.625	0.518	0.903
másodlagos CA kém. elt.	0.969	0.712	0.866	0.705	0.688	0.903
másodlagos CB kém. elt.	0.915	0.389	0.967	0.561	0.970	0.587
N-H S2	0.201	0.488			0.258	0.878

4.5. táblázat. *RMSD és korreláció értékek a MYO VI SAH sokaság kiválasztott és visszazámolt NMR paramétereire*

Ezekkel a paraméterekkel 5000 szerkezetet generáltunk számos próbálkozás után és különböző kiválasztásokkal kísérleteztünk a CoNSEnsX⁺ webszerverrel. A kiválasztott végső sokaságot a következő szelekciós beállításokkal kaptuk: RMSD alapján a következő relatív súlyokkal: CA kémiai eltolódás 10, ³J_{HNHA} skaláris csatolások 5, N-H, N-C és C-H RDC-k (100 mM NaCl-ban mérve) 1. Ezek empirikusan választott súlyparaméterek, melyek azt írják le, hogy egy 1-től 10-ig terjedő skálán mennyire vesszük figyelembe az adott paramétert az alsokaság kiválasztásakor.

Ahogy a 4.13. ábrából és a 4.5. táblázatból is látszik, a szelekció minden paraméterben javulást eredményezett, még az itt nem ábrázolt CB kémiai eltolódásokban is. Ahogy várható volt, a sokaság eltérést mutat a teljesen egyenes α -hélixhez képest. A 4.13. ábrán a 28-42 aminosavakra illesztettük a sokaságot, mivel Barnes és munkatársai NMR relaxációs adatok alapján ezt a szegmenst írták le a legmerevebbnek. Azt meg kell jegyezni, hogy ilyen rúd-szerű szerkezeteknél a szerkezetek megjelenítése nagyban függ attól, hogy mely régiót illesztjük. A DSSPcont analízis azt mutatja, hogy konzisztensen helikálisak a szerkezetek az egész szegmensben. Emellett egyes esetekben a "T", azaz kanyar állapot is megjelenik, mutatva, hogy itt-ott megtörik a folyamatos helicitás, azaz a hidrogén-kötések folytonossága a sokaságban, a szerkezet meghajlását eredményezve. Az 55. aminosavtól csökken a helicitás, mely megfelel a másodlagos kémiai eltolódásoknak és az alkalmazott beállításainknak. A főkomponens-analízisből az látszik, hogy a szelektált alsokaság csak viszonylag kis részét fedi le a felépített modellek által reprezentált konformációs térnek, mely azt jelenti, hogy a kísérleti adatoknak megfelelő sokaság generálásához valóban érdemes volt a lokális preferenciákat igazítani, hiszen enélkül feltehetően még több, nem releváns szerkezet is bekerülne a kezdeti sokaságba.

A legfontosabb kérdés az, hogy a szelektált sokaságunk mennyire reprezentálja a miozin VI fehérje SAH szegmensének belső mozgásait, különösen is összehasonlítva a PDB adatbázisba bekerült NMR szerkezettel, melynek adatai átfednek a vizsgálatunkban szereplő adatokkal. Arról se feledkezzünk el, hogy ennek a 6obi azonosítójú szerkezetnek nem az

a szerepe, hogy a szegmens belső dinamikáját tükrözze, ez az aspektus ott csak az RDC adatok skálázásában jelenik meg. Ezen kívül a mi sokaságunkban nem használtunk NOE kényszerfeltételeket az RDC adatokon kívül. A nem skálázott RDC adatokkal dolgoztunk, a szelekciót pedig másodlagos CA kémiai eltolódásokra és $^3J_{HNHA}$ skaláris csatolásokra végeztük. Ráadásul a kiválasztott sokaságunk 37 szerkezetet tartalmaz, szemben a 10 szerkezetet felvonultató 6obi PDB szerkezettel, és több szerkezetre átlagolva könnyebb elérni jobb megfelelést. Legérdekesebb megfigyelésünk az, hogy az S^2 paraméterek korrelációit tekintve kiválóan megfelel a szelektált sokaságunk, viszont a visszaszámolt értékek lényegesen alacsonyabbak, mint a kísérletből származók. Azt is meg kell jegyeznünk, hogy egy ilyen pálcaszerű szerkezetnél a rendparaméterek becslése nagyban függ a szerkezetek illesztési módjától [Best and Vendruscolo, 2004]. Mi minden aminosavat felhasználtunk ehhez a számoláshoz. Összefoglalásként úgy gondoljuk, hogy a szelektált sokaságunk jól tükrözi a szegmens szerkezeti átrendeződésének természetét, mely azt jelenti, hogy a régió nagy részében a helikális szerkezet helyi megszakadása történik, itt-ott meghajlást okozva a szerkezetben. A hélix kitékeredése csak a 10-15 aminosavból álló C terminális részen figyelhető meg. Barnes és munkatársai azt a konklúziót vonták le, hogy az S^2 rendparaméterek és az RDC adatok ugyanolyan típusú belső mozgásokat írnak le, ami – ha elfogadjuk az S^2 paraméterekkel való korrelációt – megjelenik a mi sokaságunkban is. Azonban ez azt is jelenti, hogy a gyors mozgások kisebb amplitúdóval jelennek meg, mint a mi sokaságunkban, hiszen a kísérletből származó S^2 rendparaméterek magasabbak, mint a mi szerkezetünkéből visszaszámoltak.

5. fejezet

Diszkusszió és tézisek

5.1. A gasztrotropin fehérje vizsgálata

A gasztrotropin fehérje többek között epesavak szállításában vesz részt. Ligandumkötésének atomi szintű leírása még mindig nem teljes. E kérdést vizsgáltuk *in silico* módszerekkel molekuladinamikai szimulációkat végezve a mások által kísérletesen meghatározott szerkezeti modelleken a kísérleti adatokból származó S^2 és NOE paraméterek felhasználásával, illetve azok nélkül is.

Felmerülhet az olvasóban a kérdés, hogy miért nem végeztünk hosszabb szimulációkat a gasztrotropin fehérje esetében. A MUMO protokoll alkalmazása nagyon különleges algoritmust igényel, ami egészen más, mint amilyen algoritmusokat a legtöbben használnak molekuladinamikai szimulációknál, és amelyekben az utóbbi években drasztikus hardveres gyorsítás következett be, mint például a grafikus kártyák alkalmazhatósága által. Sajnos a MUMO protokollra ezek nem alkalmazhatók, így sokkal több időt igényelnek, ezért nem volt lehetőség még több szimuláció elvégzésére. Fontos megjegyezni, hogy az ezzel a módszerrel és programverzióval végzett szimulációk sebessége nem vethető össze a manapság rutinszerűen futtatott, GPU alapú gyorsítást használó számításokkal. Ennek több, technikai jellegű oka van. Egyrészt a jelenlegi GPU alapú gyorsítások tipkusan úgy működnek, hogy egyetlen replika esetében adott típusú számításokat végez a GPU. Több replika szimulációja esetében párhuzamosan több CPU szükséges, és ebben az esetben nem triviális a megfelelő részsámítások GPU-ra való kihelyezése minden replika esetében, tudásunk szerint ezt a GROMACS sem támogatja jelenleg. Ezen felül az S^2 és NOE megkötések számítása minden lépésben intenzív kommunikációt igényel az egyes szálak között, ami jelentős plusz számítási igényt jelent. Külön nehézségként jelentkezik, hogy a GROMACS fejlesztői egyre kevésbé érdeklődnek az NMR paraméterek beépítése iránt, mivel ezt újabban a PLUMED modullal oldják meg független fejlesztők, ami viszont nem kezel S^2 paramétereket. Ezek miatt a korábbi S^2 implementáció új GROMACS verziókba való beépítése komoly plusz erőfeszítéseket igényelne, amelytől a csoportunk inkább eltekintett. Összességében tehát a MUMO protokollal elvégzett számítások össz-szimulációs ideje alatta marad a manapság rutinnak mondható, egyedi replikákat használó molekuladinamikai futások szimulációs idejének. Ez a hátrány azonban csupán látszólagos, hiszen a kísérleti megkötéseket alkalmazó szimulációk esetében a modellezett belső mozgások

jellemző időskáláját nem a szimuláció ideje határozza meg elsősorban, hanem az, hogy az alkalmazott kísérleti paraméterek milyen időskálájú mozgásokat reprezentálnak.

Felmerült, hogy a ligandumkötés energiaszintjeinek még behatóbb tanulmányozására végezhetünk volna a különböző komplexeken szabadenergia számolásokat [Abel et al., 2017] vagy "esernyő mintavételezést" (umbrella sampling) [You et al., 2019], hogy megbízhatóbb becslést kapjunk az energiákról. Azonban ezek elvégzése a félig kitekeredett állapot miatt egyáltalán nem triviális, hiszen ezek a konformációk eleve nem egyensúlyiak, így a módszer nem működött volna. Emellett a két ligandum miatt nagyon sok kombinációt kellett volna megvizsgálni, mely túlnyúlt volna e munka keretein.

