

# **Az információs entrópia felhasználása polipeptidek konformációs sokaságainak és belső dinamikájának jellemzésére**

**doktori dolgozat**

Gyimesi Gergely

Témavezetők:

Dr. Závodszy Péter  
Dr. Szilágyi András

Külső konzulens:

Dr. Hegedűs Tamás

Pázmány Péter Katolikus Egyetem  
Információs Technológiai Kar

Budapest, 2011.

## KÖSZÖNETNYILVÁNÍTÁS

Szeretném megköszönni mindazon emberek támogatását, akik nélkül e munka nem jöhetett volna létre. Mindenekelőtt a munkámat befogadó intézmények vezetőinek, Dr. Závodszy Péternek (MTA SzBK Enzimológiai Intézet), Dr. Sarkadi Balázsnak (MTA-SE Membránbiológiai Kutatócsoport) és Dr. Kellermayer Miklósnak (SE Biofizikai és Sugárbiológiai Intézet), hogy a munkámnak háttérrel biztosítottak. Köszönettel tartozom a PPKE-ITK részéről Dr. Roska Tamásnak és a PPKE-ITK doktori iskola vezetőségének, hogy lehetővé tették a doktori iskolában való részvételemet. Szeretném megköszönni témavezetőimnek, Dr. Szilágyi Andrásnak és Dr. Hegedűs Tamásnak türelmes és kitartó támogatásukat. Végül, de nem utolsósorban köszönettel tartozom kollégáimnak, akik inspiráló légkört nyújtottak a munkához.

## Tartalomjegyzék

1. Bevezetés.....	4
1.1. Entrópiaszámolás .....	6
1.1.1. A statisztikus fizikai entrópia .....	6
1.1.2. Belső és Descartes-féle koordinátarendszerek .....	10
1.1.3. Abszolút konfigurációs entrópiát számoló módszerek.....	11
1.2. ABC fehérjék.....	19
1.2.1. ABC exporterek felépítése és konformációi .....	19
1.2.2. ABC exporterek működési mechanizmusa .....	23
1.2.3. Molekuláris dinamikai számítások ABC fehérjékkel.....	25
2. Célkitűzések .....	27
3. Módszerek .....	28
3.1. Molekuláris dinamikai szimulációk .....	28
3.2. Monte Carlo mintavételezés .....	32
3.3. A teljes konformációhalmaz generálása és az egzakt entrópia számolása.....	32
3.4. Molekuláris sokaságok előállítása.....	34
3.5. A mintahalmaz centrálása .....	35
3.6. Egyéb entrópiabecslő módszerek implementálása.....	36
3.7. Fehérjemodellek .....	38
3.8. ABC fehérjék molekuláris dinamikai szimulációja.....	39
3.9. A kölcsönös információ .....	45
4. A konformációs entrópia becslése Gauss-keverék függvények használatával .....	47
4.1. Eredmények.....	47
4.1.1. A Gauss-keverék módszer.....	47
4.1.2. Molekuláris tesztrendszerek generálása .....	50
4.1.3. Az egzakt konfigurációs entrópia és állapotösszeg kiszámítása .....	51
4.1.4. Monte Carlo mintavételezett sokaságok entrópiája.....	52
4.1.5. Molekuláris dinamikai sokaságok entrópiája .....	54
4.1.6. A Gauss-keverék módszer összehasonlítása más módszerekkel.....	56
4.2. Diskusszió: komplexitás, bővíthetőség, alkalmazási lehetőségek.....	58
5. ABC exporterek konformációinak dinamikai összehasonlítása.....	60
5.1. Eredmények.....	60
5.1.1. A szimulációs rendszerek felépítése .....	60
5.1.2. A lipid kettősréteg viselkedése.....	62
5.1.3. ATP/ADP csere hatása a korrelált mozgásokra .....	63
5.1.4. A holo és az apo szerkezetek közötti konformációs átmenet jellemzése.....	67
5.1.5. Az „alul nyitott” szerkezet instabilitása .....	71
5.2. Diskusszió.....	76
6. Összefoglalás.....	82
7. Függelék.....	84
7.1. A Gibbs-eloszlás származtatása .....	84
7.2. A konfigurációs és kinetikus entrópiajárulékok szétválasztása .....	85
8. Publikációk jegyzéke.....	87
9. Irodalomjegyzék.....	88

## 1. Bevezetés

Az elméleti fehérjevizsgálati módszerek közül különösen a korai molekuláris dinamikai szimulációknak fontos szerepe volt annak a paradigmaváltásnak a megalapozásában, miszerint a fehérjék nem merev szerkezetűek, hanem dinamikusan változó rendszerek [1]. Azóta a molekuláris szimulációkat széles körben alkalmazzák különböző biomolekuláris rendszerek és folyamatok vizsgálatára. A módszer egyik legfőbb előnye, hogy atomi felbontású betekintést ad a rendszerek időbeli és dinamikai működésébe, aminek segítségével a modellrendszerrel kapcsolatos specifikus kérdések sokszor könnyebben megválaszolhatóak, mint kísérleti módszerekkel [1]. Munkámban egyrészt biomolekuláris rendszerek entrópiájának szimulációk segítségével történő becslésével foglalkozom, másrészt a szimulációs módszerek egy konkrét alkalmazását mutatom be az ABC fehérjék alloszterikus mechanizmusának vizsgálatára.

A nyitott fizikai rendszerekben lejátszódó spontán folyamatok irányát a szabadenergia (vagy szabadentalpia) mennyiségének változása határozza meg. Szintén ez a mennyiség az, amelyek a rendszer egyensúlyi tulajdonságaival összefüggésbe hozható. A szabadenergia megváltozása a termodinamika alaptételei szerint egyrészt entalpikus, másrészt entropikus járulékokból származtatható. Amíg az entalpikus tag értelmezhető a rendszer átlagenergiájának megváltozásaként, az entropikus járulék az energia különböző szabadsági fokok közötti eloszlásának megváltozásával kapcsolatos, és elméleti úton sokkal nehezebben megragadható. Számításos szempontból az entalpia megbecslése gyakorlatilag a rendszer átlagenergiájának megfelelő pontosságú közelítésén múlik, az entropikus tag kiszámítása során viszont a nehézség a konfigurációs entrópia, vagyis a rendszer számára rendelkezésre álló konformációs állapotok számának megbecslésében rejlik. Az entrópiaváltozás által hajtott folyamatok esetén ezért nem csak prediktív, hanem interpretatív értéke is lehet azoknak a számításos módszereknek, amelyekkel különböző fizikai rendszerek entrópiája megbecsülhető.

Az egyes állapotok entrópiájának megismerése azért is érdekes, mert folyamatok esetén az entrópia viselkedése gyakran betekintést ad a változás mechanizmusába és a folyamatot molekuláris szinten irányító legfőbb hatásokba. A fehérjék felgombolyodása során például az egyik fő vezérlő erő a hidrofób kölcsönhatás, amelynek alapja a víz entrópiaváltozása [2]. A konfigurációs entrópia változása a molekulák flexibilitásváltozásán keresztül a fehérje-fehérje kölcsönhatások mechanizmusának megértését is segíti [3]. Hasonlóan, kismolekulák fehérjéhez kötődésekor (ligandum-receptor rendszerekben) is sok

esetben a kötődés termodinamikáját befolyásoló entrópiaváltozás a konformációs entrópia nagymértékű megváltozásával magyarázható [4, 5]. Az automatizált, számításos módszerek egyúttal megteremtik a kötődési entrópiaváltozás *in silico* megbecslésének lehetőségét farmakológiailag érdekes vegyületek esetében [4], amely elősegítheti hatékonyabb gyógyszerek célzott tervezését. Egyensúlyi állapotok vizsgálata mellett az entrópia kiszámításával akár kötődési reakciók sebességére is lehet következtetni [6]. A konfigurációs entrópia kiszámítása elméleti modellek és molekuláris szimulációk segítségével ezért igen fontos, de nagy kihívást jelent.

A molekuláris szimulációk fontos jellemzője, hogy a szimulációs rendszerek és körülmények teljes mértékben az irányításunk alatt állnak, így tetszőleges modellrendszerek létrehozhatók, azok összehasonlíthatók. [1] Ezzel szemben a vizsgált rendszerek a valóságban átmeneti jellegűek, rövid életidejűek lehetnek, vagy kísérletileg nehezen hozzáférhetők [1].

Az integráns membránfehérjék kísérleti módszerekkel történő szerkezeti és funkcionális vizsgálatát több tényező is hátráltatja. Az egyik a membránfehérjék többségének az alacsony termelődési szintje, expressziója. Emiatt rendszerint *overexpressziós* rendszert kell létrehozni, amely azonban a jelenlegi stratégiák esetén sok finomhangolást igényel, mert a kapott fehérje mennyisége és minősége nehezen megjósolható. Szintén sok nehézség adódik a membránfehérjék amfipatikus jellegéből, ami miatt különösen érzékenyek a biokémiai környezetükre, és kezelésük, tisztításuk rendkívüli nehézségekkel jár. E problémák miatt a membránfehérjéről viszonylag kevés atomi felbontású szerkezetet sikerült megoldani (jelenleg 77882 szerkezetből 1509 membránfehérje a Protein DataBank és a PDBTM adatbázis alapján [7]), ahhoz képest, hogy a szekvenált genomokban felfedezett gének majdnem harmada feltehetően transzmembrán fehérjét kódol. A rendelkezésre álló, kristályosítással kapott szerkezetek esetén is a fehérjéket membrán nélkül, detergensekkel stabilizálják [8], ezért kérdéses, hogy az így megoldott szerkezet mekkora jósággal írja le a fehérje *in vivo*, membránbeli szerkezetét. Ezen kísérleti nehézségek miatt előtérbe kerülnek azok az elméleti módszerek, amelyekkel egyrészt az egyes membránfehérjék szerkezetére, másrészt a meglévő szerkezetek alapján azok működésére lehet plauzibilis modelleket felállítani.

Az alloszterikus jelenségek tanulmányozása esetén a molekuláris dinamikai szimulációk különösen hasznosnak bizonyultak, hiszen a különböző, statikus szerkezetek ismeretében rendszerint arra vagyunk kíváncsiak, hogy egy ligandum bekötődése hogyan indukálja a fehérje átmenetét az egyik állapotból a másikba. Molekuláris szimulációk

segítségével bizonyos esetekben maga az átmenet is direkt módon nyomkövethető, de a ligandum bekötődése által okozott dinamikai hatások, például a korrelált mozgásokban bekövetkező változások is megfigyelhetők [9]. Ezek a jelenségek kísérleti módszerekkel nehezen, vagy csak korlátozottan vizsgálhatók. Rendszerint a kísérletek nem adnak atomi felbontású képet ezekről a dinamikai változásokról.

Munkám első részében egy új elméleti módszert írok le, amely Gauss-keverék függvények használatával képes molekulák konformációs entrópiáját megbecsülni. A módszert összehasonlítom más, az irodalomban leírt és gyakran alkalmazott módszerekkel. A második részben ABC transzporter fehérjék dinamikai leírásával foglalkozom. Az ABC transzporterek mind integráns membránfehérjék, mind allosztérikusan működő rendszerek, hiszen az ATP kötődése és hidrolízise, és az annak hatására bekövetkező szubsztráttranszport egymástól fizikailag távoli helyeken megy végbe. Munkám egyrészt betekintést ad az ATP hidrolízise során a korrelált mozgást végző aminosavak mintázatában bekövetkező változásokba, másrészt felhívja a figyelmet a kísérletileg meghatározott szerkezetek értelmezése során felmerülő néhány veszélyre.

## 1.1. Entrópiaszámolás

### 1.1.1. A statisztikus fizikai entrópia

A statisztikus fizikában az entrópia mennyiségét a rendszer által az adott egyensúlyi makroszkopikus állapotban bejárt mikroállapotok számából származtatják. Ha a rendszer által egy adott állapotban elérhető mikroállapotok száma  $\Delta\Gamma$ , akkor a makroállapot entrópiája [10]

$$S = k_B \ln \Delta\Gamma \quad (1)$$

ahol  $k_B$  a Boltzmann-állandó. Legegyszerűbb esetben, ha a rendszer izolált és energiája állandó, az entrópia kiszámítható az adott energiájú állapotot megvalósító mikroállapotok számával,

$$S(E) = k_B \ln \Gamma(E) \quad (2)$$

ahol  $\Gamma(E)$  az  $E$  energiájú állapotok száma. Az ilyen rendszereket nevezzük mikrokanonikus rendszereknek. A statisztikus fizikában valószínűségi változóként fogható fel az, hogy a rendszer éppen melyik mikroállapotban tartózkodik. Az imént tárgyalt izolált rendszerek esetén, amikor az energia állandó, nincs okunk semmilyen *a priori* feltételezésre arról, hogy a rendszer melyik mikroállapota valósul meg. Ezért mindegyik, az  $E$  energiát megvalósító

mikroállapot egyformán valószínű, és az információk hiánya a rendszer állapotáról maximális. Egy  $E$  energiájú állapot megvalósulásának valószínűsége ezért

$$w(E) = \frac{1}{\Gamma(E)} \quad (3)$$

A biológiai rendszerek általában nem izoláltak, hanem egy náluk lényegesen nagyobb rendszerrel, az ún. termosztáttal termikus kölcsönhatásba léphetnek. Ekkor a részrendszer energiája változhat, és nem tudjuk a (2) és (3) összefüggéseket alkalmazni. Az viszont ekkor is feltételezhető, hogy a biológiai részrendszer és környezete, a termosztát, együttesen egy izolált rendszert alkotnak. A részrendszer által bejárt állapotok száma és az egyes állapotok előfordulási valószínűsége így megbecsülhető. Ha a teljes rendszer energiája  $E = E_1 + E_2$ , ahol  $E_1$  az általunk vizsgált részrendszer,  $E_2$  pedig a termosztát energiája, akkor a részrendszer entrópiája

$$S_1 = k_B \ln \Delta\Gamma = -k_B \ln w(\langle E_1 \rangle) \quad (4)$$

ahol  $w(E_1)$  egy  $E_1$  energiájú állapot előfordulásának valószínűsége a részrendszerben. Ez utóbbiról megmutatható, hogy

$$w(E_1) = Z^{-1} e^{-\frac{E_1}{k_B T}} = Z^{-1} e^{-\beta E_1} \quad (5)$$

ahol bevezettük a  $\beta = 1/(k_B T)$  jelölést (a **levezetések a 7.1. fejezetben található**). Ez az ún. Gibbs-eloszlás, amelynek normálási faktora az állapotösszeg, vagy más néven partíciós függvény  $Z$ :

$$Z = \sum_k e^{-\frac{E_k}{k_B T}} = \sum_k e^{-\beta E_k} \quad (6)$$

Ezek alapján a vizsgált részrendszer entrópiája a következő alakba írható:

$$S_1 = -k_B \ln w(\langle E_1 \rangle) = k_B (\ln Z + \beta \langle E_1 \rangle) \quad (7)$$

Általános esetben is megmutatható, hogy az entrópia additivitása miatt az  $\ln w(E)$  csak az energiának lineáris függvénye lehet [10]. Ebben az esetben a (7) összefüggés átírható

$$S_1 = -k_B \ln w(\langle E_1 \rangle) = -k_B \langle \ln w(E_1) \rangle \quad (8)$$

alakba, vagyis a részrendszer entrópiája kiszámítható az eloszlásfüggvény logaritmusának negatív átlagaként. Az izolált rendszer esetén kapott  $w(E)$  eloszlást mikrokanonikus eloszlásnak, az állapotok olyan halmazát pedig, amelyben az egyes állapotok  $w(E)$  gyakorisággal fordulnak elő, mikrokanonikus sokaságnak nevezzük [10]. Termosztált rendszer esetén a  $w(E)$  Gibbs-eloszlást kanonikus eloszlásnak, az ilyen gyakorisággal előforduló állapotok halmazát kanonikus sokaságnak nevezik [10].

A statisztikus fizikai entrópia (2) és (8) formája sok hasonlóságot mutat az információelméletben használt Shannon-entrópiával, ami egy adott  $p_k$  diszkrét eloszlásra a következő alakban definiálható:

$$H = -\sum_k p_k \ln p_k \quad (9)$$

A Shannon-entrópia értelmezhető úgy, mint egy adott valószínűségi változó „véletlenszerűségét” mérő mennyiség, vagy az átlagosan hiányzó információ mértéke, ami ahhoz lenne szükséges, hogy megjósoljuk a valószínűségi változó értékét. Az entrópia kifejezésében szereplő  $I_k = -\ln p_k$  ez alapján a  $k$ . állapot által hordozott információ mennyisége, avagy a „meglepődésünk mértéke”, amikor megtudjuk, hogy a rendszer a  $k$ . állapotban van [11]. Mikrokanonikus rendszerekre az információs entrópia a (9) és (3) kifejezések segítségével kiszámítható:

$$H = -\sum_{E_k=E} w(E) \ln w(E) = -\sum_{E_k=E} \frac{1}{\Gamma(E)} \ln \frac{1}{\Gamma(E)} = \ln \Gamma(E) = S(E) / k_B \quad (10)$$

mivel  $\sum_{E_k=E} 1 = \Gamma(E)$  éppen az  $E$  energiájú állapotok száma. Izolált rendszer esetén tehát a rendszer statisztikus fizikai és információs entrópiája konstans szorzó erejéig megegyezik. A statisztikus fizikai entrópia általános esetben is közvetlenül megfeleltethető az információs entrópiával, hiszen

$$S_1 = -k_B \langle \ln w(E_1) \rangle = -k_B \sum_{E_k=E_1} w(E_k) \ln w(E_k) = k_B H_1 \quad (11)$$

Klasszikus rendszerek esetén az állapottér folytonos, és a rendszer egy mikroállapotának az egymással kanonikusan konjugált koordináták és impulzusok egy  $(p, q)$  értéke felel meg. A diszkrét állapotokra bevezetett entrópia fogalma itt közvetlenül nem értelmezhető. Egy adott makroállapotot megvalósító mikroállapotok  $\Delta\Gamma$  számának klasszikus rendszerben egy  $\Delta p \Delta q$  fázistérfogat feleltethető meg, amely azonban nem tiszta szám, **hanem  $3N$  koordináta esetén a  $h$  Planck-állandó  $3N$ -edik hatványával megegyező dimenziójú mennyiség.** Ezek alapján a tisztán klasszikus formalizmus szerint definiált entrópia csak a mértékegységek választásától függő additív állandó erejéig meghatározott. Bizonyos egyszerű rendszerekre az ún. kváziklasszikus és a kvantummechanikai leírás összehasonlításával megmutatható, hogy a rendszer mozgása során bezárt fázistérfogat csak  $h$  egységekben növekedhet [10]. Ennek az összefüggésnek az általánosításával klasszikus esetben a  $\Delta p \Delta q$  fázistérfogatban található állapotok számát a

$$\Delta\Gamma = \frac{\Delta p \Delta q}{h^{3N}} \quad (12)$$



formában definiálhatjuk [10]. A diszkrét eloszlások analógiájára bevezethető a  $\rho(p, q)$  fázissűrűség, amely a rendszer egy adott  $(p, q)$  állapot körüli tartózkodásának valószínűsége. Klasszikus esetben így a (8) összefüggés analógiájára egy részrendszer entrópiája megadható az

$$S = -k_B \langle \ln(h^{3N} \rho) \rangle = -k_B \int \rho \ln(h^{3N} \rho) dp dq \quad (13)$$

formában [10]. Termosztált rendszerek esetén az előzőkkel teljesen analóg módon származtathatjuk a részrendszert leíró fázistérbeli Gibbs-eloszlást, amely

$$\rho(p, q) = \rho(E) = Z^{-1} e^{-\beta E(p, q)} \quad (14)$$

Ebben már a klasszikus állapotintegrál

$$Z = \int e^{-\beta E(p, q)} dp dq \quad (15)$$

szerepel. A termosztált részrendszer entrópiája ezek alapján a következő formába írható:

$$S = -k_B \int Z^{-1} e^{-\beta E(p, q)} \ln(h^{3N} Z^{-1} e^{-\beta E(p, q)}) dp dq = k_B (\ln(h^{3N} Z) + \beta \langle E \rangle) \quad (16)$$

A Shannon-féle információs entrópia szintén kiterjeszthető folytonos eloszlásokra az ún. differenciális entrópia fogalmával, amelyet a

$$H_{diff} = - \int f(x) \ln f(x) dx \quad (17)$$

kifejezés definiál egy folytonos  $f(x)$  valószínűségi sűrűség-függvényre. Látható, hogy ez nem feleltethető meg közvetlenül a (13)-ben kapott klasszikus entrópia kifejezésével, mert elhanyagolja a sűrűségfüggvény mértékegység-függését. A klasszikus statisztikus fizikai entrópia tehát a fázistérbeli eloszlás információs entrópiájától egy additív konstanssal különbözik, ami viszont függ a rendszer dimenziójától ( $3N$ ). A differenciális információs entrópia másik fontos tulajdonsága, hogy a fizikai entrópiával szemben negatív értéket is felvehet. Egyre keskenyedő eloszlás esetén a differenciális entrópia a mínusz végtelen felé divergál. Ez a matematikai divergencia a statisztikus fizikai entrópia klasszikus definíciója esetén is előfordulhat. Fizikai rendszerek esetén ez a divergencia azzal magyarázható, hogy mérhetetlenül több információt ad a rendszer fázistérbeli  $(p, q)$  koordinátáinak egzakt ismerete, mint az, ha csak egy véges  $\Delta p \Delta q$  fázistérbeli térfogatban ismerjük biztosan az előfordulását. A klasszikus entrópia efféle nem-fizikai viselkedése arra hívja fel a figyelmet, hogy ebben a tartományban a klasszikus fizikai állapotok definíciója nem illeszkedik jól a makroszkopikusan tapasztalt fizikai entrópia fogalmához. Az entrópia negatív értéke statisztikus fizikai értelemben azt jelentené, hogy a rendszer számára egynél kevesebb állapot érhető el a fázistérben, vagyis az általa bejárt  $\Delta p \Delta q$  fázistérfogat kisebb, mint  $h$ . Ekkor azonban a kváziklasszikus közelítés, amely alapján a (12) összefüggést származtattuk, már

nem érvényes. Szintén a klasszikus kép ezen határait mutatja, hogy a Heisenberg-féle  $\Delta p \Delta q \approx h$  határozatlansági összefüggés szerint a rendszer klasszikus állapotának egzakt megismerése a  $h$  méretű fáziscellán belül már nem lehetséges.

### 1.1.2. Belső és Descartes-féle koordinátarendszerek

Az atomi rendszerek leírására a legáltalánosabban használt koordináták az atomi háromdimenziós Descartes-koordináták és az atomi impulzusvektorok halmaza. Ebben a koordinátarendszerben a megfelelő tér- és impulzuskoordináták kanonikusan konjugáltak, és a rendszer mozgási energiája egyszerű formát ölt. Molekuláris rendszerek esetén a konformációs entrópia kiszámítására azonban gyakran célszerűbb az ún. BAT („bond-angle-torsion”) koordinátarendszert használni, amely az atomok relatív pozícióját a kötéshosszak, kötésszögek és a torziós szögek segítségével írja le. Ennek a koordinátaválasztásnak az előnye, hogy benne a rendszer potenciálisenergia-függvénye gyakran egyszerű alakot vesz fel, és az egyes szabadsági fokoknak megfelelő energia- és entrópiajárulékok könnyen szétválaszthatóak. A  $3N$  általános koordináta közül biztosan lesz 6, amely a rendszer teljes translációjával és rotációjával kapcsolat szabadsági fokoknak felel meg. Általánosságban a maradék  $3N - 6$  koordinátát nevezhetjük belső koordinátáknak, mert értékük már csak az atomok egymáshoz viszonyított pozíciójától függ. A továbbiakban azonban a molekuláris konformáció leírására használt kötéshosszakat, kötésszögeket és torziós szögeket **(a BAT koordinátákat)** tekintem belső koordinátáknak.

A kötéshosszak és a kötésszögek gyakran igen szűk tartományban ingadoznak, ami miatt az ebből adódó magas frekvenciájú vibrációk közel harmonikusnak tekinthetők. Az ilyen koordinátákat kemény szabadsági fokoknak nevezzük. A kemény szabadsági fokok vibrációi legtöbbször abba a frekvenciatartományba esnek, ahol már szükséges a kvantummechanikai leírásmód alkalmazása. A közel harmonikus mozgás miatt azonban ezeknek a szabadsági fokoknak az entrópiajáruléka akár analitikusan is kiszámítható. Konformációs entrópia számolásakor a kemény szabadsági fokok vibrációit több módszer tekinti a molekula konformációjától független, és ezért elhagyható additív konstansnak. Ezzel szemben más belső koordináták, pl. a molekula elforgatható kötése mentén levő torziós szögek, széles tartományban változhatnak, és gyakran nem harmonikus mozgást végeznek, ezek a rendszer ún. puha vagy lágy szabadsági fokai. A lágy szabadsági fokok mentén a molekula rendszerint klasszikusnak tekinthető mozgást végez. Mivel az ilyen koordináták járuléka az entrópiához nehezen becsülhető az anharmonicitás miatt, ezért az entrópiaszámoló módszerek gyakran fordítanak erre fokozott figyelmet.

A Descartes-koordinátákról belső koordinátákra való áttérés során a  $dq$  elemi térfogat transzformációja miatt a konfigurációs integrálban (lásd (84), 7.2. fejezet) megjelenik a transzformáció Jacobi-determinánsa. A Jacobi-determinánsról megmutatható, hogy a kötéshosszak és a kötésszögek függvénye, és a Herschbach és munkatársai által megadott módszerrel [12] konstruktívan is kiszámolhatóak. Mivel a Jacobi-determináns csak kemény szabadsági fokok értékétől függ, ezért a molekula konformációjától függetlennek tekinthető. Számításaim során a Jacobi-determinánst a konfigurációs entrópiához járuló additív konstansnak tekintem, és nem veszem figyelembe.

### 1.1.3. Abszolút konfigurációs entrópiát számoló módszerek

A szabadenergia számítására leggyakrabban használt módszerek a rendszer két egyensúlyi állapota közötti szabadenergia-különbséget számolják, ilyenek pl. a termodinamikai integráció (TI, [13]), a szabadenergia-perturbáció (FEP, [13, 14]) és a hisztogramanalízis módszerek [15]. Ezeknél a módszereknél azonban a megfelelő pontosság elérése érdekében szükséges, hogy a rendszert egy kvázisztatikus útvonalon át vigyük egyik végállapotból a másikba. Az átmeneti útvonal definiálásakor ügyelni kell arra, hogy kellően sima legyen, és ne legyenek benne nagy energetikai ugrások. A két állapot közötti nagymértékű szerkezeti különbség gyakran megnehezíti a számítást, néha az ilyen szimulációk nem is kivitelezhetőek [16].

Egy lehetséges megoldás az integrációs útvonal felállításának elkerülésére, ha olyan módszert alkalmazunk, amellyel kiszámíthatjuk a két végállapot abszolút szabadenergiáját. A rendszer belső energiája kiszámítható valamilyen molekuláris mechanikai erőter (*force-field*) alapján, oldószer jelenléte esetén pedig a solvatációs szabadenergia valamilyen implicit oldószermodellből (pl. a gyakran használt MM/PBSA [17]) származtatható. A rendszer teljes szabadenergiájának kiszámításához ezen kívül meg kell becsülnünk az oldott molekula entrópiáját is. **Az entrópia kinetikus és konformációs járulékot is tartalmaz, amelyek szétválaszthatóak, és a kinetikus járulék analitikusan is kiszámítható (lásd 7.2. fejezet). A konformációs járulék azonban függ a molekulán belüli atomok közötti kölcsönhatásoktól, kiszámítása nem triviális feladat.** A megoldás kulcsa ebben az esetben tehát egy hatékonyan működő módszer, amely alkalmas a konformációs entrópia becslésére például molekuláris dinamikai (MD) vagy Monte Carlo (MC) egyensúlyi szimulációk alapján.

