

Két, idegi jelátvitelben is szerepet játszó fehérjeszerkezeti elem bioinformatikai elemzése

DISSZERTÁCIÓ



Pázmány Péter Katolikus Egyetem
Információs Technológiai és Bionikai Kar

Készítette
Dudola Dániel

Konzulens
dr. Gáspári Zoltán

Budapest, 2020

Tartalomjegyzék

Kivonat	3
Abstract	4
Bevezető	6
1. Irodalmi áttekintés	8
1.1. A magányos α -hélix (SAH) motívum	8
1.1.1. Az α -hélixek és szuperhélixek	8
1.1.2. A magányos α -hélix (SAH) szerkezete	9
1.1.3. A magányos α -hélix vélt szerepei	9
1.1.4. Algoritmusok a magányos α -hélixek detektálására	10
1.2. A PDZ domének	12
1.2.1. A PDZ domének szerkezeti jellegzetességei	13
1.2.2. A PDZ domének ligandum kötési tulajdonságai	14
1.3. Szerkezeti sokaságok és szerepük a fehérjeműködés megértésében	15
1.3.1. NMR és a fehérjék belső dinamikája	16
1.3.2. A fehérjeszerkezetek modelljei	17
2. Célkitűzés	19
3. Módszerek	21
3.1. A meglévő SAH detektálásra használt módszerek	21
3.2. UniProt adatbázis, Gene Ontology	22
3.3. A CoNSEnsX+ fejlesztéshez használt programnyelvek és fejlesztőeszközök	23
3.4. PDZ szerkezetek elemzése: molekuladinamikai szimulációk, adatbázisok . .	25
4. Eredmények	29
4.1. Az FT_CHARGE algoritmus újraparaméterezése és helicitás szűrő beépítése	29
4.2. Az újraparaméterezett SAH-detektáló eljárással azonosított humán SAH-tartalmú fehérjék elemzése	30
4.2.1. Citoszkeletális fehérjék	32
4.2.2. Idegi jelátvitelben résztvevő fehérjék	35
4.3. CoNSEnsX+ webszerver elkészítése	36
4.3.1. Visszaszámolható paraméterek	39

4.3.2.	Oldallánc S2 rendparaméterek	40
4.3.3.	A megfelelések jellemzése	41
4.3.4.	Szelekció beépítése és értékelése	42
4.3.5.	A szelekció tesztelése molekuladinamikai szimulációval előállított sokaságokkal - Diskusszió	43
4.4.	PDZ domének dinamikus szerkezeti sokaságainak összehasonlító elemzése .	47
4.4.1.	A sokaságokban megfigyelt dinamika elemzése	48
4.4.2.	A sokaságokban megfigyelt állapotok	49
4.4.3.	A PSD95 PDZ doménjeinek összehasonlítása	51
4.4.4.	Szabad és kötött formában lévő PDZ domének összehasonlítása . .	51
4.4.5.	A PDZ3 sokaságokban megfigyelt mozgások – Diskusszió	52
	Köszönetnyilvánítás	57
	Irodalomjegyzék	63

Kivonat

Doktori munkám során két, általam továbbfejlesztett bioinformatikai eljárás segítségével vizsgáltam olyan fehérjeszerkezeti motívumokat, amelyek szerepet játszanak idegrendszeri jelátviteli folyamatokban. Míg a PDZ doménekről régóta ismert, hogy fontosak a posztszinaptikus denzitás szerveződésében, a magányos α -hélixet tartalmazó szinaptikus fehérjéket a kutatás során azonosítottam.

A magányos α -hélix egy ritka szerkezeti motívum, melyet a töltéssel rendelkező aminosavak szabályosan ismétlődő váltakozása stabilizál és jellemez. A domének pontos szerepe nem tisztázott, megismerésüket nehezíti, hogy csak kis számú kísérleti úton megállapított SAH domén ismert. A meglévő szerkezet predikációs algoritmusok a SAH doméneket többnyire – tévesen – szuperhélixként azonosítják. A domének funkcióinak jobb megismeréséhez elengedhetetlen ezek pontos algoritmikus, *in silico* felismerése. Az új kísérleti eredmények rávilágítottak arra, hogy a csoport ezzel a céllal létrehozott algoritmusainak érzékenysége és specifitása a paraméterezés finomhangolásával és egy helicitás szűrő alkalmazásával tovább növelhető. Az új paraméterezés segítségével több, korábban nem felismert SAH motívumot azonosítottam szinaptikus folyamatokban (is) szerepet játszó fehérjékben.

A fehérjék dinamikus tulajdonságait is reprezentáló, kísérletileg megállapított paramétereket tükröző szerkezeti sokaságai egyre nagyobb szerepet kapnak a fehérjék funkcióinak megismerésében. Az ilyen sokaságok gyakran molekuladinamikai szimulációval kerülnek előállításra, melyben a kísérletileg megállapított értékek megszorító paraméterként kerülnek alkalmazásra. Mivel a dinamikát csak a sokaságon lehet értelmezni, szükség volt egy eszközre, mely a sokaságokra vonatkozó dinamikai paraméterek ellenőrzését és elemzését lehetővé teszi. Ebből a célból modernizáltam és új funkciókkal láttam el a CoNSEnsX szervert. A legfontosabb új, általam hozzáadott funkció a szelekciós algoritmus, ami a felhasználó által megadott paraméterezéssel egy alsokaságot generál, a kísérletileg megállapított paramétereknek történő megfelelés maximalizálása mellett.

A PDZ (PSD-95/Discs-large/ZO1) domének a fehérje-fehérje felismerésben és interakcióban vesznek részt. Az interakció során jellemzően más fehérjék C-terminálisával alakítanak ki kötést, illetve közreműködnek nagyobb, az ingerületátvitel során fontos szerepet betöltő fehérje komplexek létrehozásában. A domének belső dinamikájának – mint például a ligandum kötés során fellépő konformációs entrópia változások vagy az allosztérikus moduláció – vélhetőleg fontos szerepe van a molekula biológiai funkciójában. A PDZ domének dinamikus szerkezeti sokaságainak részletes összehasonlító elemzésével a domének közötti szerkezeti különbségek és belső mozgások új aspektusait ismertem fel.

Abstract

Throughout the course of my PhD work I investigated protein structural motifs that play a role in neural signaling processes, using two bioinformatics methods on which I have carried out further development. While PDZ domains have long been known to be important in the organization of postsynaptic density, I identified synaptic proteins containing single α -helices in the research.

The single α -helix is a rare structural motif that is stabilized and characterized by a regularly repeating alternation of charged amino acids. The exact role of these domains remains unclear, the small number of experimentally determined SAH domains makes their study further complicated. Existing structure prediction algorithms usually – and incorrectly – identify SAH domains as coiled-coil regions. To better understand the functions of this domain, their accurate algorithmical, *in silico* detection is essential. The new experimental results highlighted that the sensitivity and specificity of the algorithms created by the group for this purpose can be further increased by fine-tuning their parameterization and applying a helicity filter. With the help of the improved parameterization, I have identified several previously unrecognized SAH motifs in proteins, some of them also involved in synaptic processes.

Structural protein ensembles which also represent the dynamic properties acquired from experimental parameters are playing an increasingly important role in understanding the functions of proteins. Such ensembles are often generated by molecular dynamics simulation, with experimentally determined values used as restraining parameters. Since dynamics can only be interpreted in an ensemble-based manner, a specialized tool was needed to assess the correspondence of such ensembles to experimental data. To achieve this end, I have modernized and added new features to the CoNSEnsX server. The most important new feature added by me is the selection function, which generates a sub-ensemble with the parameterization provided by the user, while maximizing compliance with the experimental parameters.

The PDZ (PSD-95/Discs-large/ZO1) domains are involved in protein recognition and interaction. During interaction the PDZ domain usually binds the C-terminus of the ligand protein. The domain also contributes to the formation of larger protein complexes that play an important role in neural signaling. The internal dynamics of the domains, such as conformational entropy changes during ligand binding or allosteric modulation are thought to play an important role in the biological function of the molecule. By comparative analysis of the dynamic structural ensembles of PDZ domains I have identified new aspects of structural differences and internal movements between domains.

A dolgozat alapjául szolgáló saját közlemények

- I D. Dudola, B. Kovács, and Z. Gáspári. Evaluation and Selection of Dynamic Protein Structural Ensembles with CoNSEnsX. *Methods Mol Biol*, 2112:241–254, 2020.
- II Á. Kovács, D. Dudola, L. Nyitray, G. Tóth, Z. Nagy, and Z. Gáspári. Detection of single alpha-helices in large protein sequence sets using hardware acceleration. *J. Struct. Biol.*, 204(1):109–116, 10 2018.
- III D. Dudola, G. Tóth, L. Nyitray, and Z. Gáspári. Consensus Prediction of Charged Single Alpha-Helices with CSAHserver. *Methods Mol. Biol.*, 1484:25–34, 2017.
- IV D. Dudola, B. Kovács, and Z. Gáspári. CoNSEnsX+ Webserver for the Analysis of Protein Structural Ensembles Reflecting Experimentally Determined Internal Dynamics. *J Chem Inf Model*, 57(8):1728–1734, 08 2017.
- V D. Dudola, A. Hinsenkamp, and Z. Gáspári. Ensemble-based analysis of the dynamic allostery in the PSD-95 PDZ3 domain in relation to the general variability of PDZ structures. *Int J Mol Sci*, 21(21), 2020

Bevezető

A fehérjék szerkezeteinek és funkciójának tanulmányozása során egyre nagyobb szerepet kapnak a szerkezetek dinamikai tulajdonságai, mivel felismerték a szerkezetek belső mozgásainak funkcionális szerepét. A szerkezetmeghatározáshoz használt kísérleti módszerek régebben nem tették lehetővé a szerkezetek mozgásainak részletes megfigyelését, amiből az a téves következtetés született, hogy a fehérjék egy jól körülhatárolható, rigid szerkezettel rendelkeznek. Az egyik első szerkezetmeghatározásra használt kísérleti módszer a röntgenkrisztallográfia, ami – ahogy a módszer nevében is megjelenik – egy fehérje egykristály diffrakciós vizsgálata. A kristályosított – meglehetősen életszerűtlen – állapotban a fehérjék eleve képtelenek a mozgásra, így a szerkezetek dinamikai tulajdonságaira csak következtetni lehet.

A mágneses magrezonancia spektroszkópia (nuclear magnetic resonance spectroscopy - NMR) az első kísérletes módszer, amivel a szerkezetek atomi szintű meghatározása mellett azok dinamikai tulajdonságairól is információ nyerhető. A fehérjeszerkezetek dinamikus szerepe a fehérjék modellezésében is megjelent a dinamikus fehérjeszerkezeti sokaságok formájában, mivel egy modellel nem írható le a szerkezet által bejárt konformációs tér. Ebben a konformációs térbeli állapotok között egy dinamikus egyensúly figyelhető meg, mely egyensúly eltolódása is hatással lehet a fehérje biológiai tulajdonságaira.

A fehérjeszerkezetek sokaságként való reprezentálása mellett szól az, hogy egyes rendszerek esetében az egy konformerre épülő modellek nem reprezentálják jól az adott fehérje tényleges struktúráját. Jó példát adnak erre a rendezetlen fehérjék (intrinsically disordered proteins – IDPs), melyeknek a globuláris fehérjékhez viszonyított nagyfokú hajlékonysága hamar rávilágított a tényre, hogy az ilyen szerkezeteket csak sokaságként lehet jellemezni. Az ilyen fehérjeszerkezetek erősen dinamikus szerkezete sok, eddig ismeretlen biokémiai mechanizmus felfedezésében segített, bár ezek atomi szintű jellemzése gyakran nehézkes.

A manapság leginkább gyümölcsözőnek tekintett megközelítés a kísérletek és a számítások kombinációja, ahol a mérésekből származó paramétereket megfelelő térbeli interpretáció segítségével kényszerfeltételként használjuk fel molekuladinamikai számításokban. Ez a megközelítés nagyban hasonlít a hagyományos szerkezetmeghatározáshoz, az eltérés a kísérleti eredmények felhasználásában rejlik. A dinamikát tükröző sokaságok előállítása során ugyanis a paraméterek értelmezése ténylegesen is sokaságalapú, azaz ahelyett, hogy minden egyes konformertől elvárnánk, hogy minden kísérleti adatnak megfeleljen, a megfelelést a teljes sokaságon értelmezzük és várjuk el.

A fehérjék tipikusan egymásba átalakuló/átalakítható konformerek formájában van-

nak jelen. Mind a hasonló energiájú állapotok, mind az ezen állapotokat elválasztó gátak nagy variabilitást mutatnak az egyes fehérjék közt. Ezek az állapotok felelősek azért, hogy a fehérjemolekulák változatos és specifikus feladatokat tudnak ellátni. Előforduló mozgásaik az időskálán 14 nagyságrendet fednek le, melyek együttes vizsgálata csak több kísérleti eljárás alkalmazásával valósítható meg. A molekulák térbeli változásai révén eltérő affinitást mutatnak különböző ligandumok felé, ezért egy specifikus ligandum megléte/hozzáadása el tudja mozdítani a rendszer dinamikus egyensúlyi helyzetét a magasabb affinitású konformerek irányába, melyek az adott ligandum hiányában csak nagyon kis hányadban vannak jelen.

Bár a molekuladinamikai szimulációban felhasznált módszerek folyamatos fejlesztés alatt állnak, ezek a szimulációk általános esetben nem adják vissza a kísérleti úton megállapított dinamikai jellemzőket. Az egyik lehetőség a számított és a mért paraméterek közötti eltérések csökkentésére, hogy a szimuláció során olyan megkötéseket teszünk a sokaságra, melyek nem engedik, hogy a szimulált mozgások nagyban eltérjenek a megfigyelt mozgásoktól. A sokaság nagyságának növekedésével meglehetősen bonyolulttá válik a paraméterek sokaság alapú átlagolása, illetve a modell túl- és az alulhatározottságának elkerülése. Problémát jelenthet, hogy a sokaság méretének növelésével az átlagolás sok erőforrást foglal le (pl. párhuzamos multireplika szimulációk esetén), illetve hogy az egyes paraméterek esetén a túlillesztés elkerülésére más effektív sokaságméretet kell figyelembe venni. Egy alternatív megközelítés a sztochasztikus optimalizáció alkalmazása, melynek során sok konformert készítenek el szimulációval, majd iteratívan szelektálják a sokaságban lévő konformereket, amíg a sokaság megfelel a kívánt paramétereknek. A módszert többnyire funkcionálisan rendezetlen fehérje szegmensek kiszámításában alkalmazzák, ahol nagy konformációs tér áll rendelkezésre, így hatékonyan lehet alkalmazni az iteratív optimalizációs eljárást.

A munka során felmerült az igény egy eszközre, amivel az ilyen, molekuladinamikai szimulációval előállított dinamikus szerkezeti sokaságok feldolgozása és elemzése rutinszerűen megoldható. Ezzel a céllal továbbfejlesztettem és új funkciókkal láttam el a csoport által előzőleg implementált CoNSEnsX szervert, amely dinamikus szerkezeti sokaságok és a kísérletileg megállapított paraméterek keresztvalidációjára szolgál. A továbbfejlesztett verzióval történt a következőkben bemutatott valamennyi szerkezeti sokaság elemzésre, illetve a szelektált alsokaságok esetén ezek előállításra.

1. fejezet

Irodalmi áttekintés

1.1. A magányos α -hélix (SAH) motívum

1.1.1. Az α -hélixek és szuperhélixek

A fehérjeszerkezetek egyik elsőként felfedezett másodlagos szerkezeti eleme az úgynevezett α -hélix. Az α -hélix egy szabályosan tekeredő, spirál formájú szerkezet, melyet a hélixet alkotó aminosavak másodlagos kötése, hidrogénkötések stabilizálnak. Az előforduló és vizsgált esetek többségében azonban az α -hélix fennmaradásához oldatfázisban nem elégséges a másodlagos szerkezetből fakadó kötések megléte, hanem a polipeptidlánc egyéb részeivel, transzmembrán fehérjék esetében a lipid kettősréteggel kialakított további kölcsönhatások stabilizálják a tekerést. Az α -hélixek gyakran szuperhélixek (coiled-coil) formájában találhatók meg, amiket két vagy több egymásba csavarodott α -hélix alkot, melyek az α -hélixekben periodikusan ismétlődő hidrofób régiók mentén kapcsolódnak.

Becslések szerint a humán proteom (a genomban kódolt és kifejezett fehérjék összessége) 11%-ában megtalálhatók ezek a szuperhélixek. Az α -hélixek menetmagassága 3,6 aminosav, viszont a feltekeredett állapotban a periodikusan ismétlődő apoláris régiók egymástól 3,5 aminosav távolságra vannak, tehát 7 aminosavakként kerülnek átfedésbe. Az jobbmenetes hélixek átfedésben lévő apoláris régiók maguk is egy balmenetes hélixet alkotnak. A szuperhélix dimerizációja során az átfedést alkotó apoláris régiók a molekula közepén vannak, ami stabilizálja az így kialakult szerkezetet. A dimereken kívül előfordulnak még trimer, sőt háromnál is több hélixből felépülő szuperhélix szerkezetek. A láncok lefutása lehet paralel vagy antiparalel, illetve több hélix esetén vegyes.

A szuperhélix szerkezetet felvevő szekvenciákra ismétlődő motívumok megléte jellemző. A leggyakoribb az ún. heptád ismétlődés, amikor hét aminosav alkot egy egységet. Az *abcdefg* jelölést alkalmazva az *a* és *d* pozíciókban hidrofób aminosavak, jellemzően leucin, izoleucin, valin találhatók, az *e* és *g* pozíciókban pedig töltéssel rendelkezők, mint arginin, lizin, glutaminsav. Előbbiek a szuperhélix belseje felé fordulva a hidrofób varrat kialakításában vesznek részt, utóbbiak a külső oldalon sóhidakat alkothatnak, melyeknek további stabilizáló szerepe van. A hidrofób és a hidrofil felszínnek ilyen elkülönülése miatt az ilyen hélixeket amfipatikus hélixeknek is szokás nevezni.

A szuperhélix szerkezetet alkotó szekvenciaszakaszok predikciója tipikusan a heptád ismétlődés felismerésén alapul. Manapság számos módszer létezik, a legkorábbiak ún.

pozícióspecifikus pontozómátrixot (position specific scoring matrix - PSSM) használnak, a későbbiek között vannak rejtett Markov modellt (hidden Markov model, HMM) vagy akár mélytanulást (deep learning) használó algoritmusok. Mindezek ellenére a szuperhélix szakaszok felismerése ma sem tekinthető teljesen megoldottnak, az algoritmusok érzékenysége meglepően alacsony [47, 45]. Ezen felül a heptád ismétlődésekre jellemző aminosaveloslás, főleg a nagyobb hidrofób oldallánccal rendelkező aminosavak ritkasága miatt a szuperhélix szakaszokat számos arra specializált program funkcionálisan rendezetlenként azonosítja [47].

1.1.2. A magányos α -hélix (SAH) szerkezete

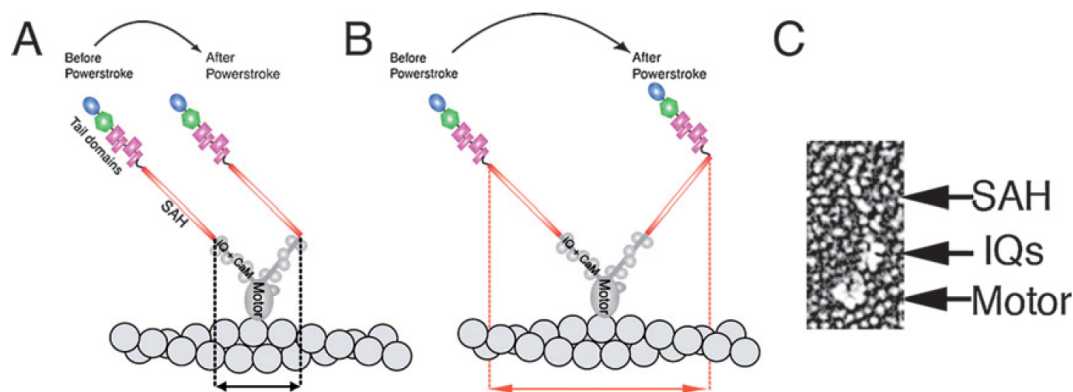
Annak ellenére, hogy általában az egyszálú α -hélixek önmagukban nem stabilak, megfigyelhető, hogy specifikus aminosav szekvenciák megléte esetén (pl. négy glutaminsav (E) és négy pozitív töltésű aminosav, az arginin (R) és lizin (K) váltakozó ismétlődése) stabil α -hélixek alakulhatnak ki a harmadlagos-negyedleges szerkezet stabilizáló hatása nélkül. Ez az ER/K motívum számos élőlény számos, különböző funkciókat betöltő fehérjében megtalálható [46]. A magányos α -hélixek (single alpha helix - SAH) megtalálását nehezíti, hogy az algoritmusok által keresett töltéssel rendelkező régiók a szuperhélixekben is megtalálhatók, bár a töltéssel rendelkező aminosavak aránya a SAH doménekben magasabb. Emiatt az algoritmusok által detektált szuperhélixek egy része vélhetően SAH domén, de ezek egyértelmű megállapításához további vizsgálatok szükségesek [39]. A SAH domén kialakulása akkor valószínűsíthető, ha negatív és a pozitív töltésű aminosavak is nagy számban vannak jelen, viszont hiányzik a hidrofób varrat (a szuperhélixek esetén a szuperhélixet alkotó α -hélixek átlapolódási helyei), valamint minél több hélixen belüli E-K, R-K interakció jön létre.

1.1.3. A magányos α -hélix vélt szerepei

Habár a SAH domének sokkal elterjedtebbek, mint amilyenek a felfedezésükkor vélték őket, a domének pontos szerepe nem tisztázott, viszont a jelenlétük több esetben jól valószínűsíthető fizikai jelentőséggel bírhat. Ezen funkciók egyike a miozin molekula karjának meghosszabbítása, mely megnöveli az egy erő kifejtés/fordulat alatt a miozin molekula aktin szálon megtett távolságát. Mérésekkel igazolták, hogy a miozin egyes fajtaiban található SAH domének elég merevek ezen kar-meghosszabító funkció betöltéséhez [6].

A SAH domének a miozin család molekuláin kívül is gyakoriak. A domén fizikai tulajdonságai, merevsége és rugalmassága, alkalmassá teszik molekula szerkezeti elemeinek fix, de rugalmas összekötésére. Ezen felül a SAH domén kis húzó erő hatására képes letekeredni, miközben a hélix által generált erő nem változik, akárcsak egy állandó erejű rugó esetén [55]. Ez a jelenség a SAH domént tartalmazó miozin molekulák karjaiban is szerepet játszhat, mivel a SAH domén letekeredésével elnyelt energiának hatása lehet a teljes motor molekula működésére. Molekuladinamikai szimulációk alapján a letekeredett SAH részek teljesen és gyorsan visszanyerik eredeti szerkezetüket a húzó erő megszűnésével.

A letekeredés által elnyelt energia a motor fehérjék működése során elősegítheti, hogy



1.1. ábra. SAH domén a miozin 10 molekula karján [39]. A: a SAH domén a miozin 3 kalmodulint kötő IQ motívumból álló karjához folytatódólagosan csatlakozik. A fekete nyíllal jelzett távolság az, amit a miozin molekula egy lépésben megtesz az aktin filamentumon a SAH domén nélkül. B: Ha a SAH domén meghosszabbítja a miozin molekula karját, egy ciklus alatt nagyobb távolságot (piros nyíl) tud megtenni az aktin filamentumon. C: A miozin 10 molekula elektronmikroszkópos képe.

a motor fehérje a vontatott szállítmány elakadása esetén is állandó erő kifejtés mellett az aktin filamentumok által kijelölt pályáján tudjon maradni, amíg a szállítmány újra ki nem szabadul. Ezt a feltételezést alátámasztják azok a felismerések, miszerint a fehérjéknek egy igen zsúfolt közegben kell végezniük munkájukat: Egy emlős sejt térfogatának kb. 20 százaléka fehérjékből áll, ami sejttenként 10000-20000 külön fehérjét jelent [6].

Emellett a korai elemzések rámutattak arra is, hogy a SAH motívum az RNS-kötő fehérjékben is gyakoriak [18]. Ennek pontos jelentősége nem ismert, ugyanis a SAH jelen ismereteink szerint nem vesz közvetlenül részt az RNS-molekulákkal való kölcsönhatások kialakításában. A sejtmagban RNS-eket megkötő paraspeckle nevű képletet alkotó PSPC és NONO fehérjékből álló multimer modellezése azt sugallja, hogy ebben az esetben a SAH az RNS-kötő RRM domének közötti megfelelő távolságtartást biztosítja [13].