5.2. Első téziscsoport

1.1.A. Tézis:

Előállítottam S^2 és NOE paraméterekkel végzett megkötött molekuladinamikai szimulációkkal a MUMO módszer segítségével a gasztrotropin 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 (4.1. 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 4.6. 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 főkomponens analízis segítségével 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, 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.

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 aminosavcsoporttal.

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 félig kitekeredik.

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. Ebben azonban az erőternek is szerepe lehet.

5.3. Fehérjeszerkezet modellezés a DIPEND program segítségével

Rendezetlen fehérjék szerkezeti modellezésére fejlesztettük ki a DIPEND programot, mely nyílt forráskódú, és maga is nyílt forráskódú programokra (vagyis a ChimeraX, SCWRL4 és GROMACS) épül. Ez növeli a program átláthatóságát és működésének érthetőségét. A nyílt forráskód azért is előnyös, mert ezeket a programokat folyamatosan fejlesztik és bárki szabadon tehet hozzájuk újabb funkciókat vagy javíthat ki hibákat. Ezen előnyök nagyobb súllyal esnek latba, mint a külső programok használatából eredő futásidő növekedés.

A DIPEND program fontos sajátossága, hogy csak a szerkezeti sokaság generálás lépést teszi meg, a kísérleti adatok alapján lévő szelekció lépését pedig a kutatócsoportunkban fejlesztett CoNSEnsX⁺ webserver segítségével végzi, melynek fejlesztésében nem vettem részt. A felhasználó által a szerkezetgenerálásakor megadható szerkezeti preferenciák biztosítják azt, hogy a generálás átlátható, nem fekete dobozként működik. A szerkezetgenerálás és sokaságkiértékelés lépések szétválasztása azt az előnyt is magában hordozza, hogy így többféle szerkezet generálási módszerből származó sokaságokat is egyesíteni lehet, például a DIPEND által előállított és molekuladinamikai szimulációkból származó szerkezeteket. Ráadásul így a kísérleti adatoknak való megfelelés vizsgálatához is lehet többféle módszerből választani, vagy akár egyszerre többel megvizsgálni az eredményül kapott sokaságot és konszenzust keresni közöttük. Emellett a kiértékelési lépés nagyon függ attól, hogy milyen célra készítjük a sokaság szerkezeti modellt, vagy hogy egyáltalán milyen kísérleti paraméterekkel dolgozunk. Gyengesége a programnak, hogy megbízhatóan csak legfeljebb 100 aminosavas szerkezetekre tud működni a sztérikus ütközések miatt a kicsomózási modul működése nélkül. Azonban a kicsomózási modult érdemes használni már 75 aminosav hosszú vagy annál hosszabb szakaszokon is, azonban annál rövidebbeknél nem jelentős az előny, ha e modult használjuk. Amennyiben túl sok a prolin aminosav a szekvenciában, akkor rövid szekvencia esetén is nagyobb valószínűséggel fordulhatnak elő sztérikus ütközések. E probléma megoldása is egy lehetséges jövőbeli programfejlesztési terv lehet. A későbbiekben cél lenne a futásidő csökkentése azáltal, hogy beleépítünk

a programba egy párhuzamosító részt, mellyel több szerkezetet lehet egyazon időben generálni és a felépítés sikerességét vizsgálni. Természetesen ez "kézzel" ebben a formában is megoldható, vagyis több különböző könyvtárban el lehet indítani a programot több, de kisebb számú sokaságot generálva. Emellett tervezünk készíteni egy saját scriptet, ami összefűzi a sikeres szerkezeteket egy sokasággá.

5.4. Második téziscsoport

2.1. Tézis:

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

A program a szomszédsági valószínűségi modellt alkalmazva megoldja azt a problémát, hogy az adott fehérjeszerkezet biológiailag releváns konformációi körül egy jó konformációs tér mintavételezést hajtson végre. A súllyal hozzáadható plusz eloszlással pedig tükrözhető 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.

2.2. Tézis:

A DIPEND programot a humán Cd3 ϵ 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 sokaság szelektálható, mint az egyéb módszerrel generált konformációs könyvtárakból.

A generált sokaságból szelektált alsokaság jobban megfelelt a kísérleti adatoknak, mint a molekuladinamikai számításból származó sokaság.

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 jól lehet egy kísérleti adatoknak megfelelő alsokaságot kiválasztani.

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

5.5. Kitekintés

Noha a célkitűzésben felvázolt feladatokat megoldottam, még jónéhány feladat, megválaszolatlan kérdés maradt a területen. Ilyenek a gasztrotropin fehérje ligandumkötésének pontosabb energiaviszony feltérképezése. A CD3 ϵ fehérjénél újabb kísérleti adatok beépítésével még pontosabb modell felépítése, majd e sokaság összeillesztése a fehérje többi részéről már elérhető szerkezeti modellekkel, így modellezve a teljes fehérjét. Hasonló feladat merül fel a miozin fehérje modellezésében is. Mindemellett nagyon sok olyan fehérje van, amely szerkezeti modellépítésre vár, hogy jobban megértsük működését. Ha pedig egy adott teljes fehérje modellezése megtörtént, akkor e fehérjemodellek egymáshoz való dokkolásával fehérjekomplexek modellezhetők, melyekből idővel akár egy teljes sejt atomi szintű modellje is felépíthetővé válna, a dinamikát is figyelembe véve. A lehetőségek háttartalnok, azonban e lépték megvalósulása még valószínűleg nagyon sokáig várat majd magára.

A fehérjék atomi szerkezetének és belső dinamikájának vizsgálata szerkezeti sokaságok segítségével, vagyis a mozgás felvázolása elkapott pillanatképekkel közelebb visz a fehérjék funkciójának jobb megértéséhez, mely a biológiai folyamatok molekuláris feltérképezéséhez visz közelebb egy lépéssel. Fehérjék szerkezetének és belső dinamikájának megértése segíthet abban is, hogy a későbbiekben az adott fehérje gyógyszermolekulákkal célozható legyen. Egy másik aspektusa a dolognak, hogy az atomi szintű szerkezet és dinamika megértése bizonyos mutációk fehérjeműködésre gyakorolt hatását is feltérképezheti, mely szintén egy érdekes terület.

6. fejezet

Köszönetnyilvánítás

Munkám során nagyon sokan segítettek abban, hogy eljussak ideig. Köszönet illeti a következőket:

- Legelsősorban témavezetőmnek, Dr. Gáspári Zoltánnak azért, hogy végtelen türelemmel vezette munkámat, hogy mindig rendelkezésre állt, emellett megértéséért és biztatásáért!
- Köszönöm az együttműködő partnereknek: Tőke Orsolyának, Szabó András Lászlónak és Dudola Dánielnek sok-sok munkájukat és támogatásukat!
- A kutatócsoporttársak támogatását és segítségét is nagyon köszönöm, különösen is Dudola Dánielnek és Kovács Bertalannak!
- Köszönöm Alessandro Bottaronak a BME program futtatásában adott segítségét!
- Köszönöm Reguly Istvánnak, hogy használhattam a számítógépklasztert és hathatós segítségét is annak használatában!
- Köszönöm Vida Tivadarné Katinka néninek is az adminisztratív ügyekben való segítségét és támogatását!
- Az Alapítvány a Magyar Peptid-és Fehérjekutatásért ösztöndíja is eszenciális volt ahhoz, hogy munkámat be tudjam fejezni, köszönöm anyagi támogatásukat!
- Köszönöm a Doktori Iskolának és a Pázmány Péter Katolikus Egyetem Információs Technológiai és Bionikai Karának is, hogy lehetőséget kaptam a munkára és hogy anyagilag és sok egyéb módon is támogattak!
- Köszönöm családomnak, kedvesemnek és barátaimnak is, hogy mellettem álltak és támogattak, különösen is Édesapámnak az élénk szakmai beszélgetéseket és kedvesemnek, Szuromi Tamásnak a (nyelvi) lektorálást!

Emellett mindenki másnak is, akit esetleg nem említettem meg, nagyon köszönöm a segítséget, támogatást!