A konformációs entrópia kiszámítására vonatkozó egyik legkorábbi módszer a kváziharmonikus közelítés [18], amely a molekula egyensúlyi állapotában bejárt bonyolult szerkezetű energiefelszín egy egyszerű, harmonikus potenciállal közelíti. Ez a megközelítés

abból az ötletből származik, hogy a fehérjék és más, hasonlóan jól strukturált biológiai makromolekulák az egyensúlyi állapotukban várhatóan kis fluktuációkat végeznek, miközben a szabadsági fokok közötti korreláció várhatóan nagy mértékű [18]. A konfigurációs térbeli sűrűségfüggvényt ezért egy olyan függvénnyel érdemes közelíteni, amely világos kapcsolatot létesít a rendszer entrópiája és az egyes szabadsági fokok fluktuációi és korreláltságuk mértéke között. Az ezeknek a kritériumoknak megfelelő legegyszerűbb és legkézenfekvőbb függvényalak a többváltozós Gauss-függvény

$$\rho(q) = \frac{1}{\sqrt{2\pi^N \det \sigma}} \exp\left(-\frac{1}{2}(q - \langle q \rangle)^T \sigma^{-1} (q - \langle q \rangle)\right) \quad (18)$$

ahol  $N$  a koordináták száma,  $q$  az eredeti leírás szerint a rendszer „fontos” (belső) koordinátáit jelöli,  $\sigma$  pedig a koordinátákhoz tartozó kovarianciamátrix, amelyet a

$$\sigma_{ij} = \langle (q_i - \langle q_i \rangle)(q_j - \langle q_j \rangle) \rangle \quad (19)$$

kifejezés definiál. A többdimenziós Gauss-eloszlás paraméterei a koordináták kovarianciamátrixából származtathatóak, és a konfigurációs entrópia ezekből analitikusan kiszámítható az alábbi összefüggéssel [18]:

$$S_{conf} = \frac{1}{2} k_B N + \frac{1}{2} k_B \ln((2\pi)^N \det \sigma) \quad (20)$$

A Gauss-függvény alakú eloszlás mögött egyúttal az a feltételezés rejlik, hogy a rendszer egyetlen, lokálisan minimális energiájú állapot körül harmonikus fluktuációkat végez. A fluktuációkat meghatározó effektív potenciálfüggvény a koordináták kvadratikus függvényeként megadható:

$$E_{eff}(q) = \frac{1}{2} k_B T \left( (q - \langle q \rangle)^T \sigma^{-1} (q - \langle q \rangle) \right) \quad (21)$$

bár az eredeti leírás is megjegyzi, hogy a  $\sigma$  korrelációs mátrix elemei valószínűleg erősen nem-triviális függést mutatnak a hőmérséklettől. A kovarianciamátrix diagonalizálásával a (21) kifejezés  $N$  tagra esik szét, amiből látszik, hogy a kváziharmonikus módszer a rendszert valójában  $N$  egymástól lineárisan függetlenül mozgó harmonikus oszcillátornak tekinti.

A (18) összefüggés elvileg tetszőleges koordinátaválasztás esetén felírható. Az eredeti Karplus-módszer a belső koordináták készletét használta, de ezeknek az azonosítása nagy rendszerek esetén bonyodalmas lehet, és gyakran nem egyértelmű. Egyszerűbben kivitelezhető és kézenfekvőbb a Descartes-koordináták használata, ebben az esetben azonban megmutatható, hogy a rendszer translációs és forgatási invarianciája miatt a  $\sigma$  korrelációs mátrix szinguláris lesz [19], ami a rendszer túlhatározottságára utal. A Schlitter által

bevezetett módszer feloldja a szingularitás problémáját anélkül, hogy kilépne a kváziharmonikus közelítés keretei közül [19].

A Schlitter-módszer a kváziharmonikus módszerhez hasonlóan a rendszert  $N$  független oszcillátorként modellezi, amelyek entrópiáját azonban kvantummechanikai leírás alapján becsli a kvantummechanikai oszcillátor entrópiáját leíró

$$S_{qm} = -k_B \ln(1 - e^{-\alpha}) + k_B \frac{\alpha}{e^{\alpha} - 1} \quad (22)$$

összefüggés segítségével, ahol  $\alpha = \beta \hbar \omega$  az oszcillátorra jellemző dimenziótlans mennyiség, és  $\omega$  az oszcillátor frekvenciája. A kvantummechanikai entrópia a klasszikus határesetben ( $\hbar \rightarrow 0$ ) visszaadja a klasszikus oszcillátor entrópiáját. A nagy frekvenciájú kemény módusok esetén a kvantummechanikai kezelésmód megszünteti az entrópia klasszikus divergenciáját a negatív végtelen felé, és a kifagyott szabadsági fokok ( $\omega \rightarrow \infty$ ) entrópiájáruléka is nullához fog tartani. A (22) kifejezés azonban numerikusan nehezen kezelhető, ezért helyette Schlitter az

$$S_{Sch} = \frac{k_B}{2} \ln \left( 1 + \frac{e^2}{\alpha^2} \right) \quad (23)$$

kifejezést vezeti be, amely a kvantummechanikai entrópia egy *ad hoc* felső becslése [19]. A módosítás előnye a numerikus egyszerűsödés mellett, hogy az új kifejezés könnyebben általánosítható többdimenziós rendszerekre. Ehhez az oszcillátor frekvenciáját Schlitter a megfelelő koordináta varianciájából származtatta a klasszikus határesetben érvényes ekvipartíciós tétel alapján, amely szerint

$$m\omega^2 \langle x^2 \rangle = k_B T \quad (24)$$

majd ezt az entrópia (23) kifejezésébe behelyettesítve kapta az

$$S_{Sch} = \frac{k_B}{2} \ln \left( 1 + \frac{k_B T e^2}{\hbar} m \langle x^2 \rangle \right) \quad (25)$$

összefüggést. Az  $m \langle x^2 \rangle$  kifejezés a koordináták tömeggel súlyozott varianciája, így a kézenfekvő kiterjesztés több dimenzióra ennek a tömeggel súlyozott kovarianciamátrixszal való helyettesítése. Az ebből származó

$$S_{Sch} = \frac{k_B}{2} \ln \det \left( 1 + \frac{k_B T e^2}{\hbar} M^{1/2} \sigma M^{1/2} \right) \quad (26)$$

kifejezésben  $M$  az atomi koordinátákhoz tartozó tömegek mátrixa,  $\sigma$  pedig a kovarianciamátrix. A determináns nem változik ortogonális transzformáció során, ezért a kifejezés tovább egyszerűsíthető az

$$S_{Sch} = \frac{k_B}{2} \ln \det \left( 1 + \frac{k_B T e^2}{\hbar} M \sigma \right) \quad (27)$$

alakra [19]. Ha olyan koordinátarendszert választunk, amelyben a  $\sigma$  kovarianciamátrix diagonális, akkor a (27) kifejezés több, (25) alakú kifejezés összegére esik szét, amelyek az egyes főkomponensek mentén történő fluktuációk entrópiajárulécai. A Schlitter-módszer ugyan egy *ad hoc* képlettel közelíti a rendszer entrópiáját, de megmutatható, hogy a kvantummechanikai entrópia  $\alpha$  szerinti sorfejtésével a (23) képlethez hasonló formulák családjá állítható elő [20], és az  $\omega$  frekvenciák kiszámításával az entrópia akár egzaktul is megadható [20]. A Schlitter-módszerrel kapcsolatban fontos megjegyezni, hogy a kiindulásként használt (22) és az eredményként kapott (27) kifejezés a kinetikus és konfigurációs entrópiát egyaránt tartalmazza, emiatt a kapott entrópiaértékek tekintetében közvetlenül nem hasonlítható össze a pusztán konfigurációs entrópiát számoló módszerekkel. Ez a különbség a módszerek között gyakran feledésbe merül, a módszerek tárgyalásakor keveset hangsúlyozzák. A kinetikus entrópia, mivel nem függ az atomi konfigurációtól, ugyanazon rendszer két állapota közötti entrópiakülönbség számolásakor kiesik és nem ad járulékot, ezért a gyakorlati alkalmazásban a jelenléte nem okoz gondot. Annak ellenére, hogy az abszolút entrópiaértékek mégsem összehasonlíthatók a kétféle típusú módszer esetén, informatívnak tartottam a Descartes-koordinátákat használó módszereket is szerepeltetni a különböző módszerek összehasonlításában, mert egyszerűségük miatt igen elterjedt a használatuk.

A kváziharmonikus közelítés jól használható abban az esetben, ha a molekuláris rendszer valóban egyetlen, minimális energiájú állapot körüli harmonikus fluktuációt végez. A legtöbb rendszer viszont ennél bonyolultabb felépítésű, a termikus fluktuációk során több, lokálisan minimális energiájú állapotot is bejár, és ezek környezetében sem feltétlenül érvényes a harmonikus közelítés. Megmutatták, hogy ugyan az anharmonicitásból eredő hibák kicsik [21], a módszer jelentős hibát követ el azzal, hogy a kváziharmonikus módusok lineárisnál magasabb rendű korrelációit elhanyagolja [21]. Azokra a rendszerekre, amelyek több lokális energiaminimumot is meglátogatnak a termikus fluktuációk következtében, a kváziharmonikus módszer jelentősen túlbecsülheti a konfigurációs entrópiát. Ilyen rendszerek tipikusan a fehérjék, amelyek natív állapotukban flexibilis hurkokat vagy szabadon forgó felszíni oldalláncokat is tartalmaznak. Ezek mozgása nem csak a rendszer háttérben rejlő energiafüggvény anharmonicitását tükrözi, de több szabadenergia-minimum jelenlétét is

mutatja a rendszerben. Emiatt a konformációs entrópia becslésének pontosításához az állapotterbeli sűrűség pontosabb közelítésére van szükség.

A kváziharmonikus módszer megjelenése óta számos, más elven működő módszer látott napvilágot, amelyek a kváziharmonikus módszer hibáit igyekeztek kiküszöbölni. A koordináták közötti lineárisnál magasabb rendű korreláció figyelembe vehető az entrópia kölcsönös információon alapuló kifejtésével. A kölcsönös információ definíciója két változó esetén

$$I_2(q_1, q_2) = S_1(q_1) + S_1(q_2) - S_2(q_1, q_2) \quad (28)$$

ahol  $S_2$  és  $S_1$  az együttes és az egyváltozós entrópia [22]. Független változók esetén az együttes entrópia a marginális, egyváltozós entrópiák összegére esik szét, ezért ekkor a kölcsönös információ értéke nulla, egyéb esetben pedig pozitív érték. A kölcsönös információ és a marginális entrópiák ismeretében a változók együttes entrópiája a (28) kifejezés átrendezésével megadható:

$$S_2(q_1, q_2) = S_1(q_1) + S_1(q_2) - I_2(q_1, q_2) \quad (29)$$

Kettőnél több változó esetén a kölcsönös információ definíciója a (28) kifejezéssel analóg módon kiterjeszhető, három változó esetén

$$\begin{aligned} I_3(q_1, q_2, q_3) &= \\ &= S_1(q_1) + S_1(q_2) + S_1(q_3) \\ &- S_2(q_1, q_2) - S_2(q_1, q_3) - S_2(q_2, q_3) + S_3(q_1, q_2, q_3) \end{aligned} \quad (30)$$

illetve több változó esetén általánosan [22]

$$I_k(q_1, \dots, q_k) = \sum_{s=1}^k (-1)^{s+1} \sum_{l_1 < \dots < l_s} S(q_{l_1}, \dots, q_{l_s}) \quad (31)$$

A (29) kifejezéssel analóg módon a rendszer teljes entrópiája több változó esetén is előállítható a kölcsönös információk segítségével, a (31) egyenletek átrendezésével

$$S(q_1, \dots, q_N) = \sum_{i=1}^N S_1(q_i) - \sum_{i < j} I_2(q_i, q_j) + \sum_{i < j < k} I_3(q_i, q_j, q_k) - \dots \quad (32)$$

alakban. A (32) összefüggés jelentősége, hogy kapcsolatot teremt az együttes entrópia és a marginális  $S_1$  entrópiák között. Ha a koordinátarendszert úgy vesszük fel, hogy a koordináták kovarianciamátrixa diagonális legyen, akkor látható, hogy a kváziharmonikus módszer csak a marginális entrópiákat, vagyis a (32) képlet jobb oldalának első tagját veszi figyelembe. Az elhagyott, kölcsönös információ jellegű tagok a rendszer magasabb rendű korrelációiról tartalmaznak információt, ami a következő gondolatmenettel belátható.

A kölcsönös információt ekvivalens módon definiálhatjuk [22] az

$$I_2(q_1, q_2) = \int \rho_2(q_1, q_2) \ln \frac{\rho_2(q_1, q_2)}{\rho_1(q_1)\rho_1(q_2)} dq_1 dq_2 \quad (33)$$

alakban is, ahol  $\rho_2$  az együttes,  $\rho_1$  pedig a marginális sűrűségfüggvények. A (33) összefüggés nem más, mint a két változót függetlennek tekintő  $\rho_1(q_1)\rho_1(q_2)$  eloszlás és az együttes  $\rho_2(q_1, q_2)$  eloszlás közötti Kullback-Leibler divergencia, vagyis a kölcsönös információ a független modell és a valódi eloszlás közötti információkülönbséget méri.

A (33) kifejezés általánosítható  $k$  változóra olyan módon, hogy a független modell helyett a  $k$  változó közös eloszlását az alacsonyabb rendű sűrűségfüggvényekből származtatjuk. Ezek alapján a  $k$  változó kölcsönös információját [22]

$$I_k(q_1, \dots, q_k) = (-1)^k \int \rho_k(q_1, \dots, q_k) \ln \frac{\rho_k(q_1, \dots, q_k)}{\hat{\rho}_k(q_1, \dots, q_k)} \prod_{i=1}^k dq_i \quad (34)$$

Egyezést a kölcsönös információ korábbi, (31) definíciójával akkor kapunk, ha a független modellt az ún. általánosított Kirkwood-szuperpozíció közelítéssel (GKSA) [23] előállított  $\hat{\rho}_k$  sűrűségfüggvénnyel helyettesítjük, amely a  $k$  változós sűrűségfüggvényt az alacsonyabb rendű sűrűségfüggvények segítségével becsli a

$$\hat{\rho}_k(q_1, \dots, q_k) = \prod_{s=1}^{k-1} \left( \prod_{l_1 < \dots < l_s} \rho_s(q_{l_1}, \dots, q_{l_s}) \right)^{(-1)^{s+k-1}} \quad (35)$$

alakban, amely a két változóra érvényes  $\hat{\rho}_2(q_1, q_2) = \rho_1(q_1)\rho_1(q_2)$  összefüggés általánosítása. A GKSA által adott sűrűségfüggvényről megmutatható, hogy az eggyel alacsonyabb rendű sűrűségfüggvények ismeretében a legjobb közelítést adja a valódi sűrűségfüggvénynek [24]. A  $k$ . rendű kölcsönös információ tehát a (34) alapján azt az információmennyiséget méri, amely már nem állítható elő a  $k-1$ . rendű eloszlások segítségével, vagyis a koordináták közötti  $k$ . rendű korrelációk mértékét.

A kölcsönös információn alapuló kifejtést (MIE) számos entrópiabecslő módszer felhasználja a másod- és magasabbrendű korrelációk figyelembevételére. Az egyik ilyen a Numata és munkatársai által létrehozott módszer, amely a kváziharmonikus módszerrel kapott entrópiát kiegészíti a módusok anharmonicitásáért és a közöttük fellépő másodrendű korrelációkért felelős korrekciós tagokkal [25]. A Numata-módszer által számolt entrópia

$$S_{Numata} = \sum_{i=1}^N S_1(q_i) - \sum_{i < j} I_2(q_i, q_j) \quad (36)$$

amelyből a kölcsönös információt a módszer a (28) képlet alapján számolja, ezáltal csak egyváltozós és kétváltozós együttes entrópia számolására van szükség. Az eloszlások anharmonicitásának figyelembevételére a Numata-módszer a nem-paraméteres  $k$ -legközelebbi



szomszéd ( $k$ -NN) entrópiabecslést használja [26, 27]. A  $k$ -legközelebbi szomszéd módszer az egyes adatpontok helyén ad becslést a mintavételezett sűrűségfüggvényre az adatpont  $k$ . legközelebbi szomszédjától való távolsága alapján, az

$$\hat{f}(x_i) = \frac{k}{n} \frac{1}{V_s(R_{i,j})} \quad (37)$$

formula segítségével, ahol

$$V_s(R_{i,k}) = \frac{\pi^{s/2} R_{i,k}^s}{\Gamma\left(\frac{1}{2}s + 1\right)} \quad (38)$$

az  $s$  dimenziós,  $R_{i,k}$  sugarú gömb térfogata,  $R_{i,k}$  az  $i$ . adatpont és a  $k$ . legközelebbi szomszédjának távolsága,  $n$  pedig az adatpontok száma [26]. A sűrűségfüggvény alapján a teljes eloszlás entrópiája becsülhető a

$$\lim_{n \rightarrow \infty} \left( -\frac{1}{n} \sum_{i=1}^n \ln \hat{f}(x_i) \right) = \lim_{n \rightarrow \infty} \left( \sum_{i=1}^n \ln \frac{n\pi^{s/2} R_{i,k}^s}{k\Gamma\left(\frac{1}{2}s + 1\right)} \right) = H + L_{k-1} - \ln k - \gamma \quad (39)$$

kifejezés segítségével, amelyben  $H$  az egzakt információs entrópia, a korrekciós tagok pedig  $L_0 = 0$ ,  $L_j = \sum_{i=1}^j 1/i$ ,  $j \geq 1$ , és  $\gamma = 0,5772\dots$  az Euler-konstans [28]. Az eloszlás entrópiája ezek alapján

$$H_{NN} = \frac{s}{n} \sum_{i=1}^n \ln R_{i,k} + \ln \frac{n\pi^{s/2}}{\Gamma\left(\frac{1}{2}s + 1\right)} - L_{k-1} + \gamma \quad (40)$$

A legközelebbi szomszéd entrópiát számoló módszerek előnye, hogy a mintát generáló eloszlás függvényalakjára nem tesznek *a priori* feltételezéseket, és nem használnak paramétereket. Hátrányuk viszont, hogy a konvergencia eléréséhez nagyszámú ( $n \sim 10^6$ ) mintára van szükség [29]. A minta mögött álló folytonos eloszlás megbecslésére más módszereket is alkalmaznak, pl. paraméteres becslést Fourier-sorok segítségével [27], vagy anizotróp magfüggvényekkel történő nem-paraméteres becslést [30], amelyek a legközelebbi szomszéd módszerhez hasonlóan lassú konvergenciát mutatnak. Egészen más megközelítést alkalmaznak Meirovitch és munkatársai, akik egy komplex rekonstrukciós algoritmust használnak az egyes konformációk valószínűségének kiszámítására [31-33]. Ennek a módszernek viszont a megvalósítása körülményes, és igen nagy a számításigénye.

A Wang és Brüschweiler-féle 2D entrópia módszere Gauss-magfüggvényeken alapuló sűrűségbecslést végez a konformációs térben, majd a differenciális entrópiát a (17) definíciója

alapján numerikus integrálással számolja [34]. A módszer a torziós szögeket komplex koordinátákká alakítja, amivel kiküszöböli a periodicitás problémáját, majd a számolást a

$$q^{(i)} = (q_1^{(i)}, \dots, q_N^{(i)}) = (e^{i\varphi_1^{(i)}}, \dots, e^{i\varphi_N^{(i)}}) \quad (41)$$

komplex adatpontokkal végzi, ahol  $\varphi_j^{(i)}$  az  $i$ . adatpont  $j$ . torziós koordinátája. Az adatpontokat ezután a komplex kovarianciamátrix

$$C_{ij} = \langle q_i \bar{q}_j \rangle - \langle q_i \rangle \langle \bar{q}_j \rangle = \langle e^{i\varphi_i} e^{-i\varphi_j} \rangle - \langle e^{i\varphi_i} \rangle \langle e^{-i\varphi_j} \rangle \quad (42)$$

diagonalizálásával kapott  $m_k$  főkomponensekre vetíti. A vetített

$$c_k^{(i)} = q^{(i)} \cdot \bar{m}_k \quad (43)$$

adatpontok entrópiáját a módszer főkomponensenként számolja ki. A  $k$ . főkomponens esetén a  $c_k^{(i)}$  adathalmazt a módszer a komplex sík feletti folytonos eloszlássá alakítja egy előre meghatározott  $\sigma$  varianciájú Gauss-függvénnyel történő konvolúcióval, és az így kapott

$$\hat{f}_k(z) = \frac{1}{n} \sum_{i=1}^n \frac{1}{2\pi\sigma^2} e^{-\frac{(z-c_k^{(i)})(\overline{z-c_k^{(i)}})}{2\sigma^2}} \quad (44)$$

sűrűségfüggvénynek az entrópiáját a (17) definíció alapján a

$$S_{2D,k} = -k_B \int \hat{f}_k(x+iy) \ln \hat{f}_k(x+iy) dx dy \quad (45)$$

kifejezés numerikus integrálásával kapja. Az egy állapothoz tartozó  $S_{ref} = k_B \ln(2\pi e \sigma^2)$  entrópiáértéket a módszer referenciaként kezeli, és levonja az egyes főkomponensekre számolt entrópiáértékekből. A teljes entrópia ezután az

$$S_{2D} = \sum_k (S_{2D,k} - S_{ref}) \quad (46)$$

összefüggés alapján, az egyes főkomponensekre kapott értékek összegzésével kapható. A 2D entrópia módszere már képes figyelembe venni az egyes főkomponensek mentén történő fluktuációk anharmonicitását, azonban tartalmaz egy, a sűrűségbecslés sávszélességét meghatározó paramétert ( $\sigma$ ), amely nem vihető át különböző rendszerek között [34, 35].

## 1.2. ABC fehérjék

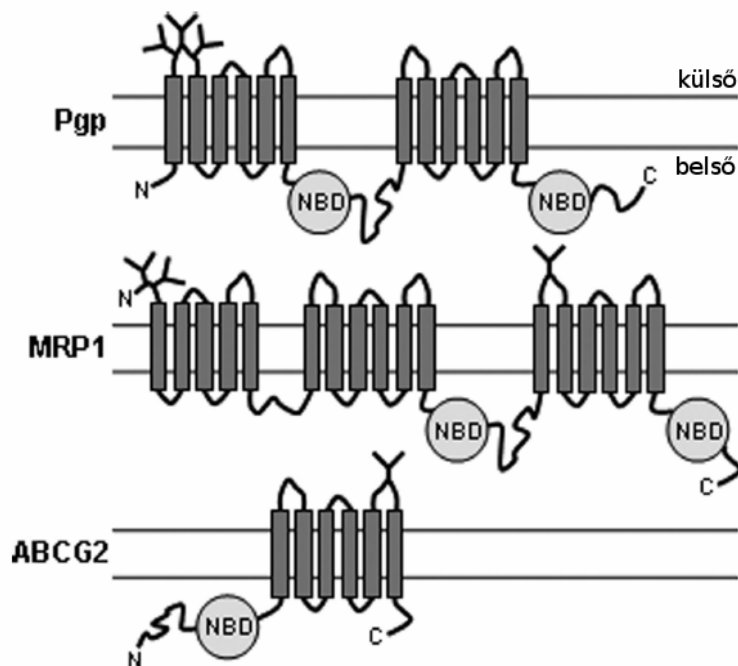
Az ATP-kötő kazetta (ATP Binding Cassette, ABC) fehérjék egy főként membránfehérjéket tartalmazó, minden organizmusban jelen lévő fehérjecsalád, amelyek fontos szerepet játszanak különféle vegyületeknek biológiai membránokon való átjuttatásában [36]. Az ABC fehérjék megtalálhatók prokariótákban, a sejt számára alapvető tápanyagok (cukrok, aminosavak, fémkomplexek) felvételére szolgáló importerek és a káros anyagok eltávolítására is képes exporterek formájában. Eukariótákban csak ABC exporter fehérjék ismertek. Az emberi genomban eddig 49 ABC fehérjét azonosítottak [37], amelyek számos különböző szubsztrát transzportjában vesznek részt (pl. foszfolipidek, szterolok, gyógyszerek, exotoxinok) [36]. Az ABC fehérjéknek akár funkciójuk elvesztése, akár túlzott működésük okozhat patológiás állapotot. Az ABC fehérjék többségében található mutációk különféle betegségeket okozhat, mint amilyen a cisztás fibrózis [38, 39], a Dubin-Johnson szindróma [40], a kettes típusú diabetes mellitus [41] vagy az adrenoleukodisztrófia [42].

A daganatok kemoterápiás kezeléssel szemben mutatott széleskörű ellenállóképességének, multidrog-rezisztenciájának (MDR) egyik fő mechanizmusáért is ABC fehérjék felelősek [36, 43, 44]. A tumorsejtek rezisztenciájának leggyakoribb oka a sejtből gyógyszerek kipumpálására alkalmas ABC transzporterek megnövekedett expressziója. Ezeknek a transzportereknek (pl. P-glikoprotein/multidrog-rezisztencia fehérje 1/MDR1, MRP1, BCRP/ABCG2) a szubsztrátspecificitása alacsony, ezért változatos kémiai tulajdonságú vegyületeket tudnak eltávolítani a sejtből, és így az egyes gyógyszerek sejten belüli koncentrációját a hatásos szint alá csökkenthetik. A kemoterápiás drogok és más, pl. idegrendszeri gyógyszerek hatékonyságának a növeléséhez kívánatos lenne a drogrezisztenciát okozó ABC fehérjék működésének modulálása kis molekulatömegű gátlószerekkel. Ehhez azonban az ABC exporterek szerkezetének és működésének a részletes ismerete szükséges.

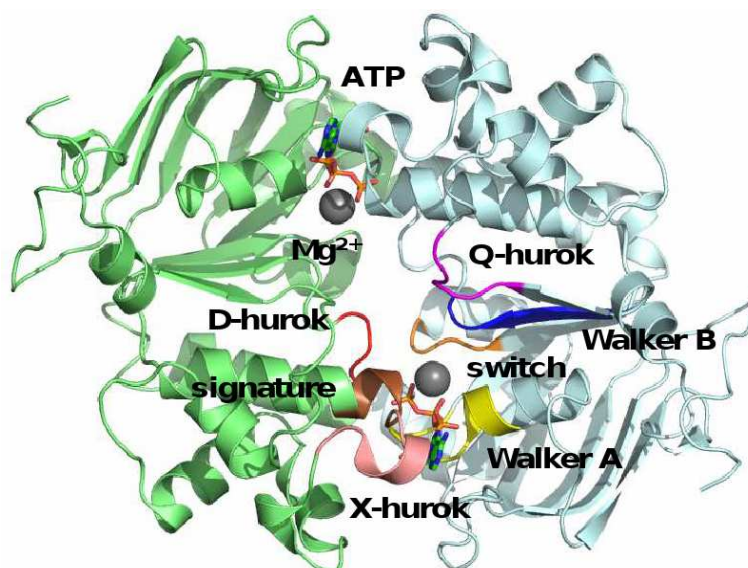
### 1.2.1. ABC exporterek felépítése és konformációi

Az ABC exporterek membránba ágyazott fehérjék, amelyek a citoplazmatikus oldalon két nukleotidkötő domént (NBD), a membránon belül pedig két transzmembrán domént (TMD) tartalmaznak [45, 46]. Bizonyos exporterek esetén egy fehérjeláncon egy NBD és egy TMD található („féltranszporterek”), más esetben egy láncon található mind a négy domén (1. ábra). A „féltranszporterek” homo- vagy heterodimerként működnek [36]. Az exportfolyamatot az ATP nukleotidkötő doménekhez történő kötődése és/vagy hidrolízise

biztosítja. Az NBD-k két aldóménre oszthatók, amelyek közül a Rec A típusú aldóménben [47] helyezkednek el az ATP kötőhelyét alkotó ún. Walker A és B motívumok [48], az  $\alpha$ -helikális aldóménben pedig az ún. *signature* szekvencia található (2. ábra) [37]. A katalízis szempontjából nélkülözhetetlen oldalláncok találhatóak a Walker B és az ún. *switch* motívumban (más néven H-hurok) [49]. Biokémiai és szerkezeti vizsgálatok megmutatták, hogy az ATP a Walker motívumokhoz és a szemközti NBD-ben található *signature* szekvenciához kötődik, és a bekötődés az NBD-k szoros asszociációját („dimerizációját”) teszi lehetővé (2. ábra) [50, 51]. A katalízishez végül a Walker B motívumban levő Glu oldallánc koordinálja a nukleofil támadást végző vízmolekulát, a H-hurokban található His oldallánc pedig a  $\gamma$ -foszfát pozícionálásában vesz részt [52]. Az ATP-kötött holo állapotban a transzmembrán domének citoplazmatikus szakaszai is kontaktusba lépnek egymással, és egy ún. „alul zárt” konformációt alkotnak (3.A. ábra) [53, 54]. Nukleotidok hiányában a röntgendiffrakciós szerkezetekben az NBD-k eltávolodást mutatnak egymástól (3.B. és 3.C. ábra) [54-56].