Ezen feltételezett funkciók vélhetően csak egy ízelítőt adnak a SAH domén által betöltött/betölthető funkciókról. A domén egyéb szerepeinek megismeréséhez további, SAH domént tartalmazó fehérjék elemzésére lenne szükség. Mivel az aminosav szekvencia alapján a SAH domén sokszor szuperhélixként kerül detektálásra, fontos a SAH detektáló algoritmusok specifikusságának fejlesztése, ezek megbízhatóbbá tétele.

1.1.4. Algoritmusok a magányos α -hélixek detektálására

A magányos α -hélixek detektálására több algoritmus is készült, mivel a szerkezeti tulajdonságaiból kifolyólag egy SAH detektálása jól algoritmizálható. Az egyik első erre a célra készült program a SCAN4CSAH [48], amely többek közt a magányos α -hélixeket jellemzően stabilizáló sóhidakat keres. Az algoritmus súlyozott, pozitív vagy negatív pontszámokkal látja el a szekvenciát, annak függvényében, hogy a lokális szekvencia alapján mennyire valószínűsíthető SAH kialakulása. A pontozott, α -hélixet stabilizáló tulajdonságok a következők:

1. egymástól 3-4 aminosav távolságra található, ellentétes töltésű aminosavak (pl. KxxxE)
2. egymástól egyenként 3-4 távolságra lévő három, váltakozó töltöttségű aminosavak (pl. ExxxRxxxE)
3. az ellentétesen töltött aminosavak egymáshoz viszonyított orientációja

Az α -hélix kialakulását hátráltató fő tényezők azok az ionos vonzó és taszító kölcsönhatások, amelyek destabilizálnák a hélixet, mint például az egymás melletti ellentétes töltésű oldalláncok vagy az azonos töltésű, egymástól 3-4 aminosav távolságra lévő aminosavak. Az így pontozásra kerülő potenciális SAH szegmensek ezután a hosszuk alapján normalizálásra kerülnek, hogy a hosszú, de gyengén töltött szekvenciák ne jelenjenek meg a találatok között.

Az FT_CHARGE [48] program egy Fourier transzformáció alapú eljárást implementál, ahol a szekvenciában található töltések távolságának függvényét ábrázolja a kiszámított spektrum. A következő aminosavak kapnak hozzárendelt töltés értéket: Asp és Glu: -1; His: +0,5; Arg és Lys: +1, minden más aminosav töltés értéke 0. Ezen értékekre a Fourier transzformáció már közvetlenül is alkalmazható. A szekvenciális távolságok kiemelésére egy töltés korreláció függvény is definiálható:

$$R(n) = \sum_{i=1}^{m-n} c(i)c(i+n)$$

ahol $c(i)$ az i . aminosavhoz rendelt töltés és m a szekvencia hossza. Ez a függvény érzékenyebb a töltési mintázatok változására, mintha csak az aminosavakhoz rendelt töltöttség értékeket direkt módon vizsgálnánk. Mivel az Fourier spektrumban megfigyelt amplitúdók függenek a szekvencia töltéssűrűségétől, nem lehet egy fix határértéket definiálni, ami fölött valószínűleg SAH-t talált az algoritmus. A probléma megoldható egy statisztikai módszer alkalmazásával: az eredeti szekvenciából származó Fourier transzformáció amplitúdói az ugyanebből a szekvenciából készült, randomizált szekvenciák Fourier spektrumbeli amplitúdóival kerülnek összevetésre. Ezen összehasonlítás alapján megállapítható, hogy a megfigyelt nagy amplitúdók a szekvencia általánosan nagyobb töltéssűrűségének a következményei, vagy valóban SAH specifikus töltöttség mintázatot talált az algoritmus. Ennek hátránya, hogy a robusztussághoz nagyszámú random szekvencia szükséges, ami egy konkrét szekvenciából történő generálás esetében nagyon időigényes. Az algoritmus egy későbbi változatában [18] a szerzők előre elkészített random szekvenciákból származtatott eloszlást használtak a szignifikancia eldöntésére. Háromféle, csupán alanint, arginint és glutaminsavat előre meghatározott, 10%-os lépésekben változó arányokban tartalmazó szekvenciákat – gyakoriságonként és hosszúságonként 5000 darabot – állítottak elő, és az ezeken detektált mintázatra ún. extrémérték-eloszlást (extreme value distribution - EVD) illesztettek. A valós szekvenciák vizsgálatakor az összetételben ahhoz legközelebb álló random szekvenciákból kapott eloszlásból kapott P-értéket használják fel, így a szignifikancia az egyedi szekvenciák randomizálásánál jobban reprodukálható.

SAH régiók detektálására a szuperhélix kereső algoritmusok is felhasználhatóak, ebben az esetben a SAH kialakulását elősegítő tényezőket és a SAH-ra jellemző aminosav mintázatokat kell keresni. A Waggawagga [44] webszerver összehasonlító elemzést kínál hét szuperhélix kereső (Marcoil, Multicoil, Multicoil2, Ncoils, Paircoil, Paircoil2) és három oligometrizációs állapotot becselő algoritmus (Scorer, ProCoil és LOGICOIL) kimenetéből. Az algoritmusok kimeneteinek vizualizálásával a hélixen belüli interakciók szemrevételezésével azonosíthatóak a potenciális SAH domének (hidrofób varrat hiánya). Emellett a szerver egy SAH pontszámot is számít a SAH kialakulását elősegítő interakciók és a vizsgált szekvencia szakasz hosszának függvényében. A SAH detektálás / pontszám számítás során a szerver a SAH definícióját is alkotó ER-K aminosav kapcsolatokat veszi csak figyelembe, az ettől eltérő szerkezetű monomer α -héliceket a szerver nem detektálja.

A SAH régiók detektálásának nehézsége a motívum ritkaságából adódik. Egyelőre nagyon kevés kísérletes adat áll rendelkezésünkre részletesen jellemzett SAH szakaszokról. Bár a töltésmintázat szerepe általánosságban ismertnek mondható, a pontos biofizikai részletek, a motívum kialakításához szükséges töltéssűrűség és a töltések váltakozásának szabályszerűsége egyelőre nem megfelelő részletességgel ismert feltételek. Az első, szisztematikusnak mondható kísérleti munkát csak 2015-ben publikálták [5]. Ezért az első algoritmusok kidolgozása során azok szerzői nagy hangsúlyt fektettek a predikciós eredmények specifikusságára, ami természetszerűleg az érzékenység rovására ment. Az FT_CHARGE és a SCAN4CSAH eljárások kombinálásának éppen az volt a célja, hogy egy szigorú kritériumrendszer segítségével a hamis pozitív találatok számát minimalizálja. Így bár elérték, hogy a mindkét eljárás által magas pontszámmal detektált SAH szakaszok kísérletileg igazolhatóan stabil monomer hélixet alkottak [48], számos SAH domént nem tudtak azonosítani. A két eljárás konszenzusával talált nagy mennyiségű szekvencia elemzése során ugyanakkor az is kiderült, hogy léteznek nagy töltéssűrűségű és szabályos ismétlődést mutató, de egyúttal prolinban gazdag szekvenciák is, amelyek várhatóan nem képesek α -helikális szerkezet kialakítására a prolin aminosav 'hélix-törő' tulajdonsága miatt. Az eredeti FT_CHARGE implementáció lassúsága miatt fejlesztések kezdődtek az algoritmus FPGA platformra történő átültetésére, mely lehetővé tette a korábbiaknál jóval nagyobb méretű adathalmazok gyors feldolgozását, ezáltal a vizsgálható tesztesetek számának radikális növelését. Így mind a szükséglet, mind a lehetőség adott volt az algoritmus újraparaméterezéséhez és proteom-szintű adatok újraelemzéséhez.

1.2. A PDZ domének

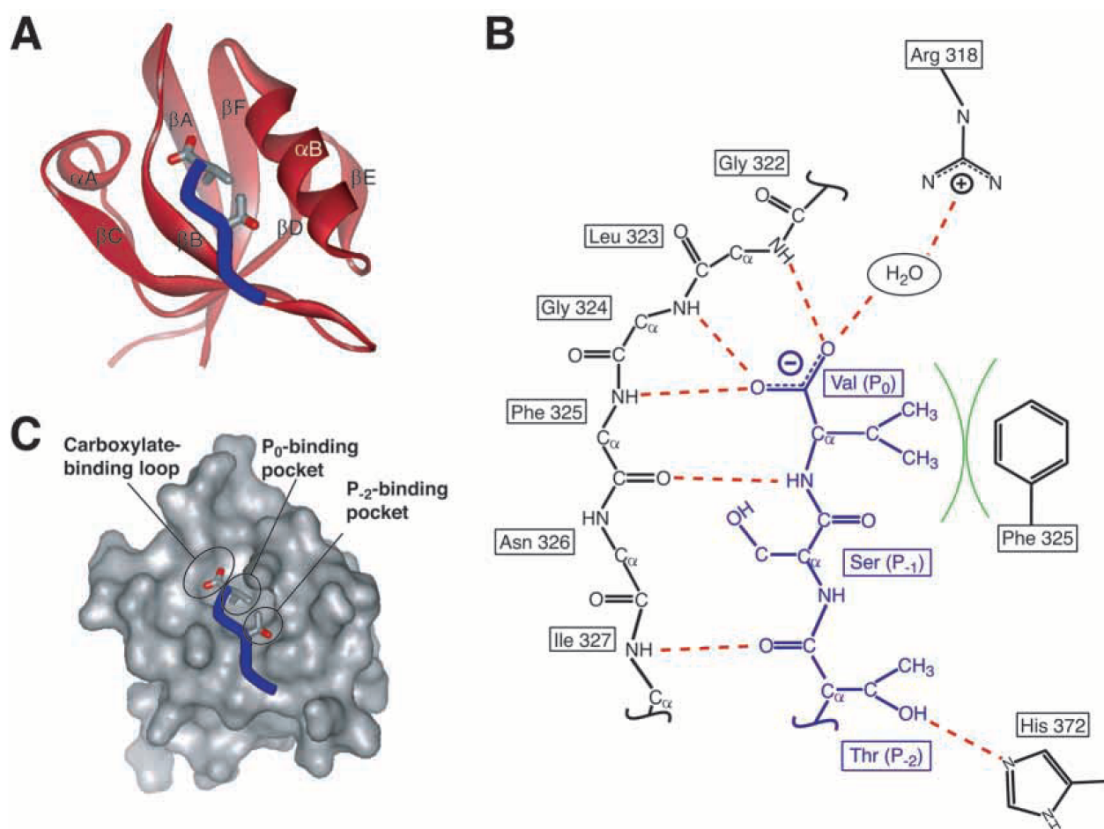
A PDZ domének a fehérje-fehérje felismerésben és interakcióban vesznek részt, mellyel hozzájárulnak többek közt a jel-/ingerületátviteli útvonalakhoz nélkülözhetetlen fehérje komplexumok létrehozásához. A PDZ doméneket is tartalmazó fehérjéket kódoló gének mutációja az emberben örökletes (és jelenleg gyógyíthatatlan) idegrendszeri megbetegedésekkel hozható összefüggésbe, mint például a 3. típusú Usher-szindróma (fokozatosan kialakulható látás- és halláscsökkenést okozó) vagy a Dejerine–Sottas betegség (a perifériás idegek károsodása által kiváltott progresszív izomsorvadás).

A körülbelül 90 aminosavból álló szekvencia igen elterjedt, megtalálható többek között

a muslica (*Drosophila*) fotoreceptoraiban és az emlősök szinapszisaiban, és a megszekvenált genomokban az egyik leggyakrabban előforduló domén [22]. Maga az elnevezés egy mozaikszó, amely abból a három fehérje nevéből áll össze, melyekben először kimutatták a PDZ domént, ezek a posztzinaptikus PSD-95, a muslicában található szoros csomópontokban (tight junction) lévő Discs-large és a ZO-1 fehérjék.

1.2.1. A PDZ domének szerkezeti jellegzetességei

A fehérje-fehérje kölcsönhatásokban betöltött szerepe a fehérje szerkezeti adottságaira vezethető vissza. A peptid ligandumok kötőhelye β B redő és az α B hélix határolta kötőzseb, melyben a kötő fehérje β -redője hozzáadódik a kötőzseb β -redőjéhez [21]. A kötés során a kis kötőzsebbe a ligandum peptid körülbelül 5 aminosav hosszúságú C-terminálisa ékelődik be, a zseb töltöttsége a hidrofób aminosavakkal (valin, izoleucin, leucin) végződő szekvenciák kötődésének kedvez. A C-terminálisokon kívül az ehhez hasonló belső motívumok is részt vehetnek a PDZ doménnel kialakított kötésben.



1.2. ábra. A peptid felismerés mechanizmusa a PDZ doménben. A: A PSD-95 3. PDZ doménjének szalagdiagramos ábrázolása (piros) kötött peptiddel (kék). A kötött peptid C-terminálisának oldalláncai bottiagram formában ki vannak emelve. B: A peptid kötőzseb diagramja. A szaggatott piros vonalak a hidrogén kötések jelölik. C: Az (A) struktúra felületi reprezentációja a kötőzsebek megnevezésével. [21]

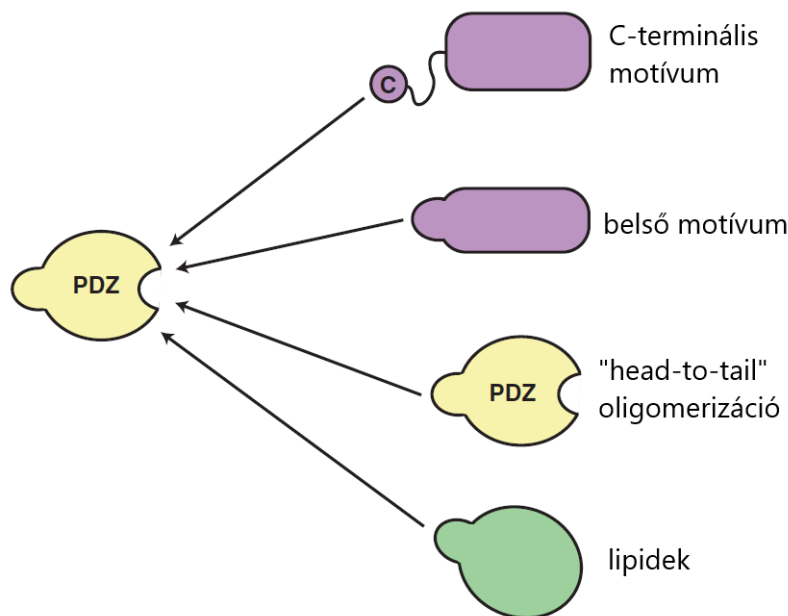
Mivel a PDZ domén rövid, flexibilis C-terminálisokat és az ehhez hasonló szerkezetű belső motívumokat köt, ezért gyakorlatilag szinte minden fehérjével képes kötés kialakítani, különösen a membrán fehérjék iránt mutat nagyobb affinitást, mint például az

ion csatornák, amik jellemzően rövid C-terminálissal rendelkeznek. A kötés a ligandum tekintetében is "kímélő", a kötött ligandum szerkezete nem módosul a szabad állapothoz képest [22], a tranzienst kötődés miatt az egyensúly gyorsan és könnyen átrendeződik.

1.2.2. A PDZ domének ligandum kötési tulajdonságai

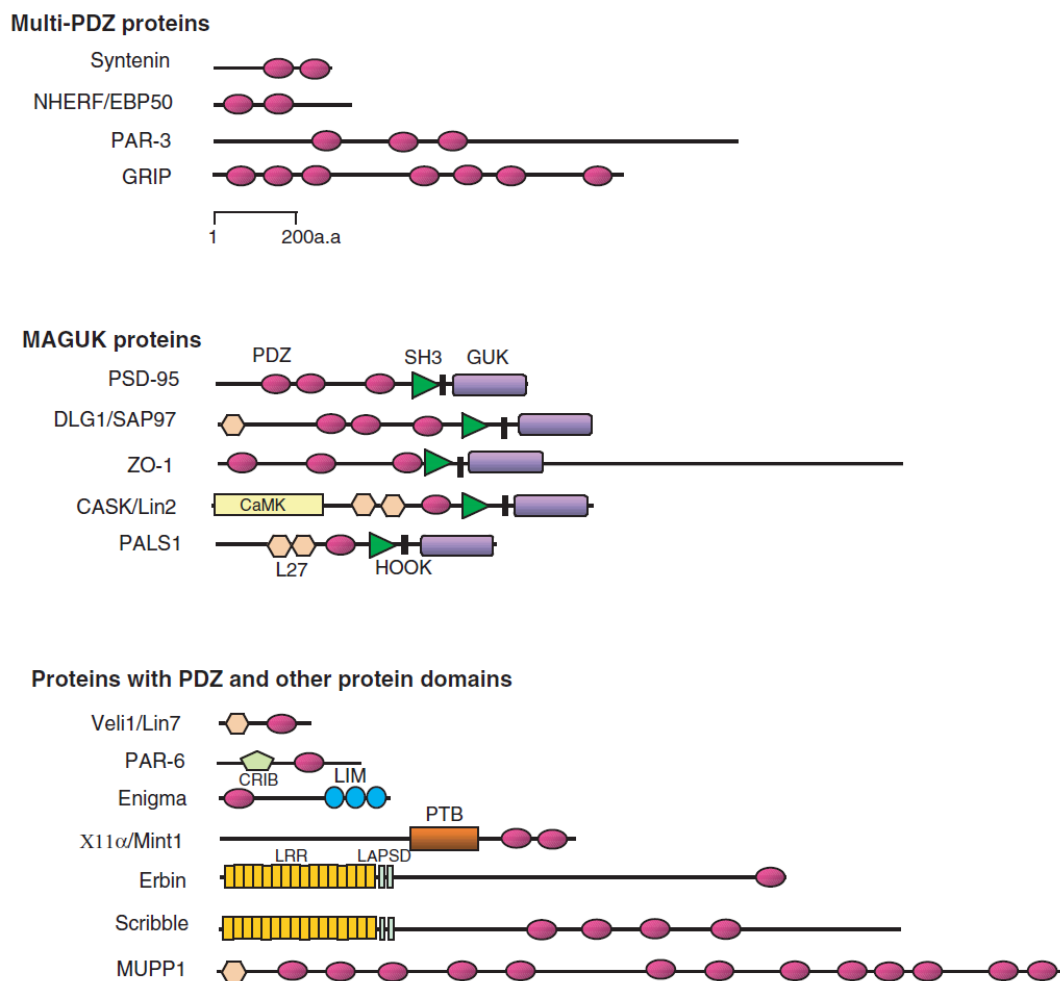
A PDZ domének egy fehérjén belül csoportokban is megfigyelhetők – bár ennek jelentősége nem tisztázott – valószínűleg a ligandumokkal történő kötési affinitással függ össze. Emellett a belső doménekhez való kapcsolódással a PDZ doméneket tartalmazó fehérjék képesek fehérjeszerkezetek összeláncolására (scaffolding), amivel több doménből álló fehérjekomplexelek jönnek létre.

A PDZ domének kötési preferenciáik alapján (legalább) három osztályba sorolhatóak, ezek rendre a $S/T - X - \Phi - COOH$, $\Phi - X - \Phi - COOH$ és a $X - X - C - COOH$ motívumokat ismerik fel, ahol Φ egy hidrofób és X egy tetszőleges aminosavat jelöl. A kötött helyzetben a β -redők stabilizálják a ligandum helyzetét, melyben a ligandum kötésben részt vevő oldallánca egyenesen a PDZ domén kötősebe felé mutat. Így a ligandum β -redőjének és a polarizált oldallánc helyzetére vezethető vissza a domén kötési specifikitása.



1.3. ábra. A PDZ domén legalább négy fajta interakcióban vehet részt: C-terminális motívum felismerés, belső szerkezeti motívum felismerés, PDZ-PDZ dimerizáció és lipidek felismerése. [37]

A rövid C-terminálisokra vonatkozó kötési affinitás mellett a PDZ domén képes felismerni belső fehérje motívumokat is, amennyiben azok térszerkezeti szempontból hasonlóak a lánc végződéshez. A kötés ezen fajtája jellemzően csak akkor jön létre, ha a ligandum peptid C-terminálisa a ligandum szerkezetéből kifolyólag nehezen elérhető, térbeli helyzete megnehezíti vagy ellehetetleníti a kötés kialakítását. A domén kötési specifikitása miatt ilyen, a ligandum belső motívumával kialakított kötés vélhetően ritkábban jön létre, mivel a domén kötési helyének geometriája és polarizáltsága kiemelkedően specifikus.



1.4. ábra. PDZ domének (rózsaszín) moduláris szervezetség szerint kategorizálva (multi-PDZ fehérjék, MAGUK fehérjecsalád, PDZ és egyéb doméneket is tartalmazó fehérjék). Fontosabb rövidítések: SH3 - Src-homology 3 domén, GUK: guanilate-kinase-like domén, CaMK - calmodulin kináz domén [37]

1.3. Szerkezeti sokaságok és szerepük a fehérjeműködés megértésében

Az egyik első kísérleti módszer, amivel megoldható a molekulák atomi szintű szerkezetének meghatározása a mára már évszázados múltra visszatekintő röntgenkristallográfia. Habár a mérés során használt kristályosított fehérjéből következtetni lehet azok belső szerkezeti rendezettségére, a szerkezetmeghatározás előfeltétele a homogén szerkezetű molekulákból előállított fehérje egykristály. Ezen kísérletek eredményei alapján a fehérjéket sokáig rigid háromdimenziós struktúrákként kezelték. A fehérjék azonban vizes közegben, így a sejten belül is, dinamikus viselkedést mutatnak, számos konformer tart egymással egyensúlyt. A mágneses magrezonancia (NMR) spektroszkópia fejlődésével lehetővé vált a fehérjeszerkezetek oldat fázisban történő mérése, ez jelenleg a legalkalmasabb kísérleti módszer a molekulászerkezetek belső mozgásainak több időskálán történő, atomi szintű vizsgálatára. Mivel a mérés oldatfázisban történik, a fehérje molekuláknak lehetőségük van mozogni, illetve a természetes környezetükben feltételezett konformációs állapotokat megtartani (a fehérjék kristályosítása a konformációs variabilitás szempontjából egy

komoly beavatkozás). Ezzel fény derült a fehérjeszerkezetek addig részleteiben nem ismert - nem mérhető dinamikus jellegére, ami jól illeszthető az addigra már feltérképezett statikus modellekhez.

A biomakromolekulák, mint a fehérjék is, tipikusan egymásba átalakuló/átalakítható konformerek formájában vannak jelen. [61] Mind a hasonló energiájú állapotok, mind az ezen állapotokat elválasztó gátak nagy variabilitást mutatnak az egyes fehérjék közt. Ezek az állapotok felelősek azért, hogy a fehérjemolekulák változatos és specifikus feladatokat tudnak ellátni. Előforduló mozgásaik az időskálán 14 nagyságrendet fednek le és ezek egy kísérlet keretében történő együttes vizsgálatára jelenleg nincs módszer. [19] A molekulák térbeli változásaik révén eltérő affinitást mutatnak különböző ligandumok felé, ezért egy specifikus ligandum megléte/hozzáadása el tudja mozdítani a rendszer dinamikus egyensúlyi helyzetét a magasabb affinitású konformerek irányába, melyek az adott ligandum hiányában csak nagyon kis hányadban vannak jelen.

A fehérjeszerkezetek kutatásában az utóbbi években egy régi észrevétel kapott újra nagyobb figyelmet, miszerint a fehérjéket nem egy adott szerkezet jellemzi, melyet adott esetben valamilyen kísérleti eljárással felderítettek, hanem sokkal inkább a fehérjére jellemző konformerek sokasága. Ennek ellenére a mai napig kevés olyan atomi szintű modell létezik, mely a folyamatot részleteiben jellemezné. [60] Bár vannak eljárások, melyek keretében kiszámíthatóak konformációs sokaságok, melyek különböző időskálákon is jól jellemzik az adott molekula belső dinamikáját [27], ezek még nem általánosan elfogadott módszerek, és bár egyre több kutatócsoport foglalkozik ezekkel, alkalmazásuk messze nem tekinthető rutinszerűnek.