Irodalomjegyzék

- [Abel et al., 2017] Abel, R., Wang, L., Mobley, D. L. and Friesner, R. A. (2017). A Critical Review of Validation, Blind Testing, and Real- World Use of Alchemical Protein-Ligand Binding Free Energy Calculations. *Curr Top Med Chem* 17, 2577–2585.
- [Abraham et al., 2015] Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B. and Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1-2, 19 – 25.
- [Alrefai and Gill, 2007] Alrefai, W. A. and Gill, R. K. (2007). Bile acid transporters: structure, function, regulation and pathophysiological implications. *Pharm Res* 24, 1803–1823.
- [Andersen et al., 2002] Andersen, C. A., Palmer, A. G., Brunak, S. and Rost, B. (2002). Continuum secondary structure captures protein flexibility. *Structure* 10, 175–184.
- [Ángyán et al., 2010] Ángyán, A. F., Szappanos, B., Perczel, A. and Gáspári, Z. (2010). CoNSEnsX: an ensemble view of protein structures and NMR-derived experimental data. *BMC Structural Biology* 10, 39.
- [Ayers et al., 2007] Ayers, S. D., Nedrow, K. L., Gillilan, R. E. and Noy, N. (2007). Continuous nucleocytoplasmic shuttling underlies transcriptional activation of PPARgamma by FABP4. *Biochemistry* 46, 6744–6752.
- [Bakan et al., 2011] Bakan, A., Meireles, L. M. and Bahar, I. (2011). ProDy: protein dynamics inferred from theory and experiments. *Bioinformatics* 27, 1575–1577.
- [Baker and Hubbard, 1984] Baker, E. and Hubbard, R. (1984). Hydrogen bonding in globular proteins. *Progress in Biophysics and Molecular Biology* 44, 97–179.
- [Banaszak et al., 1994] Banaszak, L., Winter, N., Xu, Z., Bernlohr, D. A., Cowan, S. and Jones, A. T. (1994). Lipid-binding proteins: a family of fatty acid and retinoid transport proteins. *Adv Protein Chem* 45, 89–151.
- [Barnes et al., 2019] Barnes, C. A., Shen, Y., Ying, J., Takagi, Y., Torchia, D. A., Sellers, J. R. and Bax, A. (2019). Remarkable Rigidity of the Single α -Helical Domain of Myosin-VI As Revealed by NMR Spectroscopy. *Journal of the American Chemical Society* 141, 9004–9017.

- [Berman et al., 2000] Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. and Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Res* 28, 235–242.
- [Best and Vendruscolo, 2004] Best, R. B. and Vendruscolo, M. (2004). J Am Chem Soc-Determination of protein structures consistent with NMR order parameters. *J Am Chem Soc* 126, 8090–8091.
- [Boehr et al., 2009] Boehr, D. D., Nussinov, R. and Wright, P. E. (2009). The role of dynamic conformational ensembles in biomolecular recognition. *Nature Chemical Biology* 5, 789–796.
- [Bottaro et al., 2020] Bottaro, S., Bengtsen, T. and Lindorff-Larsen, K. (2020). Integrating Molecular Simulation and Experimental Data: A Bayesian/Maximum Entropy Reweighting Approach. *Methods Mol Biol* 2112, 219–240.
- [Cai et al., 2012] Cai, J., Lücke, C., Chen, Z., Qiao, Y., Klimtchuk, E. and Hamilton, J. A. (2012). Solution structure and backbone dynamics of human liver fatty acid binding protein: fatty acid binding revisited. *Biophys J* 102, 2585–2594.
- [Chao and Byrd, 2020] Chao, F.-A. and Byrd, A. R. (2020). Protein Dynamics revealed by NMR Relaxation Methods. *Emerg Top Life Sci* 2, 93–105.
- [Cogliati et al., 2010] Cogliati, C., Ragona, L., D’Onofrio, M., Günther, U., Whittaker, S., Ludwig, C., Tomaselli, S., Assfalg, M. and Molinari, H. (2010). Site-specific investigation of the steady-state kinetics and dynamics of the multistep binding of bile acid molecules to a lipid carrier protein. *Chemistry* 16, 11300–11310.
- [Czajlik et al., 2017] Czajlik, A., Kovács, B., Permi, P. and Gáspári, Z. (2017). Fine-tuning the extent and dynamics of binding cleft opening as a potential general regulatory mechanism in parvulin-type peptidyl prolyl isomerases. *Sci Rep* 7, 44504.
- [David and Jacobs, 2014] David, C. C. and Jacobs, D. J. (2014). Principal component analysis: a method for determining the essential dynamics of proteins. *Methods Mol Biol* 1084, 193–226.
- [Dobson, 2003] Dobson, C. M. (2003). Protein folding and misfolding. *Nature* 426, 884–890.
- [Dudola et al., 2017] Dudola, D., Kovács, B. and Gáspári, Z. (2017). CoNSEnsX+ Web-server for the Analysis of Protein Structural Ensembles Reflecting Experimentally Determined Internal Dynamics. *Journal of Chemical Information and Modeling* 57, 1728–1734.
- [Dunker et al., 2001] Dunker, A., Lawson, J., Brown, C. J., Williams, R. M., Romero, P., Oh, J. S., Oldfield, C. J., Campen, A. M., Ratliff, C. M., Hipps, K. W., Ausio, J., Nissen, M. S., Reeves, R., Kang, C., Kissinger, C. R., Bailey, R. W., Griswold, M. D., Chiu, W., Garner, E. C. and Obradovic, Z. (2001). Intrinsically disordered protein. *J. Mol. Graph. Model.* 19, 26–59.

- [Dániel Dudola, 2020] Dániel Dudola, Bertalan Kovács, Z. G. (2020). Evaluation and Selection of Dynamic Protein Structural Ensembles with CoNSEnsX. *Methods Mol Biol* *2112*, 241–254.
- [Eliseo et al., 2007] Eliseo, T., Ragona, L., Catalano, M., Assfalg, M., Paci, M., Zetta, L., Molinari, H. and Cicero, D. O. (2007). Structural and dynamic determinants of ligand binding in the ternary complex of chicken liver bile acid binding protein with two bile salts revealed by NMR. *Biochemistry* *46*, 12557–12567.
- [F. Ángyán and Gáspári, 2013] F. Ángyán, A. and Gáspári, Z. (2013). Ensemble-Based Interpretations of NMR Structural Data to Describe Protein Internal Dynamics. *Molecules* *18*, 10548–10567.
- [Feldman and Hogue, 2002] Feldman, H. J. and Hogue, C. W. (2002). Probabilistic sampling of protein conformations: New hope for brute force? *Proteins: Structure, Function, and Bioinformatics* *46*, 8–23.
- [Fizil et al., 2015] Fizil, Á., Gáspári, Z., Barna, T., Marx, F. and Batta, G. (2015). "Invisible" conformers of an antifungal disulfide protein revealed by constrained cold and heat unfolding, CEST-NMR experiments, and molecular dynamics calculations. *Chemistry* *21*, 5136–5144.
- [Friesner et al., 2006] Friesner, R. A., Murphy, R. B., Repasky, M. P., Frye, L. L., Greenwood, J. R., Halgren, T. A., Sanschagrin, P. C. and Mainz, D. T. (2006). Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J Med Chem* *49*, 6177–6196.
- [Gabriel et al., 2004] Gabriel, E., Fagg, G. E., Bosilca, G., Angskun, T., Dongarra, J. J., Squyres, J. M., Sahay, V., Kambadur, P., Barrett, B., Lumsdaine, A., Castain, R. H., Daniel, D. J., Graham, R. L. and Woodall, T. S. (2004). Open MPI: Goals, Concept, and Design of a Next Generation MPI Implementation. In *Recent Advances in Parallel Virtual Machine and Message Passing Interface*, (Kranzlmüller, D., Kacsuk, P. and Dongarra, J., eds), pp. 97–104, Springer Berlin Heidelberg, Berlin, Heidelberg.
- [Gáspári and Perczel, 2010] Gáspári, Z. and Perczel, A. (2010). Chapter 2 - Protein Dynamics as Reported by NMR. In *Annual Reports on NMR Spectroscopy*, (Webb, G. A., ed.), vol. 71, of *Annual Reports on NMR Spectroscopy* pp. 35–75. Academic Press.
- [Han et al., 2011] Han, B., Liu, Y., Ginzinger, S. W. and Wishart, D. S. (2011). SHIFTX2: significantly improved protein chemical shift prediction. *J Biomol NMR* *50*, 43–57.
- [Hess et al., 1997] Hess, B., Bekker, H., Berendsen, H. J. and Fraaije, J. G. (1997). LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry* *18*, 1463–1472.