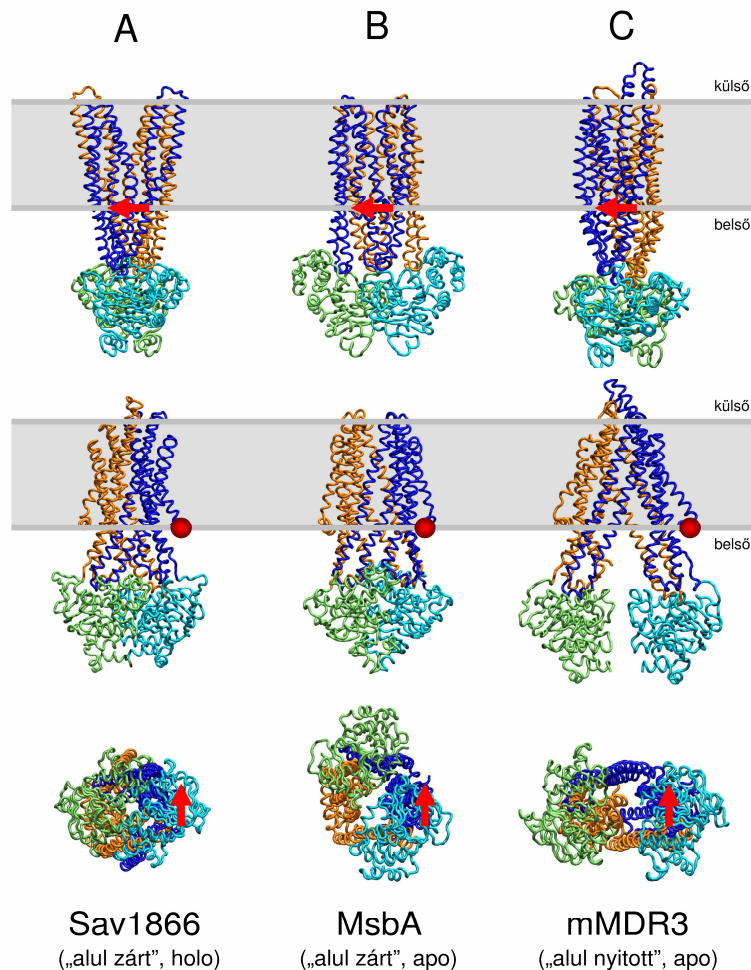
**1. ábra.** A multidrog-rezisztenciában szerepet játszó ABC exporter fehérjék felépítése. [36] A P-glikoprotein (P-gp) és az MRP1 fehérje esetén a működéshez szükséges domének egy polipeptidláncon találhatóak. Az ABCG2 fehérje ún. „féltranszporter”, amely homodimerként működik. A nukleotidkötő domének (NBD) mellett az ábrán a transzmembrán hélixek topológiája is látható. A TMD-k egyenként 6 transzmembrán hélixet tartalmaznak (szürke téglalapok). Az MRP1 esetén az N-terminálison további 5 transzmembrán hélix található, amely azonban a transzporthoz nem szükséges [57]. A glikozilációs helyeket Y alakú jelek mutatják a fehérjeláncokon.



**2. ábra. ATP-kötött NBD dimer szerkezete. (PDB ID: 2IXF)** Az ábra az extracelluláris irányból nézve mutatja az NBD dimer fontosabb szerkezeti elemeit. A Walker A, B (sárga és kék) és a *signature* (barna) motívumok az ATP kötésében vesznek részt, a *switch* motívum (narancs) és a D-hurok (piros) az ATP hasításáért felelősek. Az X-hurok (rózsaszín) és a Q-hurok (lila) feltehetően NBD és a TMD közötti allosztérikus kommunikációban vesznek részt. A Q-hurok tartalmazza az ATP és ADP megkülönböztetésére alkalmas Gln oldalláncot. Az ATP kötött konformációját pálcikamodell mutatja, a kötött  $Mg^{2+}$  ionok szürke gömbökként láthatóak.

Az NBD-k helyzetétől függően megkülönböztethetünk egy ún. „alul zárt” (3.B. ábra) és egy „alul nyitott” (3.C. ábra) apo szerkezetet. Az NBD-k közötti távolság a jelenleg ismert három apo szerkezetek alapján széles tartományban mozoghat, az egér MDR3 fehérje esetében (PDB ID: 3G5U) az NBD-k tömegközéppontjának távolsága 44,8 Å, az MsbA bakteriális exporter „alul zárt” apo szerkezetében (PDB ID: 3B5X) 42,3 Å, az „alul nyitott” szerkezetében (PDB ID: 3B5W) még nagyobb, 74,8 Å. A megoldott szerkezetek sokfélesége és különböző droggötődési vizsgálatok is rávilágítottak arra, hogy az ABC transzporterek nagyon flexibilisek [58], ezért egyetlen statikus konformáció, mint egy röntgendiffrakciós szerkezet, valószínűleg nem elegendő a fehérje szerkezetének leírására és a droggötőhelyek azonosítására. Az ABC fehérjék dinamikai jellemzőit ezért kísérleti és elméleti módszerekkel egyaránt sokat vizsgálják, azonban a kísérletekből kapott eredmények nincsenek összhangban egymással és a kristályosítás során kapott szerkezetekkel. Az utóbbi időszak egyik gyakran alkalmazott kísérleti módszere az elektron paramágneses rezonancia (EPR) spektroszkópia [59-62] és az EPR jelenségen alapuló dupla elektron-elektron rezonancia (DEER) spektroszkópia [63], amellyel megmérhető két spinjelölt aminosav távolsága a célfehérje különböző konformációiban. Az MsbA fehérjén végzett DEER mérések azonban azt mutatták, hogy az NBD-k átlagos eltávolodása meghaladja azt, ami az „alul nyitott” röntgendiffrakciós szerkezeten is mérhető (PDB ID: 3B5W) [63]. Két adott oldallánc közötti

távolság kémiai keresztkötéssel (*cross-linking*) is megmérhető. Humán MDR1 fehérjén végzett keresztkötési vizsgálatok azt mutatják, hogy ha a TMD-k citoplazmatikus szakaszainak távolságát kisméretű kémiai reagenssel rögzítjük, akkor is aktív marad [64]. A működéshez ez alapján az NBD-k túlzott eltávolodása sem szükséges, így tehát nem egyértelmű, hogy fiziológiai körülmények mellett az NBD-k az apo állapotban mennyire távolodnak el egymástól. Ugyanakkor mindkét kísérleti módszernek megvannak a maga korlátai, és nem adnak atomi szintű információt a fehérjében történő mozgásokról. A legtöbb esetben ezek a kísérleti módszerek csak a fehérjében levő két pont közötti távolságról adnak információt különböző körülmények mellett (pl. hőmérséklet, ATP vagy szubsztrát jelenléte vagy hiánya), míg a két pont elmozdulásának irányát nem fedik fel. Az atomi felbontású számítógépes módszerek ebben az esetben segíthetnek a kísérleti adatok értelmezésében.



**3. ábra. ABC exporterek konformációi.** A jelenleg megoldott röntgendiffrakciós szerkezetek alapján az ABC exportereknek háromféle konformációja ismert. Nukleotid jelenlétében megfigyelhető az „alul zárt” holo forma (PDB ID: 2HYD) (A), amelyben az NBD-k szorosan közrefogják a kötött nukleotidokat. Az apo forma esetén az „alul zárt” apo szerkezetben (PDB ID: 3B5X) (B) az NBD-k nincsenek kontaktusban egymással, de távolságuk nem jelentős. Az „alul nyitott” apo szerkezetben (PDB ID: 3G5U) (C) az NBD-k eltávolodása jelentős, és a transzmembrán domének citoplazmatikus oldalán nagyméretű nyílás figyelhető meg. Az ábrán a Sav1866, és MsbA **bakteriális** és az egér MDR3 fehérjék láthatóak, háromféle orientációból. Referenciaként a membránnal párhuzamos ún. „könyökhélixre” fektetett piros nyíl szolgál. A membrán helyét szürke sáv jelöli.

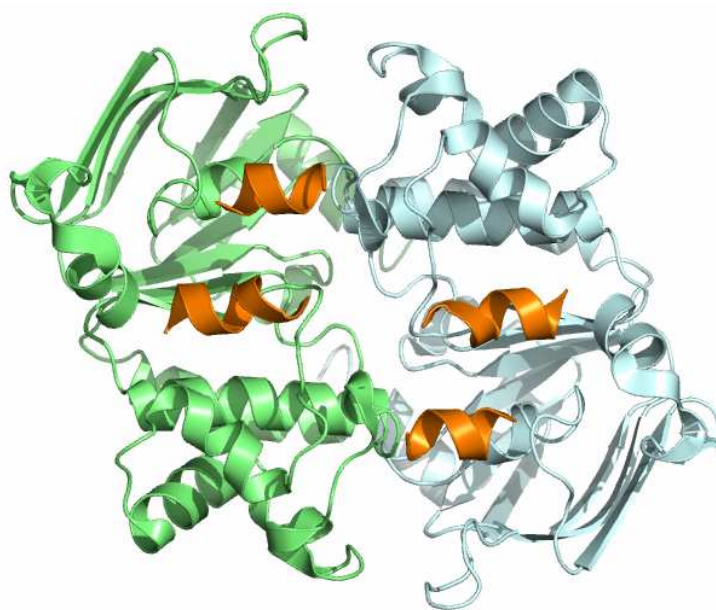
## 1.2.2. ABC exporterek működési mechanizmusa

Az ATP bekötődése során a két nukleotidkötő domén szoros kölcsönhatásba lép egymással, amely megteremti az ATP hasításának a lehetőségét. Mivel mindkét NBD tartalmaz nukleotidkötő motívumokat, egy teljes transzporterben összesen két ATP bekötődésére és hasítására van lehetőség. A két ATPáz hasítóhely működési mechanizmusára többféle modell született [65]. Az ún. *alternating sites* modell [66] szerint a transzport során a két nukleotidkötő-hely felváltva kerül egy olyan, az ATP-t erősen kötő állapotba, amely lehetővé teszi a nukleotid hidrolízisét. A modell szerint a hidrolízis után (valószínűleg a keletkezett töltések taszító hatására [50]) az ATP kötőhely meggyengül és felnyílik, amely lehetővé teszi a nukleotid kicserélődését az adott kötőhelyen. A szubsztrát transzportja a membrán túlsó oldalára a nukleotid hidrolízisével egyidőben történik. A hidrolizált nukleotid kicserélődése közben a szemközti kötőhelyen létrejöhethet az ATP-t erősen kötő konformáció, és egy újabb nukleotid elhasításával egy következő szubsztrát transzportja is bekövetkezhet. Az viszont nem egyértelmű, hogy a felnyílás a két NBD teljes disszociációjával jár-e, és hogy egy szubsztrát transzportjához mindkét ATP elhasadása szükséges-e [66, 67]. Az irodalomban leírt másik modell a transzportmechanizmusra az ún. *processive clamp* működés [68-70]. Ez a modell a két ATP egymást követő elhasadását jósolja, amely után az NBD-k disszociálnak, és a transzporter az „alul nyitott” konformációt veszi fel. A szerkezetekben megfigyelt aszimmetria miatt a két nukleotid hidrolízise valószínűleg nem történik egyidejűleg. A *processive clamp* modellt azonban izolált NBD-k vizsgálatával kapott adatok alapján állították fel, ezért elképzelhető, hogy nem tükrözi helyesen a teljes transzporterben végbemenő eseményeket.

Multidrog exporterekben a hidrolitikus aktivitást a transzportálandó szubsztrát bekötődése rendszerint fokozza, de a hidrolízis valószínűleg szubsztrát nélkül is végbemehet, ami létrehoz egy ún. alapaktivitást. Mivel a transzportot az ATP bekötődése és hidrolízise hajtja, az ATP kötőhely és a transzmembrán domének között kell, hogy legyen valamilyen kommunikáció. Az NBD és TMD közötti interfészben több olyan hurkot és oldalláncot azonosítottak, amelynek szerepe lehet a transzporterben belüli kommunikáció megvalósításában. Az interfész egyik fontos régióját képezik az intracelluláris hurkok rövid szakaszai, az ún. *coupling* hélixek (4. ábra) [53]. Ezek a mindössze két fordulatból álló hélixek a membránnal párhuzamos állásúak, és két antiparallel hélixet kötnek össze, amelyek a transzmembrán hélixek intracelluláris meghosszabbításai. A *coupling* hélixek az NBD-ken a Rec A-szerű és a helikális aldomén közötti árokban foglalnak helyet, ezáltal létesítve

kapcsolatot az NBD-k és a TMD-k között [53]. Az NBD-k oldalán a *coupling* hélixekkel kölcsönható egyik fontos régió az ún. Q-hurok (2. ábra), amely az ATP  $\gamma$ -foszfátjával és az ATP-hez kötött fémmionnal egy konzervált Gln oldallánc segítségével létesít kapcsolatot [53]. Ez az oldallánc képes a nukleotidkötő-helyen levő ATP és ADP megkülönböztetésére [71], ezért vélhetően a domének közötti kommunikációban is fontos szerepet játszik. Az NBD-k és TMD-k közötti kapcsolódási felület másik fontos régiója a nukleotidkötő doménekben a *signature* szekvencia N-terminális oldalán található ún. X-hurok (2. ábra), amely az ABC fehérjék B és C családjában konzervált GERG szekvenciamotívumot tartalmazza. Az X-hurok a szemközti NBD Rec A-szerű aldoménhez kapcsolódó *coupling* hélixszel lép kölcsönhatásba, és fontos szerepet játszik az ATP hidrolízis és a transzport csatolásában [72]. Szintén fontos szerkezeti elem az „alul zárt” apo szerkezetben található ún. tetrahélix-köteg, amely mindkét TMD-ből a TM3 és TM4 transzmembrán hélixek citoplazmatikus szakaszaiból áll [73]. Ez a négy hélix a transzporter hossz tengelye mentén egy szoros hélixköteget alkot, amely elzárja a citoplazmatikus oldalon a transzporter belső csatornáját. Célzott (*targeted*) molekuláris dinamikai szimulációk azt mutatják, hogy a tetrahélix-köteg szétesése a transzporter citoplazmatikus felnyílásának egy lényeges lépése, amely az NBD-k disszociációja után következik be [73].

Az ATP kötődésének és hidrolízisének a TMD konformációváltozásaihoz és a szubsztrát transzlokációjához való csatolódásának megértése nagymértékben segítené olyan drogok tervezését, amelyekkel a transzporter működése szabályozható lenne.



**4. ábra. A *coupling* hélixek elhelyezkedése.** Az ábrán a két NBD (zöld és cián) és a velük kölcsönhatásba lépő *coupling* hélixek (narancs) látható az extracelluláris irányból nézve a Sav1866 fehérje szerkezete (PDB ID: 2HYD) alapján. A *coupling* hélixek a transzmembrán domének intracelluláris végeit átkötő  $\alpha$ -hélixek, amelyek fontos szerepet játszanak az NBD-k és a TMD-k közötti allosztérikus kommunikációban.



### 1.2.3. Molekuláris dinamikai számítások ABC fehérjékkel

Az ABC fehérjék belső mozgásainak atomi felbontású felderítésére széles körben alkalmaznak molekuláris dinamikai (MD) szimulációkat. Több számításos vizsgálat szól az ATP stabilizáló hatásáról csak izolált NBD dimereket tartalmazó rendszerekben [67, 74-78]. Az NBD dimerekkel vagy monomerekkel ATP vagy ADP jelenlétében végzett szimulációk fényt derítenek a hidrolízis által indukált lehetséges konformációs változásokra. Annak ellenére, hogy nincs egyértelmű utalás arra, hogy az egyik NBD-ben történt hidrolízis elősegíti vagy gátolja-e a szemközti (*trans*) kötőhelyen a hidrolízist [77, 78], a legtöbb szimulációs eredmény és félig nyitott röntgendiffrakciós szerkezet azt jelzi, hogy a helikális aldómén a *trans* NBD-ben kifelé fordul az ATP hidrolízise után [67, 74], ami elősegítené a nukleotid kicserélődését. Ez a kifordulás a *trans* NBD-ben a Q- és X-hurkok együttmozgását jelentené, amikről kísérleti módszerekkel megmutatták, hogy fontos szerepet játszanak a TMD felé történő kommunikációban [72, 79].

Kevés tanulmány számol be teljes hosszúságú ABC fehérjékkel lipid környezetben végzett szimulációkról, mivel a biológiailag releváns időskálákon való mintavételezés ilyen nagy rendszerek esetén magas számítási teljesítményt kíván. A bakteriális B12 vitamin importer (BtuCD) rendszert kiterjedten vizsgálták MD szimulációk segítségével. Lipid kettősrétegben végzett 15 ns időtartamú szimulációkkal megmutatták, hogy az ATP kötődés az NBD-kben és a TMD-kben is indukál konformációs változásokat, ami azt jelzi, hogy az ATP kötődés maga lehet a katalitikus ciklus hajtóereje [80]. Elasztikus hálózati analízist és irányított molekuláris dinamikai szimulációkat szintén alkalmaztak a B12 vitamin transzport lehetséges mechanizmusának feltárására [81]. A BtuCD rendszert részletesen vizsgálták Ivetac és munkatársai [82], akik molekuláris dinamikai számítások és főkomponens-analízis segítségével aszimmetrikus záródást mutattak ki az ATP-kötött transzporter nukleotidkötő doménjeiben, ami az ABC transzporterek alternáló hidrolízis mechanizmusát [66] támasztja alá. A teljes Sav1866 transzporterrel végzett szimulációk segítségével sikerült rámutatni olyan oldalláncokra a Q- és X-hurkokban, amelyek egymástól függően több különböző konformációs állapotot is felvehetnek, és ezáltal molekuláris kapcsolóként továbbíthatják a konformációs változásokat az NBD és a TMD között [71]. Mindmáig azonban nem történt szisztematikus vizsgálat az NBD-ben eredő és a TMD felé továbbított konformációs változások teljes útvonaláról.

Molekuláris dinamikai vizsgálatokat használtak az ABC fehérjék globális stabilitásának és szerkezeti integritásának felmérésére is. A mostanra visszavont MsbA

röntgendiffrakciós szerkezet, amely a megfelelő számú transzmembrán hélixet tartalmazta, de rossz irányultságban és topologikus elrendezésben, még rövid szimulációk során is jelentős torzulásokat szenvedett [83].

## 2. Célkitűzések

A munkám első részében egy új elven alapuló módszert dolgozok ki molekuláris sokaságok konformációs entrópiájának kiszámítására. Az új módszer alapja a konformációs térbeli sűrűségfüggvény Gauss-keverékekkel történő közelítése, amit a sokaságbeli atomi konformációkra történő illesztéssel hozok létre. A kapott folytonos sűrűségfüggvény információs entrópiája megbecsülhető, amelyből a konfigurációs entrópia kiszámítható. A módszer tesztelésére öt kis peptidrendszert választok ki, amelyek egzakt konfigurációs entrópiája a teljes konfigurációs tér feletti numerikus integrálással kiszámítható. Az egzakt entrópiaértéket használva referenciaként, az új módszert az irodalomban található elterjedt módszerekkel hasonlítom össze.

A munka második részében egy információs entrópia jellegű mennyiség, a kölcsönös információ segítségével jellemzem az ABC exporter fehérjék dinamikáját. Az ATP hidrolízise által okozott dinamikai változások feltérképezésére a Sav1866 **bakteriális** transzporter nukleotidkötött szerkezetét használom fel. A hidrolízis előtti és utáni állapot modellezésére létrehozok két, lipid kettősrétegbe ágyazott szimulációs rendszert, melyből az egyik két kötött ATP molekulát, a másik pedig egy ATP és egy ADP molekulát tartalmaz. Ezeket a rendszereket az ATP hidrolízise előtti és utáni állapotot modellezésére hozom létre. A több, hosszú molekuláris dinamikai szimulációból kapott szerkezetek alapján kiszámolom a fehérje konformációját leíró torziós szögek közötti kölcsönös információt, amelynek segítségével egymással korrelált mozgást végző aminosavakat keresek. A korrelált aminosavak mintázatában a nukleotidcsere hatására bekövetkezett változásokat elemzem és értelmezem.

Az „alul nyitott” apo és az „alul zárt” holo forma dinamikai összehasonlítására és a közöttük vezető lehetséges átmeneti útvonalak leírására létrehozom a hMDR1 fehérje apo konformációjának homológiamodelljét. A két szerkezetet lipid kettősrétegbe ágyazva hosszú időtartamú molekuláris dinamikai szimulációkat végzek, és a kapott trajektóriákban olyan kollektív mozgásokat keresek, amelyek a két konformáció közötti átmenet irányába mutatnak.

## 3. Módszerek

### 3.1. Molekuláris dinamikai szimulációk

A molekuláris dinamika (MD) módszere az atomi rendszerek időfejlődésének számítógépes szimulációja. Emellett egy fontos és nagyon kézenfekvő statisztikus fizikai mintavételezési módszer, hiszen az ergodikus hipotézis alapján egy kellően hosszú szimuláció során kapott szerkezetekből, az ún. trajektóriából, a rendszer egyensúlyi tulajdonságai kiszámíthatóak.

Bár atomi rendszerek esetén a használt térbeli felbontás megkövetelné a kvantummechanikai tárgyalásmódot, a gyakorlatban azonban molekuláris biológiai rendszerek esetén ez bonyolulttá és körülményessé, gyakran kivitelezhetetlenné válhat. Ezért a legtöbb, biomolekulákra végzett MD szimuláció az atomokat klasszikus, töltéssel rendelkező tömegpontokként kezeli, és mozgásukat a klasszikus (Newton-féle) mozgásegyenletek megoldásával számolja:

$$m_i \frac{d^2 \mathbf{r}_i}{dt^2} = \mathbf{F}_i(\mathbf{r}) = -\frac{\partial}{\partial \mathbf{r}_i} U(\mathbf{r}) \quad (47)$$

ahol  $m_i$  és  $\mathbf{r}_i$  az  $i$  atom tömege és pozíciója,  $U(\mathbf{r})$  pedig az atomi kölcsönhatásokat leíró potenciálfüggvény. Az  $U(\mathbf{r})$  kölcsönhatási potenciálban a kvantummechanikai hatásokat effektív potenciálok bevezetésével veszik figyelembe. Ezek leírására rendszerint empirikus összefüggéseket használnak, és a modell paramétereit úgy állítják be, hogy a lehető legjobban reprodukálja a makroszkopikus rendszer tulajdonságait, és hogy minél több rendszerre transzferálható legyen. Az így kapott, különböző molekulák leírására használható modellt nevezzük *force-field*nek (erőtérnek). A klasszikus leírásmód ellenére a szobahőmérsékleten fellépő mozgások egy része, pl. a hidrogénatomok mozgásai, bizonyos kötéhosszak és kötéhosszok vibrációi már nem tekinthetők klasszikusnak, inkább egy oszcillátor alapállapotú rezgésének felelnek meg. Ezek a mozgások tehát igen kötöttek, ezért a hidrogénatomok relatív pozícióit vagy a kovalens kötések gyakran **kényszerfeltételként** (*constraint*) veszik figyelembe a szimulációban, és állandó értéken tartják. A potenciálfüggvény konstruálásakor alkalmazott másik fontos elhanyagolás, hogy párkölcsönhatások összegeként épül fel, emiatt többtest-kölcsönhatásokat nem képes figyelembe venni. Az egyik legfontosabb ilyen jelenség az atomi polarizáción keresztül történő kölcsönhatás, amit így effektív párkölcsönhatásként vesznek figyelembe.

A mozgásegyenletek megoldására elvben bármilyen numerikus integrálási módszer használható, az egyik gyakori választás a Verlet-integrátorok családja (pozíció Verlet, sebesség Verlet és Leapfrog algoritmusok [84-86]). Az általam használt GROMACS [87, 88] programcsomag a Leapfrog algoritmust használja. A Verlet-család tagjai ún. másodrendű szimplektikus integrátorok, emiatt jellemzőjük, hogy a rendszer mozgásállandóit (energia, impulzus, perdület) megtartják, és segítségével oszcilláló mozgások is stabilan számíthatók [89]. Az integrálás egyik kritikus paramétere az időlépés ( $\Delta t$ ), amelyet úgy kell beállítani, hogy a rendszer legnagyobb frekvenciájú mozgásainak karakterisztikus idejénél kisebb legyen. Biomolekuláris rendszerekben ezt rendszerint a hidrogénatomok mozgásai határozzák meg, ez alapján a leggyakoribb választás az időlépésre a  $\Delta t = 1$  fs. A hidrogénatomok és a kötéshosszak rögzítésével az időlépés kitolható akár 2 fs-ig is, és ezzel a szimulációk felgyorsíthatóak.

A kényszerfeltételek mellett végzett mozgás szimulálására különböző algoritmusokat dolgoztak ki, amelyek rendszerint az időlépés megtétele után, utólag korrigálják az atomi koordinátákat, hogy azok megfeleljenek a kényszerfeltételeknek. Az egyik ilyen a SHAKE algoritmus [90], amely a koordináták iteratív korrekciójával elégti ki a kényszerfeltételeket. Egy gyakran használt speciális, vízmolekulákra létrehozott változata a SETTLE [91]. A szimulációkban általam használt módszer a LINCS, amely nem iteratív és gyorsabb, stabilabb működést mutat, mint a SHAKE, azonban csak kötéshosszak és szabad kötésszögek állandó értéken tartására használható [92].

A molekuláris dinamikában használt *force-fields* függvényalakjai nagyon hasonlóak egymáshoz. A kölcsönhatási energia tagjai rendszerint kötő és nemkötő kölcsönhatásokra oszthatók. A kötő kölcsönhatások felelősek a molekula geometriájának, és így a kötéshosszak, kötésszögek és egyes torziós szögek fluktuációinak leírásáért. A kötő kölcsönhatások ezért legtöbbször harmonikus rezgéseket írnak le. A nemkötő kölcsönhatások az egymással geometriai kényszerben nem levő atomok között lépnek fel, felelősek a töltéstöltés (Coulomb) és a van der Waals kölcsönhatások leírásáért, de szükség esetén más tagokat is tartalmazhatnak. A paraméterek származtatása különböző módszerekkel történhet, egyeseket kvantumkémiai számításokból származtatnak, másokat pedig kismolekulák valamilyen makroszkopikus tulajdonsága (pl. szolvatációs szabadenergia) alapján illesztnek.

A szimuláció elindításakor a kezdeti molekuláris szerkezetben előfordulhatnak nem optimális geometriájú kötések, vagy közeli atomi kontaktusok, amelyek nagy potenciális energiával rendelkeznek. Ezek a szimulációban instabilitást és nem várt jelenségeket okozhatnak, ezért a kiindulási szerkezetet energiainimalizálásnak szokták alávetni. A

gyakorlatban ilyen előfordulhat, a szerkezetek előállítása során az alacsony felbontás vagy a módszer sajátosságai miatt keletkezhetnek közeli atomi kontaktusok, a szerkezetfinomításkor használt *force-field* pedig más geometriát preferálhat, mint a dinamikai szimulációkhoz használt. Az oldószer és lipidmolekulák, illetve behelyezett kofaktorok (pl. ATP) atomjai és a fehérjeatomok közötti esetleges ütközéseket, a hidrogénkötések nem optimális geometriáját is kijavíthatjuk. Az energiainimalizálási módszerek rendszerint a legközelebbi lokális minimumhelyet keresik meg. Erre többféle matematikai módszer is használható, a legegyszerűbb a gradiens (*steepest descent*) módszer, azonban a minimumhely környékén gyorsabb konvergenciát mutat a konjugált gradiens (*conjugate gradients*) módszere [88]. Nagyobb rendszerek esetén is hatékonyan használható, és a konjugált gradiens módszernél esetenként jobb konvergenciát mutat az L-BFGS (*limited-memory Broyden-Fletcher-Goldfarb-Shanno quasi-Newtonian minimizer*) módszer, amelyet munkámban is alkalmaztam, és a GROMACS programcsomagban megtalálható [93].

A mozgásegyenletek zárt rendszer esetén konzerválják a rendszer teljes energiáját, ezért a (47) egyenlet megoldásával kapott trajektória mikrokanonikus sokaságot eredményezne. A biológiai rendszerek azonban rendszerint állandó hőmérsékleten és állandó nyomás mellett működnek. A hőtartállyal való csatolás megvalósítására különböző számítási módszerek léteznek. A legegyszerűbb módszer az ún. gyenge csatolás (Berendsen termosztát) [94], amely a kinetikus energia időfejlődésére az alábbi, elsőrendű kinetikájú egyenletet írja fel

$$\Delta E_{kin} = (E_{kin,0} - E_{kin}) \frac{\Delta t}{\tau} \quad (48)$$

ahol  $E_{kin,0} = (3/2)Nk_B T_0$  a célhőmérséklethez ( $T_0$ ) tartozó kinetikus energia,  $\tau$  a csatolás paraméter,  $\Delta t$  az integrálás időlépése. A kinetikus energia megváltoztatását a módszer az atomi sebességek

$$\lambda = \left( 1 + \frac{\Delta t}{\tau} \left( \frac{T_0}{T} - 1 \right) \right)^{1/2} \quad (49)$$

faktorral történő átskálázásával éri el. A módszer előnye, hogy egyszerűen megvalósítható, a hőmérséklet elsőrendű kinetikával tart a célhőmérséklethez, és nem okoz fluktuációkat. A fenti faktorral történő átskálázásról azonban megmutatható, hogy egyensúly esetén nem kanonikus eloszlást hoz létre az állapotok között, és emiatt az atomi sebességek eloszlása is torzul. A kanonikus sokaság helyreállítására a kinetikus energia időfejlődését leíró (48) egyenlethez egy sztochasztikus tagot adnak hozzá, amely garantálja a kanonikus eloszláshoz

való konvergenciát [95]. A szimulációkban ezt a módosított Berendsen-termostátot használtam (a GROMACS programban `velocity-rescale` csatolás).