1.3.1. NMR és a fehérjék belső dinamikája

Az NMR spektroszkópiás mérések számos, ezeket a belső mozgásokat jellemző paramétere a sokaság tulajdonságait jellemző átlag, bár az átlagolási eljárások paraméterenként eltérőek lehetnek. Az S^2 rendparaméterek eleve a sokaságra vonatkoznak, az NOE távolság értékek r^{-6} , a többi paraméter aritmetikai átlagolással kerül kiszámításra. [3] Ezen információk felhasználása az atomi szintű szimulációban megköveteli a mögöttes fizikai interpretáció helyességét. A három leggyakrabban alkalmazott NMR paraméter a nukleáris Overhauser effektus (NOE), a maradvány dipoláris csatolás (RDC) és a Lipari-Szabó féle általános rendparaméter (S^2). Bár a módszer lehetővé teszi egy vizsgálat keretében a több időskálán történő megfigyelést, az így meghatározott, a molekula dinamikáját jellemző paramétereket gyakran csak kiegészítő információként mellékelik a statikus strukturális modellekhez. Fontos szempont az NMR spektroszkópiával előállított modellek esetén, hogy a meghatározott paraméterek a sokaságot jellemzik, nem pedig konkrét konformereket. A sokaság a fehérjék az adott időskálán történő belső mozgékonyaságáról ad információt, segítve ezzel a belső mozgások biológiai szerepének megértését.

A sokaságok elemzésekor tekintetbe kell vennünk, hogy pontosan milyen paramétereket vettünk figyelembe az előállításuk során, és azok milyen időskálán történő mozgásokat reprezentálnak. A sokaságok értékelésekor lényeges az ún. túllillesztés elkerülése, ugyanis több konformáció átlaga a több szabadsági fok miatt könnyebben megfeleltethető a mé-

résekből származó adatoknak, mint egyetlen szerkezet. Nagyon fontos a keresztvalidáció, amikor a sokaság előállításához nem felhasznált paraméterekkel való egyezést vizsgáljuk, a kiindulási szerkezetektől való eltérések mértékét az egyes konformerek esetében, valamint más, ismert tulajdonságú sokaságokkal való összevetés is szükséges lehet. A keresztvalidáció egyik célja lehet annak ellenőrzése, hogy a molekuladinamikai szimulációval előállított sokaságban a kísérleti paramétereknek történő megfelelés javult vagy romlott a kiindulási sokaságokhoz képest. A sokaságok összehasonlításában alkalmazott módszerek közös alapja a sokaság valószínűség- eloszlásának megbecslése, valamint ezen eloszlások későbbi összehasonlítása. De nemcsak a szerkezetnek, hanem a szerkezetek változásait jellemző dinamikának is jelentős szerepe van a funkció meghatározásában. Erre a dinamikára a sokaságokban előforduló szerkezetek és ezek eloszlása enged következtetni.

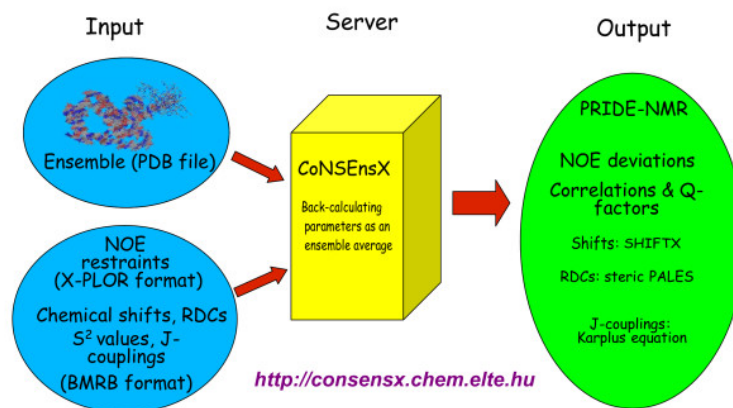
1.3.2. A fehérjeszerkezetek modelljei

A szerkezeti biológiában a természettudományos megközelítésnek megfelelően minden térszerkezet modellnek tekintendő, a sokaság-alapú megközelítés új modellek felállítását jelenti. A modellnek igazodnia kell az adott feladathoz, mivel nem reális elvárás, hogy minden modell minden feladathoz jól felhasználható legyen. A modellek pontossága csak keresztvalidáción keresztül igazolható, mivel az modellek felállítása során különböző hibák is a modellbe kerülnek. Lehetséges validációs módszer ugyanazon molekula szerkezetének kísérleti megállapítása több módszerrel, például NMR spektroszkópiával és röntgenkristallográfiával is. Ha a modell felállítása jól sikerült, akkor a kísérleti úton meghatározott paramétereknek jó egyezést kell mutatniuk a modell paramétereivel.

Általános módszer az NMR mérésekből származtatott paraméterek felhasználására, hogy először egy nagy számosságú konformer sokaság készül, melynek szimulálása során a kísérletileg meghatározott paraméterek megszorító paraméterként kerülnek alkalmazásra. A szimuláció célja, hogy a generált konformerek minél jobb egyezést mutassanak a kísérleti úton megállapított paraméterekkel. Ezt egy kiválasztási folyamat követi, mely után csak az egymásra legjobban hasonlító, az alkalmazott megszorításoknak legkevésbé ellentmondó konformerek maradnak meg. Ezzel szemben a sokaság-alapú megközelítés nem követeli meg, hogy minden egyes számított konformer önmagában megfeleljen minden mért paraméternek. A kísérletileg megállapított paraméterekkel történő egyezést a sokaság egészére nézve várjuk el.

A szerkezetek sokaságként való reprezentálását indokolja az is, hogy egy tipikus NMR mintában $10^{16} - 10^{17}$ nagyságrendű molekula van, ezért minden mért paraméter sok lehetséges konformer átlagából származik. Ez gyakorlatilag azt is jelenti, hogy nem feltétlenül létezik ténylegesen olyan konformer, amelyik maradéktalanul eleget tesz a megszorító paraméterek elvárásainak, mivel ezek a paraméter készletek is sok molekula átlagából származnak. Ezért a sokaságokkal szemben az az elvárás, hogy a sokaságra vonatkozó átlagolt NMR spektroszkópiás paraméterek tegyenek eleget a kísérletileg meghatározott megszorításoknak.

Tudomásunk szerint az első, kifejezetten dinamikus szerkezeti sokaságok elemzésére szolgáló webszervert csoportunk hozta létre, ez volt az 'eredeti' CoNSEnsX szerver [61].



1.5. ábra. Az 'eredeti' CoNSEnsX szerver [61] működésének sematikus ábrázolása.

Ennek egyik fő filozófiája volt, hogy az adatbázisokból (PDB, BMRB) letöltött fájlok-
kal közvetlenül tudjon dolgozni (1.5. ábra). Ugyanakkor a mögötte álló (Perl nyelvű)
kód egyedi megoldásokat tartalmazott, nehezen volt fejleszthető, a grafikus kimenet kor-
látozott volt és számos funkció hiányzott, pl. a több különböző RDC készlet egyidejű
elemzésére való képesség. A legnagyobb hiányosság az esetleges túlillesztés elemzésére és
a paramétereknek jobban megfelelő alsokaságok előállítására alkalmas szelekciós eljárás
hiánya volt.

2. fejezet

Célkitűzés

A SAH szakaszok detektálására rendelkezésre álló FT_CHARGE és SCAN4CSAH algoritmusok konszenzusa nem kellőképpen érzékeny, elsősorban az FT_CHARGE nagyon szigorú kritériumai miatt. Ugyanakkor a pontos töltésismétlődést mutató, de várhatóan nem alfa-helikális prolingazdag szakaszokat SAH-ként azonosítja. Ezért szükségessé vált az FT_CHARGE algoritmus érzékenységének növelése és egyúttal a prolingazdag szekvenciák kiszűrése a találatok közül. Kovács Ákos (akkor) MSc hallgató Nagy Zoltán témavezetésével elkészítette az FT_CHARGE algoritmus FPGA alapú implementációját, mely lehetővé tette nagy, több tízmillió szekvenciát tartalmazó adatkészletek (teljes UniProt adatbázis) néhány óra alatt történő feldolgozását. Így lehetővé vált a teljes humán proteom újraelemzése a módosított FT_CHARGE paraméterezéssel.

A munkám kezdetekor rendelkezésre álló eredeti CoNSEnsX webszerver, bár jól demonstrálta a dinamikus szerkezeti sokaságok elemzésének létjogosultságát, nem rendelkezett számos fontos funkcióval, többek között szelekciós lehetőséggel. Ezért szükségessé vált ennek újrainplementálása korszerű, technikai oldalról is jól menedzselhető módon, valamint további funkciók, elsősorban a szelekció hozzáadása.

Csoportunkban több PDZ domén dinamikus szerkezeti sokaságait is elkészítették munkatársaim irodalmi adatok alapján. A különböző domének összehasonlító vizsgálata, kiegészítve további doménekkal és sokaságokkal fontos összefüggésekre világíthat rá, melyek felismerésére az egyedi domének vizsgálata nem alkalmas.

A fentiek alapján az alábbi célkitűzéseket fogalmaztam meg:

1. A SAH detektálásra alkalmas FT_CHARGE algoritmus újraparaméterezése az érzékenység növelése érdekében, valamint egy szűrő beépítése a várhatóan nem helikális szerkezetet felvevő találatok kiküszöbölésére
2. Az újraparaméterezett SAH-detektáló eljárással azonosított humán SAH-tartalmú fehérjék elemzése, különös tekintettel korábban nem azonosított fehérjékre
3. Felhasználóbarát webszerver kifejlesztése, mely alkalmas dinamikus fehérjeszerkezeti sokaságok elemzésére és adott paraméterek alapján alsokaság szelekciójára
4. PDZ domének dinamikus szerkezeti sokaságainak összehasonlító elemzése. További PDZ sokaságok szelektálása és a PDB adatbázisban elérhető további szerkezetek

bevonása. A domének közötti szerkezeti eltérések és a doméneken belüli, elsősorban a ligandumkötés hatására történő szerkezeti átrendeződések vizsgálata.

3. fejezet

Módszerek

3.1. A meglévő SAH detektálásra használt módszerek

A CSAH szerverben két, belső fejlesztésű SAH domén prediktáló program konszenzusa által történik a SAH domének felismerése, ezek a SCAN4CSAH és az FT_CHARGE. [18] Az alapjaiban különböző algoritmusok eltérő számú SAH domént detektálnak, a szigorúbb, Fourier transzformáció alapú FT_CHARGE a SCAN4CSAH által SAH doménnek vélt esetek 10%-át jelzi ki. A SwissProt 54.2 adatbázist vizsgálva megállapítható, hogy a mindkét algoritmustól jó pontszámot kapó, eddig fel nem ismert SAH domének ténylegesen SAH-t alkotnak oldat fázisban, amiknek létezése CD (circular dichroism) spektroszkópiával igazolható. [48]

Mivel a kísérletileg is igazolt SAH domének száma még mindig nagyon alacsony, ezért a SAH domén detektálására egyelőre nem került kidolgozásra egy átfogó módszer. A meglévő módszerek együttes alkalmazása és finomhangolása viszont segíthet a hamis pozitív találatok csökkentésében. Az új, kísérletileg feltérképezett SAH domének megjelenése rávilágított viszont az algoritmusok gyengeségeire is, miszerint az algoritmusok nem detektálták azokat a SAH doméneket, melyen megléte kísérleti úton bizonyításra került, szükségessé téve az algoritmusok paraméterezésének felülvizsgálatát [III].

Az FT_CHARGE algoritmus új paraméterezésének beállításához 9, kísérletileg igazolt SAH domént és az azokat tartalmazó fehérjéket, valamint negatív kontrollokat használtunk fel (3.1. és 3.2. táblázat). A kontroll szerkezetek kiválasztását megnehezíti, hogy nincsenek igazán jól használható negatív esetek, illetve eleve nagyon kevés azonosított SAH domén ismert. Az algoritmus két legfontosabb paramétere a Fourier transzformáció után a spektrumban kapott legnagyobb amplitúdó, illetve az ehhez tartozó, előzetesen generált szekvenciák adatai alapján számított P érték, ezek pontosabb meghatározása / beállítása volt a cél.

A beállításához a paraméterek adott kombinációival lefutattam az FT_CHARGE eljárást a tesztben szereplő fehérjékre, és megvizsgáltam, hogy a SAH szakaszokba eső aminosavak mekkora hányadát sikerült az egyes beállításokkal azonosítani a konszenzus (SCAN4CSAH + FT_CHARGE) eljárással, illetve csak az FT_CHARGE használatával. A kapott eredmények vizsgálata alapján a javasolt új beállítások: minimális amplitúdó 7, a hozzá tartozó P-érték pedig 0.05. Fontos megjegyezni, hogy a pontos szakaszhatá-

Kísérletileg igazolt SAH domének				
Név	UniProt ID	Uniprot AC	SAH kezdete	SAH vége
Caldesmon	A0A1L1RXH5	A0A1L1RXH5_CHICK	196	252
GCP60	Q9H3P7	GCP60_HUMAN	183	238
INCENP	P53352	INCE_CHICK	503	715
MAP4K4	O95819	M4K4_HUMAN	417	480
MFAP1	P55081	MFAP1_HUMAN	267	344
Myosin 7	P97479	MYO7A_MOUSE	866	935
Myosin 10	Q9HD67	MYO10_HUMAN	813	909
Myosin 6	Q9UM54	MYO6_HUMAN	915	980
Snu23	G0S6R0	G0S6R0_CHATD	131	164

3.1. táblázat. A paraméteroptimalizáláshoz felhasznált kísérletekből megállapított SAH domének

Negatív kontroll szekvenciák				
Név	UniProt ID	Uniprot AC	Megjegyzés	
Myoglobin	P02144	MYG_HUMAN	globuláris, all- α	
GCN4	P03069	GCN4_YEAST	coiled coil	
Myosin 2	Q9UKX2	MYH2_HUMAN	coiled coil	
Kinesin-related protein 12	Q6S000	KIF12_DICDI	coiled coil	
P53	P02340	P53_MOUSE	rendezetlen	
Golgi SNAP receptor complex member 1	Q62931	GOSR1_RAT	coiled coil	
RNA-binding protein 12B	Q8IXT5	RB12B_HUMAN	korábban SAH-nak jósolt, Pro-gazdag régió	
Zinc finger protein 609	O15014	ZN609_HUMAN	rendezetlen	
Calpastatin	P20810	ICAL_HUMAN	rendezetlen	

3.2. táblázat. A paraméteroptimalizálás ellenőrzéséhez használt negatív kontroll szekvenciák

rok predikciója/megállapítása nem reális, hiszen kísérletes oldalról nem várható, hogy aminosavak egyenkénti hozzáadásával nagy számú konstrukciót állítsanak elő és elemezzenek, és várhatóan még ebben az esetben sem lenne egyszerű és egyértelmű pl. CD-spektroszkópiával ezekre kapott α -helikális jelleg esetében az egyes konstrukciókra kapott értékek között határt vonni. Így a pontos tól-ig pozíciók meghatározása a SAH szegmensek esetében nem feltétlenül életszerű, az átmenet feltehetően többé-kevésbé folytonos a rendezetlen és/vagy szuperhélix szakaszok felé a szekvencia mentén. A helicitás valószínűsége akár folyamatosan változhat a szekvencia mentén, ezért a SAH szegmensek meghatározásához mindenképpen önkényes határt kell meghúzni. A szekvencia távolabbi részeinek stabilizáló hatása miatt a tényleges helicitás CD-spektroszkópiával sem mindig vizsgálható. NMR spektroszkópiával egy átfogóbb kép kapható a szekvencia tényleges helicitásáról, viszont ilyen adatok egyelőre még csak ritkán állnak rendelkezésre.

3.2. UniProt adatbázis, Gene Ontology

Az UniProt adatbázisban több mint 120 millió, annotációkkal ellátott aminosav szekvencia kereshető [2]. Az adatbázisban többek között két fontos al-adatbázis, a Swiss-Prot

és a TrEMBL adatai érhetőek el. A Swiss-Prot adatbázis manuálisan annotált, kísérletileg is igazolt, míg a TrEBML DNS szekvenciák automatikus, in silico translációjával meghatározott szekvenciákat tartalmaz és maga az annotáció is automatikusan történik. A szekvenciákhoz tartozó annotáció tartalmazza többek közt a szekvencia variációkat, a fehérjék szerkezeti és funkcionális elemeit, ideértve a doméneket és motívumokat, illetve kereszthivatkozásokat számos, többek között szerkezeti adatbázisra. A SAH motívumok szempontjából fontos, hogy a UniProt annotáció feltünteti a szuperhélix szakaszokat. Egyes jobban jellemzett esetekben már a SAH régiók annotációja is megjelenik, pl. a miozin 10 (Q9HD67) leírásában. Ez azonban korántsem jelenti azt, hogy minden SAH-tartalmú fehérje esetében ez az információ megtalálható.

További releváns információkat tartalmaz a GO (gene ontology - gén ontológia) adatbázis, melyben génekre és fehérjékre vonatkozó, standardizált kifejezésekkel jellemzik a fehérjék funkcióját. A kategorizálás egyik alapja, hogy a gének nagy része, mely alapvető biológiai funkciókat definiál, valamennyi eukarióta fajban megegyezik [1], ezért az egyik faj esetén felismert és a génhez rendelt biológiai funkció más fajok génjeihez is hozzárendelhető. A GO adatbázis annotációi egy funkcionális szótárat képeznek, mellyel az ilyen, gének és géntermékek közötti összefüggések/hasonlóságok kifejezhetőek. Az annotáció három fő aspektust fed le: a lokalizáció (cellular localization), a biológiai folyamat, amelyben a fehérje részt vesz (biological process), illetve a molekuláris funkció (molecular function). A GO egyik fő felhasználási területe az egy adott kísérlet során azonosított szekvenciák és ezek biológiai szerepének egymáshoz rendelése. Ehhez ún. feldúsulási elemzést (enrichment analysis) szokás végezni, melynek során összevetik a vizsgált fehérjékészlet GO annotációját a háttér, legtöbb esetben az adott organizmus teljes proteomjának GO annotációjával, és meghatározzák azokat a GO kulcsszavakat, amelyek gyakorisága a két fehérjékészletben statisztikailag jelentősen eltér. A korábban azonosított SAH-tartalmú fehérjék esetében is a GO adatbázis segítségével azonosították az RNS-kötő funkció gyakori előfordulását [18]. Ilyen vizsgálat elvégzéséhez pl. az AmiGO webes eszköz használható (<http://amigo.geneontology.org/amigo>).

Az egyik, az idegi átvitelre specializált adatbázis a SynaptomeDB, melyben a szinapszis elemeinek génjei és fehérjéi, illetve a hozzájuk tartozó annotációk találhatóak, mint például a neurotranszmitterek és receptoraik, citoskeletális fehérjék, állványfehérjék és membránfehérjék [41]. Az adatbázis egyik fő célja az ezekre a génekre és géntermékekre vonatkozó információ integrálása.

3.3. A CoNSEnsX+ fejlesztéshez használt programnyelvek és fejlesztőeszközök

A CoNSEnsX szerverrel folytatott munkát a könnyebb fejlesztés és karbantarthatóság jegyében a projekt teljeskörű újrainplementálásával kezdtem, az eredeti kódbázis [61] referenciaként került csak felhasználásra. A bioinformatikai témakörökben jelenleg az egyik legelterjedtebb programozási nyelv a Python, szinte valamennyi általánosabb jellegű feladat elvégzésére már megbízható könyvtárak állnak rendelkezésre. A könyvtárak felhasználásával javul a kód minősége és több idő/erőforrás marad a tényleges algoritmikus

problémák megoldására.

A CoNSEnsX szerver valamennyi általános funkciója (NMR paraméterek beolvasása, megfelelési mérőszámok számítása, grafikonok készítése) elérhető a szerver saját könyvtáraiból, így szükség esetén szükség esetén más szkriptekben is fel lehet használni. A webszerver funkciókat a Django keretrendszer segítségével implementáltam. A szerverhez adatbázis is tartozik, melyből a már lefutott számítások visszakereshetők.

A maradvány dipoláris csatolások és a kémiai eltolódások számításához külső szoftverek is szükségesek, ezek a Pales [59] és a ShiftX [36]. A PDB fájlok kezeléséhez és írásához a prody [4] könyvtárat, az mért NMR adatokat tartalmazó STAR-NMR fájlok beolvasásához pedig nmrpystar könyvtárat használtam. Minden, a számítások eredményeit bemutató oldalon a matplotlib könyvtárral készített vektorgrafikus ábrák találhatóak.

```
version: '3'
```

```
services:
```

```
  consensx-server:
```

```
    image: derpuntigamer/consensx
```

```
    ports:
```

```
      - 8080:8000
```

```
    depends_on:
```

```
      - consensx-db
```

```
    links:
```

```
      - consensx-db
```

```
    volumes:
```

```
      - media:/usr/src/app/consensx.itk.ppke.hu/media
```

```
  consensx-db:
```

```
    image: postgres:11.1
```

```
    environment:
```

```
      - POSTGRES_PASSWORD=password
```

```
      - POSTGRES_USER=consensxuser
```

```
      - POSTGRES_DB=consensx
```

```
    volumes:
```

```
      - postgresql:/var/lib/postgresql/data
```

```
volumes:
```

```
  media:
```

```
    driver: local
```

```
  postgresql:
```

```
    driver: local
```

1. forráskód szegmens. *A CoNSEnsX+ szerver Docker Compose konfigurációja*

A szerverhez szükséges könyvtárak és külső szoftverek, illetve az adatbázis installá-

lása és konfigurálása körülményes feladattá vált. Ezen felül a külső szoftverek egy része csak Linux alapú operációs rendszeren működik és a forráskód hiányában nem is lehet lefordítani más rendszerekre. A szerver installálását és frissítését egyszerűsítendő, Docker konténereket alakítottam ki a CoNSEnsX szerver és a hozzá tartozó adatbázis számára (2. forráskód szegmens). A CoNSEnsX szerver Docker képfájl készítése kihasználja a Docker "multistage build" funkcióját, az első lépésben csak a könyvtárakat tartalmazó képfájl réteg kerül elkészítésre, ez csak akkor kerül a továbbiakban módosításra, amennyiben a könyvtár függőségek megváltoztak. A képfájlok szabadon elérhetőek a Docker Hub tárhelyről. A képfájlok a forráskód kezelő rendszerben történő változás során automatikusan, a Docker Hub által nyújtott ingyenes infrastruktúrán készülnek el. A szerver könyvtár függőségeit a Pipenv környezetben rögzítettem, az installáláskor használt (kipróbált) verziók a Pipfile.lock fájlban rögzítésre kerülnek, így a teszteléskor is használt környezet a forráskódból pontosan helyreállítható.

A Docker Compose eszköz segítségével a hozzá tartozó YAML konfigurációs fájl megléte mellett egy paranccsal el lehet indítani a szervert és a hozzá tartozó adatbázist. A szerver szükség esetén automatikusan frissíti magát és a hozzá tartozó adatbázis táblákat (1. forráskód szegmens). A konténerizált kialakítás lehetővé teszi, hogy a szervert bármilyen, a Docker futtatására képes operációs rendszeren használható legyen, illetve könnyen telepíthető privát vagy publikus felhőbe is. Szükség esetén növelhető a párhuzamosan futtatott CoNSEnsX szerverek száma a Docker környezetben, amennyiben ezt a számítási igény megkívánja.

3.4. PDZ szerkezetek elemzése: molekuladinamikai szimulációk, adatbázisok

A PDZ domének összehasonlító elemzéséhez dinamikus szerkezeti sokaságokat generáltunk molekuladinamikai szimulációval. Kiindulási alapnak a PSD-95 fehérje PDZ3 doménjének szabad és kötésben lévő állapotait használtuk fel, annak teljes és $\Delta 7CT$ formáival. A generált sokaságok kísérletileg megállapított paraméterekkel való egyezését a CoNSEnsX⁺ szerverrel ellenőriztem. A sokaságokkal leírt konformációs tereket megvizsgáltuk abból a szempontból, hogy milyen dinamikus szerkezeti változások következnek be, amikor a domén egy ligandummal kötésbe lép.

A PDZ3 szerkezeti sokaságait Hinsenkamp Anett állította elő és a részletes jellemzősüket is ő végezte. Alább azért fejtem ki ennek részleteit, mert ez szorosan kapcsolódik az általam elvégzett széleskörű PDZ elemzéshez, ezekkel egy közleményben szerepelnek. A PDZ elemzésnél a sokaságok előállítását a 1BFE és 1BE9 PDB szerkezeteket felhasználva kezdtük, melyek a patkányban található PSD-95 PDZ3 domén szabad és kötött állapotait tartalmazzák [14]. Ezeket a modelleket a referenciaként használt szerkezethez [40] kellett igazítani, ami oldalláncok kisebb módosításait (Asp322 kiegészítése és Ile328 oldallánc cseréje Val-ra) jelenti, illetve a szerkezetek végének levágását a 402. aminosavtól. Az 1BFE modell N-terminálisnál hiányzó aminosavait (301-306) a kötött forma (1BE9) alapján kipótoltuk. Az előkészített szerkezeteket a GROMACS [49] 4.5.5-ös verziójával energiaminimalizáltuk. A szerkezetekhez tartozó molekulagerinc és oldallánc S²

rendparamétereket Dr. Andrew Lee bocsájtotta rendelkezésünkre.