- [Higman et al., 2011] Higman, V. A., Boyd, J., Smith, L. J. and Redfield, C. (2011). Residual dipolar couplings: are multiple independent alignments always possible? *Journal of Biomolecular NMR* *49*, 53–60.
- [Hodsdon and Cistola, 1997] Hodsdon, M. E. and Cistola, D. P. (1997). Discrete backbone disorder in the nuclear magnetic resonance structure of apo intestinal fatty acid-binding protein: implications for the mechanism of ligand entry. *Biochemistry* *36*, 1450–1460.
- [Horváth et al., 2019] Horváth, G., Egyed, O., Tang, C., Kovács, M., Micsonai, A., Kardos, J. and Toke, O. (2019). Ligand entry in human ileal bile acid-binding protein is mediated by histidine protonation. *Scientific Reports* *9*, 4825.
- [Horváth et al., 2012] Horváth, G., Király, P., Tárkányi, G. and Toke, O. (2012). Internal Motions and Exchange Processes in Human Ileal Bile Acid Binding Protein As Studied by Backbone ^{15}N Nuclear Magnetic Resonance Spectroscopy. *Biochemistry* *51*, 1848–1861.
- [Horváth et al., 2016] Horváth, G., Bencsura, A., Simon, A., Tochtrop, G. P., DeKoster, G. T., Covey, D. F., Cistola, D. P. and Toke, O. (2016). Structural determinants of ligand binding in the ternary complex of human ileal bile acid binding protein with glycocholate and glycochenodeoxycholate obtained from solution NMR. *The FEBS journal* *283*, 541–555.
- [Horváth et al., 2017] Horváth, G., Biczók, L., Majer, Z., Kovács, M., Micsonai, A., Kardos, J. and Toke, O. (2017). Structural insight into a partially unfolded state preceding aggregation in an intracellular lipid-binding protein. *FEBS J* *284*, 3637–3661.
- [Horváth et al., 2014] Horváth, G., Egyed, O. and Toke, O. (2014). Temperature dependence of backbone dynamics in human ileal bile acid-binding protein: implications for the mechanism of ligand binding. *Biochemistry* *53*, 5186–5198.
- [Huang and Wange, 2004] Huang, Y. and Wange, R. L. (2004). T Cell Receptor Signaling: Beyond Complex Complexes. *Journal of Biological Chemistry* *279*, 28827–28830.
- [Humphrey et al., 1996] Humphrey, W., Dalke, A. and Schulten, K. (1996). VMD: visual molecular dynamics. *J Mol Graph* *14*, 33–38.
- [Isaksson et al., 2013] Isaksson, L., Mayzel, M., Saline, M., Pedersen, A., Rosenlöv, J., Brutscher, B., Karlsson, B. G. and Orekhov, V. Y. (2013). Highly Efficient NMR Assignment of Intrinsically Disordered Proteins: Application to B- and T Cell Receptor Domains. *PLOS ONE* *8*, 1–8.
- [Jenkins et al., 2002] Jenkins, A. E., Hockenberry, J. A., Nguyen, T. and Bernlohr, D. A. (2002). Testing of the portal hypothesis: analysis of a V32G, F57G, K58G mutant of the fatty acid binding protein of the murine adipocyte. *Biochemistry* *41*, 2022–2027.

- [Jenny Gu (Editor), 2009] Jenny Gu (Editor), P. E. B. E. (2009). Structural Bioinformatics, 2nd Edition. 2 edition, J. Wiley & Sons.
- [Jorgensen et al., 1983] Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. and Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* 79, 926–935.
- [Kony et al., 2002] Kony, D., Damm, W., Stoll, S. and Van Gunsteren, W. (2002). An improved OPLS-AA force field for carbohydrates. *J Comput Chem* 23, 1416–1429.
- [Kouvatsos et al., 2007] Kouvatsos, N., Thurston, V., Ball, K., Oldham, N. J., Thomas, N. R. and Searle, M. S. (2007). Bile acid interactions with rabbit ileal lipid binding protein and an engineered helixless variant reveal novel ligand binding properties of a versatile beta-clam shell protein scaffold. *J Mol Biol* 371, 1365–1377.
- [Krivov et al., 2009] Krivov, G. G., Shapovalov, M. V. and Dunbrack, R. L. (2009). Improved prediction of protein side-chain conformations with SCWRL4. *Proteins* 77, 778–795.
- [Krzeminski et al., 2013] Krzeminski, M., Marsh, J. A., Neale, C., Choy, W.-Y. and Forman-Kay, J. D. (2013). Characterization of disordered proteins with ENSEMBLE. *Bioinformatics* 29, 398–399.
- [Kurz et al., 2003] Kurz, M., Brachvogel, V., Matter, H., Stengelin, S., Thüring, H. and Kramer, W. (2003). Insights into the bile acid transportation system: the human ileal lipid-binding protein-cholytaurine complex and its comparison with homologous structures. *Proteins* 50, 312–328.
- [Lazar et al., 2021] Lazar, T., Martínez-Pérez, E., Quaglia, F., Hatos, A., Chemes, L. B., Iserte, J. A., Méndez, N. A., Garrone, N. A., Saldaño, T. E., Marchetti, J., Rueda, A. J. V., Bernadó, P., Blackledge, M., Cordeiro, T. N., Fagerberg, E., Forman-Kay, J. D., Fornasari, M. S., Gibson, T. J., Gomes, G. W., Gradinaru, C. C., Head-Gordon, T., Jensen, M. R., Lemke, E. A., Longhi, S., Marino-Buslje, C., Minervini, G., Mittag, T., Monzon, A. M., Pappu, R. V., Parisi, G., Ricard-Blum, S., Ruff, K. M., Salladini, E., Skepö, M., Svergun, D., Vallet, S. D., Varadi, M., Tompa, P., Tosatto, S. C. E. and Piovesan, D. (2021). PED in 2021: a major update of the protein ensemble database for intrinsically disordered proteins. *Nucleic Acids Res* 8, D404–D41.
- [Levinthal, Cyrus, 1968] Levinthal, Cyrus (1968). Are there pathways for protein folding? *J. Chim. Phys.* 65, 44–45.
- [Lin et al., 1990] Lin, M. C., Kramer, W. and Wilson, F. A. (1990). Identification of cytosolic and microsomal bile acid-binding proteins in rat ileal enterocytes. *J Biol Chem* 265, 14986–14995.
- [Lindorff-Larsen et al., 2005] Lindorff-Larsen, K., Best, R. B., Depristo, M. A., Dobson, C. M. and Vendruscolo, M. (2005). Simultaneous determination of protein structure and dynamics. *Nature* 433, 128–132.

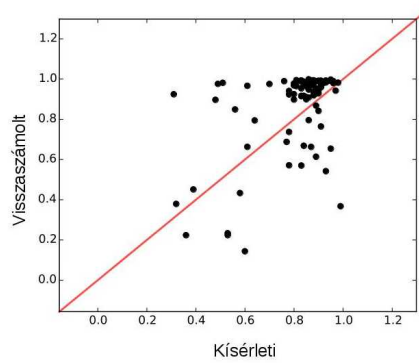
- [Lipari and Szabo, 1982] Lipari, G. and Szabo, A. (1982). Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *Journal of the American Chemical Society* *104*, 4546–4559.
- [Lücke et al., 2000] Lücke, C., Zhang, F., Hamilton, J. A., Sacchettini, J. C. and Rüterjans, H. (2000). Solution structure of ileal lipid binding protein in complex with glycocholate. *Eur J Biochem* *267*, 2929–2938.
- [Mészáros et al., 2018] Mészáros, B., Erdős, G. and Dosztányi, Z. (2018). IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Research* *46*, W329–W337.
- [Nakahara et al., 2005] Nakahara, M., Furuya, N., Takagaki, K., Sugaya, T., Hirota, K., Fukamizu, A., Kanda, T., Fujii, H. and Sato, R. (2005). Ileal bile acid-binding protein, functionally associated with the farnesoid X receptor or the ileal bile acid transporter, regulates bile acid activity in the small intestine. *J Biol Chem* *280*, 42283–42289.
- [Neal et al., 2003] Neal, S., Nip, A. M., Zhang, H. and Wishart, D. S. (2003). Rapid and accurate calculation of protein ¹H, ¹³C and ¹⁵N chemical shifts. *J Biomol NMR* *26*, 215–240.
- [Ozenne et al., 2012] Ozenne, V., Bauer, F., Salmon, L., Huang, J.-R., Jensen, M. R., Segard, S., Bernadó, P., Charavay, C. and Blackledge, M. (2012). Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. *Bioinformatics* *28*, 1463–1470.
- [Pakhrin et al., 2021] Pakhrin, S. C., Shrestha, B., Adhikari, B. and Kc, D. B. (2021). Deep Learning-Based Advances in Protein Structure Prediction. *Int J Mol Sci* *22*.
- [Pettersen et al., 2004] Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. and Ferrin, T. E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* *25*, 1605–1612.
- [Pettersen et al., 2021] Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., Morris, J. H. and Ferrin, T. E. (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci* *30*, 70–82.
- [Pronk et al., 2013] Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M. R., Smith, J. C., Kasson, P. M., van der Spoel, D., Hess, B. and Lindahl, E. (2013). GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* *29*, 845–854.
- [Quaglia et al., 2021] Quaglia, F., Lazar, T., Hatos, A., Tompa, P., Piovesan, P. D. and E., T. S. C. (2021). Exploring curated conformational ensembles of intrinsically disordered proteins in the protein ensemble database. *Curr Protoc* *1*, e192.