A nyomáscsatolás megoldása a hőmérsékleti csatoláshoz hasonló módon történik, a koordináták és a szimulációs doboz méretének átskálázásával. A gyenge csatolás módszerét [94] itt is lehet alkalmazni, ez esetben a nyomás megváltozását a

$$\Delta p = (p_0 - p) \frac{\Delta t}{\tau_p} \quad (50)$$

egyenlet fogja megadni, és a nyomás itt is elsőrendű kinetikával fog a célértékhez ( $p_0$ ) tartani.

A fenti folyamathoz a skálázási mátrixot a

$$\mu_{ij} = \delta_{ij} - \frac{\Delta t}{3\tau_p} \beta_{ij} (p_{0,ij} - p_{ij}(t)) \quad (51)$$

összefüggéssel definiálják, ahol a nyomást a háromdimenziós nyomástenzorral ( $p_{ij}$ ) veszik figyelembe. Ennek megfelelően  $p_{0,ij}$  a nyomástenzor célértéke, amely rendszerint csak diagonális elemeket tartalmaz,  $\beta_{ij}$  a kompresszibilitás tenzora,  $\tau_p$  a csatolás relaxációs ideje,  $\delta_{ij}$  pedig a Kronecker-delta. A nyomáscsatolás tehát lehet anizotróp, részlegesen izotróp (semiisotropic csatolás, csak az  $x$  és  $y$  irányokban egyező), illetve a (51) módosításával létrehozható olyan csatolás is, amellyel az egyes nyomáskomponensek helyett az  $x$ - $y$  síkra vonatkozó felületi feszültség tartható állandó érték közelében.

Az oldószer jelenlétét a legtöbb szimulációban explicit, atomi felbontású vízmolekulákkal veszik figyelembe. Bizonyos esetekben ez azonban mégsem kívánatos, például, ha a rendszer túl kicsi, és aránytalanul sok szimulációs időt vennének igénybe a vízmolekulákkal kapcsolatos számítások. Az oldószer jelenlétét más módszerekkel is figyelembe vehetjük, például az oldószer molekuláival való ütközés felfogható véletlen eseménynek, és a molekula dinamikája sztochasztikus folyamatként. Az ilyen, ún. sztochasztikus dinamika mozgásegyenlete a következő alakban írható fel [88]:

$$m_i \frac{d^2 \mathbf{r}_i}{dt^2} = -m_i \xi_i \frac{d\mathbf{r}_i}{dt} + \mathbf{F}_i(\mathbf{r}) + \sqrt{2m_i \xi_i k_B T} \mathbf{R}_i(t) \quad (52)$$

ahol  $\xi_i$  a súrlódási együttható és  $\mathbf{R}_i(t)$  egy véletlen zaj, amire  $\langle \mathbf{R}_i(t) \mathbf{R}_j(t+s) \rangle = \delta(s) \delta_{ij}$ . A sztochasztikus dinamika jól használható ún. implicit oldószer modellekkel, amelyekben az oldószer csak effektív elektrosztatikus, termikus és entrópiikus hatásain keresztül veszik figyelembe. Munkámban a sztochasztikus dinamikát használtam kisméretű molekulák kanonikus sokaságainak előállítására az oldószer explicit figyelembe vétele nélkül.

### 3.2. Monte Carlo mintavételezés

A Monte Carlo (MC) módszerek kifejezés általánosságban az algoritmusok olyan családját jelöli, amelyek véletlen mintavételezésen alapulnak. A Metropolis-féle Monte Carlo módszer segítségével fizikai rendszerek konfigurációs sokaságát állíthatjuk elő, amelyben az egyes konfigurációk kanonikus (Boltzmann-) eloszlást követnek [96]. A Metropolis MC módszer két lépésből áll: az első lépés az ún. előterjesztés (*proposal*), amely során a rendszer eddigi  $q$  konfigurációjából egy  $q'$  új konfigurációt állítunk elő véletlenszerűen. Az új konfiguráció előállítása rendszerint egy előre meghatározott transzformációkészletből, az ún. mozgáskészletből történő véletlenszerű választással történik. Második lépésben az ún. Metropolis-kritérium alapján az új konfigurációt elfogadjuk  $p = \min\{1, e^{-\Delta E/k_B T}\}$  valószínűséggel (*acceptance*), ahol az exponenciális kifejezés az új és a régi konfiguráció Boltzmann-valószínűségének hányadosa:

$$\frac{e^{-E(q')/k_B T}}{e^{-E(q)/k_B T}} = e^{-\Delta E/k_B T} \quad (53)$$

ahol  $\Delta E = E(q') - E(q)$  az új és a régi konfiguráció közötti energiakülönbség. Munkámban az MC mintavételezési módszert molekuláris konformációk kanonikus sokaságainak előállítására használtam, a  $q'$  konfigurációt egyenletes valószínűséggel választva a konfigurációs térből.

### 3.3. A teljes **konformációhalmaz** generálása és az **egzakt entrópia számolása**

Az egyes molekuláris rendszerekre a konfigurációs tér uniform mintavételezésével létrehoztam az ún. teljes **konformációhalmazt**, amelyet az egzakt entrópia kiszámításához és a Monte Carlo (MC) mintavételezett sokaságok előállításához is felhasználtam. Ehhez előállítottam a konfigurációs tér  $\varphi = (k_1\varphi_0, k_2\varphi_0, \dots, k_d\varphi_0)$  rácspontjaiban levő szerkezeteket, ahol  $k_i$  egész számok, a  $\varphi_0$  pedig egy rögzített mintavételezési távolság (rácsállandó). Csak azokat a  $k_i$  értékeket használtam fel, amelyekre  $k_i\varphi_0$  a  $[0^\circ, 360^\circ)$  intervallumba esik. **A molekulák kezdő konformációját a SYBYL 7.3 programcsomag segítségével állítottam elő. A geometria optimalizálására a szerkezeteken energiminimalizálást hajtottam végre az L-BFGS módszerrel. Leállási kritériumnak azt választottam, amikor a fellépő maximális atomi erő (emtol)  $50 \text{ kJ mol}^{-1} \text{ nm}^{-1}$  alá csökken. A rácspontoknak megfelelő szerkezeteket ezután a belső koordinátákhoz tartozó torziós szögek mentén történő forgatással állítottam elő, amihez saját készítésű Python szkripteket használtam. A  $\varphi_0$  értékét úgy választottam meg, hogy a teljes konformációhalmazban 40–60 millió szerkezetet kapjak.**



Az egzakt konfigurációs entrópiát numerikus integrálással számoltam ki, amelyhez alappontokként a teljes **konformációhalmazt** használtam fel. Így a konfigurációs állapotösszeget, az átlagos potenciális energiát és a konfigurációs entrópiát a (83)–(85) összefüggésekkel számoltam ki (lásd 7.2. fejezet), az alábbi formában:

$$Z_{conf} = \int e^{-\beta U(q)} dq \approx \sum_k e^{-\beta \varepsilon_k} \varphi_0^d = \varphi_0^d Z_d \quad (54)$$

$$\langle E_{pot} \rangle = \frac{1}{Z} \int U(q) e^{-\beta U(q)} dq \approx \frac{1}{Z} \sum_k \varepsilon_k e^{-\beta \varepsilon_k} \varphi_0^d = \frac{1}{Z_d} \sum_k \varepsilon_k e^{-\beta \varepsilon_k} = \langle E_d \rangle \quad (55)$$

$$S_{conf} = k_B (\ln Z_{conf} + \beta \langle E_{pot} \rangle) \approx k_B (\ln Z_d + \beta \langle E_d \rangle) + k_B d \ln \varphi_0 \quad (56)$$

ahol  $\varepsilon_k$  a  $k$  **konformáció** energiája,  $d$  pedig a konfigurációs tér dimenziója. A számítás során keletkező  $Z_d = \sum_k e^{-\beta \varepsilon_k}$  mennyiséget konfigurációs állapotösszegnek nevezzük. Látható, hogy a számolt entrópia függeni fog a  $\varphi_0$  rácsállandótól és annak mértékegységétől. Ez a függés a differenciális entrópia definíciójából ered, és az egyszerűség kedvéért a továbbiakban a  $\varphi_0$  mértékegységét nem vettem figyelembe. **Az egyes konformációk energiájának kiszámításához és az energiainimalizáláshoz is a GROMACS 4.0.7 programcsomagot és a GROMOS96 G53a6 *force-fieldet* használtam.**

Fontos megjegyezni, hogy az így számolt energiaértékek természetesen befolyásolják az ezek alapján számolt entrópiaértékeket is. Ez azt is jelenti, hogy a kiszámolt entrópia értéke nem lesz pontosabb, mint az energiaszámoláshoz használt *force-field*. **Az egyes *force-field*ek pontosságának tárgyalása azonban meghaladja a dolgozat kereteit, és munkám szempontjából kevésbé releváns, mivel a számolt entrópiaértékek kísérleti adatokkal nem, csak az ugyanazon sokaságokból számolt értékekkel lettek összehasonlítva. Az egységesség végett az egyes sokaságok generálásakor és a szerkezetek energiájának direkt kiértékelésekor is ugyanazt a *force-fieldet* használtam.**

Az egzakt entrópia számítását Python és Perl szkriptek formájában valósítottam meg. Az állapotösszeg ( $Z$ ) számításakor az egyes összeadandó tagok között több nagyságrendnyi különbség is lehet, ami az összegzéskor numerikus problémákat okozhat, és pontosságvesztéssel járhat. Ennek minimalizására a tagokat növekvő sorrendben adtam össze az `mpmath` beépített Python modul segítségével, 100 jegy pontossággal. A programok futtatásához a Python 2.6.6-os verzióját használtam. Az egzakt entrópia kiszámítását végül Perl szkriptekkel, a `Math::BigFloat` Perl modul és a GMP numerikus könyvtár felhasználásával, szintén 100 jegy pontossággal végeztem.

### 3.4. Molekuláris sokaságok előállítása

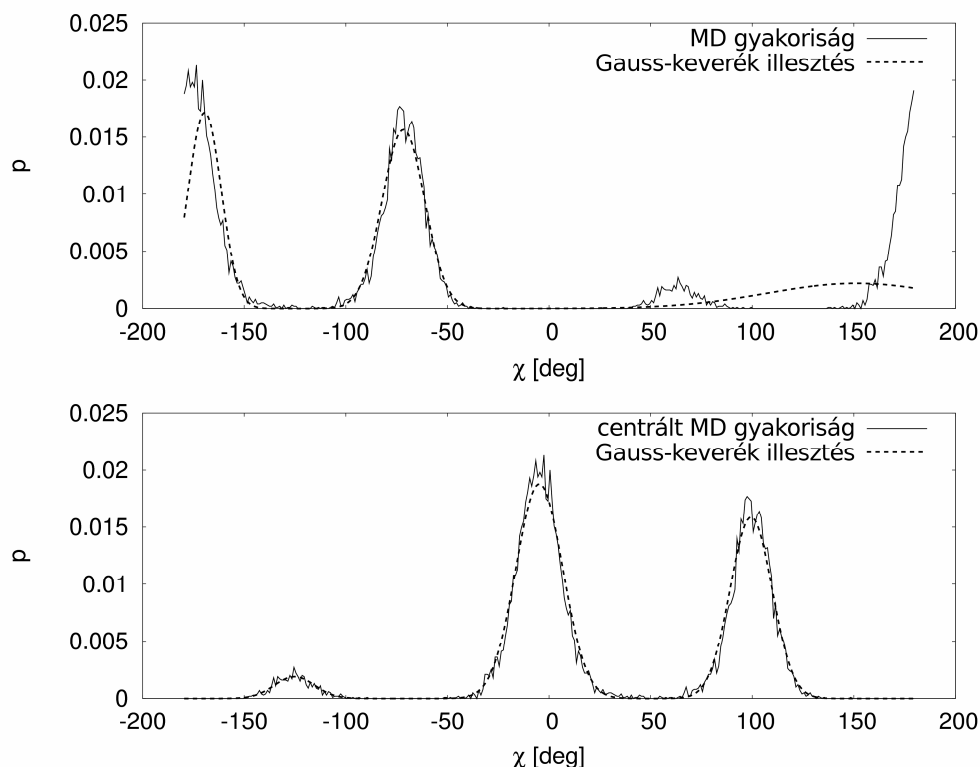
Az előző fejezetben leírtak alapján kapott teljes konformációhalmazból Monte Carlo mintavételezéssel mindegyik tesztrendszerre előállítottam molekuláris sokaságokat (a továbbiakban MC sokaságok). A sokaság elemeinek a generálásakor a következő konformációt mindig egyenletes eloszlással választottam a teljes konformációhalmazból, és azt a Metropolis-kritérium alapján elfogadtam vagy elvetettem. Az első elemet egyenletes eloszlással, véletlenszerűen választottam. A generálást addig folytattam, amíg  $n = 100500$  elfogadott elemet sikerült generálni, amiből az első 500 elemet elvettem a kezdeti konformációtól való függés minimalizálására, így a konformációs sokaságok minden molekula esetén 100000 konformációt tartalmaztak. Ezzel a módszerrel felhasználhatók a teljes konformációs halmaz generálásakor kiszámolt energiaértékek, így a generálás gyorsan végezhető, és Boltzmann-sokaságot eredményez. A kapott minták között azonban duplikátumok is lehetnek, egy konformáció többször is szerepelhet a minták között.

A gyakorlatban molekuláris sokaságok generálására gyakran alkalmaznak molekuláris dinamikai (MD) szimulációkat, ezért az entrópiabecslő módszereket kétféle módon generált MD sokaságon is teszteltem. Az eredeti tervem szerint az MC sokaságokhoz hasonlóan itt is olyan konformációkat szerettem volna generálni, amelynél a kötéshosszak és kötésszögek rögzítettek, és csak a belső koordináták változhatnak. A LINCS kényszermegoldó algoritmus azonban erre nem alkalmas, a SHAKE használata esetén pedig nagyfokú numerikus instabilitást tapasztaltam, ami miatt a kapott adatok nem voltak megbízhatóak. Ezért az első MD sokaságot minden rendszerre végül rögzített kötéshosszak és flexibilis kötésszögek alkalmazásával hoztam létre („MD/LINCS” sokaságok). Az ilyen sokaságokra számolt konfigurációs entrópia nem lesz összehasonlítható az MC sokaságokra kapott értékekkel, mert a rendszerek különböző szabadsági fokokkal rendelkeznek, hiszen az MD szimulációkban a kötésszögek is változhattak. A szabadsági fokok ilyen bővítése miatt bekövetkező entrópiaváltozás megfigyelésére informatívnak tartottam olyan adatkészlet létrehozását is, ahol mind a kötéshosszak, mint a kötésszögek flexibilisen változhattak („MD/Unc” sokaságok). A konformációkat sztochasztikus dinamikai szimulációval generáltam, oldószer nélkül, 2 fs lépésközzel összesen 10000 konformációt hoztam létre mindegyik rendszerre. A hőmérsékletet  $T = 300$  K-re állítottam, a Coulomb-erők számításához és a szomszédsági lista generálásához  $9 \text{ \AA}$  levágási értéket, a van der Waals-kölcsönhatások számításakor pedig  $12 \text{ \AA}$ -öt alkalmaztam. A számításokat a GROMACS programcsomag 4.0.7 verziójával és a G53a6 *force-field* felhasználásával végeztem.

### 3.5. A mintahalmaz centrálása

A belső koordináták által reprezentált torziós szögek periodikus változók, és rendszerint a  $[0^\circ, 360^\circ)$  intervallumon értelmezettek. Ez gondot okozhat olyan módszereknél, amelyek ezt nem képesek figyelembe venni, hanem a belső koordinátákat egy nyílt, végtelen intervallumon értelmezett mennyiségként kezelik. Az egyik jellegzetes probléma, hogy az értelmezési intervallum két végpontja környékén levő adatpontok a körkörös tartományban egymáshoz közel helyezkednek el, azonban a periodicitást figyelmen kívül hagyó módszerek ezeket távolinak veszik. A probléma megoldására az egyes belső koordináták nullpontját eltoltam olyan módon, hogy lehetőleg minél kevesebb adatpont helyezkedjen el a periodikus határon. Ennek érdekében mindegyik belső koordináta esetén elkészítettem a koordinátaértékek hisztogramját, amelyben megkerestem azt a leghosszabb összefüggő tartományt, ahol a hisztogramértékek a globális minimumukat felveszik. A nullpontot ezután úgy transzformáltam, hogy a periodikus határpont az így meghatározott tartomány közepére essen, majd ezt a műveletet minden belső koordinátára megismételtem. Ez a módszer a periodikus tartományban értelmezett eloszlás entrópiáját nem változtatja meg, hiszen az entrópia invariáns az eloszlás eltolásával szemben. **A centrálási algoritmust Octave szkript formájában valósítottam meg, a futtatáshoz az Octave 3.0.1 verzióját használtam.**

A centrálási műveletet a „val” rendszer  $\chi$  belső koordinátája esetén mutatom be, amelynek adatait a kényszerek nélküli MD trajektóriából vettem (5. ábra). Az eredeti és a centrált adathalmazra is  $K = 3$  komponensből álló Gauss-keverék függvényt illesztettem, amely nem veszi figyelembe a koordináta periodicitását. A centrált adatkészlet láthatóan jobb illeszkedést biztosít, ami a modell által jósolt entrópia pontosságát növeli. Az ábrán bemutatott esetben a hisztogram módszerrel számolt egzakt entrópiaérték  $S_{exact} = 38,478 \text{ J K}^{-1} \text{ mol}^{-1}$ . A centrálás nélkül illesztett adatkészlet esetén a Gauss-keverék modellből becsült entrópia  $S_{original} = 42,384 \text{ J K}^{-1} \text{ mol}^{-1}$ , a centrált adatkészletre pedig  $S_{centered} = 38,676 \text{ J K}^{-1} \text{ mol}^{-1}$ , ami jól mutatja a centrálás fontosságát az illesztés helyes végrehajtásában.



**5. ábra. A mintahalmaz centrálásának bemutatása.** A „val” rendszer  $\chi$  belső koordinátájának eloszlásfüggvénye látható a felső grafikonon centrálás nélkül, az alsón centrálás után. Az eredeti eloszlásban  $180^\circ$  környékén csúcs található, amely a periodikus határfeltételek figyelmen kívül hagyása miatt két távoli csúcsra esik szét. A centrálási művelet a csúcsokat elmozgatja a periodikus határpontról, és ezáltal jobb illesztést biztosít, miközben a periodikus eloszlás eredeti entrópiáját nem változtatja.

### 3.6. Egyéb entrópiabecslő módszerek implementálása

Az irodalomban megtalálható más, gyakran alkalmazott módszereket is megvalósítottam a Gauss-keverék módszerrel történő összehasonlítás érdekében. A kváziharmonikus módszert a Karplus és Kushick által leírt formában valósítottam meg [18]. Mivel ez a módszer nem képes figyelembe venni, hogy a belső koordináták korlátos és periodikus tartományon értelmezhetőek, a konformációs adatpontokat centráltam a 3.5. fejezetben leírt módszerrel. Ezzel elkerülhető, hogy a periodikus határok környékén adatpontok helyezkedjenek el, ami számítási műtermékeket okozna. A kváziharmonikus módszer Schlitter-féle, Descartes-koordinátákkal működő változatát is megvalósítottam [19]. Ebben az esetben két változatot hoztam létre: egyikben a molekula összes atomját felhasználtam a számításokban (Schlitter/AA), a másik esetén csak a belső koordináták torziós szögeit meghatározó atomokat (Schlitter/TA).

A Wang és Brüschweiler által leírt 2D entrópia módszerét is megvalósítottam [34]. Ez a módszer tartalmaz egy olyan  $\sigma$  paramétert, amely a magsűrűség-bebecslésben használt Gauss-magfüggvény szélességét jelöli. A módszer különböző alkalmazásaiban különböző  $\sigma$  értékek

bizonyultak optimálisnak [34, 35] a 0,1–0,3 tartományban, amiből kiderül, hogy a paraméter optimális értéke nem átvihető az egyes rendszerek között. A módszer több különböző  $\sigma$  értékek melletti tesztelése során a kapott entrópiaértékek nagyon széles skálán mozogtak, és erősen függtek a  $\sigma$  értékétől. Az egzakt entrópiaértékekkel való legjobb egyezést a  $\sigma = 0,05$  érték esetén találtam. Az ehhez hasonló magűrűség-becslési feladatok esetén megmutatható, hogy a magűrűvény optimális sávszélessége függ többek között a mintahalmaz nagyságától ( $n$ ) a  $\sigma \sim 1/n^{0,2}$  összefüggés szerint [97]. A Li és munkatársai által végzett tanulmányban [35] a  $\sigma$  optimális értéke 0,1 és 0,2 között volt, átlagosan  $n = 4791$  elemű mintahalmazra. Ezek alapján esetemben az  $n = 100000$  elemszámú Monte Carlo és a  $n = 10000$  méretű MD mintahalmazra a  $\sigma$  optimális értéke 0,01 és 0,1 közé esik. A továbbiakban ezért a számításaimhoz a  $\sigma = 0,05$  értéket használtam. **Ezeket a módszereket Octave szkriptek formájában valósítottam meg, amelyeket az Octave 3.0.1 verziójával futtattam.**

A Numata és munkatársai által kidolgozott módszerhez tartozó programot a cikk szerzői bocsátották rendelkezésemre. Az MC sokaságokon végzett számítások azonban sok esetben rendkívüli eltérést mutattak az egzakt entrópiaértékektől. Az eltérés hátterében feltételezésem szerint az MC sokaságok sajátos, rácson mintavételezett jellege húzódik meg, amely a mintahalmazban duplikált (degenerált) mintákat eredményezhetett. A duplikált minták megzavarhatják a módszer legközelebbi szomszéd entrópiát számoló részét, hiszen ilyen adathalmazban a  $k$ -legközelebbi szomszéd nem egyértelműen definiált, ami nem várt hibákat okozhat az eredményekben. A kapott entrópiaértékek erősen függtek a módszer felbontás (*resolution*,  $r$ ) paraméterétől is, amellyel csak minden  $r$ . adatpont figyelembevétele állítható be a számítás során (pl.  $r = 4$  esetén csak minden negyedik adatpontot veszi a program figyelembe a számításakor). A minta degeneráltságának feloldására praktikus megoldásként a belső koordináta értékeihez véletlen, nullára centrált Gauss-zajt adtam, amelynek varianciáját a rácscsillapító ( $\varphi_0$ ) négyzetére állítottam. Az így korrigált adatokra kapott eredmények már nem mutattak érzékenységet az  $r$  felbontási paraméterre, ami az entrópiaértékek stabilitását és konvergenciáját mutatja. Bár a hozzáadott zaj megnöveli a sokaságokra kapott entrópiaértékeket, várhatóan nem változtatja meg jelentősen azok pontosságát, hiszen a varianciája kisebb, mint a rácson végzett mintavételezés által okozott zaj. A degenerált mintahalmaz problémája nem merült fel az MD sokaságok esetén, ahol az  $r = 1$  értéket használtam.

### 3.7. Fehérjemodellek

A kísérletileg megoldott fehérjeszerkezetek felbontása széles skálán mozog. Membránfehérjék esetén a kristályosítás nehézségei miatt különösen kevés atomi felbontású szerkezet áll rendelkezésre, más szerkezetvizsgálati módszerekkel pedig ekkora méretű rendszerek általában nem, vagy nagyon nehezen vizsgálhatóak (a membránfehérjék tipikusan nagyok, és szerkezetük nagyon érzékeny a biokémiai környezetre). A molekuláris biológia alapvető dogmája, hogy a fehérje szekvenciája meghatározza a szerkezetét, ezek alapján feltételezhető, hogy hasonló szekvenciájú fehérjék szerkezete hasonló. Amennyiben tehát van egy már megoldott szerkezetű fehérje, amely nagyfokú szekvenciális **hasonlóságot** mutat a vizsgálni kívánt fehérjénkkal, akkor az ismeretlen szerkezet felépítéséhez használhatjuk kiindulópontként a már ismert szerkezetet (ez az ún. templát), ez a homológiamodellezés [98-100]. A hasonlóság megállapításának és a szerkezet felépítésének kiindulópontja a két fehérje szekvenciaillesztése. A szekvenciaillesztés olyan biológiai probléma, amely során két vagy több fehérjében az egymásnak funkcionálisan és/vagy szerkezetileg megfelelő aminosavpárokat keressük. A legjobb egyezés megállapításához más információ hiányában az aminosavak egymáshoz való biokémiai hasonlóságát veszik alapul, ekkor az optimális illesztés megállapítása matematikai probléma. Az optimális illesztés megállapítására többféle biológiai modell és algoritmus létezik [101-104].

A homológián alapuló szerkezetépítésre szintén sokféle módszer létezik, amelyek közül munkámban az igen elterjedt MODELLER nevű programot használtam [105, 106]. A MODELLER szekvenciaillesztés alapján automatikus modellépítésre képes. A modell alapját azok az atomok képezik a templát szerkezetből, amelyeknek van megfelelőjük az ismeretlen szerkezetű fehérjében a szekvenciaillesztés alapján. A hiányzó atomok kiegészítésre kerülnek a CHARMM *force-field* által meghatározott kötéshosszak és kötésszögek alapján. Az így kapott kezdeti szerkezet ezután több lépésben optimalizálásra kerül. Az optimalizálás célja, hogy az adott bemeneti adatok (szekvenciaillesztés, templát szerkezet) alapján megtaláljuk a legvalószínűbb kapható fehérjeszerkezetet. A valószínűség kiszámításához különböző geometriai paraméterek (pl.  $C_{\alpha}$ - $C_{\alpha}$  távolságok, kötéshosszak és kötésszögek, torziós szögek) eloszlásfüggvényét használják, amelyeket feltételes valószínűségeként fejezhetünk ki a következő alakban:

$$p(x | a, b, \dots, c) \quad (57)$$

ahol  $a, b, \dots, c$  ismert,  $x$  pedig az ismeretlen paraméter. Ezeket a függvényeket a MODELLER **kényszereknek** (*restraint*) nevezi, és ismert szerkezetű, homológ fehérjepárok

és -tripletek alapján származtatja őket [107]. A felhasznált  $x$  és a megjóslásához használt  $a, b, \dots, c$  paraméterek megválasztásakor igyekeznek figyelembe venni az egyes fehérjeszerkezeteken belül található korrelációkat, és a homológ aminosavak közöttiekét is. Az  $a, b, \dots, c$  paraméterek geometriai és szekvenciális jellemzőket egyaránt tartalmazhatnak. Az összes eloszlásfüggvény kombinálásával kapott valószínűségi-sűrűség-függvény negatív logaritmus a egy effektív szabadenergia jellegű mennyiséget határoz meg,

$$F = -\ln p \quad (58)$$

amelyet a MODELLER *objective function*nek nevez. Az *objective function* minimalizálására a MODELLER először az ún. *variable target function* módszerrel (VTFM) végez több iterációt, amely során fokozatosan egyre több **kényszert** vesz figyelembe, amelyeket a konjugált gradiensek módszerrel minimalizál. Ezután az optimalizált szerkezet finomítására a MODELLER több lépésben molekuláris dinamikával szimulált hőkezelést végez, így áll elő a végleges szerkezet [107].