A kísérletileg megállapított S^2 rendparamétereket a molekuladinamikai szimuláció során megszorító paraméterként használtuk fel a GROMACS 4.5.5 módosított változatával [16]. A 4 szimuláció (szabad és kötött, teljes hosszúságú és levágott végű) az AMBER99SB erőterrel és a TIP4P víz modellel került lefuttatásra 8 párhuzamos replikával, a hozzájuk tartozó S^2 paraméterek megszorításként történő alkalmazása mellett. A 20 ns hosszú szimulációknak csak a második felét (10 ns) tartottuk meg, így mindegyik szimulációból egy 808 konformerből álló sokaság keletkezett.

A megszorító paraméterek hatásának vizsgálatára rövidebb, 4 ns hosszú szimulációkat futtattunk, csak a molekulagerincre, illetve csak az oldalláncra vonatkozó S^2 rendparaméterekkel. Ezek alátámasztották, hogy a megszorító paraméterek alkalmazása nem befolyásolja hátrányos módon a szimulációval előállított konformerek szerkezeteit. Az előállított sokaságok megfelelése a kísérletileg megállapított S^2 rendparamétereknek a CoNSEnsX⁺ szerverrel került ellenőrzésre.

A molekuladinamikai szimuláció során az S^2 rendparaméterek megszorító paraméterként történő alkalmazásának egyik alapvetése, hogy a szerkezetek egy jól definiált szerkezet körül fluktuálnak. A számításainkban viszont nem tudtunk a szerkezetekre vonatkozó NOE (Nuclear Overhauser Effect) megszorító paramétereket alkalmazni, mellyel a szimulált szerkezetet a kísérleti adatokból kiindulva szerkezetmeghatározás során meghatározott szerkezet közelében lehet tartani. Emellett az S^2 rendparaméterek megszorító paraméterként történő alkalmazása főleg a mozgások irányát korlátozza, a mozgások amplitúdójára kevésbé van hatással [11]. Ebből kifolyólag a mozgások mértéke a szimuláció során felkálázódhat a doménhez tartozó kísérletileg megállapított dinamikához képest.

A szabad állapotban lévő $\Delta 7CT$ sokaság eloszlása az első főkomponens mentén bimodális és a mozgás jellemző amplitúdója miatt felmerül, hogy a "nyitó-csukó" mozgás vélhetően egy lassabb időskálán történik, mint amelyet az S^2 rendparaméterek jellemeznek. Ennek ellenére megjegyzendő, hogy az ilyen és ehhez hasonló hurok mozgások nem feltétlen bonthatóak le nem átlapoló időskálákra. Az alkalmazott módszerrel a kísérletileg megállapított paramétereknek jól megfelelő sokaságok állíthatóak elő, így feltételezhető, hogy a sokaságok jól reprezentálják a szerkezet valós konformációs terét és dinamikáját, alkalmassá téve ezeket a szerkezeti átmenetek vizsgálatára. A megszorító paraméterekkel futtatott molekuladinamikai szimuláció mint módszer viszont magában hordozza, hogy az így előállított sokaságok eloszlása és a mozgások amplitúdói kissé eltérnek a kísérletileg megállapítottaktól.

A vizsgált sokaságok elemzése során felmerült az igény egy viszonyítási alapként is használható, a PDZ domént általánosaságban reprezentáló nagyobb sokaságra. Ennek összeállításához a már meglévő NOE alapú PSD-95 PDZ1 és PDZ2 sokaságokhoz [25] további sokaságokat kerestem, amelyekhez elérhetőek kísérletileg megállapított kémiai eltolódások (2KPL, 2LC6, 2LC7, 2LOB, 1GM1, 1OZI, 2KPK, 1VJ6, 1UM7). Az így kiválasztott sokaságokkal ezután rövid (20 ns), megszorító paraméterek nélküli molekuladinamikai szimulációkat futtattam, majd az így készült sokaságokból a CoNSEnsX⁺ szerver szelekció funkciójával a H α és C α kémiai eltolódásoknak való megfelelést maximalizálva

alsokaságokat készítettem.

Két szabad és kötött állapotban lévő párra (1GM1-1VJ6 és 2KPK-2KPL) egy replikás hosszabb, 500 ns-os szimulációkat is futtattunk a GROMACS 2020 verziójával, az AMBER99SB-ILDN erőter [26] és a spc/e víz modell alkalmazásával. Energiaminimalizálás után a szerkezetek előkészítése pozíciómegkötéses szimulációkat történt, először állandó térfogaton (NVT), majd állandó nyomáson (NPT), 1-1 ns hosszan. Az utóbbi végállapotából indultak az 500 ns hosszú, megkötések nélküli szimulációk. Ezekből és a rövidebb szimulációkkal előállított sokaságokból alsokaságokat képeztem a CoNSEnsX⁺ szerveren, a C α és a H α kémiai eltolódásokra futtatott szelekcióval (3.3. ábra).

PDB ID	Leírás	BMRB ID	Alsokaság méret
Rövid szimulációk			
2KPK	MAGI-1 PDZ1	16558	11
2KPL	MAGI-1 PDZ1 / E6CT	16559	7
1GM1	Second PDZ Domain (PDZ2) of PTP-BL	5131	9
1VJ6	PDZ2 from PTP-BL in complex with the C-terminal ligand from the APC protein	5131	12
2LC6	Solution structure of Par-6 Q144C/L164C	17599	13
2LC7	Solution structure of the isolated Par-6 PDZ domain	17600	17
1OZI	The alternatively spliced PDZ2 domain of PTP-BL	5762	11
1UM7	Solution structure of the third PDZ domain of synapse-associated protein 102	11198	21
Hosszú szimulációk			
2KPK	MAGI-1 PDZ1	16558	11
2KPL	MAGI-1 PDZ1 / E6CT	16559	14
1GM1	Second PDZ Domain (PDZ2) of PTP-BL	5131	25
1VJ6	PDZ2 from PTP-BL in complex with the C-terminal ligand from the APC protein	5131	29

3.3. táblázat. *A molekuladinamikai szimulációval előállított sokaságok és a belőlük szelektált alsokaságok mérete*

Az előállított sokaságok egy közös sokságba történő összefűzéséhez először illeszteni kellett azok szerkezeteit. A több szerkezetre alkalmazott illesztés a MAMMOTH-Mult programmal készült [29]. Ehhez az egyes sokaságok első modelljeit, illetve a több láncból álló szerkezetek első láncait használtuk fel. A tandem PDZ doméneket tartalmazó szerkezetek fel lettek darabolva úgy, hogy a PDZ doméneket külön-külön is fel lehessen használni. Mivel a MAMMOTH-Mult programmal nem lehetett egyszerre mind a 152 szerkezetet illeszteni, ezért az illesztés több körben történt, az egyes illesztések eredményeit összefűztük, azokat a részeket megtartva, amelyek valamennyi felhasznált szerkezetben megtalálhatók. Az így kapott illesztés (http://users.itk.ppke.hu/~dudda/PhD/PDZ_CORE.fst) felhasználásával minden felhasznált PDB-s modellt és sokaságot redukáltam, hogy csak az illesztés során kapott, közös aminosavakat tartalmazzák, ez lesz a további összehasonlítások egyik alapja, amit "PDZ mag"-nak neveztünk el.

```

FROM ubuntu:20.04 AS builder
LABEL maintainer="dudola.daniel@itk.ppke.hu"
ENV LANG=C.UTF-8

RUN apt-get update -q \
    && apt-get install -y -q wget python3 python3-pip wget \
    && wget users.itk.ppke.hu/~dudola/progs_consensx.tar.gz \
    && pip3 install pipenv

ENV PYROOT /pyroot
ENV PYTHONUSERBASE $PYROOT
WORKDIR /build
COPY Pipfile Pipfile.lock ./

RUN PIP_USER=1 PYTHONUSERBASE=$PYROOT pipenv install --system --deploy

# At this point, the dependencies are built in the builder container.
# Copy dependencies to final image.

FROM ubuntu:20.04
ENV PGPASSWORD=password
ENV LANG=C.UTF-8

COPY --from=builder /progs_consensx.tar.gz .

# add i386 architecture (pales dependency)
RUN dpkg --add-architecture i386 \
    && apt-get update -q \
    && apt-get install -y -q postgresql-client \
        libc6:i386 libncurses5:i386 \
        libstdc++6:i386 libx11-6:i386 python3 \
        python3-setuptools python3-six \
    && mkdir -p /usr/src/app \
    && tar -C /usr/src/app -zxf progs_consensx.tar.gz \
    && rm -rf progs_consensx.tar.gz \
    && apt-get autoremove -y \
    && apt-get clean -y \
    && rm -rf /var/lib/apt/lists/*

ENV PYROOT /pyroot
ENV PYTHONPATH $PYROOT/lib/python:$PATH
ENV PYTHONUSERBASE $PYROOT

COPY --from=builder $PYROOT/lib/ $PYROOT/lib/
COPY . /usr/src/app/consensx.itk.ppke.hu

WORKDIR /usr/src/app/consensx.itk.ppke.hu

CMD ["bash", "docker_cmd.sh"]
EXPOSE 8000

```

4. fejezet

Eredmények

4.1. Az FT_CHARGE algoritmus újraparaméterezése és helicitás szűrő beépítése

A SCAN4CSAH és az FT_CHARGE algoritmusok megtervezésekor szempont volt, hogy az algoritmusok csak SAH domént alkotó, töltéssel rendelkező aminosavakat vegyék figyelembe. A hélix kialakulásában (vagy a kialakulásának megakadályozásában) szerepe van viszont a szekvenciában található aminosavak együttes kompozíciójának is, például a magas prolin tartalmú szekvenciák általában nem képeznek hélixet. Ezen probléma kiküszöbölésére a dekektált SAH doméneken még beépítettem egy helicitás filtert a munkamenetbe a Chou-Fasman módszer alapján. A Chou-Fasman másodlagosszerkezet-predikciós eljárás az egyedi aminosavak térszerkezeti preferenciáin alapszik, azaz minden aminosavhoz hozzá van rendelve egy-egy érték, ami jellemzi a hajlandóságát (propensity) az alfa-helikális és béta-redőzött szerkezeti elemekben való részvételére [10]. Ennél természetesen ma már jóval pontosabb eljárások léteznek a másodlagos szerkezeti elemek predikciójára. A választásom azért esett mégis erre a módszerre, mert ez a modernebb eljárásokkal szemben nem használja fel hasonló fehérjék szekvenciáit, ami megítélésünk szerint nagyban befolyásolhatná az eredményeket ebben az esetben. A SAH szegmensek ráadásul ún. alacsony komplexitású szakaszok, azaz kevés típusú aminosavból állnak, és ezért kifejezetten nem informatív a hasonló szekvenciák figyelembe vétele, hiszen ebben az esetben a hasonlóság – szemben a globuláris régiók komplex szekvenciájával – mérhető valószínűséggel kialakulhat evolúciósan nem rokon fehérjeszakaszok esetében is. A filter beállításához a nem-redundáns PDB SELECT adatbázisban [17] található szekvenciák DSSP annotációját [9] használtam fel. Az itt található, legalább 15 aminosav hosszú alfa-hélixek esetében kiszámoltam az egy aminosavra jutó átlagos Chou-Fasman-féle $P\alpha$ értéket. Ezen értékekre az R program [42] segítségével EVD eloszlást illesztettem. Az ennek alapján kapott valószínűségi küszöbérték meghatározásához a SwissProt adatbázis 2015_08 kiadásán prediktált SAH szekvenciákat vizsgáltam meg, és ennek alapján a P értéket 0.5-nek választottuk, azaz csak az ennél kisebb P-értékkel rendelkező szekvenciákat vettem figyelembe. A webszerver későbbi fejlesztése során a filtert tovább finomítottam, mert észrevettem, hogy a hosszú, több szakaszból álló SAH szekvenciák esetében nem működik megbízhatóan. Ezért a filtert nem csupán a teljes SAH szakaszokra, hanem –

Név	UniProt ID	SAH kezdete	SAH vége	Javasolt paraméterek A \geq 7 P \leq 0.05	Magas érzékenység A \geq 5 P \leq 0.5	Előző értékek A \geq 10 P \leq 0.01
Caldesmon	A0A1L1RXH5	196	252	174-316 408-521	1-36 153-384 394-528	178-312 411-517
GCP60	Q9H3P7	183	238	158-274	158-283	158-268
INCENP	P53352	503	715	502-679	484-769	
MAP4K4	O95819	417	480	374-503	355-533	386-500
MFAP1	P55081	267	344	225-259	173-273	
Myosin 10	Q9HD67	813	909	773-870	765-880 903-955	
Myosin 7	P97479	866	935	848-929	829-939	
Myosin 6	Q9UM54	915	980	932-995	889-923 932-995	932-994
Snu23	G0S6R0	131	164		110-214	

4.1. táblázat. Az FT_CHARGE optimalizált paraméterezése a kísérletekből megállapított SAH domének alapján

mivel az FT_CHARGE az FFT eljárás miatt adott ablakméretbe eső szegmenseket elemez – az FT_CHARGE által azonosított minden egyes szegmensre alkalmazom, és csak a filteren 'átment' szakaszokat használom fel a teljes SAH összerakásához. Ezzel biztosítható, hogy ne forduljon elő az a helyzet, hogy egy hosszú SAH a végső szűrőn essen ki, hanem megtartható a releváns, várhatóan valóban helikális szakasza. Mindkét szűrési lépést a csahdetect.pl szkriptbe építettük be, ez a szkript végzi az FT_CHARGE által talált szakaszok összefűzését, a SCAN4CSAH eljárással való konszenzus megállapítását és így a végső kimenetként megadott SAH szakaszok pontos határainak megállapítását [II, III].

4.2. Az újraparaméterezett SAH-detektáló eljárással azonosított humán SAH-tartalmú fehérjék elemzése

Az FT_CHARGE algoritmus újraparaméterezése és a helicitás filter hatékonysága - a fent említett 9 kísérletileg azonosított SAH domént tartalmazó szekvenciákon kívül - az elegendő mennyiségű mért adat hiányában negatív kontrol bemenetekkel is tesztelhető, hiszen az érzékenység növelése mellett a hamis pozitív találatok továbbra is elkerülendők. Negatív kontrolnak használható bármilyen, illetve legjobb esetben a SAH doménre hasonlító szekvencia alkalmazható, mely nem képez helikális szerkezetet.

A kilenc kísérletileg megállapított SAH domént tartalmazó doménre az FT_CHARGE algoritmus több amplitúdóval (0-tól 20-ig, 0,5-ös lépésközzel) és P-értékkel (0,001; 0,005; 0,01; 0,05; 0,1 és 0,5) került lefutattásra, melyből kiderült, hogy a 7-es amplitúdóval és a 0,05-ös P-értékkel a legpontosabb a predikció. (4.2 ábra) A negatív kontrollként felhasznált szekvenciák esetén az algoritmus a legmegengedőbb paraméterezéssel sem talált SAH domént.

CSAH server 2.5

Detect Single α -Helices in protein sequences

Description	Example	CSAH Server	Download
-------------	---------	-------------	----------

Enter a single amino acid **sequence** (plain or FASTA):

Predict SAHs

or upload a single amino acid **sequence file** in FASTA format:

Browse... No file selected.

Reset form

Choose which **prediction method(s)** to run to detect Single α -Helices:

SCAN4CSAH & FT_CHARGE (consensus prediction)

Minimum length of SAH segment for SCAN4CSAH: 30 residues

Window size for FT_CHARGE: 32-64, minimum amplitude: 7, maximum P-value: 0.05

Questions and suggestions regarding the CSAH server should go to Zoltán Gáspári (gaspari.zoltan at itk.ppke.hu)
© ZG & GT, 2009,2015

4.1. ábra. A csahserver bemeneti űrlapja. A szekvencia megadása vagy fájlként való feltöltése után kiválasztható és paraméterezhető a predikciós módszer (SCAN4CSAH, FT_CHARGE és a konszenzus módszer)

Megjegyzendő viszont, hogy az algoritmus(ok) által prediktált SAH szegmensekben megfigyelhető uniform arginin-lizin eloszlás valamelyest ellentmond a kísérleti megfigyeléseknek. Ennek egy lehetséges értelmezése, hogy az újonnan meghatározott paraméterezés ellenére is több hamis pozitív predikció készül, valószínűleg a SAH szegmensek kialakulását hátráltató egyéb tényezők figyelmen kívül hagyása miatt. Egy további magyarázat lehet, hogy a vélt hamis pozitív találatok ténylegesen stabil magányos hélixeket alkotnak az eredeti környezetükben fellépő egyéb hatások miatt. Az ilyen jellegű tényezők felderítéséhez több, potenciálisan SAH domént tartalmazó fehérje részletes kísérleti vizsgálata lenne szükséges. Amint több, a SAH doménekre vonatkozó kísérleti adat áll rendelkezésre, lehetségessé és szükségessé válik a paraméterezés további finomítása.

A két algoritmus konszenzusából 483 humán fehérjében sikerült SAH doméneket de-

Adatbázis	Konszenzus módszer		Csak az FT_CHARGE algoritmus által prediktált	
	SAH szegmensek	SAH domént tartalmazó fehérjék	SAH szegmensek	SAH domént tartalmazó fehérjék
Új, módosított paraméterezés ($A \geq 7$, $P \leq 0.05$)				
SwissProt	1137	985	242	237
TrEMBL	121171	108455	30367	30044
Régi paraméterezés ($A \geq 10$, $P \leq 0.01$)				
SwissProt	629	555	78	77
TrEMBL	64327	57631	7753	7688

4.2. ábra. A konszenzus módszer és csak az FT_CHARGE algoritmus által prediktált, SAH domént tartalmazó humán funkcionális csoportok a módosított és az eredeti paraméterezéssel (UniProt, 2018 márciusi kiadás)

tektálni, csak az FT_CHARGE algoritmus 527-et talált, ezekből 111 az AmiGO adatbázisban is megtalálható (GO - Gene Ontology).

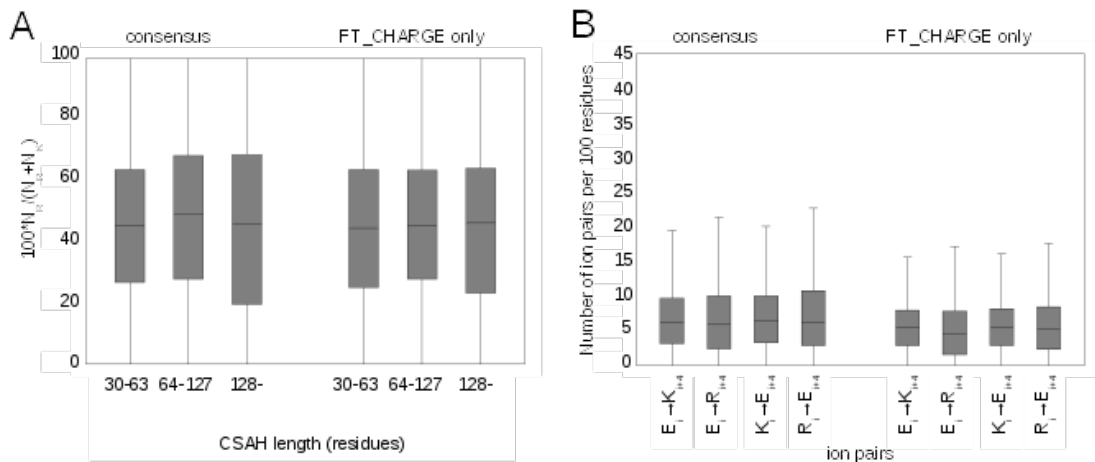
	SAH szegmensek	Fehérjék
Konszenzus	483	396
FT_CHARGE	527	473
Csak FT_CHARGE	138	128

4.2. táblázat. SAH szegmensek és SAH domént tartalmazó fehérjék a humán UniProt szekvenciákban (2018 márciusi kiadás). A "csak FT_CHARGE" értékek olyan SAH találatok, melyekben a konszenzus módszer a detektált SAH domén aminosavainak kevesebb mint felét azonosította SAH doménként.

A detektált SAH domént tartalmazó találatok "GO-term enrichment" analízise kimutatta, hogy a fehérjék többnyire a citoskeletonban vannak jelen, illetve az RNS feldolgozásában vesznek részt. A PANTHER elemzés [31] nem mutatta ki, de a találatok átnézése során feltűnt, hogy több fehérje is rendelkezik idegi differenciálódásban vagy jelátvitelben betöltött szereppel. Ezért külön figyelmet fordítottam az ilyen folyamatokban részt vevő fehérjék elemzésére is, többek között a SynaptomeDB adatbázist is felhasználva.

4.2.1. Citoszkeletális fehérjék

A konszenzus módszer 17 SAH domént tartalmazó fehérjét prediktált, melyek a citoskeletonban találhatóak. Érdekes, hogy mindhárom, a troponin komplexet alkotó troponin T fehérje (TnC, TnI és TnT) a találatok között van. A troponin komplexnek a váz- és szívizomzat (harántcsíkolt izom) összehúzódásának modulálásában van fontos szerepe. Habár a troponin fehérjék közül a szívizomzat troponin T fehérjéje a legtöbbet tanulmányozott, a SAH domént is tartalmazó régiójának szerkezetét még nem vizsgálták behatóan. A szívizomzat troponin T fehérjében a prediktált SAH domén a központi domén



4.3. ábra. Töltéssel rendelkező aminosavak és ionpárok összetevése: konszenzus módszerrel és a csak az FT_CHARGE által prediktált SAH domének (Uniprot, 2018 márciusi kiadás). A) $N_R/(N_R+N_K)$ százalékos eloszlása a SAH szegmensek hossza szerinti felosztásban B) egyes ionpárok normalizált előfordulási gyakorisága

N-terminális régiójában található, mely a tropomiozin kötésében vesz részt és feltehetőleg alfa-helikális szerkezetű [24]. A fehérje egyik deléciós mutációja, amely a SAH régióban négy egymás utáni glutaminsavból egy kiesésével jár, súlyos strukturális zavarokat okoz a szívizomzatban [32].