- [Ragona et al., 2014] Ragona, L., Pagano, K., Tomaselli, S., Favretto, F., Ceccon, A., Zanzoni, S., D’Onofrio, M., Assfalg, M. and Molinari, H. (2014). The role of dynamics in modulating ligand exchange in intracellular lipid binding proteins. *Biochim Biophys Acta* *1844*, 1268–1278.
- [Richter et al., 2007] Richter, B., Gsponer, J., Várnai, P., Salvatella, X. and Vendruscolo, M. (2007). The MUMO (minimal under-restraining minimal over-restraining) method for the determination of native state ensembles of proteins. *J Biomol NMR* *37*, 117–135.
- [Sanchez-Martinez and Crehuet, 2014] Sanchez-Martinez, M. and Crehuet, R. (2014). Application of the maximum entropy principle to determine ensembles of intrinsically disordered proteins from residual dipolar couplings. *Phys Chem Chem Phys* *16*, 26030–26039.
- [Sessler and Noy, 2005] Sessler, R. J. and Noy, N. (2005). A ligand-activated nuclear localization signal in cellular retinoic acid binding protein-II. *Mol Cell* *18*, 343–353.
- [Sivaramakrishnan et al., 2008] Sivaramakrishnan, S., Spink, B. J., Sim, A. Y. L., Doniach, S. and Spudich, J. A. (2008). Dynamic charge interactions create surprising rigidity in the ER/K α -helical protein motif. *Proceedings of the National Academy of Sciences* *105*, 13356–13361.
- [Smathers and Petersen, 2011] Smathers, R. L. and Petersen, D. R. (2011). The human fatty acid-binding protein family: Evolutionary divergences and functions. *Human Genomics* *5*, 170.
- [Tamiola et al., 2010] Tamiola, K., Acar, B. and Mulder, F. A. A. (2010). Sequence-Specific Random Coil Chemical Shifts of Intrinsically Disordered Proteins. *Journal of the American Chemical Society* *132*, 18000–18003.
- [Ting et al., 2010] Ting, D., Wang, G., Shapovalov, M., Mitra, R., Jordan, M. I. and Dunbrack, Jr, R. L. (2010). Neighbor-Dependent Ramachandran Probability Distributions of Amino Acids Developed from a Hierarchical Dirichlet Process Model. *PLOS Computational Biology* *6*, 1–21.
- [Tochtrop et al., 2003] Tochtrop, G. P., Bruns, J. L., Tang, C., Covey, D. F. and Cistola, D. P. (2003). Steroid ring hydroxylation patterns govern cooperativity in human bile acid binding protein. *Biochemistry* *42*, 11561–11567.
- [Tochtrop et al., 2004] Tochtrop, G. P., DeKoster, G. T., Covey, D. F. and Cistola, D. P. (2004). A single hydroxyl group governs ligand site selectivity in human ileal bile acid binding protein. *J Am Chem Soc* *126*, 11024–11029.
- [Tochtrop et al., 2002] Tochtrop, G. P., Richter, K., Tang, C., Toner, J. J., Covey, D. F. and Cistola, D. P. (2002). Energetics by NMR: site-specific binding in a positively cooperative system. *Proc Natl Acad Sci U S A* *99*, 1847–1852.

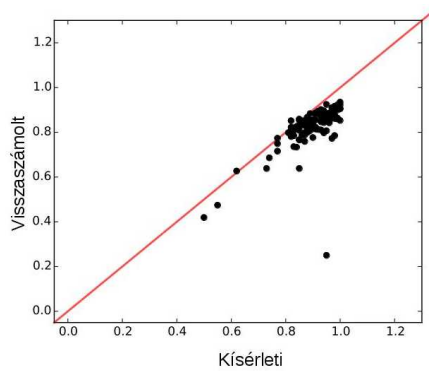
- [Toke et al., 2006] Toke, O., Monsey, J. D., DeKoster, G. T., Tochtrop, G. P., Tang, C. and Cistola, D. P. (2006). Determinants of cooperativity and site selectivity in human ileal bile acid binding protein. *Biochemistry* *45*, 727–737.
- [Tompa, 2012] Tompa, P. (2012). Intrinsically disordered proteins: a 10-year recap. *Trends in Biochemical Sciences* *37*, 509 – 516.
- [Trauner and Boyer, 2003] Trauner, M. and Boyer, J. L. (2003). Bile Salt Transporters: Molecular Characterization, Function, and Regulation. *Physiological Reviews* *83*, 633–671.
- [Turpin et al., 2013] Turpin, E. R., Fang, H.-J., Thomas, N. R. and Hirst, J. D. (2013). Cooperativity and site selectivity in the ileal lipid binding protein. *Biochemistry* *52*, 4723–4733.
- [Ulrich et al., 2007] Ulrich, E. L., Akutsu, H., Doreleijers, J. F., Harano, Y., Ioannidis, Y. E., Lin, J., Livny, M., Mading, S., Maziuk, D., Miller, Z., Nakatani, E., Schulte, C. F., Tolmie, D. E., Kent Wenger, R., Yao, H. and Markley, J. L. (2007). BioMagResBank. *Nucleic Acids Research* *36*, D402–D408.
- [UniProt, 2020] UniProt, T. C. (2020). UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research* *49*, D480–D489.
- [Van Der Spoel et al., 2005] Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E. and Berendsen, H. J. C. (2005). GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry* *26*, 1701–1718.
- [Veerkamp and Maatman, 1995] Veerkamp, J. H. and Maatman, R. G. (1995). Cytoplasmic fatty acid-binding proteins: Their structure and genes. *Progress in Lipid Research* *34*, 17–52.
- [Wang and Bax, 1996] Wang, A. C. and Bax, A. (1996). Determination of the Backbone Dihedral Angles ϕ in Human Ubiquitin from Reparametrized Empirical Karplus Equations. *Journal of the American Chemical Society* *118*, 2483–2494.
- [Xu et al., 1997] Xu, D., Tsai, C. J. and Nussinov, R. (1997). Hydrogen bonds and salt bridges across protein-protein interfaces. *Protein Engineering, Design and Selection* *10*, 999–1012.
- [You et al., 2019] You, W., Tang, Z. and Chang, C.-e. A. (2019). Potential Mean Force from Umbrella Sampling Simulations: What Can We Learn and What Is Missed? *J Chem Theory Comput* *15*, 2433–2443.
- [Zweckstetter and Bax, 2000] Zweckstetter, M. and Bax, A. (2000). Prediction of Sterically Induced Alignment in a Dilute Liquid Crystalline Phase: Aid to Protein Structure Determination by NMR. *Journal of the American Chemical Society* *122*, 3791–3792.

[Zwicker and Agellon, 2013] Zwicker, B. L. and Agellon, L. B. (2013). Transport and biological activities of bile acids. *The International Journal of Biochemistry & Cell Biology* 45, 1389–1398.

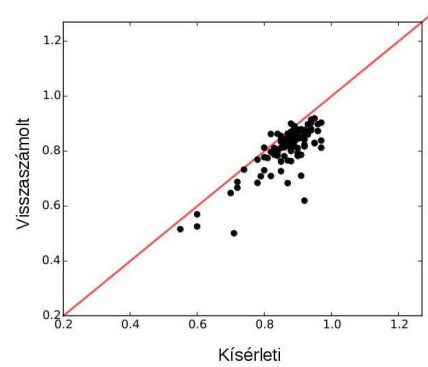
Függelék



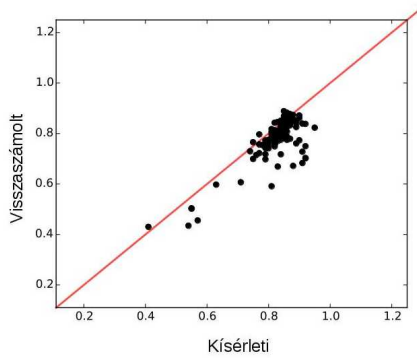
101U PDB 313 K



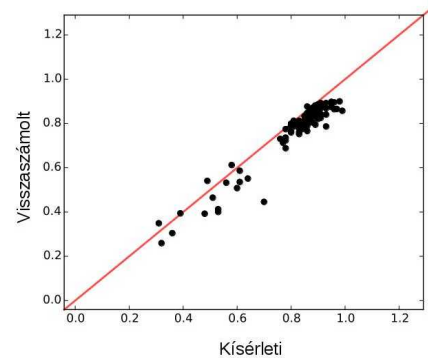
Apo 6 ns 283 K



Apo 6 ns 291 K

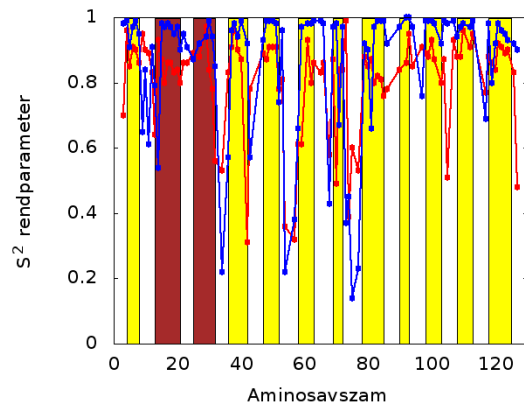


Apo 6 ns 298 K

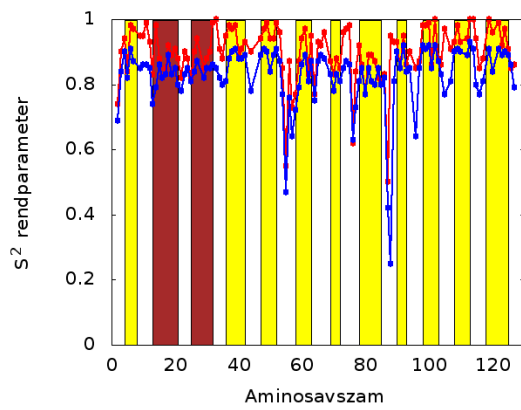


Apo 6 ns 313 K

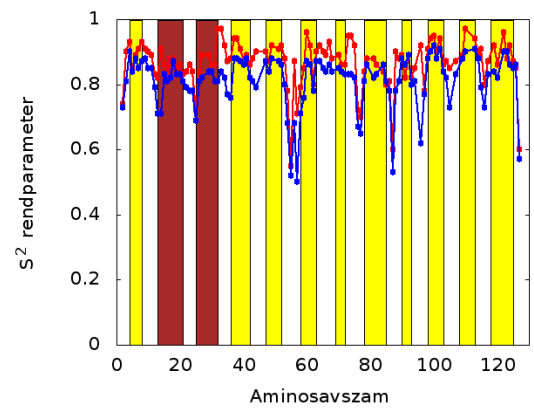
.0.1. ábra. A kísérleti és a visszaszámolt S^2 értékek korrelációja az apo 6 ns sokaságban