### 3.8. ABC fehérjék molekuláris dinamikai szimulációja

Az ATP kicserélés hatásának vizsgálatára végzett szimulációkhoz a Sav1866 **bakteriális** ABC exporter fehérje „alul zárt”, ADP kötött konformációban kristályosított röntgendiffrakciós szerkezetéből indultam ki (PDB ID: 2HYD). **Ez a szerkezet, bár bakteriális eredetű fehérjét ír le, mégis fontos, mert az egyetlen rendelkezésre álló szerkezet teljes ABC exporterek nukleotidkötött konformációjáról.** Az ATP előállításához az MJ0796 NBD dimer szerkezetéből (PDB ID: 1L2T) a vele együtt kristályosított ATP molekula  $\gamma$ -foszfát csoportját és a megkötött  $Mg^{2+}$  iont vettem át a Sav1866 szerkezetbe, a fehérjék egymásra illesztése után. Ezzel a módszerrel állítottam elő a dupla ATP kötött Sav1866 ATP/ATP szerkezetet és az egyik kötőhelyen ATP-t, másikon ADP-t tartalmazó Sav1866 ATP/ADP szerkezetet is. Mindkét rendszeren energiainimalizálást hajtottam végre az ATP és a fehérje–ATP kontaktusok geometriájának optimalizálására. A minimalizálást a *steepest descent* módszerrel végeztem, és konvergálnak tekintettem, amikor a fellépő maximális atomi erő (emt<sub>ol</sub>)  $500 \text{ kJ mol}^{-1} \text{ nm}^{-1}$  alá csökkent [88]. A membránba ágyazáshoz a későbbiekben az így kapott szerkezeteket használtam.

Az apo és holo szerkezetek dinamikai összehasonlításához olyan fehérjére volt szükség, amelyről mindkét konformációban elérhető atomi felbontású szerkezet. Jelenleg ilyen szerkezetek nem érhetők el, ezért a humán MDR1 fehérje mindkét konformációban felépített homológiamodelljeit használtam. A holo állapot modellje a Sav1866 **bakteriális** fehérje szerkezete alapján készült [108], és Wiese és munkatársai bocsátották

rendelkezésemre. Az apo modellt a MODELLER 9v7 program segítségével az egér MDR3 szerkezet (PDB ID: 3G5U) alapján építettem. A szekvenciaillesztést a ClustalW 2.0.10 program [101] alapbeállításával végezve a szekvenciális azonosság 88,4% volt a két fehérje között. A fehérjék szekvenciáit a 3G5U PDB modellből, illetve a Wiese és munkatársai által rendelkezésre bocsátott fehérjeszerkezetből vettem. Ez utóbbi megfelelt az UniProt adatbázisban szereplő MDR1\_HUMAN szekvencia 36–630 és 696–1275 közötti régióinak. A két ABC alegység közötti kb. 60 aminosav hosszúságú összekötő (*linker*) régió a modellben nem szerepelt. A modellépítéshez használt szekvenciaillesztés a 6. ábrán látható. Egy-egy aminosavnyi eltérés a két szekvencia között egyenletesen elszórtan található, a nagyobb különbségek pedig főleg az extracelluláris hurkok területeire esnek (hMDR1 fehérjében az UniProt alapján ezek a 73–119, 210–215, 318–325, 732–756, 875–936, 958–973 pozíciók). A MODELLER az automatikus modellépítés közben több modellt is létrehoz, amelyek közül végleges modellként az *objective function* által legjobbnak mondott szerkezetet választottam. Az apo modellben csak azokat az oldalláncokat építettem meg, amelyeket a holo modell is tartalmazott.

```

3G5U/33-1270      33 V S V L T M F R Y A G W L D R L Y M L V G T L A A I I H G V A L P L M M L I F G D M T D S F A S V G N V S - - - K N S T N M S E A D K R A M F A 101
mdr1_human/36-1275 36 - - V F S M F R Y S N W L D K L Y M V V G T L A A I I H G A G L P L M M L V F G E M T D I F A N A G N L E D L M S N I T N R S D I N D T G F F M 105

3G5U/33-1270      102 K L E E M T T Y A Y Y T G I G A G V L I V A Y I Q V S F W G L A A G R Q I H K I R O K F F H A I M N Q E I G W F D V H D V G E L N T R L T D 173
mdr1_human/36-1275 106 N L E E D M T R Y A Y Y S G I G A G V L V A A Y I Q V S F W G L A A G R Q I H K I R K Q F F H A I M R Q E I G W F D V H D V G E L N T R L T D 177

3G5U/33-1270      174 D V S K I N E G I G D K I G M F F Q A M A T F F G G F I I G F T R G W K L T L V I L A I S P V L G L S A G I W A K I L S S F T D K E L H A Y A K 245
mdr1_human/36-1275 178 D V S K I N E G I G D K I G M F F Q S M A T F F T G F I V G F T R G W K L T L V I L A I S P V L G L S A A V W A K I L S S F T D K E L L A Y A K 249

3G5U/33-1270      246 A G A V A E E V L A A I R T V I A F G G Q K K E L E R Y N N L E E A K R L G I K K A I T A N I S M G A A F L L I Y A S Y A L A F W Y G T S L V 317
mdr1_human/36-1275 250 A G A V A E E V L A A I R T V I A F G G Q K K E L E R Y N K N L E E A K R I G I K K A I T A N I S I G A A F L L I Y A S Y A L A F W Y G T T L V 321

3G5U/33-1270      318 I S K E Y S I G Q V L T V F F S V L I G A F S V G Q A S P N I E A F A N A R G A A Y E V F K I I D N K P S I D S F S K S G H K P D N I Q G N L E 389
mdr1_human/36-1275 322 L S G E Y S I G Q V L T V F F S V L I G A F S V G Q A S P S I E A F A N A R G A A Y E I F K I I D N K P S I D S Y S K S G H K P D N I K G N L E 393

3G5U/33-1270      390 F K N I H F S Y P S R K E V O I L K G L N L K V K S G Q T V A L V G N S G C G K S T T V Q L M Q R L Y D P L D G M V S I D G O D I R T I N V R Y 461
mdr1_human/36-1275 394 F R N V H F S Y P S R K E V K I L K G L N L K V Q S G Q T V A L V G N S G C G K S T T V Q L M Q R L Y D P T E G M V S V D G O D I R T I N V R F 465

3G5U/33-1270      462 L R E I I G V V S Q E P V L F A T T I A E N I R Y G R E D V T M D E I E K A V K E A N A Y D F I M K L P H O F D T L V G E R G A O L S G G Q K O 533
mdr1_human/36-1275 466 L R E I I G V V S Q E P V L F A T T I A E N I R Y G R E N V T M D E I E K A V K E A N A Y D F I M K L P H K F D T L V G E R G A O L S G G Q K O 537

3G5U/33-1270      534 R I A I A R A L V R N P K I L L L D E A T S A L D T E S E A V V Q A A L D K A R E G R T T I V I A H R L S T V R N A D V I A G F D G G V I V E Q 605
mdr1_human/36-1275 538 R I A I A R A L V R N P K I L L L D E A T S A L D T E S E A V V Q V A L D K A R K G R T T I V I A H R L S T V R N A D V I A G F D D G V I V E K 609

3G5U/33-1270      606 G N H D E L M R E K G I Y F K L V M T Q T ----- 626
mdr1_human/36-1275 610 G N H D E L M K E K G I Y F K L V T M Q T ----- 630

3G5U/33-1270      684 ----- L D E D V P P A S F W R I L K L N S T E W P Y F V V G I F C A I I N G G L O P A F S V I F S K V V G V F T N G G P P E T Q R Q N S 749
mdr1_human/36-1275 696 ----- S F W R I M K L N L T E W P Y F V V G F C A I I N G G L O P A F A I I F S K I I G V F T R I D D P E T K R Q N S 753

3G5U/33-1270      750 L F S L L F L I L G I I S F I T F F L Q G F T F G K A G E I L T K R L R Y M V F K S M L R Q D V S W F D D P K N T T G A L T T R L A N D A A Q V 821
mdr1_human/36-1275 754 L F S L L F L A L G I I S F I T F F L Q G F T F G K A G E I L T K R L R Y M V F R S M L R Q D V S W F D D P K N T T G A L T T R L A N D A A Q V 825

3G5U/33-1270      822 K G A T G S R L A V I F Q N I A N L G T G I I I S L I Y G W Q L T L L L A I V P I I A I A G V V E M K M L S G Q A L K D K K E L E G S G K I A 893
mdr1_human/36-1275 826 K G A I G S R L A V I T Q N I A N L G T G I I I S F I Y G W Q L T L L L A I V P I I A I A G V V E M K M L S G Q A L K D K K E L E G A G K I A 897

3G5U/33-1270      894 T E A I E N F R T V V S L T R E Q K F E H M Y A Q S L Q I P Y R N A M K K A H V F G I T F S F T Q A M M Y F S Y A A A F R F G A Y L V T Q Q L M 965
mdr1_human/36-1275 898 T E A I E N F R T V V S L T Q E Q K F E H M Y A Q S L Q V P Y R N S L R K A H I F G I T F S F T Q A M M Y F S Y A G C F R F G A Y L V A H K L M 969

3G5U/33-1270      966 T F E N V L L V F S A I V F G A M A V G Q V S S F A P D Y A K A T V S A S H I I R I I E K T P E I D S Y S T O G L K P N M L E G N V Q F S G V V 1037
mdr1_human/36-1275 970 S F E D V L L V F S A V V F G A M A V G Q V S S F A P D Y A K A K I S A A H I I M I I E K T R L I D S Y S T E G L M P N T L E G N V T F G E V V 1041

3G5U/33-1270      1038 F N Y P T R P S I P V L Q G L S L E V K K G Q T L A L V G S S G C G K S T V V Q L L E R F Y D P M A G S V F L D G K E I K R L N V Q W L R A H L 1109
mdr1_human/36-1275 1042 F N Y P T R P D I P V L Q G L S L E V K K G Q T L A L V G S S G C G K S T V V Q L L E R F Y D P L A G K V L D G K E I K R L N V Q W L R A H L 1113

3G5U/33-1270      1110 G I V S Q E P I L F D C S I A E N I A Y G D N S R V V S Y E E I V R A A K E A N I H Q F I D S L P D K Y N T R V G D K G T Q L S G G Q K Q R I A 1181
mdr1_human/36-1275 1114 G I V S Q E P I L F D C S I A E N I A Y G D N S R V V S Q E E I V R A A K E A N I H A F I E S L P N K Y S T K V G D K G T Q L S G G Q K Q R I A 1185

3G5U/33-1270      1182 I A R A L V R Q P H I L L L D E A T S A L D T E S E K V V Q E A L D K A R E G R T C I V I A H R L S T I Q N A D L I V V I Q N G K V K E H G T H 1253
mdr1_human/36-1275 1186 I A R A L V R Q P H I L L L D E A T S A L D T E S E K V V Q E A L D K A R E G R T C I V I A H R L S T I Q N A D L I V V F Q N G R V K E H G T H 1257

3G5U/33-1270      1254 Q O L L A Q K G I Y F S M V S V Q - 1270
mdr1_human/36-1275 1258 Q O L L A Q K G I Y F S M V S V Q A - 1275

```

**6. ábra. A humán MDR1 fehérje apo homológiamodelljének megépítéséhez használt szekvenciaillesztés.** Az ábrán a templátként használt egér MDR3 szekvencia (PDB ID: 3G5U) és a humán MDR1 szekvencia („mdr1\_human”) illesztése látható. A megegyező aminosavakat kék háttér jelöli. A két ABC alegység közötti összekötő régiót (hMDR1: 631–695 aminosavak) nem tartalmazta egyik szerkezet sem. A két szekvencia közötti nagyobb különbségek az extracelluláris hurkok területeire esnek.



A membránszimulációkhoz a fehérjeszerkezeteket egy egyensúlyba hozott, explicit vízmolekulákkal szolvatált lipid kettősrétegbe illeszttem. A membrán kettősréteg koordinátáit a [109] referenciához tartozó kiegészítő anyagból (*supplementary material*) vettem. Ebből *x* és *y* irányban a membrán duplázásával, *z* irányban pedig vízzel való kitöltéssel nyertem egy kb. 125×125×180 Å nagyságú membrán-víz rendszert, amelyet 20 ns időtartamú szimulációval újra egyensúlyba hoztam. A fehérje membránba eső régiót a TMDet [110] és a HMMTOP [111] módszerekkel is jólsoltam, a különböző szekvenciákra és szerkezetekre kapott eredmények az 1. táblázatban láthatóak. A fehérjeszerkezetet végül kézi pozicionálással illeszttem úgy, hogy a jólsolt transzmembrán szakaszok membránba esése a lehető legjobban teljesüljön. A beillesztésnél ügyeltem arra, hogy az N-terminális amfipatikus ún. könyökhélixek a membrán belső felületére feküdjenek fel. A fehérjével szterikusan ütköző lipid- és vízmolekulákat ezután a genbox programmal eltávolítottam.

### Sav1866

	TMH1	TMH2	TMH3	TMH4	TMH5	TMH6
HMMTOP, 2HYD	17-41	60-79	134-158	163-182	246-270	275-294
TMDet, 2HYD, „A” lánc	12-36	62-88	133-158	161-183	241-265	284-305
TMDet, 2HYD, „B” lánc	12-36	61-88	133-157	161-183	241-265	283-305

### hMDR1

Első ABC alegység	TMH1	TMH2	TMH3	TMH4	TMH5	TMH6
HMMTOP, UniProt	49-73	117-141	190-209	214-233	299-323	328-352
HMMTOP, modell	49-73	117-141	190-209	214-233	299-322	331-355
TMDet, holo modell	49-71	114-136	187-209	212-232	296-316	334-353
TMDet, apo modell	49-70	117-135	190-208	218-235	297-314	333-349

Második ABC alegység	TMH7	TMH8	TMH9	TMH10	TMH11	TMH12
HMMTOP, UniProt	710-733	754-778	833-852	857-874	935-957	974-996
HMMTOP, modell	711-735	754-778	833-852	857-874	937-954	975-999
TMDet, holo modell	709-731	752-777	830-851	855-876	938-957	978-996
TMDet, apo modell	713-733	755-773	834-853	855-873	939-957	975-993

### mMDR3

Első ABC alegység	TMH1	TMH2	TMH3	TMH4	TMH5	TMH6
HMMTOP, UniProt	48-71	115-137	186-205	210-229	295-318	323-342
HMMTOP, 3G5U	50-72	115-137	186-205	210-229	295-318	323-342
TMDet, 3G5U „A” lánc	48-69	113-131	186-204	214-232	293-310	329-345

Második ABC alegység	TMH7	TMH8	TMH9	TMH10	TMH11	TMH12
HMMTOP, UniProt	706-729	750-774	829-848	853-870	934-953	970-992
HMMTOP, 3G5U	706-729	750-774	829-848	853-870	934-953	970-992
TMDet, 3G5U „A” lánc	708-729	752-769	830-849	851-869	935-953	971-988

**1. táblázat. Különböző módszerekkel jólsolt transzmembrán régiók az egyes ABC fehérjékben.** A HMMTOP szekvenciaalapú jólsolást végez, a TMDet módszer pedig térbeli szerkezet alapján jólsol. A módszer mellett a bemeneti adatok forrása van megjelölve; UniProt: a fehérje szekvenciája az UniProt adatbázis alapján, 2HYD, 3G5U: a felhasznált fehérjeszerkezetek PDB kódjai és láncazonosítója, a hMDR1 fehérje homológiamodelljei esetén a modell szekvenciája és konformációja van feltüntetve. A kétféle módszer által adott szakaszok nagyjából egybeesnek, de a szerkezetalapú predikció a membránba illesztéskor tipikusan jobban megvalósítható volt.

A teljes rendszer töltését a töltés előjelétől függően  $\text{Na}^+$  vagy  $\text{Cl}^-$  ellenionok elhelyezésével semlegesítettem, a genion program segítségével. Az így szolvatált rendszereken először energiminimalizálást hajtottam végre a nukleotidkötött Sav1866 szerkezetek előállításánál írt módon. Ezután több lépcsőben rövid MD szimulációkat végeztem a fehérje–lipid és fehérje–víz felszínek optimalizálására. A fehérje atomjait kezdetben egy  $1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$  erőállandójú harmonikus potenciállal kötöttem a kiindulási koordinátáihoz, amelyet az egyes szimulációk között  $250 \text{ kJ mol}^{-1} \text{ nm}^{-2}$  lépésekben csökkentettem. Az első szimuláció 20 ns, a továbbiak 200 ps időtartamúak voltak. A számításokhoz felhasznált szimulációkban a fehérje már korlátozások nélkül mozoghatott.

A Sav1866 rendszerek esetén POPC (palmitoil-oleoil-foszfadilkolin) kettősréteget, a G53a6 *force-field*et és felületifeszültség-csatolást (*surface-tension*) alkalmaztam, a *z* irányú kompresszibilitás  $4,5 \cdot 10^{-5} \text{ bar}^{-1}$  volt. Mindkét rendszer esetén először két 100 ns időtartamú szimulációt végeztem, majd a mintavételezés javítására további három 50 ns időtartamút, így mindkét rendszer esetén a további számításokra összesen öt szimulációból kapott adatokat használtam fel. Az MDR1 rendszerek esetén DPPC (dipalmitoil-foszfadilkolin) kettősréteget használtam, állandó felületi feszültséget, és a *z* irányú kompresszibilitást nullára állítottam. Bár a POPC lipidekből álló kettősréteg gyakori modellje a sejtmembránnak, a DPPC-t itt a hMDR1 apo szerkezet POPC-ben mutatott instabilitása miatt választottam. A DPPC molekulák telítettek, így képesek a teljesen nyújtott konformáció felvételére. Ez merevebb membránt hoz létre, amittől a fehérjeszerkezet stabilizálását vártam, amellet, hogy eleve stabil fehérjeszerkezet esetén a kétféle membrán használata nem okoz számottevő különbséget [71]. Az apo hMDR1 rendszerrel egy 100 ns, a holo hMDR1 rendszerrel három 100 ns, az eredeti egér MDR3 kristálybeli szerkezettel pedig egy 70 ns időtartamú szimulációt végeztem. Az állandó felületi feszültséggel végzett szimulációkban nulla felületi feszültséget használtam. A hMDR1 holo szerkezettel POPC kettősrétegben, részlegesen izotróp (*semiisotropic*) nyomáscsatolással is végeztem egy 100 ns hosszú szimulációt. Ebben az esetben 1 bar környezeti nyomást és  $\tau = 4 \text{ ps}$  csatolási állandót használtam, a kompresszibilitásra a lipidréteg síkjában (*x-y*)  $4,5 \cdot 10^{-5} \text{ bar}^{-1}$  értéket, a *z* irányban pedig nullát állítottam be. Mindegyik szimuláció esetén a termosztáláshoz a korrigált sebességátskálázás (*velocity-rescale*) módszerét és  $T = 310 \text{ K}$  referencia-hőmérsékletet választottam. A távoli (*long range*) elektrosztatikus kölcsönhatásokat a PME (Particle-Mesh Ewald, [88]) módszerrel számoltam, a Coulomb- és van der Waals-erők és a szomszédsági lista (*neighbor list*) számításakor használt levágási távolságot  $10 \text{ \AA}$ -re

állítottam. A szimulációhoz periodikus peremfeltételeket alkalmaztam mindhárom ( $x$ ,  $y$ ,  $z$ ) koordináta mentén. A szolvatált rendszerek előkészítését, az energiainimalizálást és a molekuláris dinamikai szimulációkat a GROMOS96 G53a6 *force-field*del és a GROMACS programcsomag 4.0.3 verziójával végeztem a University of North Carolina (Chapel Hill) számítóközpontjában, 4x SDR InfiniBand összeköttetéssel ellátott, 2,3 GHz órajelen futó Intel E5345 négymagos processzort tartalmazó számítógépeken. Egy MD szimuláció 128 számítási magon párhuzamosítva futott, egy 100 ns időtartamú szimuláció tipikus futási ideje kb. 10 nap volt.

A lipid kettősréteg analíziséhez saját, Python nyelven írt szkripteket használtam. Az egy lipid fejcsoportra eső terület meghatározásához a lipid fejcsoportok foszforatomjainak pozícióit a membrán síkjára vetítettem, ezek jelképezték az egyes lipidmolekulák pozícióit. Ezután kiszámoltam a fehérje membránba eső részének bennfoglaló téglalapját. A membrán ezen kívül eső területének kétszeresét elosztottam az ezen területen kívül eső lipidmolekulák számával. Az így kapott értéket vettem az egy lipidmolekulára eső területnek. A kétszeres szorzóra a két lipidréteg figyelembevétele miatt volt szükség. A lipidréteg vastagságát a felső és alsó lipidrétegben levő fejcsoport-foszforatomok átlagos  $z$  koordinátáinak különbségeként számoltam ki.

A fehérje szerkezeti stabilitásának követésére a trajektóriabeli szerkezetek  $C_\alpha$  atomjainak átlagos négyzetes gyök eltérését (RMSD) számoltam ki a kezdeti szerkezethez képest az

$$RMSD = \sqrt{\frac{1}{N} \sum_{k=1}^N d_k^2} \quad (59)$$

összefüggés segítségével, ahol  $d_k$  az egymásnak megfelelő  $C_\alpha$  atomok távolsága.

A vizsgált fehérjék kollektív mozgásainak kiszámítására az ún. esszenciális dinamikai (*essential dynamics*) analízist alkalmaztam [112]. Ennek során a tömegekkel súlyozott atomi koordináták kovarianciamátrixának főkomponenseit számoljuk ki, amelyek a rendszer egy-egy, lineárisan független (nem korrelált) vibrációs módusnak megfelelő elmozdulásvektort adják meg. Több fehérje esetén is megmutatták, hogy a funkció szempontjából releváns mozgások pusztán néhány, ún. esszenciális módus által kifeszített altérben történnek [112, 113]. ABC exporterek esetén az általam vizsgálni kívánt mozgás a holo és az apo konformáció közötti átmenet, amelyet a

$$\Delta \vec{r}_i = \vec{r}_{i,apo} - \vec{r}_{i,holo} \quad (60)$$

elmozdulásvektorral jellemeztem, ahol  $\vec{r}_{i,apo}$  és  $\vec{r}_{i,holo}$  a fehérje  $i$ . aminosavjához tartozó  $C_\alpha$  atom pozíciója az apo és a holo szerkezetben. Az átmenethez szükséges és a kollektív módusokból kapott elmozdulásvektorok egybeesését az ún. átfedési (*overlap*) faktorokkal [114] jellemeztem, amelyet a  $k$ . módusra az

$$I_k = \frac{\left| \sum_{i=1}^N \vec{m}_{ki} \Delta \vec{r}_i \right|}{\left( \sum_{i=1}^N \vec{m}_{ki}^2 \right)^{1/2} \left( \sum_{i=1}^N \Delta \vec{r}_i^2 \right)^{1/2}} \quad (61)$$

összefüggés definiál, ahol  $\vec{m}_k$  a  $k$ . módushoz tartozó,  $\Delta \vec{r}_i$  pedig a konformációs átmenethez tartozó elmozdulásvektor. Elképzelhető, hogy az esszenciális alteret nem egy, hanem több mozgási módus feszíti ki. Az ilyen, kombinált módusok elmozdulásvektorát az alábbi képlettel definiáljuk:

$$m'_n = \sqrt{\lambda_1} m_1 + \sqrt{\lambda_2} m_2 + \dots + \sqrt{\lambda_n} m_n \quad (62)$$

ahol  $\lambda$  a  $k$ . módushoz tartozó sajátérték, aminek négyzetgyöke a módus amplitúdójával arányos. Az esszenciális alter meghatározásához ezután az  $n$  legnagyobb átfedéssel rendelkező módus kombinációjára is kiszámoltam az átfedési faktort, és azt a kombinációt választottam, amelyre ez maximális. Az ezen módusok kombinációja által meghatározott elmozdulás tehát az, amely leginkább egybeesik az apo–holo átmenet irányával. A humán MDR1 fehérje esetén a kovarianciamátrix kiszámítása előtt a trajektóriából kapott szerkezeteket a transzmembrán hélixek mentén egymásra illesztettem. Az illesztéshez használt szakaszok a Leu49–Gly73, Tyr117–Gly141, Ile190–Thr209, Leu214–Ala233, Ile299–Ser323, Ile328–Ile352, Pro709–Phe732, Asn753–Phe777, Arg832–Phe851, Leu857–Val874, Ala935–Cys956 és Asp973–Ala995 aminosavak voltak. **A kovarianciamátrix kiszámítását, diagonalizálását és az esszenciális módusok kiszámítását a GROMACS 4.0.3 programcsomagban megtalálható `g_covar` programmal végeztem.**

A trajektóriákból kapott konformációk másodlagos szerkezeti tartalmát a DSSP programmal becsültem meg [115]. A DSSP először a fehérje főláncán található hidrogénkötéseket keresi meg az energiájuk alapján, amelyet geometriai paraméterekből származtat. A hidrogénkötések hálózatában aztán az egyes másodlagos szerkezeti elemeknek megfelelő mintázatokat keres, amely alapján meg tudja állapítani, hogy az egyes aminosavak milyen másodlagos szerkezeti elemekben találhatóak. Domináns másodlagos szerkezeti elemnek egy aminosav esetében azt neveztem, amelyben a szimulációs idő legalább 80%-ában megtalálható. A másodlagos szerkezet analízisében a transzmembrán doméneket a

következőképp határoztam meg: Met1–Asp319 (A és B lánc) Sav1866 esetén, Val35–Asn367 és Ser692–Glu1009 mMDR3 esetén és Val36–Asn371 és Ser696–Glu1013 hMDR1 esetén.

Az oldószer által hozzáférhető felület számításához a CCP4 csomag SURFACE programját használtam [116, 117], amely egy vízmolekulát reprezentáló 1,4 Å sugarú gömböt gördít végig a fehérje felszínén, és minden aminosavra kiszámolja a kontaktfelület nagyságát a próbagömbbel. Az analízist a hMDR1 fehérjében a 141–190, 236–295, 354–376, 782–831, 881–940 és 1005–1020 aminosavak által definiált intracelluláris hurkokon végeztem.

### 3.9. A kölcsönös információ

A nukleotidkötő és a transzmembrán domének közötti lehetséges allosztérikus útvonalak felderítéséhez olyan oldalláncokat kerestem, amelyeknek a konformációja és mozgása egymással összefügg. Ha egy oldallánc konformációja egy másik oldallánc konformációjától függ, akkor a kettőjük konformációja és így mozgása is korrelált lesz. Az egymással ilyen formájú korrelált mozgást mutató oldalláncok alkalmasak arra, hogy az allosztérikus viselkedés alapját képező konformációs jeleket a fehérjén belül továbbítsák. Az oldalláncok mozgása közötti korreláció leírására a kölcsönös információ mennyiségét használtam. A munkám során a kölcsönös információt az egyes fehérjerendszerekre készített egyensúlyi trajektóriákból származtattam a MutInf módszer alkalmazásával. [118] **A módszerhez tartozó szoftvert a szerző bocsátotta rendelkezésemre.** Ez a módszer a két oldallánc közötti kölcsönös információt az őket alkotó belső koordináták (torziós szögek) közötti kölcsönös információ összegeként számolja:

$$I_r(i, j) = \sum_{\substack{k=\varphi, \psi, \chi_1, \chi_2, \dots \\ i \text{ oldalláncból}}} \sum_{\substack{l=\varphi, \psi, \chi_1, \chi_2, \dots \\ j \text{ oldalláncból}}} I_{k,l} \quad (63)$$

ahol  $i$  és  $j$  aminosavak,  $k$  és  $l$  pedig az őket alkotó torziós szögek. Az  $I_{k,l}$  két torziós szög közötti kölcsönös információ, amelyet a MutInf módszer az értékpárok hisztogramja és a (28) összefüggés alapján számol ki. A számítás továbbá tartalmaz korrekciót a véges mintaméretre vonatkozóan, és ha a kölcsönös információ nem különbözik szignifikánsan nullától, a két változót függetlennek tekinti.

Esetemben mindkét vizsgált rendszerre az 5 szimulációs trajektóriából a 30–50 ns szimulációs idő közötti 10000 szerkezet egyesítésével mindkét rendszerre 50000–50000 szerkezetet használtam a számításokhoz. Az összes aminosavpár közül csak azokat tekintettem jelentősen korreláltaknak, amelyek kölcsönös információja meghaladta a  $3k_B$  értéket. Ezt a korlátot úgy választottam, hogy a jelentős korrelációt mutató aminosavak

száma 25–50 között legyen. A számítást a Sav1866 ATP/ATP és ATP/ADP rendszerre is elvégeztem, és az így kapott aminosavak közötti dinamikai korrelációkat gráfon ábrázoltam.

## 4. A konformációs entrópia becslése Gauss-keverék függvények használatával

### 4.1. Eredmények

#### 4.1.1. A Gauss-keverék módszer

A Gauss-keverék függvényeken alapuló, konformációs entrópia becslésére használatos módszer létrehozását az motiválta, hogy létezzen egy olyan koncepcionálisan egyszerű, áttekinthető módszer, amely a nagyobb, sok szabadsági fokkal rendelkező molekulák komplex energiafelszíneiből származó komplikációkat is képes kezelni. A nehézségeket rendszerint az okozza, hogy a molekula nem egy, jól definiált energiaminimum körül fluktuál, hanem natív állapotában több, egymáshoz energiában és az állapotterben is közeli lokális energiaminimumot is bejár termikus fluktuációk következtében. Az egyes lokális minimumhelyek környéke továbbá nem minden esetben írható le harmonikus potenciálfüggvénnyel, vagyis az egyes energiaminimumok körüli fluktuáció is lehet anharmonikus.