	Cyto skele ton	Cell corte x	Paras peckle s	Exon- exon junctio ns	Golgi appar atus	RNA bindi ng	MAP KKK K aktivit y	Neur onal proce ss	Syna ptom eDB
All	31	10	3	3	18	21	3	13	17
Cytoskeleton		9		1	4	2	1	8	9
Cell cortex					2			5	6
Paraspeckles						3		1	1
Exon-exon junctions						3			
Golgi apparatus						3	1	3	
RNA binding								1	3
MAP kinase kinase kinase activity								2	2
Neuronal process									8

4.4. ábra. A konszenzus módszer által prediktált, SAH domént tartalmazó humán funkcionális csoportok

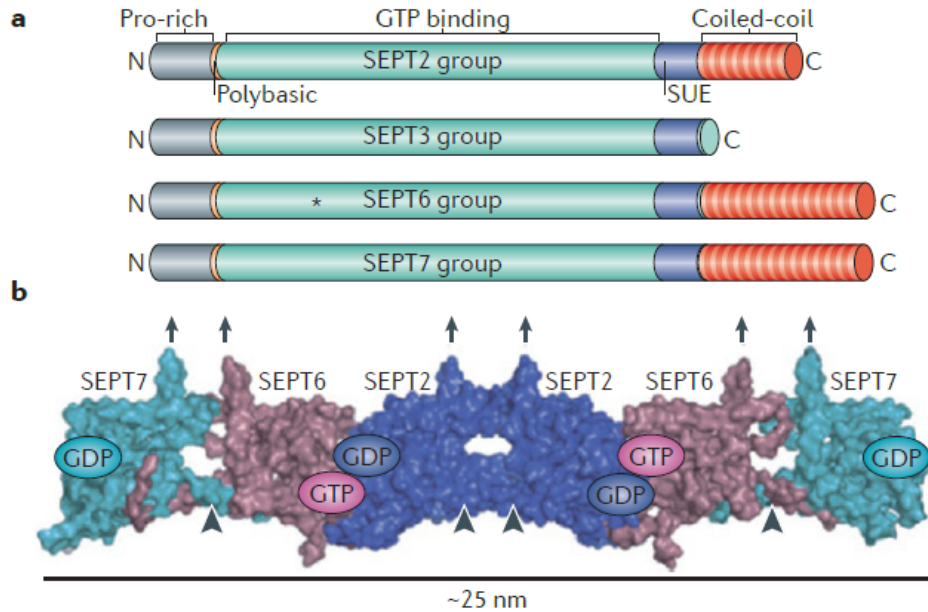
A septin 7 fehérje a citoszekeletális fehérjék egy viszonylag újonnan azonosított csa-

Fehérje	Régió	Szekcenvia
Q9HAU5 (UPF2)	48-115	EVSKAPEDKQKRLLEDDKQKCEDKERKQKDEEKVKAEESKQKKEEEEKQKHQEER KKQEEQAKRQQEE
Q9H1J1 (UPF3A)	237-303	EERRRELEKQKRLREEEKRRRREERCKKKEQTDKQKQKIAEKEVRIKLLKKPEKGE EPTTEKPKERGE
Q9BZ17 (UPF3B)	209-269	RMREEKREERRRREIERKRQREERRRKWKKEEKRRKQDIEKLLKIDRIP
P45379 (TNNT2)	142-185	AERAEQQIRNREKERQNRLAEEARREEEENRRKAEDDEARKK
P13805 (TNNT1)	112-151	EQQRFRTKEKERERQAKLAEEKMRKEEEEAKKRAEDDAKKK
P45378 (TNNT3)	116-155	EQQIRIAEKERERQNRLEAEKARREEDAKRRAEDDLKKK
Q16181 (SEPT7)	366-419	KEKVQKLDSEAEQRRHEQMKQKLEAQHKLEEKRRQFDEKANWEAQQRILE
Q15811 (ITSN1)	630-662	LKQKEQERKIIIELEKQKEEAQRRQERDKQWLE
	668-706	DEHQRPRLHEEEKLKRRESVKKKGEEKGKQEAQDKLG
Q9NZM3 (ITSN2)	671-737	KLKEIERKRLLEMQKQKLEDEAARKAKQGKENLWKENLRKEEEEKQKRLQEEKTQ EKIQEERKAEE
O95819 (MAP4K4)	378-480	EQQLREQEEYKQRLLAERQKRIEQQKEQRRRLEEQQRREREARRQEREQRRREQ EEKRRLLEELEERRRKEEERARRAEKRRVEREQEYIRRQLEEEQRHLE
Q9UKE5 (TNIK)	364-459	RSEALRRQOLEQQQRENEEHKQRLLAERQKRIEQQKEQRRRLEEQQRRKELRKQ QEREQRRHYEQMRREERARRAEHEQEYIRRQLEEEQRQLE
Q8N4C8 (MINK1)	396-467	RRIEEQKEERRRVEEQRREREQRKLQEKQRRLEDMQALRREERRQAEREQE YKRKQLEEQRQSERLQR
Q86SQ0 (PHLDB2)	1034-1097	RIEMERLLKQAHAEKTRLLESREREMEAKKRALEEKRRREILEKRLQEETSQR QKLIKEVK
Q08AD1 (CAMSAP2)	1197-1237	KQQLEAEMEHKKEETRRKTEERQKKEDEARREFFIRQYIM
Q9Y6V0 (PCLO)	1212-1251	KPLPEEKKLIPEEEKIRSEKPKLLEKKPTPEDKLLPE
Q16643 (DBN1)	173.238	KRINREQFWEQAKKEEELRKEEERKKALDEBLRFQERMEQERQEQEERERRRYE REQQIEEHRK
Q9UDT6 (CLIP2)	413-485	KLQARLLVSVRKEKVDLSNQLEEERRKVEDLQFRVEEESITKGDLETQTQLEH ARIGLEQSLLEKQAE

4.5. ábra. Néhány detektált SAH domén pozíciója és szekvenciája

ládjába, a septinek közé tartozik. A septinek szálatokat és gyűrűket képeznek, melyek állványzatként és diffúziós határolóként működnek. Részt vesznek többek közt a dendritikus tüskék kifejlődésében és a sejtosztódás folyamatában is [34]. A septin 7-ben a prediktált SAH domén a C-terminális közelében található, a fehérje szuperhélix régiójában. A septinek hexamereket alkothatnak, melyekben a septin 7 egy terminális fehérje (4.6. ábra). Feltételezhető, hogy a septin 7 SAH doménje a hexamerek egymás közötti interakcióinak irányításában játszik szerepet. A septin 7-en kívül egyik másik septin fehérjében sem találtunk SAH domént.

Az interszeptinek citoszekeletális állványfehérjék melyek az endocitózisban vesznek részt és vélhetően szerepet játszanak az autizmus és a Down-szindróma kialakulásában is [23]. A prediktált SAH domén a középső szuperhélix N-terminális felé eső részében található. Ez a szuperhélix régió többféle fehérjével is képes interakcióba lépni, ezért a feltételezett SAH domén szerepe is többféle lehet.



4.6. ábra. A: A 13 humán szeptin fehérje 4 csoportja.
 B: A SEPT2–SEPT6–SEPT7 hexamer komplex szerkezete. A prediktált SAH domén a terminális SEPT7 fehérjén található (az eredeti ábrán szuperhélixként kerül ábrázolásra). [34]

Az FT_CHARGE algoritmus a CLIP2 (CAP-Gly domain-containing linker protein 2) fehérjében is detektált egy SAH szegmenst, melyet a PANTHER analízis a citoskeletális összetevők közé sorol. A fehérje vélhetőleg az organelumok transzlokációjának agyra jellemző folyamataiban játszik szerepet [12]. A prediktált SAH domén az első szuperhélix régióban található a két CAP-Gly domént követően. Sajnos a CAP-Gly domén régiók kivételével nem találtunk a fehérjére vonatkozó, kísérletileg megállapított szerkezeti információt. Érdekes, hogy az FT_CHARGE algoritmus több más fajhoz tartozó CLIP2 fehérjében is detektál SAH régiókat, viszont az algoritmusok konszenzusában ezek nem jelennek meg.

A sejtkereg a közvetlenül a plazma membrán alatt található régió, általában aktin filamentumokat és a hozzájuk tartozó fehérjéket tartalmaz [7]. A PANTHER elemzés 4 SAH domént tartalmazó fehérjét azonosított a sejtkeregben, ezek mindegyike a citoszekeleton része is.

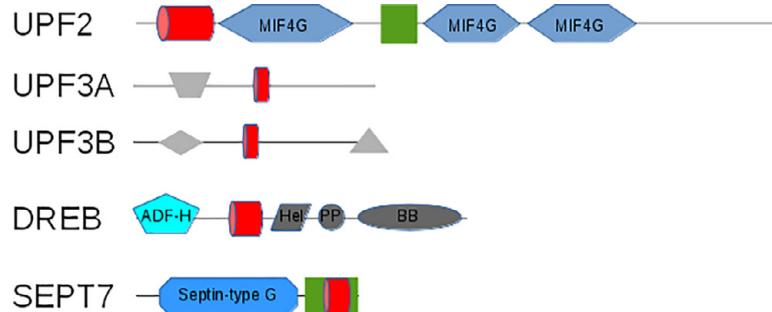
4.2.2. Idegi jelátvitelben résztvevő fehérjék

Habár a SAH domént tartalmazó fehérjék nem jelentek meg nagy számban a neurális folyamatokban résztvevő fehérjék között, egyes találatoknak mégis köze van a neurális folyamatokhoz. A citoskeletális (pl. CLIP2, miozin 6, caldesmon) és a sejtkeregben található fehérjék (piccolo, drebrin) részt vesznek a sejtfüggelékek, mint például a dendritek fejlődésében és a szinaptikus vezikulák kibocsátásában. A SAH domén keresés során 17 olyan fehérjét találtunk, amely a SynptomeDB-ben is megtalálható, illetve 13 fehérje UniProt annotációja arra enged következtetni, hogy ezek funkciója is kapcsolatba hozható a neurális fejlődéssel (4.4. ábra). Például a PHLDB2 (Pleckstrin homology-like domain

family B member 2) részt vesz a neuromuszkuláris szinapszisokban található posztszinaptikus komplex kialakításában [30], míg a CAMSAP2 (Calmodulin-regulated spectrin-associated protein 2) fehérje a mikrotubulusok lassú (mínusz) végéhez köt és részt vesz a neurális polaritás szabályozásában [56]. A SAH domén pontos szerepére vonatkozóan nem találtunk ide vonatkozó kutatási eredményeket.

A PCLO (protein piccolo) fehérje a preszinaptikus aktív zónákban található és a szinaptikus vezikulák szállításában vesz részt [15]. A prediktált SAH domén régiója szuperhélixként került azonosításra és felhasználásra a fehérje szerkezeti modelljének megalkotása során. Habár a szóban forgó szerkezeti elem megfelelt a helicitás filterben támasztott elvárásainknak, a régió magas prolin tartalma miatt megkérdőjelezhető, hogy valóban egy egyenes helikális szerkezetet képez. A régió a COILS predikció [28] alapján két szuperhélix között helyezkedik el és vélhetően zsanérként funkcionál a szuperhélixek között. Ettől a feltételezéstől eltekintve, a vélt SAH domén pontos szerkezeti szerepe nem ismert.

A drebrin egy posztszinaptikus fehérje, melynek a dendritekben található citoskeletonális állványzat formálásában van szerepe és a hosszútávú potenciáció keretében történő szinaptikus változásokkal, illetve az Alzheimer-kórral hozható összefüggésbe [43]. A feltételezett SAH domén ez esetben is egy szuperhélix régióban található. Ez a szuperhélix régió egy szomszédos helikális szegmensevel az aktin kötésében vesz részt, ezért egy drebrin molekula összeköttetést tud létesíteni két aktin filamentum között. Ebben a helyzetben a SAH doménnek vélhetően az összekapcsolt aktin filamentumok közötti távolságtartás a szerepe.



4.7. ábra. A kiválasztott SAH domént tartalmazó fehérjék sematikus ábrázolása. Piros hengerek: SAH domének, kék: UniProt annotációval rendelkező domének, szürke: egyéb régiók. A zöld színnel az annotált szuperhélixek kerültek jelölésre, melyek vagy nem vagy csak részlegesen fednek át a prediktált SAH régiókkal [11].

4.3. CoNSEnsX+ webszerver elkészítése

A mágneses magrezonancia spektroszkópia (Nuclear Magnetic Resonance - NMR) fejlődésével lehetővé vált a fehérjeszerkezetek dinamikai tulajdonságainak feltérképezése. Az adat fázisban történő mérés során a vizsgált fehérjék egy természetes állapotukhoz hasonló közegben vannak (más mérési módszerekhez képest, mint például a röntgen kristallográfia), ami lehetővé teszi a szerkezetek természetes mozgásainak több időskálán történő megfigyelését, mérését. A dinamikai paraméterek megjelenésével azonban a fehérjék szerkezeti tulajdonságairól alkotott kép is megváltozott: egy fehérje szerkezeti tulajdonságai

jellemzően nem írhatóak le jól egy adott konformációval. Sokkal inkább a lehetséges konformációs állapotok és az ezek közti állapotátmenetek adnak csak átfogó képet a vizsgált fehérje szerkezeti felépítéséről. A kísérlet során az NMR spektrométerbe helyezett hozzávetőleg 10^{16} - 10^{17} konformer kiátlagolódtott tulajdonságai kerülnek mérésre, ezért nem reális elvárás, hogy egy adott konformer teljesítse a mért NMR-es paramétereket, hiszen ezek eleve egy sokaság átlagát jelentik. Fontos ezért, hogy a dinamikát reprezentáló sokaságok esetén ne a sokaság egyes elemeivel várjunk jó megegyezést, hanem a konformereket sokaságként kezelve, ezek átlagának a megfelelését kell vizsgálni.

A dinamikus fehérjeszerkezeti sokaságok molekuladinamikai szimulációval történő előállítása alapvetően kétféleképpen történhet: csak egy kiválasztott erőter felhasználásával és az erőteret megszorító/kiegészítő, kísérletekből származó mért adatok felhasználásával. A szimuláció során használt algoritmusok gyengepontjai és a paraméterezési hibatalehetőségek miatt mindig érdemes ellenőrizni az így előállított sokaságok helyességét. Ennek egyik módja a mért adatokkal (NMR paraméterek) történő összevetés. Amennyiben egy szerkezethez több dinamikai paraméter is rendelkezésre áll, vagy a paraméterek független kísérletekből származnak, a paraméterek egy csoportja felhasználható a molekuladinamikai szimuláció során megszorító paraméterként, a másik csoportja pedig az elkészült sokaság keresztvalidációjához.

The screenshot shows the CoNSENsX+ web interface in a Mozilla Firefox browser. The page title is 'CoNSENsX+ - Mozilla Firefox' and the URL is 'consensx.itk.ppke.hu'. The main heading is 'Compliance of NMR-derived Structural Ensembles with experimental data + selection'. The interface is divided into two main sections: 'START CALCULATION' and 'INFO'.

START CALCULATION

- Ensemble: Upload PDB file... ?
- NOE restraints: Upload NMR-Star file... ?
- NMR parameters: Upload NMR-Star file... ?
- BME weights: Upload BME weights... ?
- Karplus equation parameter set: Wang & Bax 1996 ?
- Fit models in PDB
 - on range: 1-72 (optional) ?
- Use r^{-3} averaging between models (NOE) ?
- Use SVD for RDC back-calculation ?
- RDC LC model: bicelles ?

Buttons: Reset form, Ready and fire!

INFO

The server is specifically designed for structural ensembles that were generated to reflect the internal dynamics of proteins at a given time scale. Such ensembles are typically generated by restrained molecular dynamics methods or selection-based approaches. CoNSENsX+ provides an ensemble-averaged analysis of all experimental parameters recognized in the input and offers a simple greedy selection approach to identify the sub-ensemble best reflecting the parameters chosen.

USAGE

Input a multi-model PDB file, a BMRB format NOE distance restraint file (optional), V2 NMR Restraints at the Protein Data Bank and a BMRB restraint file to check the accuracy of the protein structural ensemble against experimental data. For a detailed description, see the [description page](#). For reference on accepted file formatting, please consult the [downloadable test data](#). You can run a test calculation on the test data anytime with the provided test data, to check if the server is working.

CITING

CONTACT

4.8. ábra. A CoNSENsX⁺ szerver nyitó oldala.

A kísérleti adatokkal történő összevetés során szem előtt kell tartani, hogy azok egy átlagot reprezentálnak. Ezért a vizsgált konformereket is sokaságként kell kezelni, mivel csak így van értelme a mért NMR-es paraméterekkel történő összehasonlításnak. Az ilyen

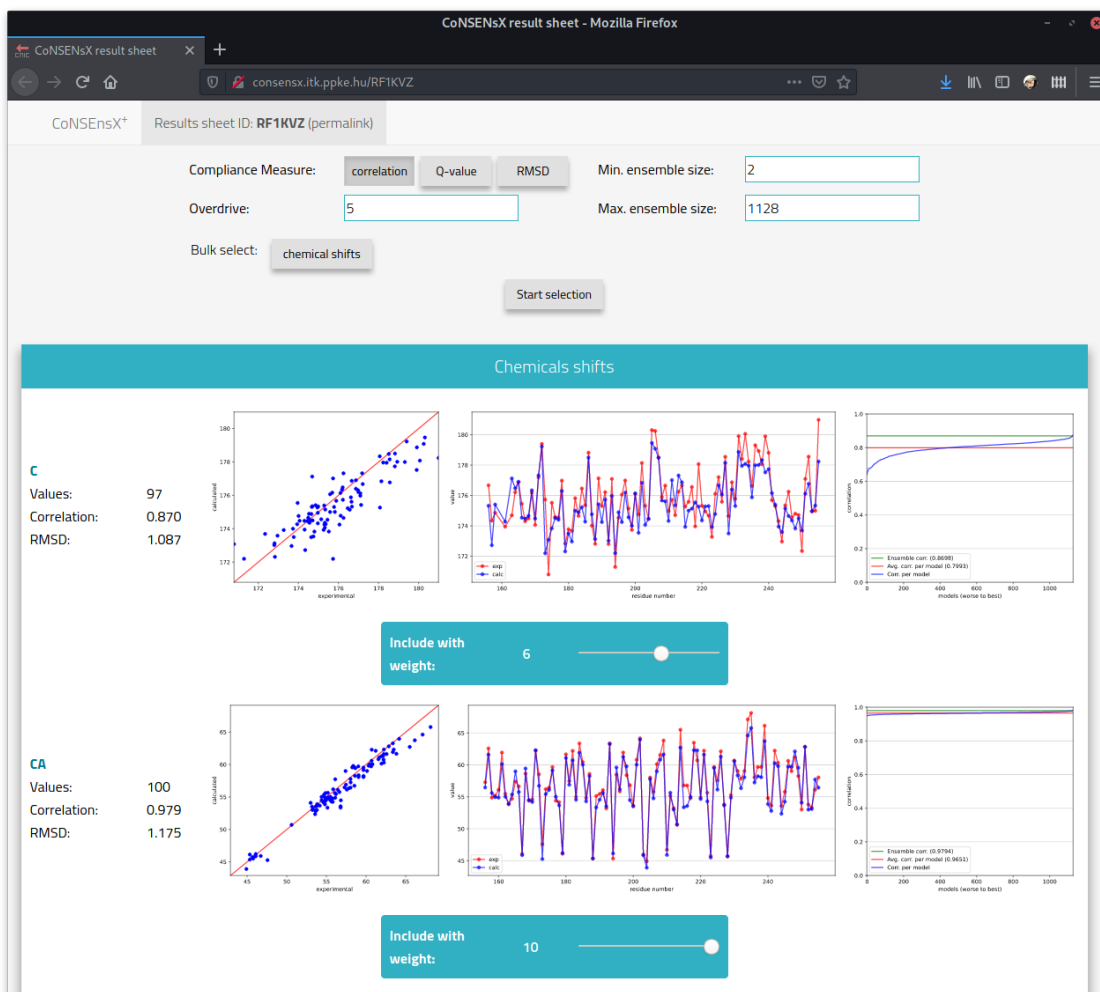
ellenőrzések rutinszerű futtatására jött létre az eredeti CoNSEnsX szerver [61], melynek továbbfejlesztésével foglalkoztam.

A hosszabb molekuladinamikai szimulációval készült sokaságoknak jellemzően a mérete (bennefoglalt konformerek darabszáma) is nagyobb. A nagyobb sokaságok kezelése technikailag is nehezebb, illetve általában egy túlllesztett állapotot reprezentálnak. Felmerült az igény ennek a túlllesztettségnek a csökkentésére, ami egyidejűleg a sokaságban foglalt konformerek számának csökkenésével és a mért NMR-es paraméterekkel való egyezés növekedésével is járhat. A szelekció tovább hangolható azzal, ha csak a számunkra érdekes NMR paraméterekkel történő megfelelést szeretnénk optimalizálni akkor is, ha a sokaság készítése során egyéb megfontolásokból más megszorító paramétereket is alkalmaztunk. A szimulációval előállított sokaságok szelekciójának kimenetele egy, az eredetinél kisebb szelektált sokaság, amely ideális esetben jobb egyezést mutat a kiválasztott NMR-es paraméterekkel. Amennyiben ez a feltétel teljesül, akkor sikerült egy olyan minimális elemszámú sokaságot előállítani, ami ugyanúgy vagy jobban reprezentálja az adott struktúra dinamikai tulajdonságait, mint a jellemzően sokkal nagyobb elemszámú kiindulási sokaságok.

A dinamikus fehérjeszerkezeti sokaságok rutinszerű elemzése magával hozta az ehhez használt saját fejlesztésű programok és szkriptek fejlesztését, integrálását. Az eredeti CoNSEnsX szerver algoritmusait kiindulási alapként felhasználva valamennyi funkciót újrainplementáltam Python könyvtárak formájában. Ezen kívül több új, technikai jellegű fejlesztés is került a szerverbe, melynek forráskódja szabadon elérhető és felhasználható (<https://github.com/PPKE-Bioinf/consensx.itk.ppke.hu>).

Az egyik ilyen technikai jellegű fejlesztés a PDB és STAR-NMR formátumú fájlok kezelése az erre a célra létrehozott Python könyvtárakkal, így a bemeneti fájlok beolvasásánál és a bennük tartalmazott információ kinyerésénél sok hibalehetőség elhárult. Valamennyi megjelenített grafikon/ábra vektorgrafikus és felhasználható publikációkban további szerkesztés nélkül. A szerver adatbázisban tárolja a lefutott elemzések paraméterezését és eredményeit, melyek így később a futtatások azonosítóján keresztül visszakereshetők, így a lefutott elemzések eredményei online megoszthatóak és újrafelhasználhatóvá váltak.

A parancssoros felületen elérhető továbbá két, a peptid kötések planaritását ellenőrző funkció, melyek a peptid kötések Ω (CA-N-C'-CA) és Ω H (H-N-C=O) torziós szögeinek eloszlását jelenítik meg. A peptid kötés planaritása egy ismert jelenség, ezért a túl nagy torziós szögek a peptid kötések mentén azt jelzik, hogy vagy már a kiindulási modell is hibás volt, vagy a molekuladinamikai szimulációban alkalmazott megszorító paraméterek túl erősen kerültek alkalmazásra és ezzel deformálták a molekula természetes(nek vélt) geometriáját. A kiszámított torziós szögek hisztogramszerűen, sávokba kategorizálva kerülnek reprezentálásra (<2, 2-5, 5-10, 10-20, >20 fok). Amennyiben a magasabb tartományokban sok előfordulás található, az jó indikátora annak, hogy a szerkezettel valami nincs rendben, megfelelő kiindulási szerkezetet feltételezve ez általában azt jelenti, hogy az alkalmazott kényszerfeltételek "kicsavarják" a planáris peptid kötések, ami valószerűtlen szerkezeteket eredményez.



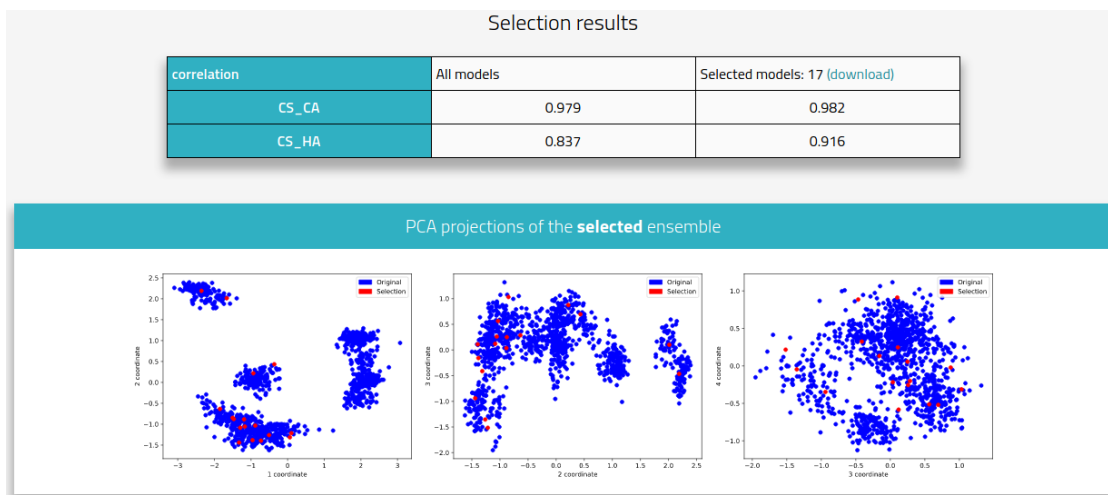
4.9. ábra. A szelekció paraméterezése CoNSENsX⁺ szerveren. Az ábrán a 2LC7 PDB szerkezettel futatott molekuladinamikai szimulációból visszszámolt kémiai eltolódás értékek látszanak (részlet).