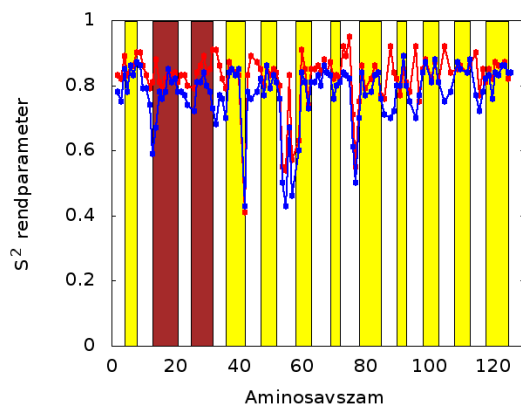
101U PDB 313 K



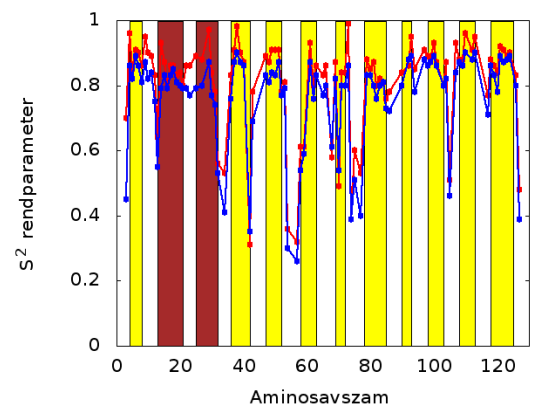
Apo 6 ns 283 K



Apo 6 ns 291 K

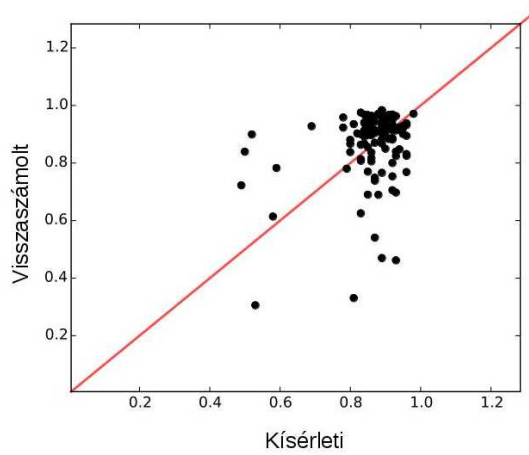


Apo 6 ns 298 K

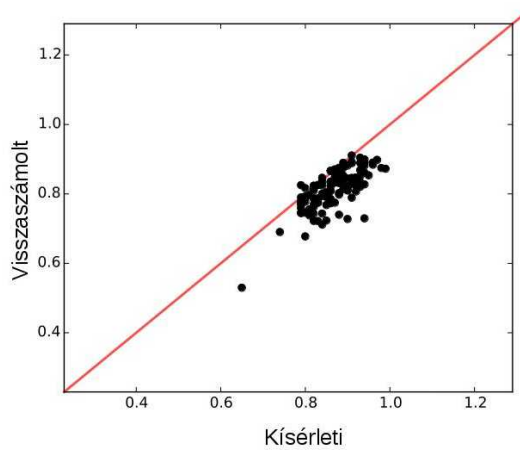


Apo 6 ns 313 K

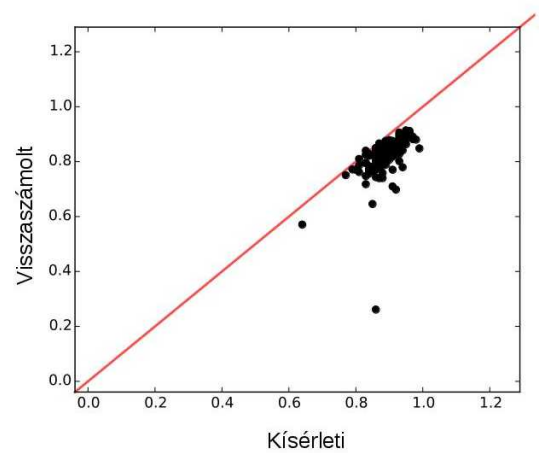
.0.2. ábra. A kísérleti és a visszaszámolt S^2 értékek az apo 6 ns sokaságban



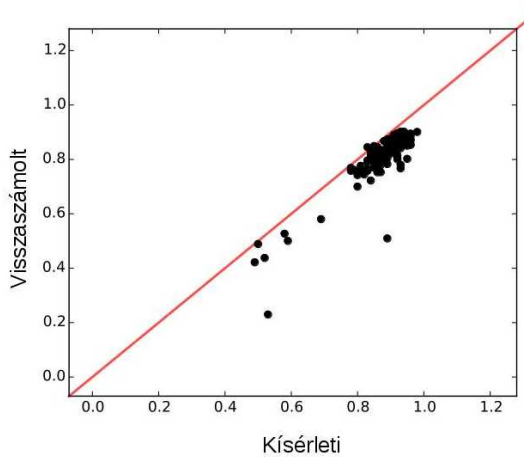
2MM3 PDB 298 K



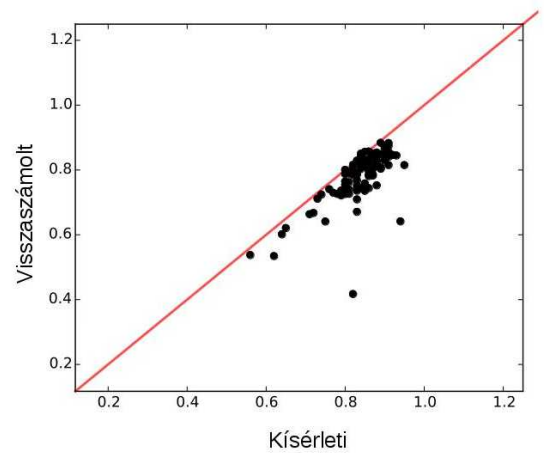
Holo 6 ns 283 K



Holo 6 ns 291 K

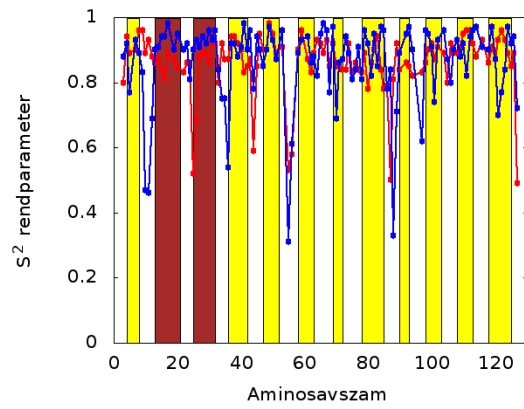


Holo 6 ns 298 K

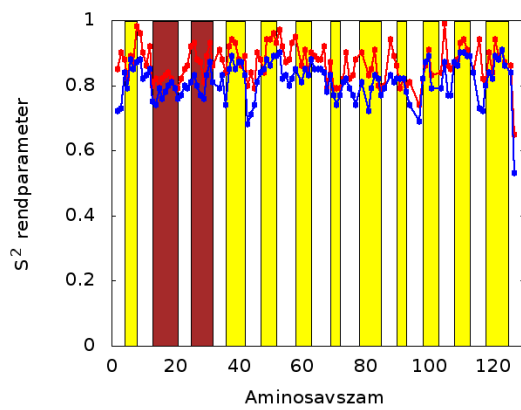


Holo 6 ns 313 K

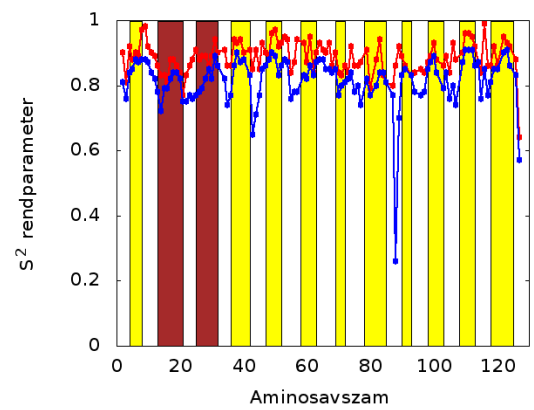
.0.3. ábra. A kísérleti és a visszaszámolt S^2 értékek korrelációja az holo 6 ns sokaságban