Az új módszer lényege, hogy Gauss-keverék függvényt használ a konfigurációs térbeli sűrűség leírására. A Gauss-keverék függvények különálló, többváltozós Gauss-függvények súlyozott összegei,

$$f(x; \theta) = \sum_{k=1}^K p_k g(x; m_k, \sigma_k) \quad (64)$$

ahol  $g(x; m_k, \sigma_k)$  az  $m_k$  középpont körüli,  $\sigma_k$  kovarianciamátrixszal jellemzett Gauss-függvény,  $\theta$  az  $m_k, \sigma_k$  paraméterek összessége,  $K$  pedig a Gauss-komponensek száma. A Gauss-keverék függvények segítségével bármilyen sima függvényt tetszőleges pontossággal meg lehet közelíteni [119]. A módszer első lépésben a vizsgált konformációs sokaságra egy Gauss-keverék függvényt illeszt, majd megbecsli az így kapott folytonos eloszlás információs entrópiáját.

A Gauss-függvény az állandó hőmérsékleten Brown-mozgást végző harmonikus oszcillátor eloszlásfüggvénye, és emiatt a Gauss-keverék használata egy természetes kiterjesztése az eredeti kváziharmonikus módszernek. A Gauss-keverék módszer közvetlenül belső koordinátákkal dolgozik, ezért nem használ főkomponens-analízist vagy a marginális valószínűségi sűrűségeket. A módszert Metropolis-féle Monte Carlo (MC) és molekuláris

dinamikai (MD) mintavételezéssel előállított kanonikus sokaságok entrópiájának kiszámolására alkalmaztam.

A vizsgált kismolekulák konformációs térbeli eloszlását (64) alakú Gauss-keverék függvényekkel közelítettem, amelynek paramétereit a konformációs mintából származtattam. Egy adott konformációs sokaság  $x_i$  elemeire ( $i = 1 \dots n$  a minták száma, az  $\{x_i\}$  halmazt  $X$ -szel jelöljük) a közelítő Gauss-keverék függvényt a legnagyobb valószínűség módszerével (*maximum likelihood estimation*, MLE) illesztettem. A  $\theta$  paraméterek valószínűségét (*likelihood*-ját) a megfigyelt adatpontok függvényében a következőképpen írhatjuk fel:

$$\Lambda(\theta | X) = \prod_{i=1}^n f(x_i | \theta) = \prod_{i=1}^n \sum_{k=1}^K p_k g(x_i; m_k, \sigma_k) \quad (65)$$

ahol  $f(x_i | \theta)$  annak a valószínűsége, hogy a  $\theta$  paraméterekkel jellemzett modell az  $x_i$  adatpontot generálja. Ha a Gauss-keverék csak egyetlen komponenset tartalmazna, akkor az illesztés viszonylag egyszerű lenne, azonban több komponens esetén a  $p_k$  súlyok ismeretének hiánya a feladatot megnehezíti. Ezért a *likelihood* maximalizálására az ún. *expectation maximization* (EM) módszert alkalmaztam [120], amelyet hiányos adatokból történő becslésre dolgoztak ki. A hiányzó adat ebben az esetben az, hogy melyik az az  $y_i$  Gauss-komponens, amelyik az adott  $x_i$  adatpontot generálta. Ez az  $\{y_i\} = Y$  adathalmaz közvetlenül nem megfigyelhető, de a  $\theta$  paraméterek és az  $x_i$  adatpontok ismeretében az eloszlásfüggvénye kiszámítható.

Az EM algoritmus során két lépcsős iterációt végzünk a modell  $\theta$  paramétereit fokozatosan finomítva. Az első lépésben kiszámoljuk a *log-likelihood* függvény várható értékét a hiányzó adatok rekonstruálásával („E”, *expectation step*) a jelenlegi  $\theta^{(j)}$  paraméterek alapján:

$$\begin{aligned} E(\ln \Lambda(\theta | X, Y) | X, \theta^{(j)}) &= \\ &= \sum_y (\ln \Lambda(X | y, \theta)) P(y | X, \theta^{(j)}) = Q(\theta | \theta^{(j)}) \end{aligned} \quad (66)$$

ahol az  $y = \{y_i, i = 1 \dots n\}$  vektor összes lehetséges értékére összegzünk. A *log-likelihood* függvény behelyettesítése után ez a kifejezés

$$Q(\theta | \theta^{(j)}) = \sum_{k=1}^K \sum_{i=1}^n (\ln p_k g(x_i; m_k, \sigma_k)) p^{(j)}(k | i) \quad (67)$$

formájúra egyszerűsíthető, ahol egy  $p^{(j)}(k | i) = P(y_i = k | x_i, \theta^{(j)})$  érték annak a valószínűsége, hogy az  $x_i$  adatpontot a  $k$ . komponens generálta. Ezek az ún. tagsági (*membership*) valószínűségek csak a megfigyelt adatoktól és az aktuális  $\theta^{(j)}$  paraméterektől függenek, és kiszámíthatók:



$$p^{(j)}(k|i) = \frac{p_k^{(j)} g(x_i; m_k^{(j)}, \sigma_k^{(j)})}{\sum_{k'=1}^K p^{(j)}(k'|i) g(x_i; m_{k'}^{(j)}, \sigma_{k'}^{(j)})} \quad (68)$$

Az EM algoritmus második lépése a maximalizáció („M”, *maximization step*), amely során kiszámoljuk az új, legvalószínűbb  $\theta$  paramétereket

$$\theta^{(j+1)} = \arg \max_{\theta} Q(\theta | \theta^{(j)}) \quad (69)$$

Ez a lokális maximumhely szintén kiszámítható a (73) és (74) összefüggések alapján:

$$m_k^{(j+1)} = \frac{\sum_{i=1}^n p^{(j)}(k|i) x_i}{\sum_{i=1}^n p^{(j)}(k|i)} \quad (70)$$

$$\sigma_k^{(j+1)} = \frac{\sum_{i=1}^n p^{(j)}(k|i) (x - m_k^{(j)})(x - m_k^{(j)})^T}{\sum_{i=1}^n p^{(j)}(k|i)} \quad (71)$$

$$p_k^{(j+1)} = \frac{1}{n} \sum_{i=1}^n p^{(j)}(k|i) \quad (72)$$

A konvergencia megállapítására a  $p_k$ ,  $m_k$ ,  $\sigma_k$  paraméterek változását követtem, és a konvergencia feltételeként azt fogadtam el, amikor két egymást követő iterációra a  $\|p_k^{(j+1)} - p_k^{(j)}\| < 10^{-5}$ ,  $\|m_k^{(j+1)} - m_k^{(j)}\| < 10^{-4}$  és  $\|\sigma_k^{(j+1)} - \sigma_k^{(j)}\| < 10^{-4}$  korlátok teljesültek minden  $k$  értékre.

Az összes számítás esetén a Gauss-komponensek száma ( $K$ ) 10 volt. Az  $m_k$  kezdeti értékeit a `kmeans++` klaszterezési algoritmus alapján választottam meg [121], a  $p_k$  kezdeti értékeit az egyenletes eloszlásnak megfelelően vettem, a  $\sigma_k$  kezdeti értékeit pedig a (71) összefüggés alapján számoltam. Az illesztés konvergenciája részletesen a 4.2. fejezetben kerül tárgyalásra.

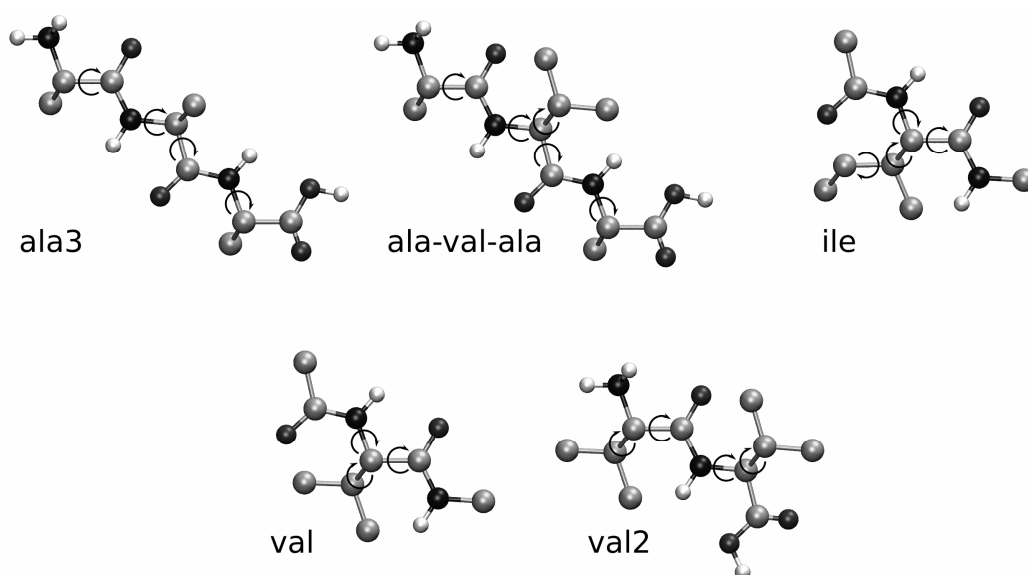
A kapott Gauss-keverék függvények entrópiáját a Huber és munkatársai által leírt módszerrel számoltam ki [122]. A Huber-féle módszer az egyes Gauss-komponensek középpontjának környékén a többi komponens járulékát egy másodrendű Taylor-sorfejtéssel közelíti, így a közelített érték analitikusan kiszámítható. Az eredeti módszer tartalmaz egy leírást a pontosság növelésére is az egyes Gauss-komponensek további komponensekre történő felbontásával, de ezt nem valósítottam meg, mert a számított entrópiaértékek pontossága enélkül is megfelelőnek bizonyult. A `kmeans++` klaszterezést, az EM

algoritmust és az entrópia közelítő kiszámítását Octave szkript formájában valósítottam meg, a futtatáshoz az Octave 3.0.1 verzióját használtam.

A módszert öt különböző, kisméretű peptidet tartalmazó rendszeren teszteltem, és a kapott eredményeket összehasonlítottam más, már létező módszerekkel kapott eredményekkel. Az összehasonlítás kedvéért az irodalomban megtalálható négy másik, az entrópia becslésére kifejlesztett módszert is megvalósítottam: az eredeti Karplus és Kushick-féle kváziharmonikus módszert [18], a Schlitter-féle kovarianciamátrix-alapú módszert [19], a Wang és Brüschweiler-féle 2D entrópia módszerét [34], és Numata és munkatársai módszerét [25].

#### 4.1.2. Molekuláris tesztrendszerek generálása

A módszer teszteléséhez, az irodalomban található más entrópiaszámoló módszerekkel való összehasonlításához olyan molekuláris tesztrendszereket kerestem, amelyek kisméretűek és energiefelszínük elég tagolt ahhoz, hogy a nagyméretű rendszerek jellegét mutassák (pl. többszörös, anharmonikus energiamedencék). A jövőbeli, fehérjékre való alkalmazhatóság előkészítésére ezek alapján ötféle, kisméretű peptidekből álló tesztrendszert hoztam létre, amelyek között szerepelnek olyan peptidek is, melyeknek oldalláncai több rotamer állapotot is felvehetnek. A vizsgálatban használt öt molekula az Ala<sub>3</sub>, Ala-Val-Ala, Ace-Ile-Nme, Ace-Val-Nme, Val<sub>2</sub> peptid volt, amelyek 4, 5, 4, 3 és 4 elforgatható kötéssel rendelkeznek (7. ábra).



**7. ábra.** A konformációs sokaságok generálásához kiválasztott kisméretű peptidek és peptidszármazékok. A kiválasztott molekulák az Ala<sub>3</sub> („ala3”), Ala-Val-Ala („ala-val-ala”), Ace-Ile-Nme („ile”), Ace-Val-Nme („val”) és Val<sub>2</sub> („val2”). A peptidek végei minden esetben semlegesek. Az egyesített szénatomokat szürke, az oxigén és nitrogénatomokat fekete, a hidrogéneket fehér szín jelöli. A szabadsági fokokat jelentő elforgatható torziós szöveget nyilak jelzik.

Összehasonlítási alapként mindegyik tesztrendszernek kiszámoltam az egzakt konfigurációs entrópiáját a statisztikus fizikában használatos definíció segítségével. A számoláshoz előállítottam mindegyik rendszer ún. teljes **konformációhalmazát** a belső koordináták terében való uniform és szisztematikus mintavételezésével. A módszerek bemeneti adataként mindegyik rendszerre három kanonikus sokaságot hoztam létre. Az elsőt a teljes **konformációhalmazból** származó Monte Carlo (MC) mintavételezéssel kaptam (a továbbiakban MC sokaság). Az MC sokaságban tehát a teljes **konformációhalmaz** előállításának megfelelően csak az elforgatható kötések torziós szögei eltérőek, míg a kötőszögek és kötőhosszak állandó értéken maradtak. Így az ebből a sokaságból számolt entrópiaértékek megfelelnek és közvetlenül összehasonlíthatóak az egzakt számításból kapott értékekkel. A gyakorlatban azonban a nagyobb méretű biomolekulákra általában molekuladinamikai (MD) szimulációval hoznak létre konformációs sokaságokat, amely során a molekulát flexibilisen kezelik, vagyis a kötőszögek és kötőhosszak változhatnak. Ez azonban növeli a rendszer szabadsági fokainak a számát. Az MC sokaságok mellett fontos, hogy az entrópiabecslő módszereket valódi MD sokaságokon is teszteljük, ezért mind az öt tesztrendszerhez generáltam két 10 ns hosszúságú MD trajektóriát, amelyek egyenként 10000 konformációt tartalmaztak. Az egyik trajektóriát kényszerek nélküli MD szimuláció segítségével készítettem („MD/Unc” sokaság), így mind a kötőhosszak, mind a kötőszögek flexibilisen változhattak. A másik szimuláció esetében a kötőhosszak nagyságát a LINCS algoritmus segítségével állandó értéken tartottam („MD/LINCS” sokaság), a kötőszögek pedig változhattak.

#### 4.1.3. Az egzakt konfigurációs entrópia és állapotösszeg kiszámítása

A különböző módszerek pontosságának értékelése érdekében referenciaértékként definiáltam a rendszerek egzakt konfigurációs entrópiáját. Ez az egzakt érték az entrópia, az átlagos energia illetve az állapotösszeg statisztikus fizikai definíciója alapján numerikus integrálással kiszámolható az (54)–(56) összefüggések segítségével. Az integrálás alappontjaiként a **konformációk** teljes halmazát használtam, amely egy adott molekulára ~40–60 millió diszkrét állapotból tevődött össze a szabadsági fokok (belső koordináták) számától függően. Az így kapott konfigurációs állapotösszeg ( $Z_d$ ) értékei több nagyságrendet is átfogtak 3,843-tól  $1,7197 \cdot 10^{16}$ -ig, az átlagos energia ( $\langle E_d \rangle$ ) pedig kisebb variációkat mutatott  $-62,75$  és  $22,97 \text{ kJ mol}^{-1}$  között. A rendszerek teljes, egzakt konfigurációs entrópiája  $107$  és  $172 \text{ J K}^{-1} \text{ mol}^{-1}$  közé esett (2. táblázat).

Rendszer	$Z_d$	$\langle E_d \rangle$ (kJ mol <sup>-1</sup> )	$S_{\text{exact}}$ (J K <sup>-1</sup> mol <sup>-1</sup> )
ala3	$1.9208 \times 10^5$	1.491124	154.668
ala-val-ala	$3.0435 \times 10^3$	3.025460	172.497
ile	$3.0666 \times 10^{15}$	-61.497609	137.602
val	$1.7197 \times 10^{16}$	-62.751832	107.415
val2	3.8430	22.956057	133.820

**2. táblázat.** A konfigurációs állapotösszeg ( $Z_d$ ), az átlagos energia ( $\langle E_d \rangle$ ) és az egzakt konfigurációs entrópia ( $S_{\text{exact}}$ ) értéke a vizsgált molekuláris rendszerekre. Az így kapott egzakt entrópia szerepelt összehasonlítási alapként a többi entrópiaszámoló módszer értékelésére.

#### 4.1.4. Monte Carlo mintavételezett sokaságok entrópiája

A teljes **konformációhalmaz** Monte Carlo (MC) mintavételezésével kanonikus sokaságokat állítottam elő minden egyes tesztmolekulára. Ezek az adatkészletek egyenként 100000 **konformációból** álltak mind az öt rendszerre. A mintavételezés konvergenciájának és lefedésének ellenőrzésére a teljes **konformációhalmaz** segítségével kiszámítottam a belső koordináták egzakt marginális eloszlását. Ez az egzakt eloszlás kiváló egyezést mutatott az MC sokaságban való előfordulási valószínűségekkel, ami azt mutatja, hogy szabályos kanonikus sokaság keletkezett a mintavételezés során. Az entrópiaszámoló módszereknek így az MC sokaságok alapján tudniuk kell reprodukálni az egyes rendszerek egzakt konformációs entrópiáját.

Az MC sokaságra a különböző módszerekkel kiszámolt és az egzakt entrópiaértékeket a 3. táblázat foglalja össze. A táblázat alapján látható, hogy a Gauss-keverék alapú módszer által számolt értékek ( $S_{\text{gaussmix}}$ ) kiváló egyezést mutatnak az egzakt értékekkel ( $S_{\text{exact}}$ ), az átlagos eltérés mindössze 2%. A legkisebb eltérés az „ala-val-ala” rendszer esetében (0,03%), a legnagyobb pedig a „val” rendszer esetében volt tapasztalható (5,76%). Az összes többi módszer által adott entrópiaérték mind lényegesen nagyobb eltérést mutat az egzakt értéktől, az átlagos eltérés 10 és 40% közé esik.

A várakozásoknak megfelelően a Karplus és Kushick-féle kváziharmonikus módszer szisztematikusan túlbecsüli a valódi értéket [19], bár az átlagos eltérése az egzakt értéktől viszonylag kicsi (10,4%).

A Schlitter-féle módszer Descartes-koordinátákat használ a belső koordináták (torziós szögek) helyett, ezt a módszert kétféle verzióban valósítottam meg. Az első esetben az összes atom koordinátája szerepelt a módszer bemeneteként ( $S_{\text{Schlitter/AA}}$ ), a másodikban viszont csak azok az atomok, amelyek a belső koordinátákat definiáló torziós szögeket alkotják ( $S_{\text{Schlitter/TA}}$ ). Azt találtam, hogy az első verzió az egzaktnál lényegesen magasabb értékeket ad

	$S_{\text{gaussmix}}$	$S_{\text{QH}}$	$S_{2D}$	$S_{\text{Numata}}$	$S_{\text{Schlitter/TA}}$	$S_{\text{Schlitter/AA}}$
<b>ala3</b> ( $S_{\text{exact}} = 155$ )						
Monte Carlo	151	157	135	162	130	229
MD/LINCS	154	167	142	392	204	523
MD/Unc.	164	168	131	418	212	536
<b>ala-val-ala</b> ( $S_{\text{exact}} = 172$ )						
Monte Carlo	173	188	144	191	146	247
MD/LINCS	193	216	170	414	267	568
MD/Unc.	130	134	75	463	256	577
<b>ile</b> ( $S_{\text{exact}} = 138$ )						
Monte Carlo	139	158	125	107	133	177
MD/LINCS	155	172	141	204	191	339
MD/Unc.	153	167	150	242	191	338
<b>val</b> ( $S_{\text{exact}} = 107$ )						
Monte Carlo	101	114	96	77	82	121
MD/LINCS	114	125	103	182	128	275
MD/Unc.	112	127	99	189	136	289
<b>val2</b> ( $S_{\text{exact}} = 134$ )						
Monte Carlo	135	161	95	105	175	224
MD/LINCS	155	172	124	317	236	468
MD/Unc.	156	177	131	326	255	499
átlagos eltérés $S_{\text{exact}}$ -től						
Monte Carlo	2,04%	10,36%	15,60%	17,54%	17,77%	39,97%

**3. táblázat. Az öt molekuláris tesztrendszer entrópiája ( $\text{J K}^{-1} \text{mol}^{-1}$ ) különböző módszerekkel számolva.** A Monte Carlo sokaságokra kapott eredményeken kívül a táblázatban szerepelnek az MD/LINCS sokaságra (molekuladinamika, a kötőhosszak fixen tartva a LINCS algoritmussal) és az MD/Unc sokaságra (kényszerek nélküli molekuladinamika, a kötőhosszak és kötőszögek szabadon változhatnak) kapott entrópiaértékek is. A Gauss-keveréken alapuló módszerrel számolt entrópiát  $S_{\text{gaussmix}}$  jelöli. Az  $S_{\text{QH}}$  a Karplus és Kushick-féle kváziharmonikus módszerrel számolt értéket, az  $S_{2D}$  a Wang és Brüschweiler-féle 2D entrópia értékét jelöli. Az  $S_{\text{Numata}}$  a Numata és munkatársai módszerével számolt értékeket jelöli. Az  $S_{\text{Schlitter/AA}}$  és  $S_{\text{Schlitter/TA}}$  értékek a Schlitter-féle módszerrel számolt entrópiaértéket jelentik, az első esetben minden atomot (AA), a második esetben csak az elforgatható torziós szögeket meghatározó atomokat (TA) véve figyelembe a számításkor. Az átlagos eltérések csak az MC sokaságokra lettek kiszámolva, mert a többi sokaság egzakt entrópiája nem ismert.

(átlagos eltérése 39,97%), míg a második szisztematikusan alulbecsli azt (átlagos eltérés 17,77%).

A Wang és Brüschweiler-féle 2D entrópia módszere ( $S_{2D}$ ) esetén szintén azt találtam, hogy alulbecsli az egzakt értékeket. Ennél a módszernél a kapott értékek azonban erősen függenek a módszer részeként szereplő  $\sigma$  sáv szélesség-paramétertől (erről részletek a 3.6. fejezetben található). Az általam használt  $\sigma = 0,05$  érték tekinthető legjobb *a priori* becslésnek, és így az eredmények a 2D entrópia módszer által elérhető legjobb közelítést jelentik.

A Numata-féle módszer ( $S_{\text{Numata}}$ ) a duplikált állapotok miatti korrekció után viszonylag jó egyezést mutat az egzakt értékekkel (átlagos eltérés 17,54%). Az  $S_{\text{Numata}}$  értékek mindig

kisebbségek, mint az  $S_{\text{Schlitter/AA}}$  értékek, ami a várakozásoknak megfelelő, hiszen a Numata-módszer először  $S_{\text{Schlitter/AA}}$ -t számol, amit végül negatív értékekkel korrigál. Ez a módszer azonban ötből három esetben alulbecsli az entrópiát.

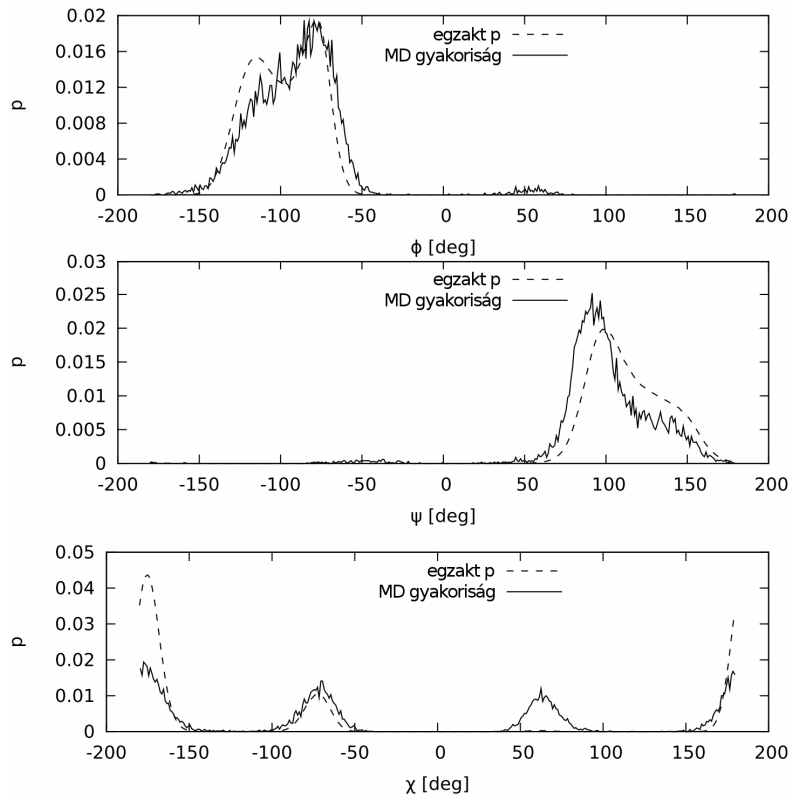
#### 4.1.5. Molekuláris dinamikai sokaságok entrópiája

MC sokaságok esetén a kötőszögek és kötőhosszak állandó értékkel rendelkeztek, ami csökkentette a szabadsági fokok számát, és így lehetőség nyílt egzakt entrópiaértékeket számolni. A gyakorlati alkalmazás lehetőségének bemutatására az egyes rendszerekhez kétféle, molekuláris dinamikai (MD) mintavételezéssel előállított sokaságot is létrehoztam. Az első MD sokaságban a kötőszögek szabadon változhattak, azonban a kötőhosszak értéke állandó volt („MD/LINCS”), míg a másik MD sokaságban mind a kötőszögek, mind a kötőhosszak szabadon változhattak („MD/Unc”). Ezen adatok segítségével megbecsülhetjük a kötőhosszak vibrációjának járulékát a rendszer entrópiájához.

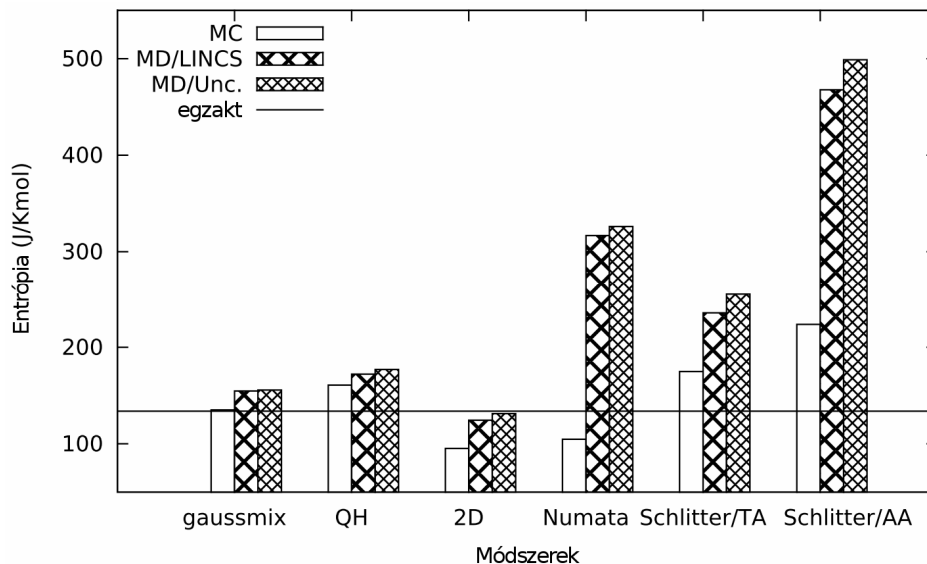
Amellett, hogy az MD sokaságok több szabadsági fokkal rendelkeznek, a belső koordináták terében mutatott eloszlásuk is különbözik az MC sokaságokétól. A kötőhosszak és/vagy kötőszögek változásának megengedése lehetővé teszi bizonyos nagyobb energiájú konformációk relaxálódását, ezáltal megváltoztatva az egyes állapotok előfordulási valószínűségét a torziós szögek terében. Ezt illusztrálja a 8. ábra a „val” adatkészletre.

Az MC sokaságokhoz hasonlóan az MD/LINCS és MD/Unc sokaságokra is kiszámoltam a konformációs entrópiát a Gauss-keveréket alkalmazó módszerrel és a négy, irodalomból átvett módszerrel is. Az eredményeket a 3. táblázat foglalja össze. Fontos, hogy változó kötőhosszak és kötőszögek esetén az összes konformáció enumerációja nem kivitelezhető, így ezekre a sokaságokra nem volt lehetőség egzakt konformációs entrópiát számolni.