4.3.1. Visszaszámolható paraméterek

A feltöltött STAR-NMR formátumú fájlokból a CoNSENsX⁺ szerver a következő paramétereket tudja az elemzéshez felhasználni:

- Maradvány dipoláris csatolások (RDC - Residual dipolar couplings)
- $^3J_{H_N H_\alpha}$, $^3J_{H_N C_\alpha}$, $^3J_{H_N C_\beta}$ és $^3J_{H_N C}$ skaláris csatolások (választható Karplus egyenlet paraméterezéssel [20, 52])
- Amid N, H, H α , C α , and C β kémiai eltolódások
- Molekulagerinc és oldallánc S² rendparaméterek
- NOE távolságértékek

A kémiai eltolódásokat a (C-ben írt) ShiftX programmal [36] számoljuk vissza. Bár a program második verziója (a Java-ban írt SHIFTX2) már rendelkezésre áll, ennek lassúsága nagyon megnyújtaná a paraméterek visszszámolásához szükséges időt. A maradvány



4.10. ábra. A szelekció kimenete CoNSEnsX⁺ szerveren. Az eredeti, 1120 szerkezetből álló sokaságból (2LC7 szim.) 17 modell kiválasztásával a megadott kémiai eltolódás korrelációk jelentősen javulnak. Az ábrákon a kiindulási és a szelektált sokaságok főkomponens analízise látható.

dipoláris csatolások visszaszámolása a PALES programmal történik, az SVD (szinguláris érték felbontás - Singular Value Decomposition) alapú illesztéssel. Ilyenkor a PALES minden egyes konformer esetében meghatározza azt az orientációt, ami az adott konformerből visszaszámolt és a kísérleti értékek legjobb illeszkedését eredményezi. Ezen eljárás mellett szól, hogy a valóságban is változó lehet az RDC-méréshez használt közegben (oriented media) az egyes eltérő szerkezetű konformációk mágneses térhez képest való orientációja. Ellenérvként megemlíthető, hogy így viszont előfordulhat, hogy ezen illesztés miatt "túl jó" megfeleléseket kapunk. Ezért ez az illesztés a szerverben kikapcsolható, ilyenkor a PALES minden egyes konformáció esetében megbecsüli a szerkezet orientációját a sztérikus faktorok figyelembevételével. A CoNSEnsX⁺ szerver felismeri a bemeneti fájlból az egy kísérlet keretében megállapított RDC paramétereket és ezeket együtt kezeli, a számítás kimenetében ezek RDC csoportokként jelennek meg. Ezen kívül az S² rendparaméterek miatt lehetséges a feltöltött sokaság illesztése, alapértelmezetten a teljes szekvencia mentén, de a felhasználó által saját tartomány is definiálható. Az illesztésre azért van szükség, mert a rendparaméterek csak akkor értelmezhetőek megfelelően, ha az egyes modellek egymáshoz képest való rotációját kiküszöböljük, más szóval, csak a molekulán belüli, lokális elmozdulásokat vesszük figyelembe.

4.3.2. Oldallánc S² rendparaméterek

Az oldalláncokra vonatkozó S² rendparaméterek visszaszámolása egy újonnan megvalósított funkciója a CoNSEnsX⁺ szervernek. A számításban szerepet kapó atompárok egy előre definiált táblázatban kerülnek eltárolásra:

```

sc_lot = {
    "VAL": {"CG1": "CB", "CG2": "CB"},
    "ILE": {"CG2": "CB", "CD1": "CG1"},
    "THR": {"CG2": "CB"},
    "LEU": {"CD1": "CG", "CD2": "CG"},
    "ALA": {"CB": "CA"},
    "MET": {"CE": "SD"},
}

```

3. forráskód szegmens. *Oldallánc rendparaméter táblázat*

tehát például egy oldalláncon található izoleucin "CG2" atom S^2 számításbeli párja ugyanazon aminosav "CB" atomja. Mivel a két metilcsoportot tartalmazó aminosavakhoz (Val, Leu, Ile) több visszazámolt érték is tartozhat a visszazámolandó rendparaméterek kategorizálatlansága miatt (valamennyi oldallánc rendparaméter érték egyszerre kerül visszazámolásra és ábrázolásra), minden, az adatok eltárolására és ábrázolására szolgáló funkcionalitást újra kellett implementálni oly módon, hogy az egy aminosavhoz tartozó különböző értékek megvalósíthatóak legyenek. A visszazámolt értékeket tartalmazó táblázatban ennek megfelelően egy aminosav pozícióhoz több oszlop, az adatok megjelenítése során egy x-tengelyen lévő aminosav pozícióhoz több y érték tartozhat. Az ugyanazon aminosavhoz tartozó értékek az ide vonatkozó grafikus ábrázoláson egymáshoz képest kis x-tengely menti eltolással kerülnek ábrázolásra.

4.3.3. A megfelelések jellemzése

A mért és a visszazámolt paraméterek egyezését a Pearson korrelációval és RMSD (négyzetes középérték - Root Mean Square Deviation) értékkel, RDC paraméterek esetén ezen felül a Q-értékkel jellemzi.

Az NOE távolságértékek megsértése távolságú alapú eloszlás grafikonon kerül vizualizálásra, a többi visszazámolt paraméterhez több ábra is tartozik:

- a mért és a visszazámolt értékek korrelációja
- a mért és a visszazámolt értékek egymás melletti összehasonlítása aminosav sorrendben
- A visszazámolt értékek modellenkénti értéke, ezek átlaga, illetve a sokaságra vonatkozó megfelelési érték

A Kresten Lindoff-Larsen csoportja által kifejlesztett BME (Bayesian Maximum Entropy) eljárás egy biomolekuláris szerkezeti sokaság tagjaira megadott egyedi paraméterek és a rendelkezésre álló kísérleti adatok segítségével egy olyan súlyozást készít, amely a bemeneti sokaság egyes tagjainak a paraméterek mért értékeihez való hozzájárulását jellemzi, és azokkal a lehető legjobb megfelelést adja. A BME eljárás azonban egy parancssorból használható, grafikus felület nélküli program, amely speciálisan előkészített bemenetet igényel. A CoNSEnsX⁺ eljáráshoz hasonló felhasználása miatt úgy gondoltam,

hogy érdemes a BME használatát megkönnyítő támogatást is beépíteni a szerverbe. A megvalósítás során a következő funkciókat építettem be:

- A beadott BMRB file alapján a szerver előállít egy olyan állományt, amely a BME számára bemenetként használható, azaz az egyedi konformerekre visszazámolt paramétereket azzal kompatibilis formában tartalmazza
- A szerver képes arra, hogy a BME kimenetként előállított, az egyes konformerek relatív súlyát tartalmazó fájlt beolvassa és a sokaság értékelése során ezeket a súlyokat figyelembe véve elemezze a sokaság megfelelését

4.3.4. Szelekció beépítése és értékelése

A molekuladinamikai szimulációval előállított sokaságok jellemzően sok konformert tartalmaznak, melyeket dinamikai sokaságként kezelünk, tehát a mért paramétereknek történő megfelelést a teljes sokaságra várjuk el. Itt érdemes megfigyelni, hogy sok esetben a sokaság megfelelése a mért paramétereknek jobb, mint a sokaságban lévő legjobb modell megfelelése. Felmerül viszont annak az esélye, hogy a számunkra érdekes dinamikát a szükségesnél sokkal több modellel írjuk le (túlillesztés), így sok olyan konformer is része a sokaságnak, ami a dinamikai leírás szempontjából nem szignifikáns, emellett a nagyobb sokaságokon az elvégzendő vizsgálatok is sokkal számításigényesebbek. Felmerült tehát az igény, hogy a szimulációval előállított, jellemzően sok modellt tartalmazó sokaságban megfigyelt mozgásokat a lehető legkevesebb modellel írjuk le úgy, hogy mellette maximalizáljuk a szelektált sokaság mért adatokkal történő megfelelését.

A CoNSEnsX⁺ szerverben implementáltam egy szelekciós algoritmust, amely lehetővé teszi a feltöltött sokaságból egy alsokaság létrehozását, a mért paramétereknek történő megfelelés szinten tartása vagy javítása mellett. A felhasználó kiválaszthatja és súlyozhatja a szelekció során a számára érdekes mért paraméteret, illetve a megfelelés mérőszámát (korreláció, RMDS, RDC paraméterek esetén Q-érték). A szelekciós algoritmus ezen a kiválasztott paraméterek mentén keresi azt a legkisebb sokaságot, ahol a kiválasztott paraméterek és súlyozás mellett maximális a kísérleti adatokkal történő megfelelés.

A szelekció paramétereit:

- a megfelelés mérőszáma (korreláció, Q-érték, RMSD)
- minimális szelektált sokaság méret
- maximális szelektált sokaság méret
- "overdrive" - ennyi iterációval több modell hozzáadását vizsgálja meg az algoritmus a lokális minimumok áthidalása céljából

A implementált szelekció egy mohó algoritmus, az első lépésben megkeresi célfüggvény szerint legjobban egyező egyedi konformert (S^2 estén a legjobb konformer párt), majd ezután megkeresi azt a következő konformert, amit ehhez a kiindulási sokasághoz adva a legjobban növekszik a célfüggvény szerinti egyezés. Ez az iteráció addig folytatódik, amíg:

- a szelektált sokaság mérete el nem éri a kívánt minimum méretet
- a szelektált sokaság mérete a kívánt maximum méreten belül van és
- a célfüggvény szerinti egyezés javul
- vagy az "overdrive" paraméterben meghatározott határértékig az algoritmus megpróbál egy lokális minimumot áthidalni
- (vagy ha az "overdrive" paraméterben meghatározott határértékig történő iteráció közben a szelektált sokaság eléri a kívánt maximális méretet)

A szelekció végeztével megjelenítésre kerülnek a szelektált sokaságra vonatkozó mérőszámok és letölthetővé válik a szelektált sokaság. Emellett a szelektált sokaság PCA analízissel is összehasonlításra kerül a kiindulási sokasággal, aminek vizualizálása a szelekció után láthatóvá válik.

A determinisztikus mohó algoritmus vélhetően nem érzékeny a konformerek pontos eloszlására a kiindulási sokaságban, amíg az kellőképpen lefedi a referencia sokaságot. Mivel az algoritmus csak a megadott megfelelési mérőszám optimalizálására törekszik, a sokaság modelljeinek konformációs térbeli eloszlásának nincs szerepe a számítás során. A szelektált sokaságról nem garantálható, hogy ez a bemeneti paramétereknek legjobban megfelelő alsokaság, hiszen ehhez valamennyi modell kombinációt figyelembe kellene venni, viszont megfelelő paraméterezés mellett az algoritmus a bemeneti paramétereknek kielégítően – a bemeneti sokaságnál általában jobban – megfelelő, kis létszámú alsokaságot generál. Ezért a generált alsokaság a bemeneti sokaság tagjaival – uniform súlyozás esetén – elérhető legjobb megfelelés alsó becslésének tekinthető.

4.3.5. A szelekció tesztelése molekuladinamikai szimulációval előállított sokaságokkal - Diskusszió

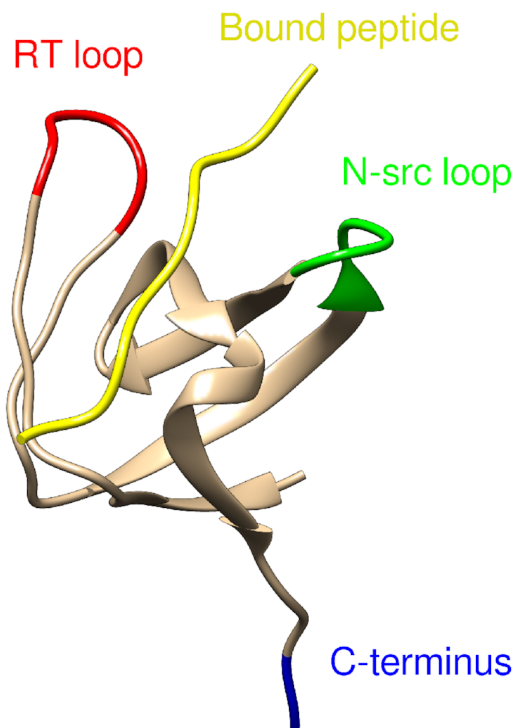
A szelekciós algoritmus tesztelése során is szem előtt tartottuk, hogy egy dinamikus szerkezeti sokaság esetén nem elvárt, hogy az egyes modellek külön-külön megfeleljenek a kísérletileg megállapított értékeknek, hanem a sokaság által reprezentált konformációs tértől várjuk el ezt a megfelelést. Ezért a tesztelést és a túlillesztettség becslését a különböző sokaságokból mintavételezett konformációs terek összehasonlításával végeztük.

A Drk SH3 domén elemzése

A Drk fehérje (PDB ID: 2A35) N-terminális SH3 doménjének felhasználásával molekuladinamikai szimulációkat futtattunk implicit oldószer modellt használva, a GBSA algoritmusmal [38]. A 100ns hosszú szimuláció 1fs-os lépésküzzel, 300K hőmérsékleten, implicit oldószeres környezetben került lefuttatásra. A szimuláció utolsó 20ns hosszú tartományából készült a referencia sokaság, melyből a teszteléshez felhasznált paraméterek (kémiai eltoldódások, 5 különböző RDC készlet és molekulagerinc oldallánc S^2 rendparaméterek) kerültek kiszámításra. A kémiai eltoldódások és az RDC készletek felhasználásával készítettünk egy szelektált sokaságot a túlillesztettség becslésére, azaz hogy meg lehet-e felelni az előbb kiszámolt paramétereknek kevesebb modell alkalmazásával is.



4.11. ábra. Az illesztett referencia sokaság



4.12. ábra. Ligandumot (sárga) kötő SH3 szerkezet (PDB ID: 4LNP)

A referencia sokaságtól függetlenül generáltunk egy nagyobb konformációs teret lefedő sokaságot, az AMD séma [54] felhasználásával, ami a csoportban a GROMACS szoftverhez előzőleg implementálásra került [16]. Négy párhuzamos, 20ns hosszú szimuláció került lefuttatásra ugyanazzal a kiindulási szerkezettel, de eltérő kezdeti atom sebességekkel, $E = 2400$ kJ/mol AMD "boost potential"-lal és $\alpha = 0,5$ gyorsítási faktoral. A szimulációk utolsó 18,75 ns-át felhasználva készült a 3044 konformerből álló "AMD sokaság". Ebből a sokaságból szelekcióval különböző alsokaságokat állítottunk elő, a referencia sokaságból visszazámolt paraméterek kombinációinak felhasználásával. A szelekciós alsokaságok által lefedett konformációs tér becslését főkomponens elemzéssel végeztük.

A referencia sokaságban (4.11 ábra) az első PCA (principal component analysis - főkomponens elemzés) módusban a $\beta 2$ - $\beta 3$ redőket összekötő hurok (N-scr hurok) hajló mozgása figyelhető meg [8], amely a kötőzsebre hajló hurkok egyike (4.12 ábra). A második módusban a C-terminális három aminosavjának heterogentiája dominál, ez funkcionálisan nem tartozik a releváns mozgások közé.

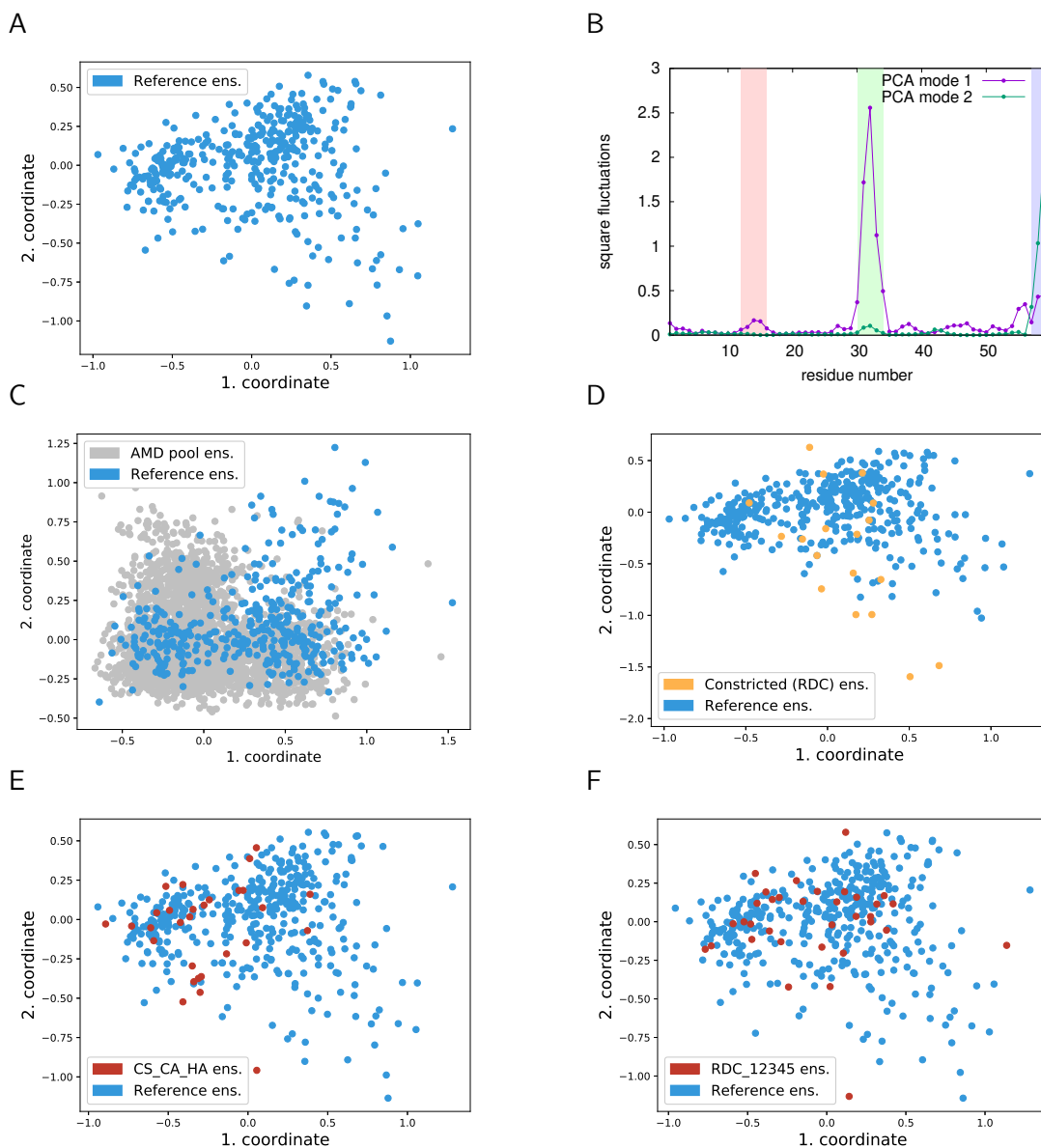
Minden szelektált sokaságban az első főkomponens a referencia sokaságban is megfigyelt első főkomponensbeli mozgás, emellett a sokaságok megfelelése a kiszámolt dinamikai paramétereknek kielégítő maradt. Eszerint tehát a referencia sokaságban megfigyelt mozgások kisebb sokaságokkal is jól reprezentálhatóak, ami egyúttal a referencia sokaság túlillesztettségére is utal, tehát a referencia konformációs teret magában foglaló konformációs térből szelekcióval elő lehet állítani a referencia sokasághoz hasonló sokaságokat (4.13 ábra C-F).

A Pin1 peptidil-prolil cisz–transz izomeráz elemzése

A szelekciós algoritmus a Pin1 peptidyl-prolyl cisz–transz izomerázzal is tesztelésre került, mivel a fehérje egyik, a szubsztrát kötő oldalán található aldoménekre ható módusa vélhetően hatással van a kötés katalizációjára és annak szabályozására [51, 11]. A 2RUC szerkezetet kiindulási alapnak véve, három különböző molekuladinamika szimulációt végeztünk, különböző AMD paraméterezéssel. Az így készült sokaságokra megvizsgáltuk a hozzájuk tartozó kémiai eltolódásokkal (BMRB ID: 11559) történő egyezést és a szelekciós algoritmussal a kémiai eltolódásokat optimalizálva alsokaságokat készítettünk. Habár mindegyik sokaság jól megfelelt a kísérletileg mért adatoknak, az AMD sokaságok megfelelései jobbak voltak, mint az edetei 2RUC sokaságé, ami alátámasztja, hogy egy-egy változatosabb sokaság könnyebben megfelel a mért paramétereknek. A szelektált sokaságokban ez az egyezés még jobb volt, úgy is, hogy a sokaságok csak 5-15 konformerből állnak, a megfeleléségi mérőszámtól függően (korreláció vagy RMSD).

Habár a különbségek kicsik, a leghosszabb AMD-s szimulációból szelektált alsokaságok megfelelése a legjobb. A főkomponens elemzés kimutatta, hogy az ebből képzett alsokaságok olyan konformereket is tartalmaznak, melyek csak ezen kiindulási sokaság konformációs terében találhatók meg. Ezen megfigyelés jelentőségének a megbecslésére megvizsgáltuk, hogy a szelektált alsokaságokban megjelenik-e valamilyen biológiailag releváns mozgás. Az alsokaságokban megjelenő mozgásokat már azonosított (rotamáz domének) mozgási módusokkal [11] vetettük össze. A szerkezeti illesztés a különböző rotamáz domé-

nekben közös 52 CA atom mentén történt. A főkomponens elemzés kimutatta, hogy a szelektált sokaságok első PCA módusa átfed a funkcionálisan legrelevánsabbnak vélt mozgással, mely a ligand kötő oldal "nyitó-csukó" mozgása. Ez a felfedezés alátámasztja a megfelelő konformációs mintavételezés jól ismert fontosságát és rámutat, hogy a kísérleti paramétereknek történő megfelelés kis javulása is releváns különbséget jelenthet. Ebben és az ehhez hasonló speciális esetekben tehát a szelektációs algoritmus – megfelelő kiindulási sokaság mellett – olyan alsokaságok előállítására képes, amikben biológiailag releváns mozgások is megjelenhetnek.



4.13. ábra. *A: a referencia sokaság PCA ábrája
 B: az első 2 PCA módus négyzetes fluktuációi
 C: a referencia és az AMD sokaság PCA ábrája
 D: a referencia és a szelektált sokaság PCA ábrája
 E: a ref. és a kémiai eltolódásra szelektált sokaság PCA ábrája
 F: a ref. és az RCK-re szelektált sokaság PCA ábrája*

Az α -Szinuklein elemzése

Az algoritmust rendezetlen (IDP - intrinsically disordered protein) fehérjeszerkezetekkel való működését a pE-DB [50] adatbázis 9AAC α -szinuklein sokaságával is teszteltük. A sokaság paramágneses relaxációs (PRE - paramagnetic relaxation enhancement) kísérleti adatokból került előállításra, melynek visszaszámolása nincs implementálva a CoNSENsX⁺ szerverben. A sokasághoz készültek viszont kémiai eltolódás mérések (BMRB ID: 16300), ami alapján el tudtuk végezni a szelekciót, ráadásul olyan kísérletileg megállapított adatok felhasználásával, amik nem kerültek felhasználásra a sokaság előállításánál. A 37 konformerből álló szelektált sokaság PCA analízise kimutatta, hogy ez a kisebb sokaság is lefedi az eredeti sokaság konformációs variabilitását. A kiindulási és a szelektált sokaság másodlagos szerkezeti elemzése alapján a helikális szegmensek helyzete jól reprezentált a szelektált sokaságban is. Ez a megfigyelés alátámasztja, hogy a szelekciós algoritmus egy rendezetlen fehérjéből álló sokaság esetén is képes a releváns tulajdonságokat megőrizve egy kisebb sokaságot előállítani.

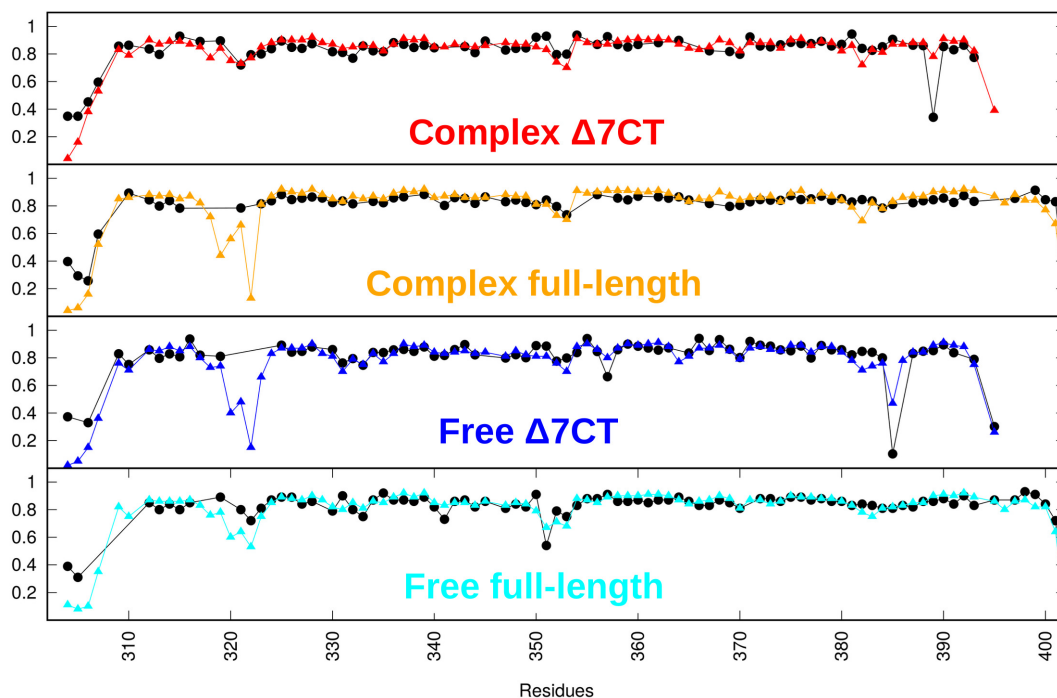
4.4. PDZ domének dinamikus szerkezeti sokaságainak összehasonlító elemzése

A domén allosztérikus tulajdonságaira fontos hatással vannak annak oldalláncai [40, 33], így ezek dinamikájának modellezése kulcsfontosságú a ligandum kötés során bekövetkező változások megértéséhez. A megszorító paraméterként használt, kísérletileg megállapított oldallánc rendparaméterek hatásának jobb megértéséhez molekuladinamikai szimulációkat futtattunk különböző megszorító paraméter kombinációkkal. Az így készült sokaságok főkomponens analízise kimutatta, hogy a csak a molekulagerinc, csak az oldallánccok illetve a kettő megszorító paraméter együttes alkalmazásból kapott sokaságok kissé eltérő konformációs teret fednek le (4.14. ábra). A kísérletileg megállapított paraméterekkel történő korreláció jelentősen alacsonyabb azoknak a sokaságoknak az esetében, ahol molekulagerinc vagy az oldallánc megszorító paramétereit elhagytuk, ezért a továbbiakban minden előállított sokaság esetén mindkét megszorító paramétert egyidejűleg alkalmaztuk.