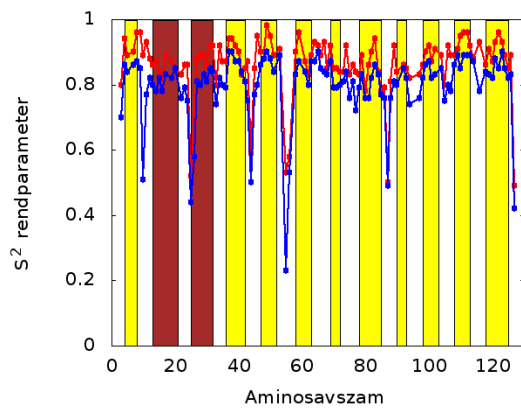
2MM3 PDB 298 K



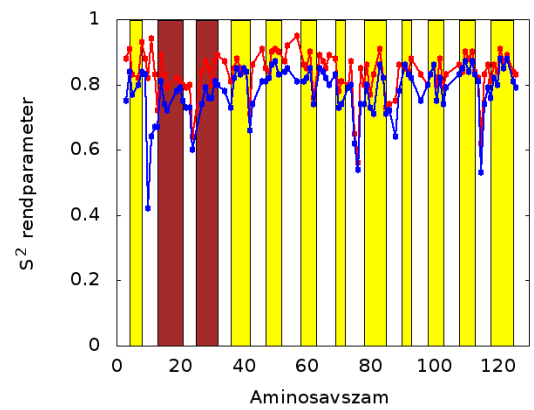
Holo 6 ns 283 K



Holo 6 ns 291 K

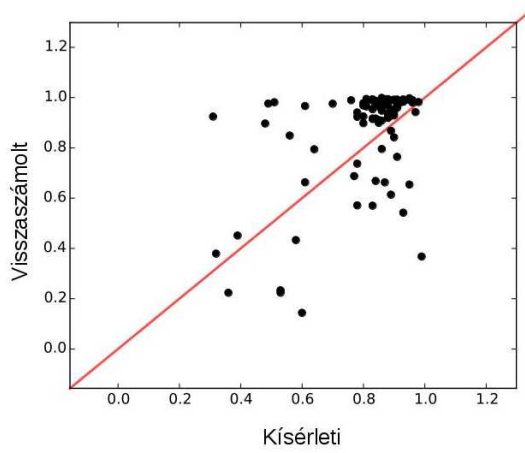


Holo 6 ns 298 K

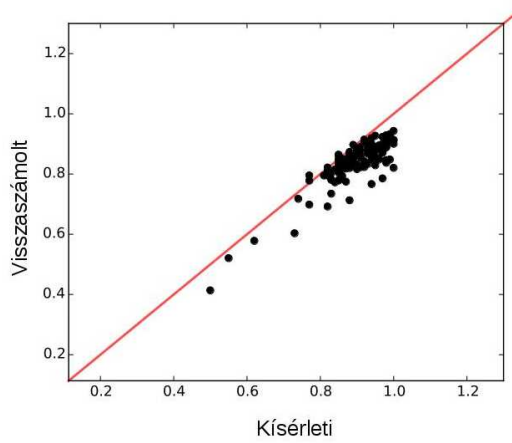


Holo 6 ns 313 K

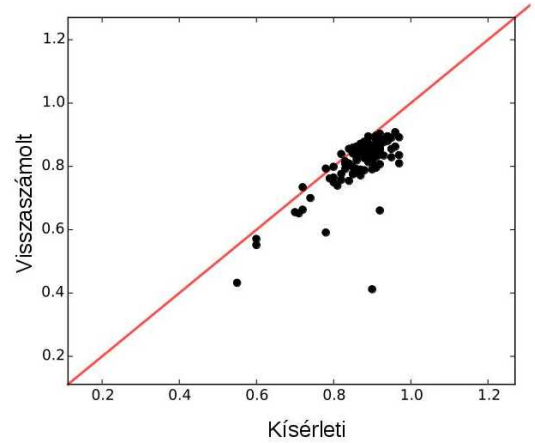
.0.4. ábra. A kísérleti és a visszaszámolt S^2 értékek a holo 6 ns sokaságban



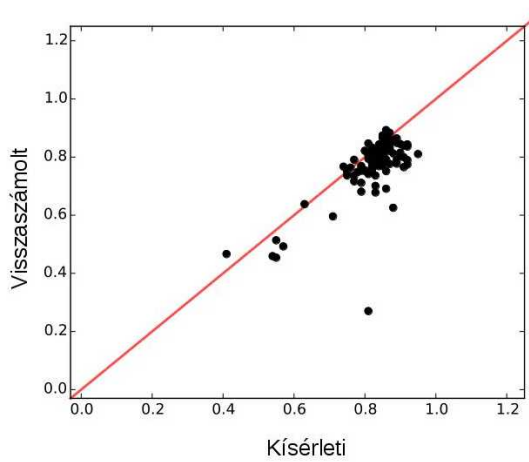
101U PDB 313 K



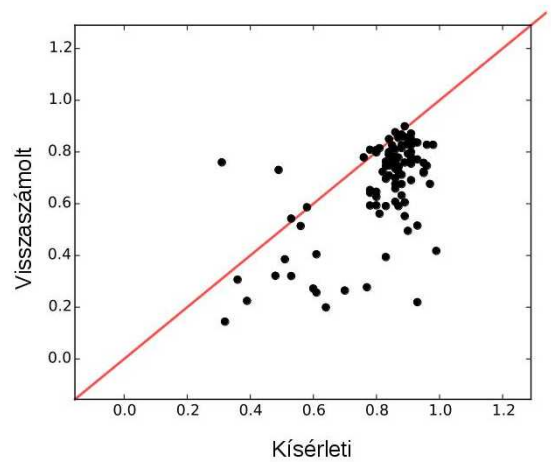
Apo 2 ns 283 K



Apo 2 ns 291 K

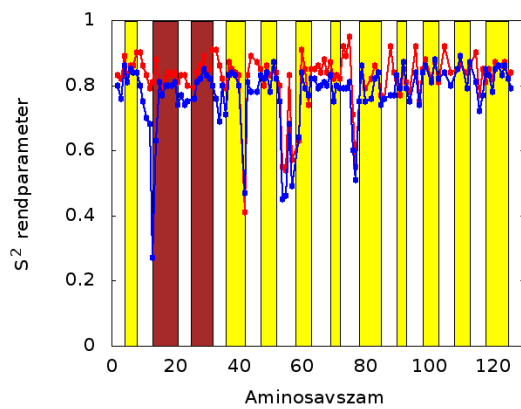
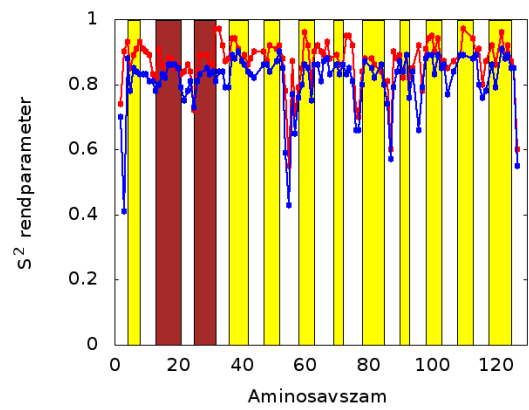
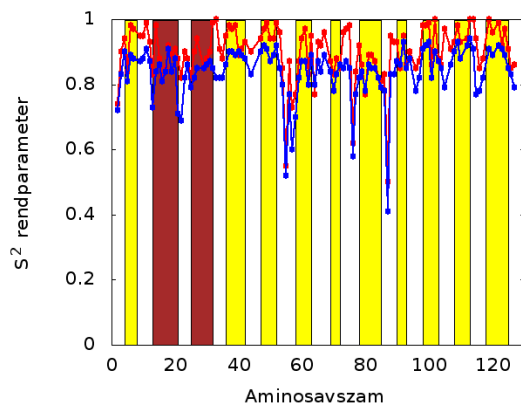
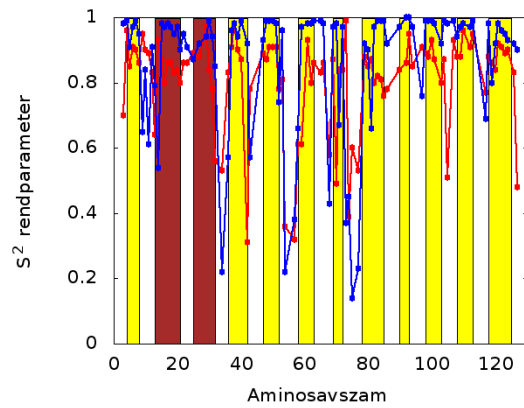