Összehasonlítva a három generált sokaságra (MC, MD/LINCS, MD/Unc) kapott értékeket mind az öt módszer esetén, több szabályszerűségekre is felfigyelhetünk. Ennek illusztrálására a 9. ábrán mutatom be a számolt entrópiaértékek nagyságát a „val2” rendszer esetére, a többi rendszerre az eredmények kvalitatívan hasonlóak.



**8. ábra.** Az egyes torziós szögek egzakt eloszlásainak és a molekuladinamikai trajektóriából kapott előfordulási valószínűségeinek összehasonlítása a „val” rendszerre. Az MD szimulációban a kötőhosszak nagyságát állandó értéken tartottam a LINCS algoritmus segítségével. Az egzakt eloszlás a Boltzmann-eloszlásból származik, amelyet a fix kötőhosszakkal és kötőszögekkel rendelkező teljes konformációhalmazból számoltam. Az MD sokaságból kapott eloszlásokban megfigyelhető a csúcsok eltolódása az egzakt eloszláshoz képest (pl.  $\psi = 90^\circ$  környékén), és több régióban az eloszlás is különbözik (pl.  $\chi = 60^\circ$  és  $\phi = 50^\circ$  körül). Ezen eltéréseket az MD szimulációkban flexibilisen kezelt kötőszögek okozzák.



**9. ábra.** A „val2” rendszer entrópiái grafikonon ábrázolva, öt különböző módszerrel számolva a három sokaságra. Az értékek megfelelnek a 3. táblázatban bemutatottaknak. Az egzakt entrópiáértéket vízszintes vonal mutatja. Az MC sokaságok esetén a legjobb egyezést az egzakt értékkel a Gauss-keveréket alkalmazó módszer adja. Az MD sokaságok több szabadsági fokkal rendelkeznek (kötőszögek és kötőhosszak), és az egzakt entrópiájuk nem ismert. Esetükben a számolt entrópia nagymértékben függ attól, hogy az egyes módszerek hogyan veszik figyelembe a kötőszögek és kötőhosszak vibrációiból adódó entrópiajárulékot.

Az MD sokaságokra kapott entrópiaértékek általánosan mindig magasabbak, mint az MC sokaságokra kapott értékek. Az MD/Unc sokaságok majdnem mindig magasabb entrópiaértéket adnak, mint az MD/LINCS sokaságok, bár a különbség viszonylag kicsi. Ez arra utal, hogy a kötőszögek vibrációinak járuléka csekély. Megfigyelhető az is, hogy a Descartes-koordinátákat használó módszerek (a Schlitter- és Numata-féle módszerek) lényegesen magasabb entrópiaértéket eredményeznek mindkét MD sokaság esetén, mint az MC sokaságokra. Ezzel ellentétben a belső koordinátákat használó módszereknél (a kváziharmonikus, a 2D entrópia és a Gauss-keverék módszerek) ez a különbség sokkal kisebb. Ez azt mutatja, hogy a kötőszögek vibrációja jelentős járulékot ad a konfigurációs entrópiához, amit a belső koordinátákat használó módszerek nagyrészt elhanyagolnak. Ugyanakkor ezek a módszerek így is tükrözik az MD sokaságok megváltozott eloszlását az MC sokaságokéhoz képest.

#### **4.1.6. A Gauss-keverék módszer összehasonlítása más módszerekkel**

A Gauss-keveréken alapuló entrópiabecslő módszer tesztelésére ötféle, komplex energiefelszínnel rendelkező peptidrendszert használtam. Az eredmények alapján az mondható, hogy az MC sokaságokra kapott értékek kiváló egyezést mutatnak (átlagos eltérés 2,04%) az egzakt, enumeráción alapuló értékkel (3. táblázat). A többi tesztelt módszer lényegesen rosszabb egyezést mutatott az egzakt értékekkel (átlagos eltérés 10–40%).

A módszerek ilyen típusú eltérő teljesítményének több oka is lehet. Az egyik az entrópia becslésére alkalmazott eltérő elvek alkalmazása. A vizsgált módszerek egyik csoportja, a kváziharmonikus és a 2D entrópia módszere a Gauss-keverékeken alapuló módszerhez hasonlóan a konfigurációs térben elhelyezkedő mintákra egy valószínűsűrség-függvényt illeszt, majd annak az információs entrópiáját becsli. Mivel a Gauss-keverékeken alapuló módszer pontosabb sűrűségbecslést alkalmaz (több Gauss-komponens, projekciók helyett többváltozós Gauss-függvények), ezért ennek megfelelően pontosabb eredményeket is ad. A szerepelt másik két módszer más elveken alapul. A Schlitter-féle módszer főkomponens-analízist végez a koordináták tömeggel súlyozott kovarianciamátrixán, majd a kapott komponenseket független, kvantummechanikai oszcillátoroknak tekinti, és ezek entrópiajárulékait számolja ki. A Numata-féle módszer is első közelítésként egy ehhez hasonló értéket számol, majd különböző korrekciókat alkalmaz, amelyek a minták sűrűségfüggvényének paraméterek nélküli becsléséből származnak. A Gauss-keverékeken alapuló módszer, a kváziharmonikus és 2D entrópia módszereihez hasonlóan, pusztán a megfigyelt konfigurációs mintákra hagyatkozik, míg a másik két



módszer az adatok háttérében rejlő fizikai rendszerről tett feltételezésekből indul ki. Ez a különbség abban is látszik, hogy a Schlitter- és Numata-módszerrel ellentétben a Gauss-keverék módszer nem használja fel az atomi tömegeket a számításokhoz. A fizikai képből kiinduló módszerekben gyakori az oszcillátorok kvantummechanikai kezelése, ez azonban csak nagyfrekvenciás mozgási módusok esetén indokolt, mint például a kötэшossz-rezgések. Ehhez hasonló korrekciókat a Gauss-keverék módszer nem vesz figyelembe.

A módszerek e kétféle csoportja abban is különbözik egymástól, hogy milyen koordinátákat használ a számításokhoz. A Gauss-keverék módszer, a kváziharmonikus és a 2D entrópia módszeréhez hasonlóan, belső koordinátákat használ, míg a másik csoportba tartozó módszerek Descartes-koordinátákkal dolgoznak. Ennek eredményeként a Schlitter- és Numata-módszerekkel számolt entrópia tartalmazza a kötэшosszak és kötэшszögek vibrációiból származó járulékokat is, szemben a másik három módszerrel, amelyek ezeket elhanyagolják. Ezeknek a vibrációs járulékoknak a fontossága attól függ, hogy mi az entrópiaszámítás célja. Karplus és Kushick részletesen vizsgálták [18], hogy az egyes belső koordináták mekkora járulékot adnak egy makromolekula két konformációja közötti entrópiakülönbséghez. Eredményeik azt mutatják, hogy a kötэшosszak vibrációiból adódó járulékok gyakorlatilag függetlenek a konformációtól, ezért entrópiakülönbségek számolása esetén elhagyhatóak. Ezzel szemben a kötэшszögek rezgése jelentős mértékű csatolást mutattak a torziós szögek fluktuációihoz, emiatt már egyetlen konformáció entrópiájának számítása esetén sem elhanyagolhatóak, és megjelennek a két állapot közötti entrópiakülönbségben is. Ez megfigyelhető azon eredményeink esetében is, ahol a kötэшszögeket impliciten figyelembe vevő Descartes-koordinátákat használó módszerek lényegesen magasabb entrópiát eredményeznek az MD/LINCS sokaság esetében az MC sokasággal szemben. A Gauss-keveréken alapuló módszer esetében a kötэшszögek nem voltak külön koordinátaként figyelembe véve a szabadsági fokok számának alacsonyan tartása érdekében, de a módszer egyszerűen kiterjeszhető ilyen formában.

A belső koordináták entrópiaszámoláshoz történő felhasználása több problémát is felvet. A belső koordináták egy független rendszerének definiálása gyakran nehéz és nem egyértelmű, figyelembe kell venni a Jacobi-determináns által okozott metrikus járulékokat, és a torziós szögek periodicitása miatt a hagyományos harmonikus oszcillátorral való közelítés kérdéses [21]. Ennek ellenére belső koordinátákat használó módszereknek jelentős előnye a Descartes-koordinátákat használó módszerekkel szemben az, hogy újabb adathalmaz létrehozása nélkül, könnyen szétválaszthatóak az egyes szabadsági fokok járuléka az entrópiához. Egyetlen, a kötэшhosszak és kötэшszögeket is kényszerek nélkül kezelő MD

trajektóriából a belső koordináták bármilyen alkalmas részhalmazára számolhatunk entrópiát. A kötésszögek entrópiajáruléka például megkapható a kötésszögeket tartalmazó és nem tartalmazó belső koordinátákból számolt entrópia különbségeként. Ha az entrópiát közvetlenül a Descartes-koordináták felhasználásával számolnánk, akkor egy új trajektóriára lenne szükség, ahol a kötésszögek értéke állandó.

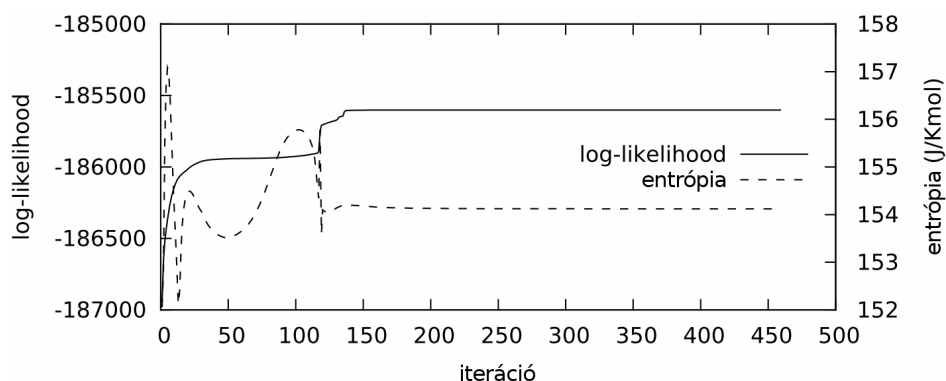
A Descartes-koordinátákról a belső koordinátákra való áttérés során a transzformáció Jacobi-determinánsa megjelenik a konfigurációs integrálban (84), és befolyásolja a kapott entrópia értékét. Ezt a hatást a belső koordinátákat használó módszerek nem veszik figyelembe arra hivatkozva, hogy a Jacobi-determináns csak a kötéhosszaktól és kötésszögektől függ, amelyek viszont csak egy konformációtól független, szűk tartományban változnak [18]. Ha elhanyagoljuk ezeknek a kemény szabadsági fokoknak az egyéb belső koordinátákhoz (pl. torziós szögekhez) való csatolódását, akkor ezek a szabadsági fokok a Jacobi-determinánssal együtt egy konstans járulékot adnak az entrópiához. A Jacobi-determinánssból adódó járulék egy újabb forrása annak, hogy a pusztán belső koordinátákat használó módszerek különböző eredményeket adnak a Descartes-koordinátákban számoló módszerektől.

## **4.2. *Diszkusszió: komplexitás, bővíthetőség, alkalmazási lehetőségek***

A Gauss-keverék módszer tesztelésére kisszámú (3–5) szabadsági fokkal rendelkező rendszereket és állandó számú (10) Gauss-komponenst használtam a Gauss-keverékekben. Ezekben az esetekben gyors konvergenciát tapasztaltam (10. ábra), és a Gauss-keverékekből kapott entrópiaértékek jó egyezést mutattak az egzakt értékekkel. Az illesztésnek és az entrópiabecslésnek a futási ideje a legtöbb szabadsági fokkal rendelkező „ala-val-ala” rendszer esetén sem haladta meg az 54 órát egy Intel Xeon E5430 számítógépen. **Az abszolút futási idő értékelésekor fontos figyelembe venni, hogy a számítást végző program interpretált szkriptnyelven íródott, és a megvalósításakor nem törekedtem a futási idő szerinti optimalizálására. Ezért pusztán a programkód optimalizálásával és más programnyelv választásával jelentős javulást lehetne elérni a futási idő tekintetében.** Nagyobb rendszerek esetén azonban egyrészt a Gauss-komponenseket leíró mátrixok mérete is növekszik ( $d^2$  mátrixelem  $d$  szabadsági fok esetén), másrészt szükség lehet a Gauss-komponensek számának ( $k$ ) növelésére is. A módszer jelenlegi megvalósításának skálázódása  $O(nd^3k)$ , ahol  $n$  az adatpontok száma. Nagyobb rendszerekre ezért várhatóan a számítások konvergenciája is nehezebben érhető el. A szabadsági fokok számának csökkentésére nagyobb rendszerek (pl.

fehérjék) esetén különböző technikákat lehet alkalmazni. Teljes korrelációs analízis (FCA, [123]) segítségével a rendszert feloszthatjuk minimálisan csatolt alterekre [30], majd adott esetben a kölcsönös információn alapuló kifejtéssel a dimenziószám tovább csökkenthető [124]. Ezután a csatolt alrendszerek entrópiája a Gauss-keverék módszerrel könnyen és gyorsan kiszámítható. A szükséges Gauss-komponensek számának megbecslésére alkalmazhatók klaszterezési módszerek (pl. *fuzzy* klaszterezés), amelyekkel a konformációs sokaságban kompakt klaszterek azonosíthatók.

A Gauss-keverék módszer kedvezőtlen skálázódásának enyhítésére felmerülhet az algoritmus párhuzamosításának és a párhuzamos architektúrákon való futtatásának lehetősége. Az illesztéshez használt (70)-(72) kifejezésekben vektor- és mátrixműveletek szerepelnek, amelyek könnyen párhuzamosíthatóak. Emellett az egyes Gauss-komponensek paramétereinek számítása is végezhető párhuzamosan. Ezzel a módszerrel a minták számától és a Gauss-komponensek számától való függést gyakorlatilag megszüntethetjük. A GPU architektúrák lehetővé teszik a masszívan párhuzamos adatfeldolgozást, amelyet  $n = 10^4$ – $10^5$  nagyságú mintahalmazon már hatékonyan ki lehetne használni.



**10. ábra.** Gauss-keverék illesztésének konvergenciája az „ala3” MD/LINCS adathalmazra. Az EM iterációk számának függvényében ábrázolva a paraméterek log-likelihood értékét és az illesztett Gauss-keverék entrópiáját látható, hogy a log-likelihood értékek monoton növekednek egy határértékhez tartva.

A jelenlegi munkám annak a bemutatására szolgál, hogy a Gauss-keverékeken alapuló megközelítés működőképes és pontos módszer nem-diffúzív molekuláris rendszerek konfigurációs entrópiájának becslésére.

## 5. ABC exporterek konformációinak dinamikai összehasonlítása

Az ABC exporter fehérjék esetén a molekulán belüli allosztérikus kommunikáció fontos szerepet játszik a működésben, hiszen a transzporthoz szükséges ATP hasítása a transzport színterétől viszonylag távol következik be. Feltételezések szerint a transzportfolyamat kezdetén a szubsztrát a fehérje „alul nyitott” konformációjában, a transzmembrán régióban kötődik [56, 125-127]. A citoplazmatikus oldalon ATP bekötődésével létrejövő nukleotidkötött konformáció lehetővé teszi a szubsztrát távozását az extracelluláris térbe. A transzportciklus e két végpontja (az „alul nyitott” és „alul zárt” konformációk) közötti átmenet mechanizmusa azonban még nem teljesen ismert. Az egér MDR3 transzporter röntgendiffrakcióval megállapított „alul nyitott” konformációja egy lehetséges szerkezeti modellje az apo fehérjének [56]. Ez a konformáció az NBD-k jelentős eltávolodását mutatja, ami több kísérleti eredménnyel sem egyeztethető össze [63, 64]. Szintén nincs egyetértés azon a téren, hogy mindkét kötött ATP hasítása szükséges-e a transzportciklus folytatásához, amiben a hidrolízis szerepe is kevésbé ismert.

Munkámban molekuláris dinamikai szimulációk segítségével összehasonlítom a transzportciklus különböző pontjait reprezentáló ABC exporter szerkezetekben megfigyelhető belső, korrelált mozgásokat. Az ATP-hidrolízis szerepét a Sav1866 bakteriális transzporter „alul zárt” holo szerkezetében, két különböző állapotban vizsgáltam. Egyik esetben mindkét nukleotidkötő helyet ATP foglalta el, míg a másik esetben az egyik kötéshelyre ATP molekulát, a másikra pedig ADP molekulát kötöttem. Továbbá a humán MDR1 fehérje szerkezeti modelljeinek segítségével megvizsgáltam a holo és az apo konformációk dinamikáját és a két konformáció közötti átmenetet.

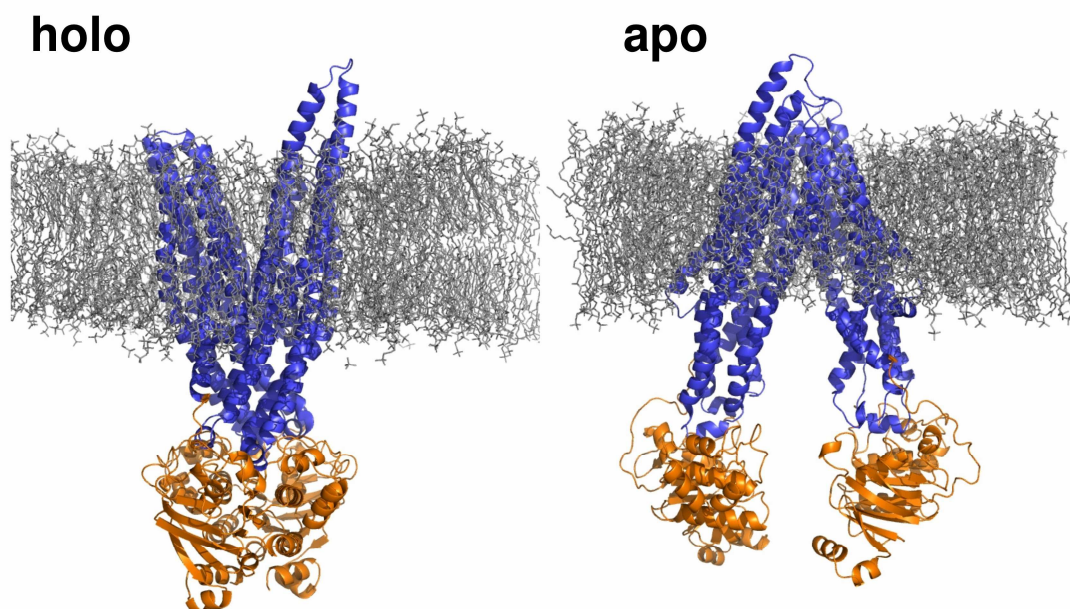
### 5.1. Eredmények

#### 5.1.1. A szimulációs rendszerek felépítése

Az ABC exporterek dinamikai jellemzését molekuláris dinamikai (MD) szimulációk alapján végeztem, amelyhez többféle fehérjeszerkezetet is felhasználtam. A jelenleg elérhető, kísérletileg meghatározott szerkezetek között kettő olyan található, amely a teljes hosszúságú fehérjét tartalmazza és elegendően nagy felbontású. Ezek a szerkezetek két eltérő konformációt mutatnak, egy „alul zárt” konformációt (Sav1866; PDB ID: 2HYD) és egy „alul nyitott” konformációt (egér MDR3; PDB ID: 3G5U). Azonban a két különböző konformáció

jellemzőinek összehasonlításához olyan szerkezetekre volt szükség a szimulációk végrehajtása során, amelyek azonos fehérjékhez (aminosav-szekvenciához) tartoznak. Ezért azon számítások során, amelyekben a két konformáció összehasonlítása volt a cél, a humán MDR1 fehérje homológiamodelljeit használtam. Ezek a modellek a fenti Sav1866 és mMDR3 szerkezetek alapján készültek. A többi szimuláció kiindulási szerkezeteként röntgen krisztallográfiával meghatározott szerkezeteket használtam (Sav1866; PDB ID: 2HYD, mMDR3; PDB ID: 3G5U).

A kiválasztott fehérjeszerkezeteket egy egyensúlyba hozott lipid kettősrétegbe helyeztem. A membránba illesztés kritikus pontja minden membránszimulációnak, mert a fehérje transzmembrán régiói nem minden esetben ismertek pontosan. A szekvencia alapú transzmembrán hélix prediktorok gyakran nem adnak racionális eredményeket ABC transzporterekre. Néhány módszer a transzmembrán hélixek számában is téved olyan esetben, amelyben azt már kísérletileg meghatározták. Ezért mind szekvencián, mind szerkezeten alapuló predikciós módszereket (HMMTOP [111] és TMDDET [110]) is használtam a szerkezetek membránba való beágyazásához (11. ábra). A membránba való beillesztés után az N-terminális amfipatikus „könyökhélixek” felfeküdtek a membrán belső felére, ami azt mutatja, hogy a beillesztés megfelelő volt.

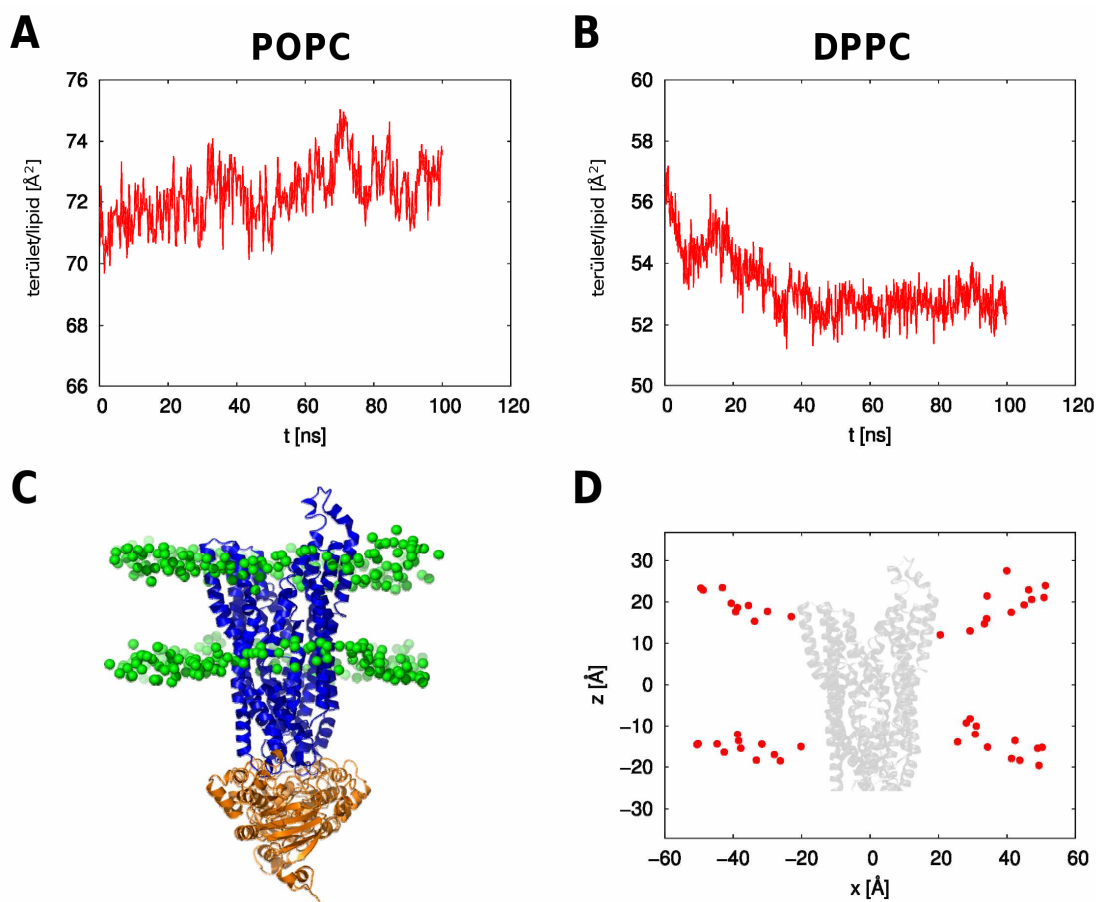


**11. ábra.** Az ABC exporterek molekuláris dinamikai szimulációkban használt két konformációja. Az ábrán a humán MDR1 „alul zárt” holo (bal oldalt) és „alul nyitott” apo (jobb oldalt) homológiamodelljei szerepelnek, lipid kettősrétegbe ágyazva.

### 5.1.2. A lipid kettősréteg viselkedése

A Sav1866 „alul zárt” rendszerek szimulációiban POPC lipidből álló kettősréteget használtam. A membrán minőségét a szimulációkban a fejcsoportokra eső átlagos felület nyomkövetésével ellenőriztem, ami POPC esetén  $67 \text{ \AA}^2$  és  $71 \text{ \AA}^2$  között változott (12.A. ábra). Ezek az értékek konzisztensek a lipidparaméterek ellenőrzésekor kapottakkal, és jól egyeznek a kísérleti értékekkel is [109]. A membrán vastagsága  $36$  és  $38 \text{ \AA}$  között változott. Mindezek az értékek azt mutatják, hogy a POPC kettősréteg a szimulációk során stabil és egyensúlyi állapotban van.

A humán MDR1 szerkezetek szimulációit DPPC kettősrétegben végeztem, hogy ezzel egy merevebb lipidekből álló környezetet hozzak létre. Ez várhatóan jobban korlátozza a fehérje mozgását, és ezzel jobban imitálja az alacsony hőmérséklet hatásait, amin a kristályosítás történt. Mivel az „alul nyitott” konformáció más források szerint is instabil a GROMOS96 *force-field*del (M. L. O’Mara, személyes konzultáció), ezért ezekhez a számításokhoz az OPLS *all-atom force-field*et használtam. **A különböző *force-field* használatának célja az volt, hogy kizárjam azt a lehetőséget, hogy az instabilitás esetleg specifikus a használt *force-field*re.** A fejcsoportra eső átlagos felület a DPPC-vel végzett szimulációk esetében kisebb volt ( $53 \text{ \AA}^2$ ), mint amit folyékony membrán esetében kellett volna kapni ( $64 \text{ \AA}^2$ ) [109], mert fiziológias hőmérsékleten a DPPC membránok gél fázisban vannak (12.B. ábra). A DPPC kettősréteg átlagos vastagsága  $34,8 \text{ \AA}$  volt. Szintén megfigyelhető, hogy a fehérje környezetében a membrán elvékonyodik (12.C. ábra). A kettősréteg vastagságát egy keresztmetszeti képen ábrázoltam a fejcsoportok koordinátáinak az  $x-z$  síkra történő vetítésével (12.D. ábra). Az ábrán jól látszik az elvékonyodás a fehérje körül, amelynek az oka az aminosav oldalláncok és a lipidmolekulák között fellépő hidrofób/hidrofil kölcsönhatás.

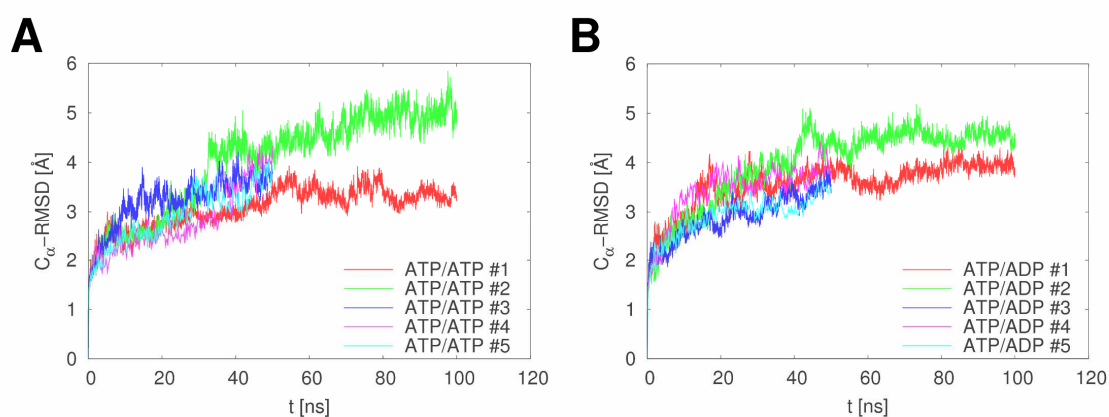


**12. ábra. A lipid kettősréteg viselkedése.** (A) A POPC membránsűrűség stabil a szimulációk során, és átlaga a várt 67 és 71 Å<sup>2</sup> közé esik (az ábrán az 1. számú Sav1866 ATP/ATP szimuláció értékei láthatóak). (B) Az átlagos DPPC sűrűség 53 Å<sup>2</sup>, mivel a szimuláció hőmérsékletén a DPPC membrán gél fázisban található (37°C, holo hMDR1 szimuláció). (C) A rendszer oldalnézeti képe mutatja a fehérje körül a membrán elvékonyodását (holo hMDR1). (D) A keresztmetszet mentén levő lipid fejcsoportok pozícióját az  $x$ - $z$  síkra vetítve látható a membrán vastagsága. Az origót a fehérje transzmembrán régiójának a középpontjába helyeztem. A  $z$  tengely merőleges a membrán síkjára.