Sokaság	Molekulagerinc S^2	Oldallánc S^2
Szabad állapot, teljes hossz (ff)	0,92	0,95
Szabad állapot, $\Delta 7CT$ (fd)	0,87	0,96
Kötött állapot, teljes hossz (cf)	0,95	0,97
Kötött állapot, $\Delta 7CT$ (cd)	0,83	0,95

4.3. táblázat. Az generált sokaságok molekulagerinc és oldallánc S^2 korrelációja

A generált sokaságok konformációs entrópiáját kiszámoltuk az oldallánc S^2 rendparaméterekre, ezen kívül a GROMACS programmal és a Schlitter-formulával is végeztünk entrópia számításokat. A számítások eltérő jellegéből adódóan nem várható el, hogy az általunk kiszámolt konformációs entrópia értékek jól megfeleljenek a kísérletesen megállapított adatoknak, mivel ezek gyakran különböző atomcsoportokat vesznek számításba. A számítások célja a kötési entrópia megállapítása, azaz a szabad és a kötésben lévő szer-



4.14. ábra. Kísérletileg megállapított (fekete) és visszszámolt (színes) S^2 értékek.

kezetek entrópia különbségének megbecslése a teljes hosszúságú és a $\Delta 7CT$ sokaságok esetén.

Valamennyi kombináció közül a szabad állapotban lévő $\Delta 7CT$ entrópiája volt a legnagyobb, amíg a többi sokaság entrópiája között egymáshoz képest nem nagyok az eltérések. Ez arra enged következtetni, hogy a $\Delta 7CT$ sokaság egy nagyobb entrópia változáson megy keresztül a kötés kialakulása során, mint a teljes hosszúságú sokaság. Az entrópia számítások eredményei megfelelnek a doménre vonatkozó kísérleti megfigyeléseknek [40]. A 20 ns-os szimulációkkal előállított sokaságok 4,1-es entrópia különbség aránya ($(S_{cd} - S_{fd}) / (S_{cf} - S_{ff})$) egy megfelelően jó egyezést mutat az izotermikus titrációs kalimetriával megállapított értékekkel (2,4) [40]. Ezekből kifolyólag a szimulációval előállított sokaságaink mind a kísérletileg megállapított rendparamétereknek, mind a konformációs entrópia változásoknak jól megfelelnek.

4.4.1. A sokaságokban megfigyelt dinamika elemzése

A szimulációval előállított 4 sokaság együttes vizsgálatához összefűztük ezeket egy közös sokasággá, amin főkomponens elemzést végeztünk. Mivel a módszer megköveteli, hogy a sokaság konformerei egyforma hosszúságúak legyenek, ezért a teljes hosszúságú sokaságot is a $\Delta 7CT$ sokaság méretéhez kellett igazítani, ami a C-terminális hélix levágását jelentette. Megfigyeltük, hogy a mozgékony N-terminális régió dominánsan megjelenik a főkomponens elemzés során, ezért ezt is eltávolítottuk a sokaságokból, hogy a vizsgált domén fontosabb belső mozgásai hangsúlyozottabban megjelenjenek. Ezután láthatóvá váltak a domén ligandum kötőhelyénél fellépő mozgások (4.15. ábra).

Az első főkomponens mentén főleg a $\beta 1$ - $\beta 2$ hurkok és az $\alpha 2$ hélix C-terminális régiójának nyitó-csukó mozgása figyelhető meg, melyek egymáshoz közel, a ligandum C-terminálisánál helyezkednek el. A második főkomponens mentén a $\beta 1$ - $\beta 2$ és a $\beta 2$ - $\beta 3$ hurkok mozgásai jelennek meg az $\alpha 2$ hélix és a $\beta 2$ redő együttes elmozdulásával (4.15. ábra narrow-wine irányok). A harmadik főkomponens mentén többek közt a $\beta 1$ - $\beta 2$ és a $\beta 2$ - $\beta 3$ hurkok mozgásai jelennek, a ligandum kötőzseb ellentétes oldalain.

Mivel a 8 replikával futtatott szimulációk 8 különböző trajektóriát eredményeznek, ezért fontos ellenőrizni, hogy az egyes replikák a nyitott és csukott állapotokat is érintik, vagy beragadnak ezen állapotok egyikébe. Megfigyeltük, hogy a szabad állapotban lévő $\Delta 7CT$ sokaság mindkét állapotot megközelíti a szimuláció során, ami alátámasztja, hogy ez a fajta mozgás valóban a domén saját dinamikus tulajdonsága.

Sokaság	Összefűzött	$\Delta 7CT$	Complex	cd	cf	fd	ff
Nyitott (Open)	53%	35%	10%	2%	18%	69%	97%
Szűk (Narrow)	54%	30%	50%	31%	69%	30%	87%
Nyitott és szűk együttes előfordulás	0,56	0,51	0,54	0,68	0,61	0,67	0,85

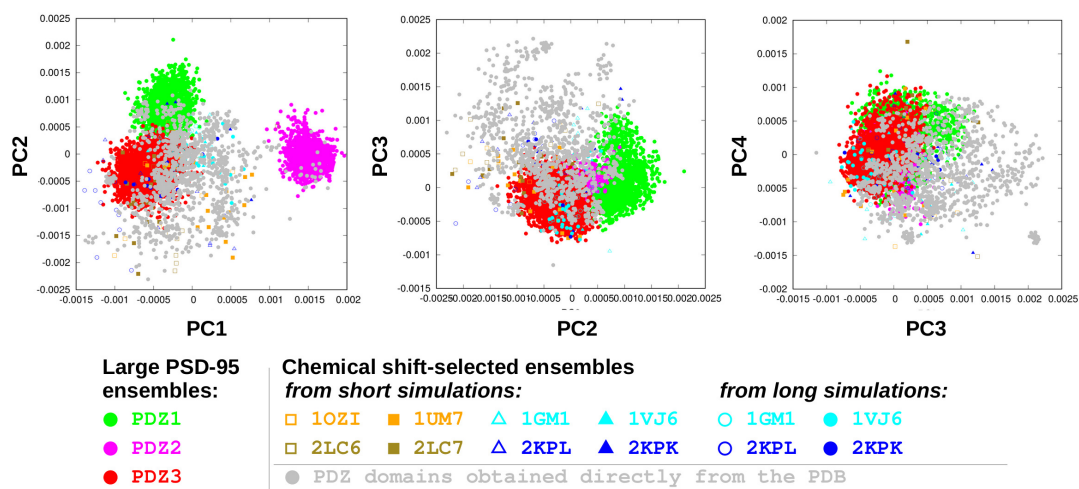
4.4. táblázat. *A sokaságok konformereinek főkomponens szerinti eloszlása. A nyitott-csukott illetve szűk-széles állapotok közötti határvonalat az első és a második főkomponens változó 0 értékénél határoztuk meg. Az együttes előfordulás értékek ezen állapotok átlapoltságát jelzik.*

Az összefűzött sokaság esetén a konformerek diverzitását szinte 40%-ban első főkomponens határozza meg (az első 10 komponenshez képest). A kötésben lévő szerkezetek többnyire a "csukott" állapotban vannak, amíg a szabad állapotban lévő sokaságok többsége a "nyitott" konformációs teret járják be. A rövidebb, $\Delta 7CT$ szerkezetek (levágott C-terminális felőli $\alpha 3$ hélix) a "nyitott" és "csukott" állapotokat egyaránt felveszik, amíg a szabad állapotban lévő teljes hosszúságú szerkezetek a "nyitott" állapotban maradnak. A "szűk" és "széles" állapotok közötti különbség látszólag a $\alpha 3$ hélix meglététől függ, a teljes hosszúságú szerkezetek többnyire "szűk" állapotban, amíg az $\alpha 3$ hélixet nélkülöző $\Delta 7CT$ szerkezetek a "széles" állapotban vannak, bár ezen különbségek nem olyan hangsúlyosak, mint a "nyitott-csukott" állapotok esetén (4.15. ábra). Habár a főkomponensek által reprezentált mozgások a főkomponens elemzés definíciója szerint egymástól függetlenek, ez nem zárja ki az egyes állapotok jellemző együttes előfordulását. Ez a jelenség figyelhető meg például a szabad állapotban lévő, teljes hosszúságú szerkezeteket tartalmazó sokaság esetén, melynek konformerei szinte kizárólag a "nyitott" és a "széles" állapotokban vannak.

4.4.2. A sokaságokban megfigyelt állapotok

Az előállított sokaságokban megfigyelt szerkezeti átrendeződések és diverzitás általánoságának ellenőrzéséhez készítettünk egy ezektől független, több PDZ domént tartalmazó sokaságot. Ehhez felhasználtuk a csoportban előzőleg molekuladinamikai szimulációval előállított [25] PSD95 PDZ1 és PDZ2 doméneket, illetve további PDZ doméneket tartalmazó sokaságot gyűjtöttünk és ezekből egy molekuladinamikai szimulálást követően

a CoNSEnsX⁺ szerverrel kémiai eltolódás szelekcióval alsokaságokat állítottunk elő, a "Módszerek" fejezetben leírtak szerint. Ezen kívül több, a PDB-ben megtalálható PDZ domén szerkezetet is felhasználtunk azok eredeti formájában, tehát röntgendiffrakcióval meghatározott egyedi és NMR-spektroszkópiával kapott sokaságokat is. Az ezen forrásokból összeállított nagy sokaság célja, hogy jól lefedje a különböző PDZ domének által meghatározott konformációs teret. Az így előállított nagy sokasággal a további vizsgálódás előfeltétele egy olyan közös szerkezeti mag létrehozása, amely minden vizsgált doménben megtalálható (vagy megfeleltethető). A közös aminosavak megtalálásához több szerkezeti illesztésre is szükség volt, valamint az illesztések alapján csak a közös C α atomokat tartottuk meg további felhasználásra.



4.16. ábra. A kiegészített PDZ domén készletek főkomponens elemzése

A kiegészített PDZ domén készletek főkomponens elemzése során feltűnt, hogy az első két főkomponens mentén a három PSD-95 PDZ domén más pozíciókat tölt be a konformációs térben, mint a "PDZ mag" sokaság (4.16. ábra). A PSD-95 sokaságok esetén is csak a "PDZ magban" található aminosavak szerepeltek a számításban, ami kizárja, hogy a különbségek a mozgékony hurok régiók meglétére vezethetők vissza. Ebben a főkomponens elemzésben az első változó (PC1) főleg a $\beta 6$ redőt és az $\alpha 2$ hélixet összekötő régió változásainak feleltethető meg. A második változó (PC2) mentén a legnagyobb különbség a $\beta 2$ és az $\alpha 2$ (hélixek) távolsága, ami a "szűk-széles" állapotokban is meghatározó. A harmadik változó (PC3) mentén az $\alpha 2$ (hélix) hintázó mozgása, míg a negyedik változó mentén (PC4) a fentebb említett "nyitó-csukó" mozgás figyelhető meg. A PSD95 PDZ2 doménje az első változó mentén nagyon elkülönül az elemzésben résztvevő többi sokaságtól, az egész kiegészített sokaságot tekintve csak két másik szerkezet jelenik meg ebben a régióban: a 2KA9 [53] szerkezet második PDZ doménje, ami egyben a PSD95-PDZ2 sokaság generálásához használt kiindulási szerkezet [25] és az 1B8Q szerkezet, amely egy a PSD-95-tel interakcióba léphető fehérje része. Amíg az első két főkomponens mentén főleg a PDZ domének közötti szerkezeti különbségek látszanak, a harmadik és a negyedik mentén a vizsgált szerkezetek egy homogén halmazt alkotnak, ezek az állapotok tehát a legtöbb PDZ domén számára elérhetőek.

4.4.3. A PSD95 PDZ doménjeinek összehasonlítása

A PSD-95 PDZ doménjeinek összehasonlításához ezekben a szerkezetekben is meg kellett keresni a közös aminosavakat, amik a további összehasonlítások alapját képezik. Az egyik esetben csak a három PSD domén közös aminosavait választottuk ki, a másik esetben a "PDZ magban" szereplő szerkezetillesztést használtuk fel, tehát csak azokat az aminosavakat kerültük kiválasztásra, amelyek valamennyi vizsgált PDZ szerkezetben jelen vannak. A főkomponens elemzés alapján a két illesztés eredménye nagyon hasonló, ami arra enged következtetni, hogy a domének közötti különbségek függetlenek az illesztés során elhagyott hurok régióktól. A főkomponens elemzések átlapolódásának vizsgálata alapján az első két főkomponens ugyanazt a mozgást jellemzi, illetve a harmadik és a negyedik főkomponens által jellemzett mozgások is nagyon hasonlóak.

Az egyes sokaságok és az összefűzött sokaság főkomponenseinek átlapolódását megvizsgálva előtűnt, hogy a domének közötti különbségeket jól definiáló első kettő főkomponens nem függ össze jelentősen az egyes doménekre jellemző mozgásokkal. Csak a harmadik és a negyedik főkomponens mentén jellennek meg a PDZ3 sokaság esetén is megfigyelt "nyitó-cukó" mozgások. Ez a PDZ1 doménre is igaz, viszont a PDZ2 domén esetében nem, ami alátámasztja azokat az előző megfigyeléseket, hogy a PDZ1 és PDZ2 domének esetén másfajta szerkezeti átrendeződés következik be a ligandum kötése során [25]. Az ilyen kivételek nem jelentik viszont azt, hogy a domének között ne lenne hasonlóság a ligandum kötés során bekövetkező szerkezeti változásokban.

4.4.4. Szabad és kötött formában lévő PDZ domének összehasonlítása

A kiegészített PDZ domén készlet esetén a szabad és kötött állapotban lévő konformerek között a negyedik főkomponens mentén látni a legnagyobb különbséget. Ez a főkomponens nagyon hasonlít a PSD-95 fehérje doménjeinek harmadik főkomponenséhez, ami ugyancsak nagyon hasonló a PDZ3 sokaságok esetén megfigyelt "nyitó-cukó" mozgást leíró első főkomponenshez (4.5. táblázat és 4.15, 4.17, 4.18. ábrák). A PSD-95 sokaságok esetén ez azt jelenti, hogy a kötött állapotban lévő szerkezetek által lefedett konformációs tér javarészt a szabad állapotban lévő szerkezetek által lefedett konformációs téren belül található és csak a negyedik főkomponens mentén jellennek meg ezek az állapotbéli különbségek.

Várakozásainkkal ellentétben a "nyitó-cukó" mozgás mentén a szabad és kötött formák eloszlása nem egységes az egyes PDZ domének esetében. A PSD-95 PDZ1 és PDZ2 esetében a szabad forma ezen mozgás mentén a teljes teret mintavételezi, a kötött forma ezen térrész zárt konformációjának megfelelő részében mozog. Ugyanakkor a PSD95 PDZ3 domén kötött formája ezzel ellentétben a nyitott formában stabilizálódik a ligandum kötött állapotban. Ez a megfigyelés igaz marad akkor is, ha az ezkehez a doménekhez tartozó, PDB-ben elérhető szerkezeteket vizsgáljuk, bár ebben az esetben a különbségek kevésbé látványosak, mivel ezek lényegesen kisebb konformációs teret reprezentálnak. Ez a megfigyelés két másik, kémiai eltolódások alapján szelektált sokaságra is igaz, valamint az ezekhez használt kiindulási PDB-s modellekről is elmondható (4.18. ábra). Az egyetlen kivétel az 1GM1-1VJ6 PDB azonosítójú sokaság pár (a PTP-BL fehérje második PDZ

doménje, szabad és kötésben lévő állapotban), ahol az eredeti sokaságok elemei csak a harmadik főkomponens mentén válnak el.

Főkomponens	Mozgás	PSD-95 PDZ domének	Csak a PSD-95 PDZ3
1	$\beta 6$ - $\alpha 2$ hurok	1	-
2	"szűk-széles"	2	-
3	$\alpha 2$ hinta	-	-
4	"nyitó-csukó"	3	1

4.5. táblázat. *Egymásnak megfeleltethető főkomponensek a kiegészített PDZ készletben és a PSD-95 PDZ doménjei között.*

Érdekesség, hogy a hosszabb szimulációkból kémiai eltolódás szelekcióval készített alsokaságok a harmadik és a negyedik főkomponens mentén közelebb helyezkednek el a eredeti sokaságokhoz, viszont a kísérletileg megállapított paramétereknek nem felelnek meg jobban, mint a rövid szimulációkból szelektált alsokaságok.

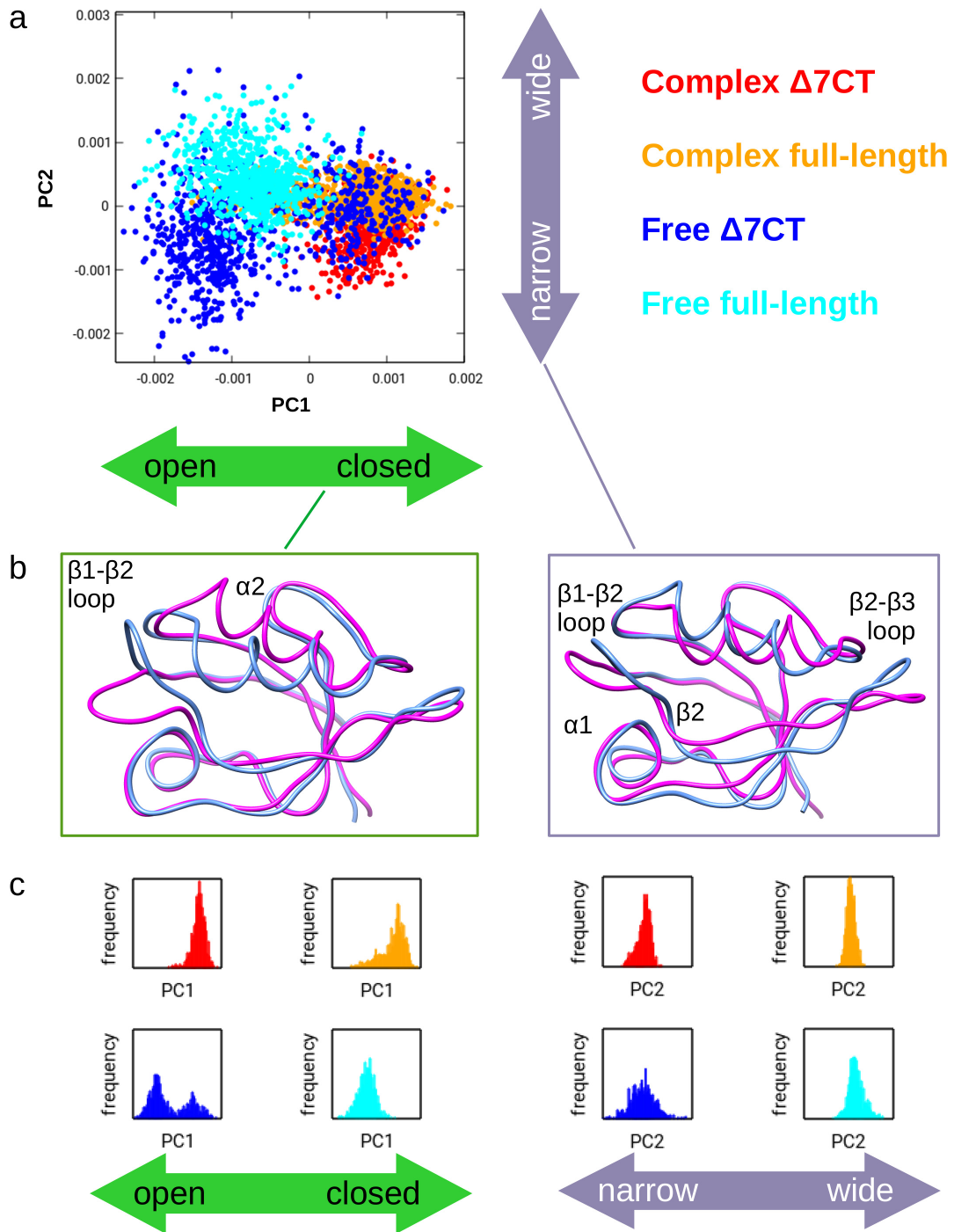
4.4.5. A PDZ3 sokaságokban megfigyelt mozgások – Diskusszió

A molekuladinamikai szimulációval előállított PDZ3 sokaságokkal kapcsolatos egyik megfigyelésünk, hogy a teljes hosszúságú szerkezet inkább a "széles", a rövidebb $\Delta 7$ CT szerkezet inkább a "szűk" állapotot veszi fel. Ez arra enged következtetni, hogy a teljes hosszúságú sokaságban meglévő $\alpha 3$ hélix okozza ezt a különbséget a konformációs egyensúlyban. Az extra hélix jelenléte a "nyitott-csukott" állapotokra is hatással van, különösen a szabad állapotban lévő szerkezetek esetén, míg kötött állapotban a ligandum stabilizálja a "csukott" állapotot. Ez a megfigyelés konzisztens a szerkezetek dinamikájára vonatkozó szakirodalommal [33], miszerint az $\alpha 3$ hélix jelenléte csökkenti a kötősebet körülvevő hurkok rugalmasságát. A modellek dinamikáját tekintve viszont valószínű, hogy az állapotok közötti populációs változások szabályozzák a ligandum kötését, ami egyezik a ligandum kötés dinamikájára vonatkozó szakirodalommal [35]. Egy peptid ligandum és az $\alpha 3$ együttes megléte egy gátló hatást vált ki, a dinamikus egyensúly eltolódik a kötés kialakulásához kedvezőtlenebb állapotok irányába. Például az $\alpha 3$ hélix Tyr397-es aminosavának foszforizációja a negyedjére csökkenti a ligandum kötés kialakulásának valószínűségét [58].