Apo 2 ns 298 K

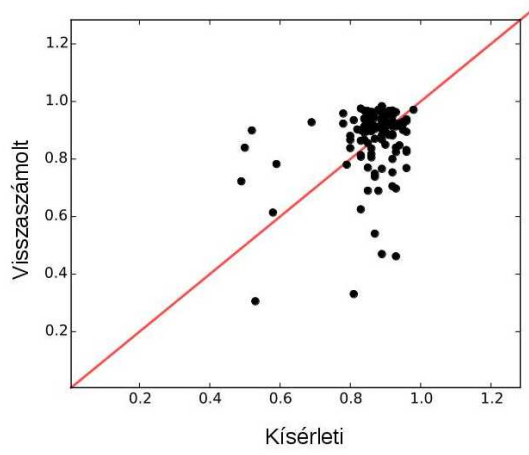


Apo 2 ns 313 K

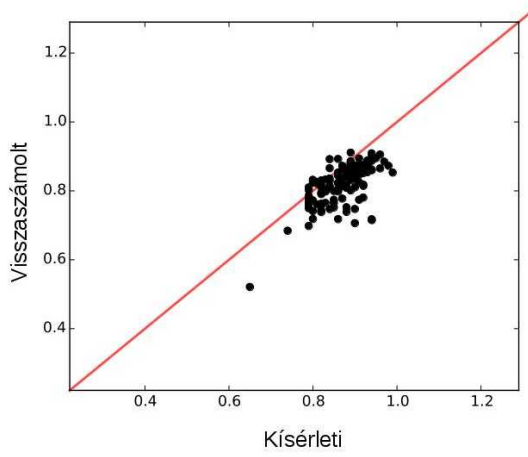
.0.5. ábra. A kísérleti és a visszaszámolt S^2 értékek korrelációja az apo 2 ns sokaságban



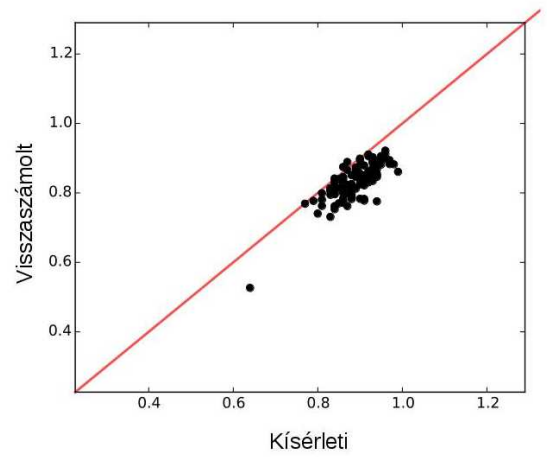
.0.6. ábra. A kísérleti és a visszaszámolt S^2 értékek az apo 2 ns sokaságban



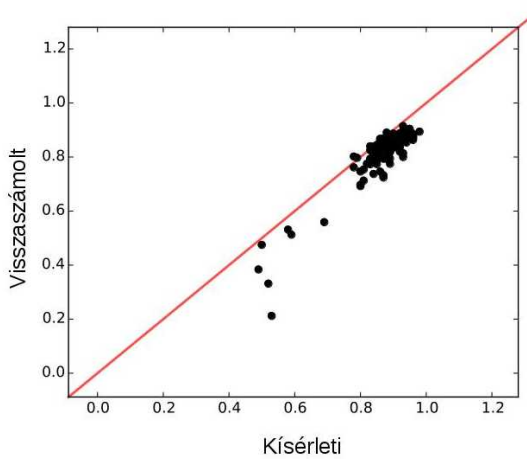
2MM3 PDB 298 K



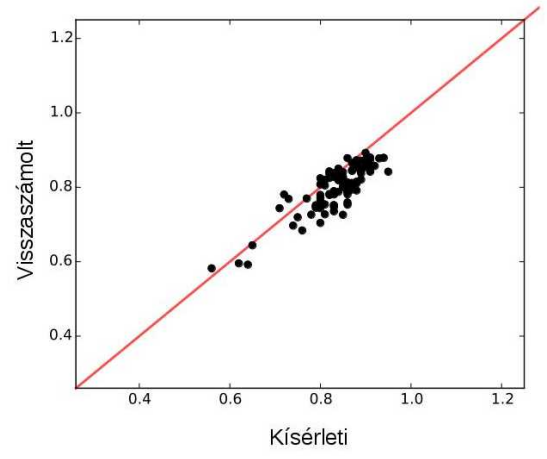
Holo 2 ns 283 K



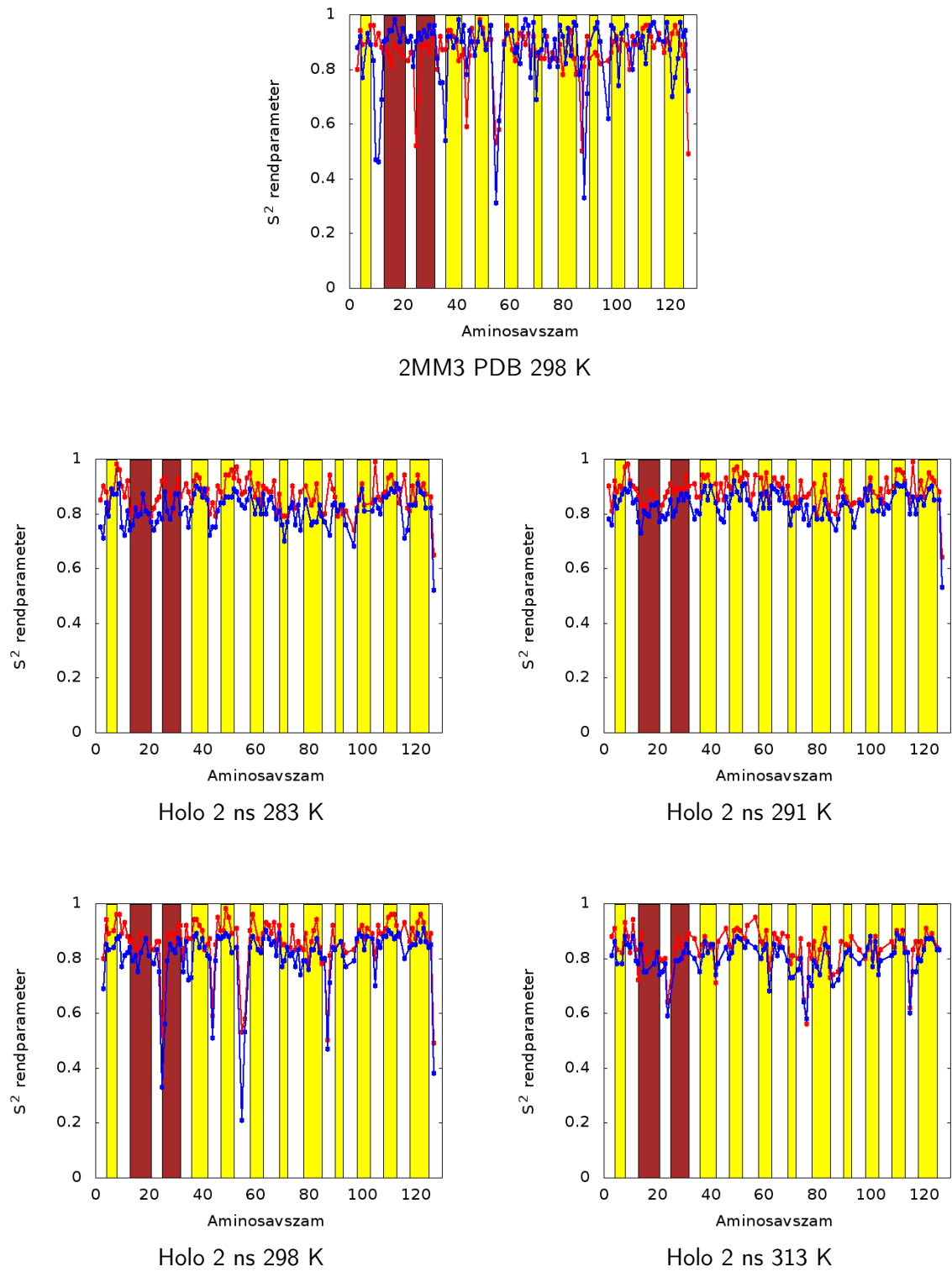
Holo 2 ns 291 K



Holo 2 ns 298 K



Holo 2 ns 313 K



.0.7. ábra. A kísérleti és a visszaszámolt S^2 értékek korrelációja a holo 2 ns sokaságban