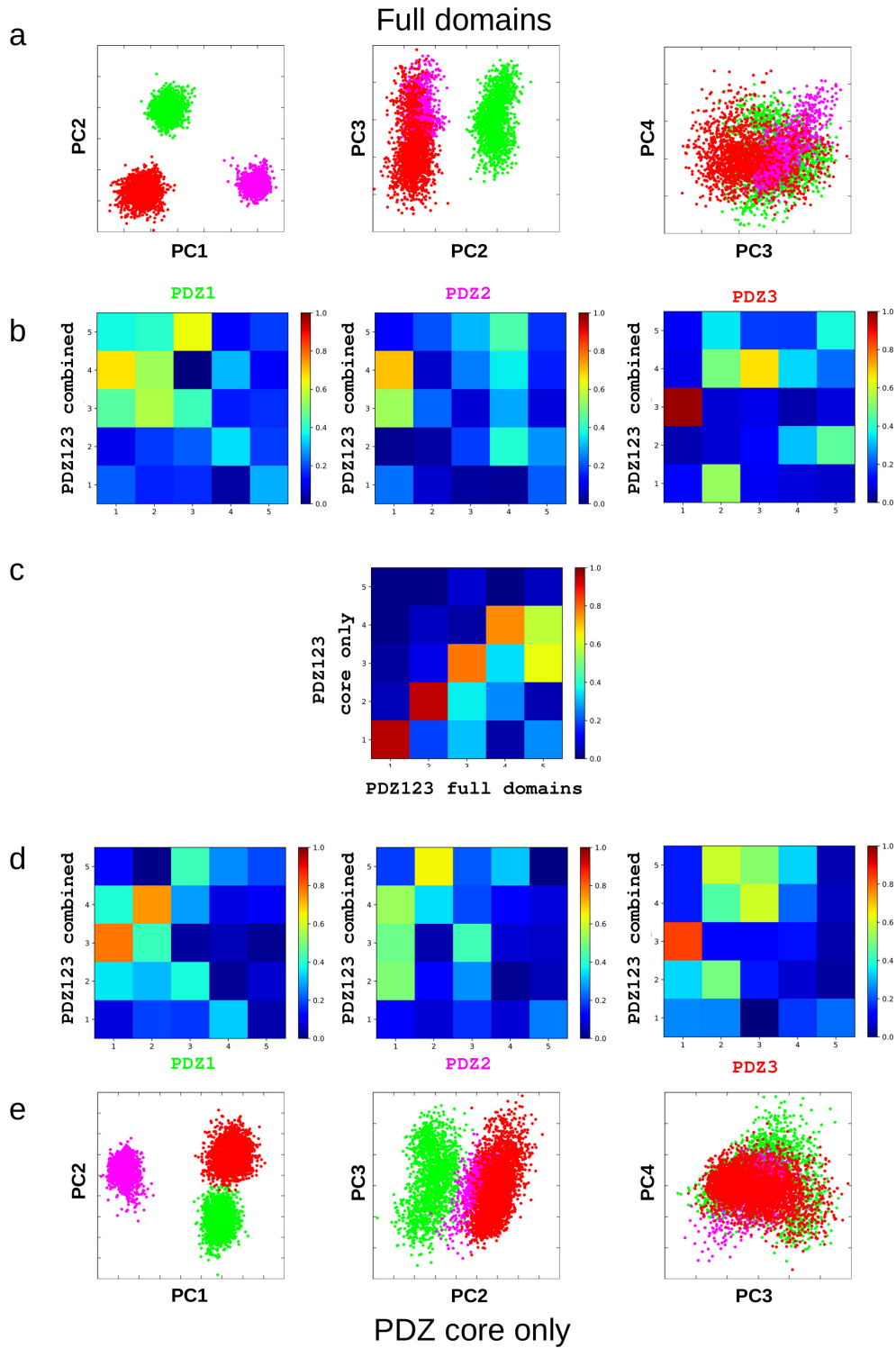
A teljes PDZ3-SH3-GK szupramodul vagy a teljes hosszúságú PSD-95 szerkezet esetén az ehhez hasonló moduláció vélhetően kiterjedtebb. A PDZ3-SH3-GK szupramodul teljes, NMR-es méréseken alapuló modellje szerint a PDZ3 "canonical peptide site"-ja és $\beta 2$ - $\beta 3$ hurokjka (gyenge) kapcsolatban van az SH3 doménnel [57], tehát a modult alkotó szomszédos szerkezeti elemek hatással vannak egymásra, ami komplex szabályozási folyamatok része. Az általunk alkalmazott módszer csak a PDZ3 szerkezeti sokaságokban megfigyelt aminosav kapcsolatokra hagyakszik. A "PDZ mag"-ra való alkalmazással megmutatható, hogy a megfigyelt mozgások a PDZ domének közös szerkezeti elemeire vonatkoznak, nem pedig egyedi, változékony hurok régiókra.

A PDZ domének nyitott-zárt konformációjának és szabad-kötött állapotának feltárt összefüggése meglepő. Technikai oldalról egybecseng a csoport korábbi megfigyelésével, miszerint a dinamikus szerkezeti sokaságok – megfelelő elemzés mellett – képesek mint-

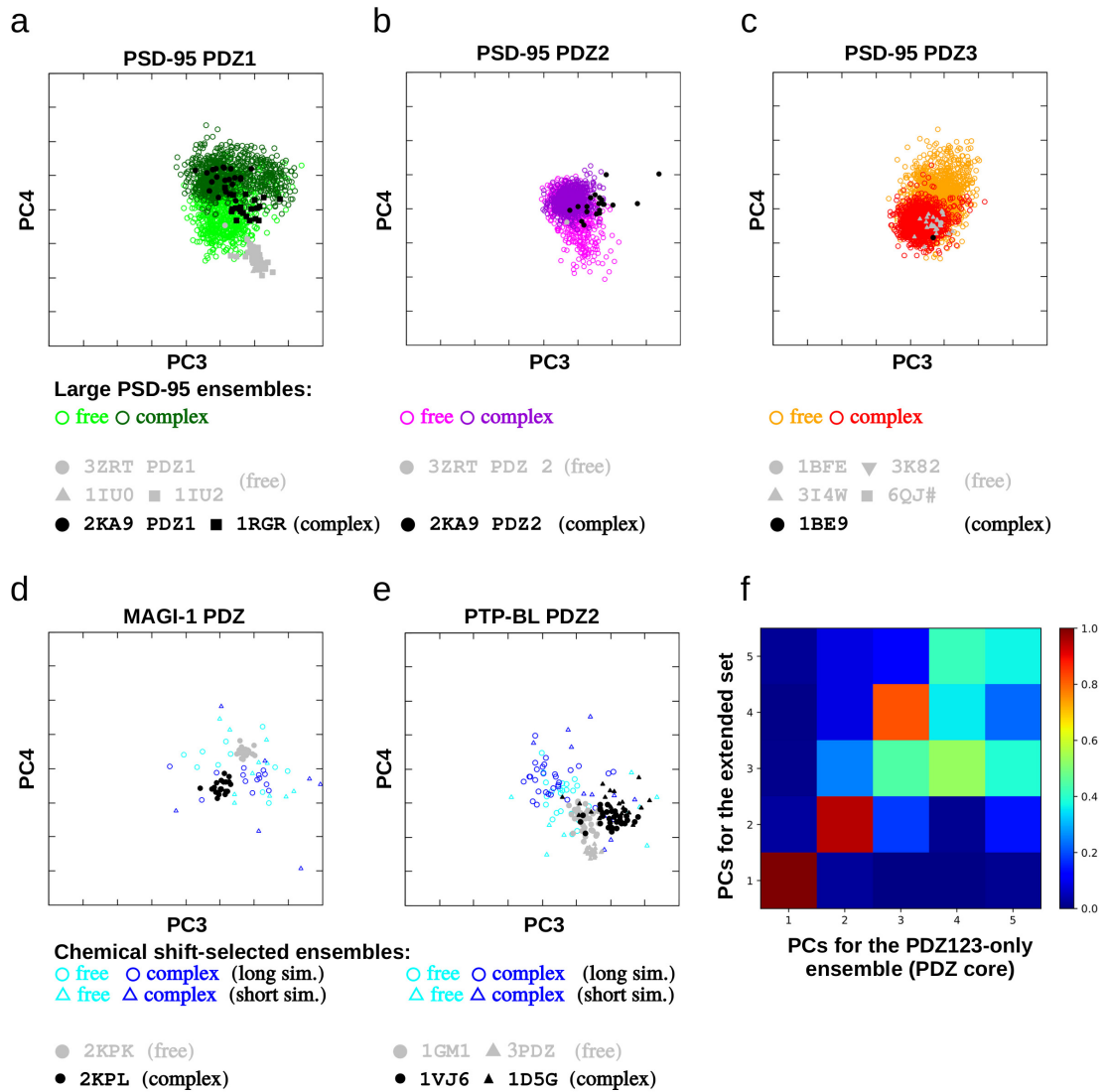
egy "felnagyítani" a funkcionálisan releváns mozgásokat, melyek a PDB szerkezetekben is megfigyelhetőek, csak nehezebben detektálhatóak [11]. Másfelől a PDZ domének korábban nem leírt aspektusára világít rá, miszerint a ligandumkötés hatására bekövetkező térszerkezeti-dinamikai változások, bár nagyrészt értelmezhetőek ugyanazon mozgásforma mentén való populációeltolódással, nem egységesen ugyanúgy jelennek meg az egyes domének esetében. A PTP-BL fehérje PDZ doménje esetében pedig nem is ez a mozgásforma dominál a szabad a kötött forma közötti szerkezeti átrendeződéseket tekintve. A PDZ domének tehát nem csupán ligandumkötési preferenciáikban, hanem szerkezeti-dinamikai viselkedésükben is változatosságot mutatnak.



4.15. ábra. A PSD-95 PDZ3 domének kombinált sokaságának főkomponens elemzése. **a:** A kombinált sokaság konformereinek eloszlása a sokaság konformációs terében az első és a második főkomponens szerint (színkód az ábrán). Az első főkomponens a kötőszeg körüli "nyitó-csukó", a második főkomponens a "szűk-széles" átmeneteknek feleltethető meg. **b:** Az azonosított mozgásokhoz rendelhető strukturális átrendeződések. A mozgások két extrém állapota kék és magenta színekkel került megrajzolásra a UCSF Chimera programmal (a mozgások mértéke a könnyebb vizualizálhatóság miatt nem arányos). A fontosabb, mozgásban résztvevő szerkezeti elemek felcímkézésre kerültek. **c:** A konformerek eloszlása az első két főkomponens mentén (színkódok az **a** ábrarészén).



4.17. ábra. PSD-95 három PDZ doménjéből készített sokaságok összehasonlítása. **a:** A sokaságok eloszlása a főkomponensek mentén: PDZ1 (zöld), PDZ2 (magenta) and PDZ3 (piros), valamint a főkomponensek átlapolódása az egyes sokaságok és a kombinált sokaság között. **b:** Főkomponensek átlapolódása az egyes sokaságok és a kombinált sokaság között. **d** és **e:** ugyanaz mint az **a** és **b** ábrarészekben, de csak a "PDZ mag" sokaságban is megtalálható közös aminosavak felhasználásával. **c:** A 2 elemzés eredményeinek átlapolódása az első 5 főkomponens mentén.



4.18. ábra. A kiegészített PDZ készlet néhány kiválasztott PDZ doménjének szabad és kötött állapotú főkomponens elemzése. Az **a-e** ábrákon a 4.16. ábra adatai láthatóak, csak a kiválasztott PDZ doménekre/konformerekre. **a-c:** A PSD-95 három PDZ doménje (1, 2, 3 sorrendben) és néhány releváns PDB-ből származó szerkezet. **d-e:** A hosszú és rövid szimulációkkal előállított MAGI-1 PDZ és PTP-BL PDZ2 sokaságokból kémiai eltolódás szelekcióval készített alsokaságok és a hozzájuk tartozó PDB-s szerkezetek. **f:** A PSD-95 PDZ domének és a kiegészített PDZ készlet főkomponenseinek átlapolása.

Köszönetnyilvánítás

A disszertáció és a disszertáció alapjaként szolgáló közlemények létrejöttéért köszönet illeti:

- Témavezetőmet, Gáspári Zoltánt, akinek a segítségére és iránymutatására a doktórandsztként eltöltött évek alatt is mindig számíthattam
- Társszerőimet és a csoportom tagjait, hiszen az itt prezentált eredmények egy része közös munka eredménye

Irodalomjegyzék

- [1] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, and G. Sherlock. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*, 25(1):25–29, May 2000.
- [2] No authors listed. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res*, 47(D1):D506–D515, 01 2019.
- [3] Richter B, Gsponer J, Várnai P, Salvatella X, and Vendruscolo M. The mumo (minimal under-restraining minimal over-restraining) method for the determination of native state ensembles of proteins. *J Biomol NMR*, 2:117–35, 2007.
- [4] Ahmet Bakan, Lidio M. Meireles, and Ivet Bahar. ProDy: Protein Dynamics Inferred from Theory and Experiments. *Bioinformatics*, 27(11):1575–1577, 04 2011.
- [5] E. G. Baker, G. J. Bartlett, M. P. Crump, R. B. Sessions, N. Linden, C. F. Faul, and D. N. Woolfson. Local and macroscopic electrostatic interactions in single α -helices. *Nat Chem Biol*, 11(3):221–228, Mar 2015.
- [6] M. Batchelor, M. Wolny, L. Dougan, E. Paci, P. J. Knight, and M. Peckham. Myosin tails and single α -helical domains. *Biochem. Soc. Trans.*, 43(1):58–63, Feb 2015.
- [7] M. Biro, Y. Romeo, S. Kroschwald, M. Bovellan, A. Boden, J. Tcherkezian, P. P. Roux, G. Charras, and E. K. Paluch. Cell cortex composition and homeostasis resolved by integrating proteomics and quantitative imaging. *Cytoskeleton (Hoboken)*, 70(11):741–754, Nov 2013.
- [8] T. V. Borchert, M. Mathieu, J. P. Zeelen, S. A. Courtneidge, and R. K. Wierenga. The crystal structure of human CskSH3: structural diversity near the RT-Src and n-Src loop. *FEBS Lett.*, 341(1):79–85, Mar 1994.
- [9] P. Carter, C. A. Andersen, and B. Rost. DSSPcont: Continuous Secondary Structure Assignments for Proteins. *Nucleic Acids Res.*, 31(13):3293–3295, Jul 2003.
- [10] P. Y. Chou and G. D. Fasman. Prediction of the secondary structure of proteins from their amino acid sequence. *Adv Enzymol Relat Areas Mol Biol*, 47:45–148, 1978.

- [11] A. Czajlik, B. Kovács, P. Permi, and Z. Gáspári. 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, 03 2017.
- [12] C. I. De Zeeuw, C. C. Hoogenraad, E. Goedknecht, E. Hertzberg, A. Neubauer, F. Grosveld, and N. Galjart. CLIP-115, a novel brain-specific cytoplasmic linker protein, mediates the localization of dendritic lamellar bodies. *Neuron*, 19(6):1187–1199, Dec 1997.
- [13] L. Dobson, L. Nyitray, and Z. Gáspári. A conserved charged single α -helix with a putative steric role in paraspeckle formation. *RNA*, 21(12):2023–2029, Dec 2015.
- [14] D. A. Doyle, A. Lee, J. Lewis, E. Kim, M. Sheng, and R. MacKinnon. Crystal structures of a complexed and peptide-free membrane protein-binding domain: molecular basis of peptide recognition by PDZ. *Cell*, 85(7):1067–1076, Jun 1996.
- [15] S. D. Fenster, W. J. Chung, R. Zhai, C. Cases-Langhoff, B. Voss, A. M. Garner, U. Kaempf, S. Kindler, E. D. Gundelfinger, and C. C. Garner. Piccolo, a presynaptic zinc finger protein structurally related to bassoon. *Neuron*, 25(1):203–214, Jan 2000.
- [16] Á. Fizil, Z. Gáspári, T. Barna, F. Marx, and G. Batta. "Invisible" conformers of an antifungal disulfide protein revealed by constrained cold and heat unfolding, CEST-NMR experiments, and molecular dynamics calculations. *Chemistry*, 21(13):5136–5144, Mar 2015.
- [17] S. Griep and U. Hobohm. PDBselect 1992-2009 and PDBfilter-select. *Nucleic Acids Res*, 38(Database issue):D318–319, Jan 2010.
- [18] Z. Gáspári, D. Süveges, A. Perczel, L. Nyitray, and G. Tóth. Charged single alpha-helices in proteomes revealed by a consensus prediction approach. *Biochim. Biophys. Acta*, 1824(4):637–646, Apr 2012.
- [19] Zoltán Gáspári and András Perczel. Protein dynamics as reported by nmr. *Annual Reports on NMR Spectroscopy*, 71:35–76, 2010.
- [20] M. Habeck, W. Rieping, and M. Nilges. Bayesian Estimation of Karplus Parameters and Torsion Angles from Three-Bond Scalar Couplings Constants. *J. Magn. Reson.*, 177(1):160–165, Nov 2005.
- [21] Baruch Z. Harris and Wendell A. Lim. Mechanism and role of pdz domains in signaling complex assembly. *Journal of Cell Science*, 114(18):3219–3231, 2001.
- [22] Albert Hung and Morgan Sheng. Pdz domains: Structural modules for protein complex assembly. *The Journal of biological chemistry*, 277:5699–702, 03 2002.
- [23] M. P. Hunter, A. Russo, and J. P. O'Bryan. Emerging roles for intersectin (ITSN) in regulating signaling and disease pathways. *Int J Mol Sci*, 14(4):7829–7852, Apr 2013.

- [24] I. A. Katrukha. Human cardiac troponin complex. Structure and functions. *Biochemistry Mosc.*, 78(13):1447–1465, Dec 2013.
- [25] B. Kovács, N. Zajácz-Epresi, and Z. Gáspári. Ligand-dependent intra- and inter-domain motions in the PDZ12 tandem regulate binding interfaces in postsynaptic density protein-95. *FEBS Lett*, 594(5):887–902, 03 2020.
- [26] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, and D. E. Shaw. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins*, 78(8):1950–1958, Jun 2010.
- [27] Kresten Lindorff-Larsen, Robert B. Best, Mark A. DePristo, Christopher M. Dobson, and Michele Vendruscolo. Simultaneous determination of protein structure and dynamics. *Nature*, 433:128–132, 2005.
- [28] A. Lupas, M. Van Dyke, and J. Stock. Predicting coiled coils from protein sequences. *Science*, 252(5009):1162–1164, May 1991.
- [29] Dmitry Lupyán, Alejandra Leo-Macias, and Angel R. Ortiz. A new progressive-iterative algorithm for multiple structure alignment. *Bioinformatics*, 21(15):3255–3263, 06 2005.
- [30] R. Madhavan and H. B. Peng. Molecular regulation of postsynaptic differentiation at the neuromuscular junction. *IUBMB Life*, 57(11):719–730, Nov 2005.
- [31] Huaiyu Mi, Xiaosong Huang, Anushya Muruganujan, Haiming Tang, Caitlin Mills, Diane Kang, and Paul D. Thomas. PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic Acids Research*, 45(D1):D183–D189, 11 2016.
- [32] R. K. Moore, L. T. Grinspan, J. Jimenez, P. J. Guinto, B. Ertz-Berger, and J. C. Tardiff. HCM-linked $\alpha^{\dagger}160E$ cardiac troponin T mutation causes unique progressive structural and molecular ventricular remodeling in transgenic mice. *J. Mol. Cell. Cardiol.*, 58:188–198, May 2013.
- [33] S. Mostarda, D. Gfeller, and F. Rao. Beyond the binding site: the role of the β_2 - β_3 loop and extra-domain structures in PDZ domains. *PLoS Comput Biol*, 8(3):e1002429, 2012.
- [34] S. Mostowy and P. Cossart. Septins: the fourth component of the cytoskeleton. *Nat. Rev. Mol. Cell Biol.*, 13(3):183–194, Feb 2012.
- [35] H. N. Motlagh, J. O. Wrabl, J. Li, and V. J. Hilser. The ensemble nature of allostery. *Nature*, 508(7496):331–339, Apr 2014.
- [36] S. Neal, A. M. Nip, H. Zhang, and D. S. Wishart. Rapid and accurate calculation of protein 1H , ^{13}C and ^{15}N chemical shifts. *J. Biomol. NMR*, 26(3):215–240, Jul 2003.

- [37] C. Nourry, S. G. Grant, and J. P. Borg. PDZ domain proteins: plug and play! *Sci. STKE*, 2003(179):RE7, Apr 2003.
- [38] A. Onufriev, D. A. Case, and D. Bashford. Effective Born radii in the generalized Born approximation: the importance of being perfect. *J Comput Chem*, 23(14):1297–1304, Nov 2002.
- [39] Michelle Peckham and Peter J. Knight. When a predicted coiled coil is really a single α -helix, in myosins and other proteins. *Soft Matter*, 5:2493–2503, 2009.
- [40] C. M. Petit, J. Zhang, P. J. Sapienza, E. J. Fuentes, and A. L. Lee. Hidden dynamic allostery in a PDZ domain. *Proc Natl Acad Sci U S A*, 106(43):18249–18254, Oct 2009.
- [41] M. Pirooznia, T. Wang, D. Avramopoulos, D. Valle, G. Thomas, R. L. Haganir, F. S. Goes, J. B. Potash, and P. P. Zandi. SynaptomeDB: an ontology-based knowledge-base for synaptic genes. *Bioinformatics*, 28(6):897–899, Mar 2012.
- [42] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013. ISBN 3-900051-07-0.
- [43] T. Shirao, K. Hanamura, N. Koganezawa, Y. Ishizuka, H. Yamazaki, and Y. Sekino. The role of drebrin in neurons. *J. Neurochem.*, 141(6):819–834, 06 2017.
- [44] Dominic Simm, Klas Hatje, and Martin Kollmar. Waggawagga: Comparative visualization of coiled-coil predictions and detection of stable single α -helices (sah domains). *Bioinformatics (Oxford, England)*, 31, 10 2014.
- [45] Dominic Simm, Klas Hatje, Stephan Waack, and Martin Kollmar. Protein function prediction in genomes: Critical assessment of coiled-coil predictions based on protein structure data. *bioRxiv*, 2019.
- [46] S. Sivaramakrishnan, B. J. Spink, A. Y. Sim, S. Doniach, and J. A. Spudich. Dynamic charge interactions create surprising rigidity in the ER/K alpha-helical protein motif. *Proc. Natl. Acad. Sci. U.S.A.*, 105(36):13356–13361, Sep 2008.
- [47] B. Szappanos, D. Süveges, L. Nyitray, A. Perczel, and Z. Gáspári. Folded-unfolded cross-predictions and protein evolution: the case study of coiled-coils. *FEBS Lett.*, 584(8):1623–1627, Apr 2010.
- [48] Dániel Süveges, Zoltán Gáspári, Gábor Tóth, and László Nyitray. Charged single α -helix: A versatile protein structural motif. *Proteins: Structure, Function, and Bioinformatics*, 74(4):905–916, 2009.
- [49] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. Berendsen. GROMACS: fast, flexible, and free. *J Comput Chem*, 26(16):1701–1718, Dec 2005.

- [50] M. Varadi, S. Kosol, P. Lebrun, E. Valentini, M. Blackledge, A. K. Dunker, I. C. Felli, J. D. Forman-Kay, R. W. Kriwacki, R. Pierattelli, J. Sussman, D. I. Svergun, V. N. Uversky, M. Vendruscolo, D. Wishart, P. E. Wright, and P. Tompa. pE-DB: a database of structural ensembles of intrinsically disordered and of unfolded proteins. *Nucleic Acids Res.*, 42(Database issue):D326–335, Jan 2014.
- [51] H. A. Velazquez and D. Hamelberg. Conformational selection in the recognition of phosphorylated substrates by the catalytic domain of human Pin1. *Biochemistry*, 50(44):9605–9615, Nov 2011.
- [52] Andy C. Wang and Ad Bax. Determination of the Backbone Dihedral Angles ϕ in Human Ubiquitin from Reparametrized Empirical Karplus Equations. *J. Am. Chem. Soc.*, 118:2483–2494, Mar 1996.
- [53] Wenning Wang, Jingwei Weng, Xu Zhang, Maili Liu, and Mingjie Zhang. Creating conformational entropy by increasing interdomain mobility in ligand binding regulation: A revisit to n-terminal tandem pdz domains of psd-95. *Journal of the American Chemical Society*, 131(2):787–796, 2009. PMID: 19072119.
- [54] Y. Wang, C. B. Harrison, K. Schulten, and J. A. McCammon. Implementation of Accelerated Molecular Dynamics in NAMD. *Comput Sci Discov*, 4(1), 2011.
- [55] M. Wolny, M. Batchelor, P. J. Knight, E. Paci, L. Dougan, and M. Peckham. Stable single α -helices are constant force springs in proteins. *J. Biol. Chem.*, 289(40):27825–27835, Oct 2014.
- [56] K. W. Yau, S. F. van Beuningen, I. Cunha-Ferreira, B. M. Cloin, E. Y. van Battum, L. Will, P. Sch?tzle, R. P. Tas, J. van Krugten, E. A. Katrukha, K. Jiang, P. S. Wulf, M. Mikhaylova, M. Harterink, R. J. Pasterkamp, A. Akhmanova, L. C. Kapitein, and C. C. Hoogenraad. Microtubule minus-end binding protein CAMSAP2 controls axon specification and dendrite development. *Neuron*, 82(5):1058–1073, Jun 2014.
- [57] J. Zhang, S. M. Lewis, B. Kuhlman, and A. L. Lee. Supertertiary structure of the MAGUK core from PSD-95. *Structure*, 21(3):402–413, Mar 2013.
- [58] J. Zhang, C. M. Petit, D. S. King, and A. L. Lee. Phosphorylation of a PDZ domain extension modulates binding affinity and interdomain interactions in postsynaptic density-95 (PSD-95) protein, a membrane-associated guanylate kinase (MAGUK). *J Biol Chem*, 286(48):41776–41785, Dec 2011.
- [59] M. Zweckstetter. NMR: prediction of molecular alignment from structure using the PALES software. *Nat Protoc*, 3(4):679–690, 2008.
- [60] Annamária F. Ángyán and Zoltán Gáspári. Ensemble-based interpretations of nmr structural data to describe protein internal dynamics. *Molecules*, 18:10548–10567, 2013.

- [61] Annamária F Ángyán, Balázs Szappanos, András Perczel, and Zoltán Gáspári. Consensx: an ensemble view of protein structures and nmr-derived experimental data. *BMC Structural Biology*, 10:39, 2010.