### 5.1.3. ATP/ADP csere hatása a korrelált mozgásokra

Az ATP hidrolízisének hatását a teljes transzporter dinamikájára a Sav1866/POPC fehérje/lipid rendszerben vizsgáltam. Kétféle rendszert hoztam létre, az egyik esetén mindkét nukleotidkötő helyen ATP szerepelt (ATP/ATP rendszer), a másik esetben a B lánc Walker A motívumához tartozó kötőhelyen az ATP-t ADP-re cseréltem (ATP/ADP rendszer). Mindkét rendszerre két 100 ns hosszú és három 50 ns hosszú molekuláris dinamikai szimulációt végeztem. A C<sub>α</sub> atomok koordinátáira számolt átlagos négyzetes gyök eltérés (RMSD) a trajektóriákban 3–5 Å között változott, ami a szimulációs rendszer stabilitását jelzi (13. ábra) [71]. A TMD-k esetén nagyobb változások észlelhetők (3,5–5 Å RMSD), mint az NBD-k esetén (2,5 Å), de a domének összességében megtartották szerkezetüket. **A számításokhoz felhasznált szakaszban (30–50 ns) látható egy állandó drift az RMSD értékekben a 13. ábrán, ami azt jelzi, hogy a rendszer ebben az időtartományban még nincs egyensúlyban, és**

vélhetően konformációs változáson megy keresztül. Az ehhez hasonló drift a Sav1866 fehérjével végzett membránbeli szimulációk esetén gyakori (lásd pl. [71, 128]). A publikációkban megjelenő RMSD grafikonokon a széthúzott ábra vagy rövidebb szimulációk miatt ez gyakran nehezen vehető észre. Az RMSD értékek folyamatos emelkedése akkor sem tűnik el, ha az RMSD számolásakor referenciaként a 30 ns (az analizált szakasz kezdetén) időpillanatban előálló szerkezetet vesszük. A trajektóriából számolt eredményekben a drift egy szisztematikus hibát okoz, azonban ennek a hatását enyhíti a több, független trajektória összesítéséből származó adatkészletet használata.



**13. ábra. A Sav1866 szerkezet stabilitása a szimulációkban.** A fehérje stabilitását a szimulációk során a  $C_{\alpha}$  atomoknak a kezdeti szerkezettől számolt átlagos négyzetes gyök eltéréssel (RMSD) jellemeztem. Ilyen méretű fehérje esetén általánosan 5–6 Å alatti RMSD értékek esetén stabilnak tekinthető a rendszer. A Sav1866 ATP/ATP (A) és ATP/ADP (B) szimulációk adatai láthatók az ábrán.

Az ATP/ATP és ATP/ADP rendszerek közötti dinamikai különbségek összesítésére a McClendon és munkatársai által létrehozott MutInf módszert alkalmaztam [118]. Ez a módszer a fehérje torziós szögei közötti kölcsönös információ kiszámításán alapul, és ezáltal a fehérjében végbemenő atomi skálájú, egyidejű konformációs változásokat képes detektálni. A két oldallánc közötti magas kölcsönös információ értéke erős dinamikai korrelációt jelent, és használható olyan oldalláncok azonosítására, amelyek allosztérikus kommunikációban vesznek részt [118]. A módszer a kölcsönös információ mennyiségét minden aminosavpár összes torziós szögére kiszámolja, majd aminosav-páronként összegzi. Ezt a számolást az ATP/ATP és az ATP/ADP rendszerre is elvégeztem. Az oldalláncok közötti korrelált mozgások bemutatására egy olyan korrelációs hálózatot alkottam, amelyen az oldalláncok a típusuk, számuk és láncazonosítójuk szerint vannak jelölve, az őket összekötő vonalak vastagsága pedig a korreláció mértékével arányos (14. ábra).

Ez a korrelációs hálózat az ATP/ATP rendszer esetében sok éllel rendelkezik, ami nagymértékű kapcsoltságot mutat (14.A. ábra). Három olyan klikket (teljes részgráfot)



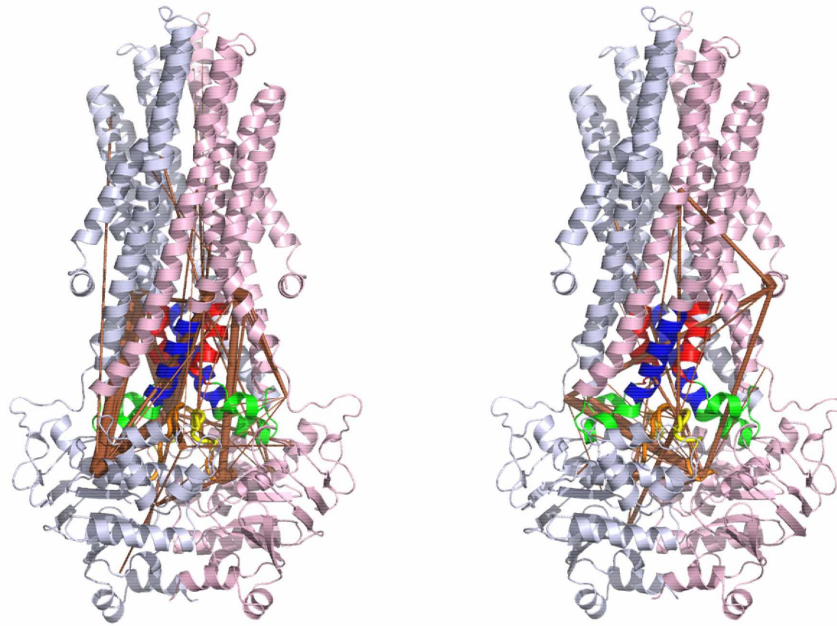
figyelhetünk meg, amely legalább három tagot tartalmaz. Ezek egyike a B láncban található, ATP jelenlétét észlelő Gln422B oldalláncot tartalmazó csoport (Gln422B-Asp323B-Met311B-Asp312B), míg a másik kettő a konzervált Glu473B oldalláncot tartalmazza (Glu473B-Ser310A-Asn494A-Met311A-Asn495A és Glu473B-Gln116A-Glu191B-Met311A-His204B). A Gln422B oldalláncot tartalmazó csoport és a Glu473B oldalláncot tartalmazó két csoport egymástól gyakorlatilag függetlennek tekinthető, közöttük csak gyenge dinamikai kapcsoltság figyelhető meg.

A hálózatban központi szerepet játszanak az ún. tetrahélix-köteg aminosavjai, amelyet a TM3 és TM4 hélixek citoplazmatikus szakaszai alkotnak az „alul zárt” szerkezet központi tengelye mentén. A tetrahélix-köteg az X-hurok oldalláncaival mutat erős együttmozgást. A 473-as pozícióban található konzervált Glu igen fontos szerepet játszik az NBD-kből a transzmembrán domének felé történő konformációs változások továbbításában [72, 79]. Ezzel összhangban a Glu473B központi helyet foglal el a korrelációs hálózatban, környezetében a Met311A, Ser310A, Asn494A, Asn495A, és Gln116A aminosavakkal. Ennek a csoportnak a szimmetrikus (másik láncbeli) párja szintén szerepel a hálózatban, de a Glu473A és B oldallánc környezetében a kapcsolódásokban aszimmetria figyelhető meg. Az X-hurok és tetrahélix-köteg közötti aszimmetrikus dinamikai csatlakozás utalhat arra az ABC területen pillanatnyilag a legelfogadottabb működési mechanizmusra, miszerint a NBD dimerek aszimmetrikusan vesznek részt az ATP hasításában [65]. Fontos megjegyezni azonban, hogy ez az eredményünk mások hasonló eredményeihez hasonlóan származhat a felhasznált Sav1866 fehérjeszerkezetben megfigyelhető eredendő aszimmetriából is. A röntgendiffrakciós szerkezetben több aszimmetrikus konformációban levő oldallánc is található a transzporter központi tengelye mentén, mint például az Asn102, Asn126, Gln200 vagy His204 aminosavak.

Az ATP/ADP rendszer korrelációs hálózata egyszerűbb szerkezetet mutat, gyengébb kötésekkel és kisebb klikkekkel, ami azt jelzi, hogy az oldalláncok között általában gyengébb a csatlakozás (14.B. ábra). A Glu473B oldallánc központi szerepét a Gln422B vette át. A Gln422A/B-Gln421A/B oldalláncok csoportja, amely az ATP-érzékelő oldalláncokat is tartalmazza, csatlakoztatá vált a Ser108A oldalláncot is tartalmazó *coupling* hélixhez, amely a nukleotidcsere színhelyéül szolgáló NBD-vel is kölcsönhatásba lép. Az ATP/ATP rendszer esetén megfigyelhető csoportok itt hiányoznak, ami az X-hurok és a tetrahélix-köteg közötti csatlakozás megszűnését mutatja. Ez a szétkapcsolódás elősegítheti a tetrahélix-köteg szétesését, amelyről célzott (*targeted*) MD szimulációk megmutatták, hogy a transzporter katalitikus ciklusának egyik fontos lépése [73]. Említésre méltó még, hogy a TM1 és TM2



**A Sav1866 ATP/ATP**      **B Sav1866 ATP/ADP**



**15. ábra. A korrelációs hálózatban megjelenő aminosavak térbeli elhelyezkedése.** Az ábrán a Sav1866 ATP/ATP (A) és ATP/ADP (B) rendszer esetén talált, korreláltan mozgó aminosavak hálózata látható a fehérje térbeli szerkezetére vetítve. Az egymással kapcsolt aminosavakat barna vonal köti össze, melyek vastagsága a kölcsönös információ nagyságával arányos (hasonlóan a 14. ábrához). A kiemelt régiók színezzése megegyezik a 14. ábrán használttal.

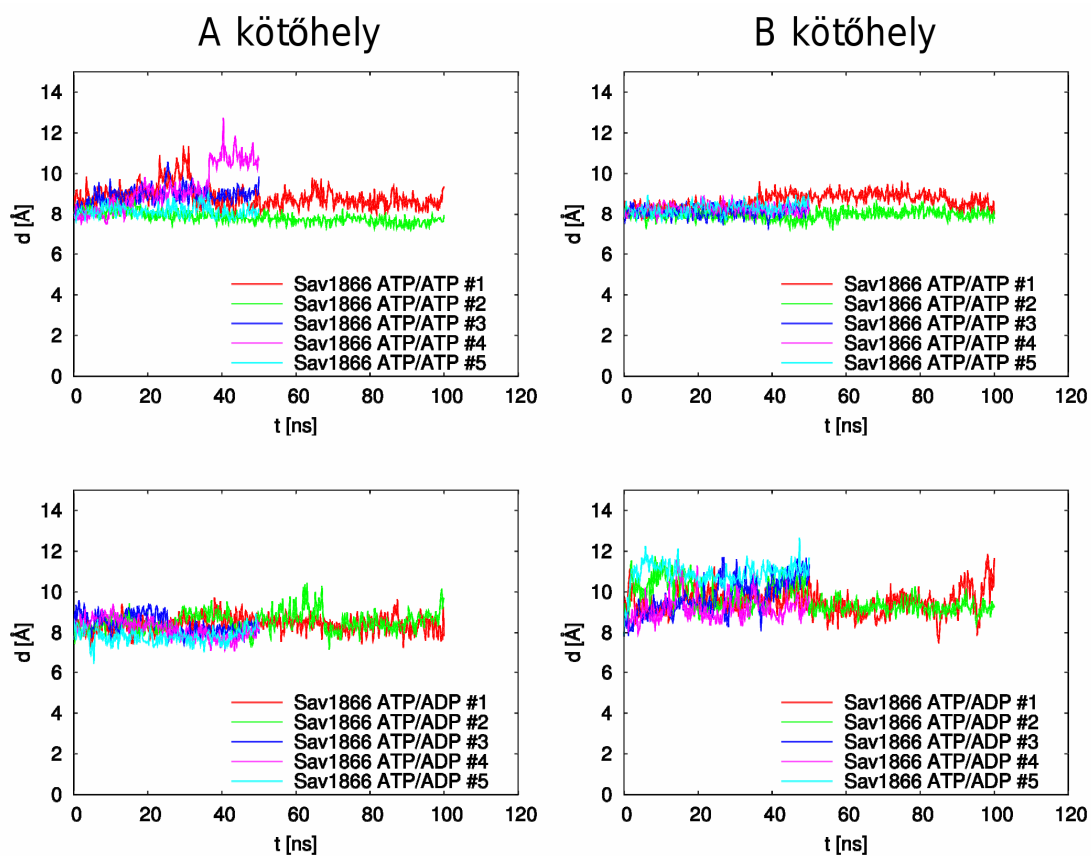
A korrelációs hálózat alapján a fenti, ismert funkcióval rendelkező aminosavak mellett más aminosavak is (pl. Met311, Asn494, Asn495, Asn252 vagy Asp323) fontos szerepet játszhatnak a dinamikailag erősen csatolt klaszterek létrehozásában, és ezáltal a fehérje funkciójának megvalósításában. A Met311 oldallánc például, amely az ATP/ATP rendszerben erősen csatolt más oldalláncokkal, abban a transzmembrán hélixben (TM6) helyezkedik el, amely a multidrog transzporterek droggötésében fontos szerepet játszik [129-131].

Az eredmények azt mutatják, hogy egyetlen nukleotid kicserélése az NBD-ken levő kötőhelyeken elegendő a teljes transzporter korrelált mozgásainak és dinamikai csatolásainak átrendeződéséhez.

#### **5.1.4. A holo és az apo szerkezetek közötti konformációs átmenet jellemzése**

Az irodalomban jelenleg nincs egyetértés arra vonatkozóan, hogy a két NBD egymástól való disszociációja megtörténik-e kötött ATP vagy ADP esetén [65]. A szimulációkban ha mindkét kötőhelyen ATP szerepel, jelentős disszociáció nem figyelhető meg (16. ábra). Ha a B lánc kötőhelyén a nukleotidot ADP-re cseréljük, mindkét kötőhely

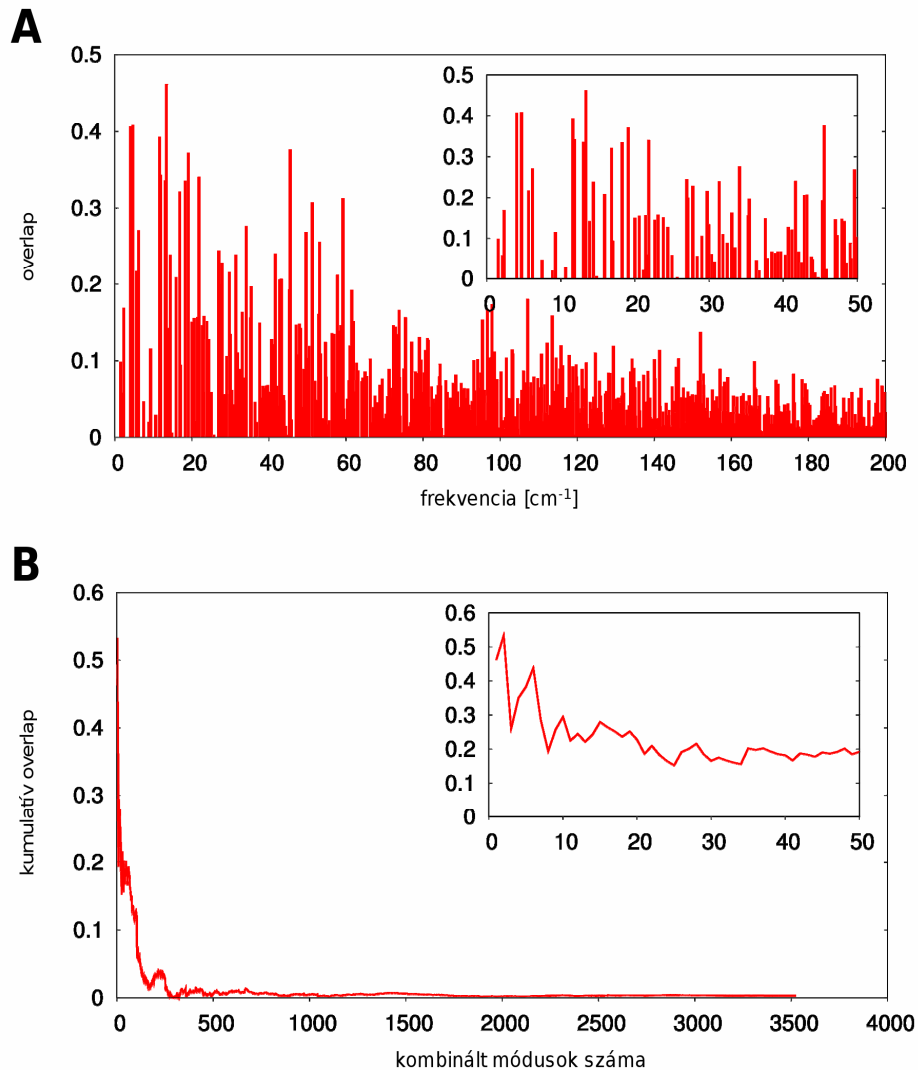
nagyobb flexibilitást mutat, és a másik kötőhely gyakrabban mutat eltávolodást a kezdeti szerkezettől. Ez jelezheti azt, hogy az ATP hidrolízise az innenső (*cis*) kötőhelyen befolyásolja a túlsó (*trans*) kötőhely viselkedését, és gyengítheti a túlsó domén nukleotidkötését. Ennek ellenére a disszociáció jelzésére használt két oldallánc között a legnagyobb eltávolodás is csak 4 Å, ami alapján nem állapítható meg egyértelműen, hogy a két domén disszociál-e az ATP/ADP rendszerben. Ennek oka lehet az, hogy az NBD-k disszociációját leíró tanulmányokkal szemben szimulációinkban nem pusztán a két nukleotidkötő domén van jelen, hanem a teljes hosszúságú fehérje. A transzmembrán domének korlátozhatják az NBD-k mozgását, ezáltal akadályozva a disszociációt.



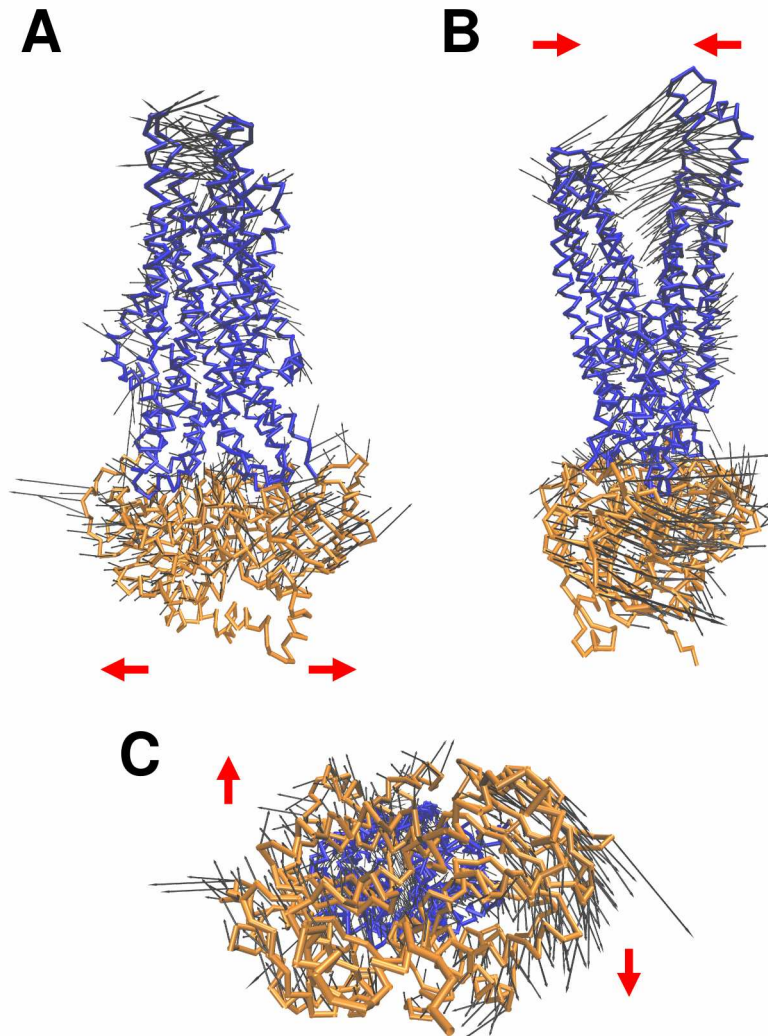
**16. ábra.** Az NBD-k nem disszociálnak jelentős mértékben sem két ATP, sem egy ATP és egy ADP jelenléte esetén. A Ser381A és Ser479B („A” kötőhely) és a Ser381B és Ser479A („B” kötőhely)  $C_{\alpha}$  atomjainak távolsága látható az ábrákon a szimulációs idő függvényében. Ez a távolság az NBD-k szétnyílását jellemzi a két nukleotidkötő helyen, a dupla ATP-kötött rendszerre (felső két panel) és az ATP/ADP kötött rendszerre (alsó két panel). Az ATP/ADP rendszerben a „B” kötőhelyen levő ATP lett ADP-re cserélve. Az eltávolodás egyik esetben sem haladja meg 2–4 Å értéket ( $d \leq 12,5$  Å), aminek oka az lehet, hogy a jelen lévő TMD-k akadályozzák az NBD-k disszociációját.

Sok fehérjében már egy rövid molekuláris dinamikai szimuláció során tapasztalhatóak olyan kollektív mozgások, amelyek a fehérje funkciójához köthető elmozdulásokat mutatnak [112, 113]. Annak érdekében, hogy az ABC exporterek holo és apo konformációi közötti

nyitó átmenethez kapcsolható esetleges belső mozgásokat jellemezni tudjam, az „alul nyitott” és „alul zárt” konformációval is végeztem szimulációkat. Mivel jelenleg nincs olyan transzporter, amelynek mindkét konformációban rendelkezésre állna atomi felbontású, kísérletileg meghatározott szerkezete, a számításokhoz a humán MDR1 két homológiamodelljét használtam. Az „alul zárt” szerkezet esetében elhagytam a nukleotidokat a kötőhelyekről, ezzel is elősegítve az „alul nyitott” forma felé vezető átmenetet. A releváns kollektív mozgások meghatározását az esszenciális dinamika (ED) módszerével végeztem, amely az atomi fluktuációk főkomponens-analízisén alapul [112]. A számított főkomponensek a rendszer egy-egy lineárisan független kollektív mozgását (módusát) írják le. Ebből kiválasztható a módusoknak egy olyan részhalmaza, amely azt az ún. esszenciális alteret feszíti ki, ahol a fehérje funkció szempontjából releváns mozgásai történnek. A hMDR1 holo konformáció esetén az apo konformáció felé történő átmenethez legjobban illeszkedő esszenciális módusokat rendszerint kevés számú kollektív mozgási módus választásával kaptam (lásd 3.8. fejezet és 17. ábra). A kiválasztott módusok szuperponálásával megkapjuk a fehérjének a szimuláció során az esszenciális alterében végzett kollektív mozgásait. A kapott összetett módus által meghatározott kollektív mozgásban a  $C_{\alpha}$  atomok kezdeti elmozdulása a 18. ábrán látható. Mindhárom szimuláció esetén az összetett módusban megfigyelhető volt egy záródó és csavarodó mozgás a transzmembrán régióban, amely ezt a régiót az „alul nyitott” szerkezet felé mozgatja, és az extracelluláris oldal záródásának felel meg (18. ábra). A 18. ábrán bemutatott móduson megfigyelhető, hogy ehhez a mozgáshoz csatolva az NBD-k oldalirányú mozgást végeztek, merőlegesen a tömegközéppontjukat összekötő tengelyre. Ez a mozgás a holo és apo szerkezetek összehasonlításából nem következik, és a holo konformációt egy „alul zárt” apo konformáció felé mozgatja. Az NBD-k ehhez hasonló, oldalirányban elcsúsztatott elhelyezkedése figyelhető meg az MsbA transzporter „alul zárt” apo (PDB ID: 3B5X) röntgendiffrakciós szerkezetében. Az „alul zárt” apo konformációban az NBD-k közötti kapcsolatok fellazulnak, de a két domén közötti távolság nem akkora, mint az „alul nyitott” apo konformációban. A megfigyelt kollektív mozgás, bár csak háromból egy szimuláció esetén fordult elő, azt mutatja, hogy a holo fehérje képes egy „alul zárt” konformáció felé tartó közvetlen átmenetre, és nem feltétlenül van szükség az NBD-k egymástól való nagymértékű eltávolodására. A másik két szimulációban nem volt megfigyelhető olyan esszenciális altér, amelyben az NBD-k elmozdulása az „alul nyitott” apo állapot felé vezetett volna.



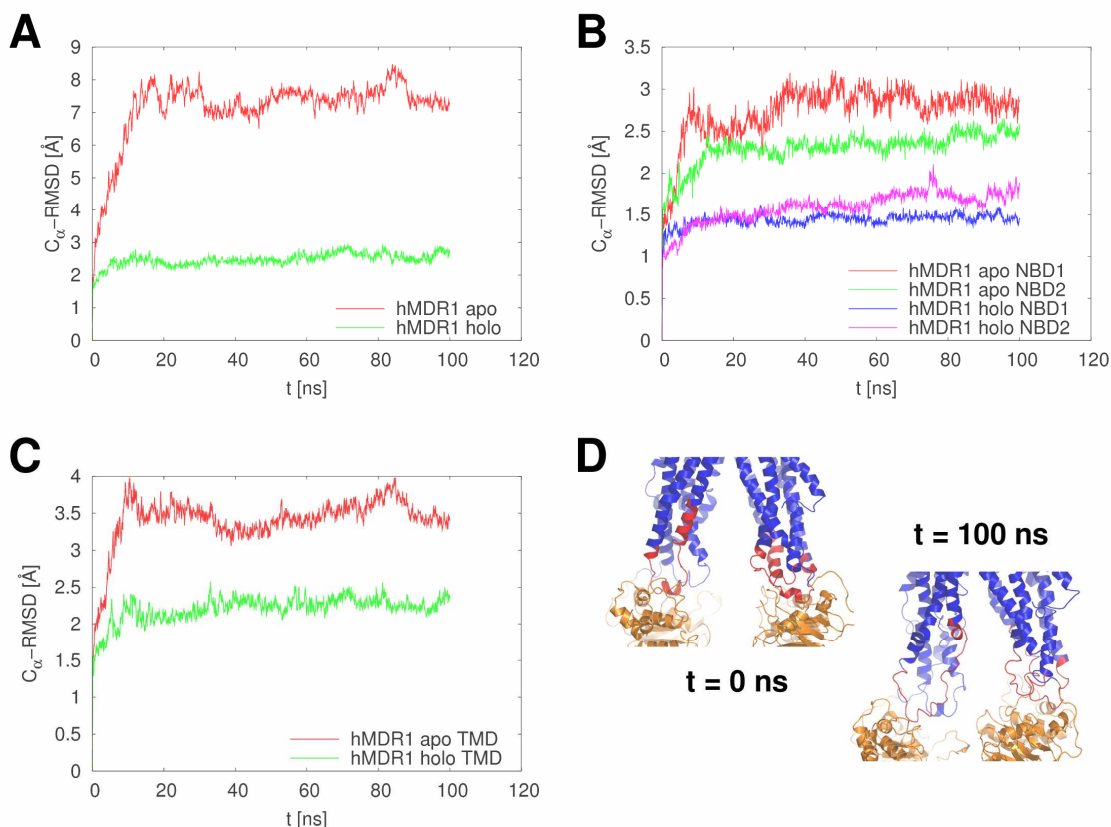
**17. ábra. A holo–apo konformációs átmenetet jellemző kollektív módusok kombinációjának meghatározása.** (A) A nukleotid nélküli holo hMDR1 rendszerre számolt alacsony frekvenciájú ( $< 200 \text{ cm}^{-1}$ ) kollektív módusok a holo–apo konformációs átmenet irányával való átfedése a módusok frekvenciájának függvényében (ld. 3.8. fejezet). Magas átfedéssel rendelkező módusok főleg az alacsony frekvenciájú tartományban találhatóak. (B) Az első  $n$ , legnagyobb átfedést mutató módus superponálásával kapott összetett módus átfedése a holo–apo konformációs átmenettel, a superponált módusok számának ( $n$ ) függvényében. A két legnagyobb átfedéssel rendelkező módus kombinációja adja a holo szerkezetből az apo szerkezet felé mutató elmozdulás legjobb közelítését. A beillesztett ábrák a grafikonok kezdeti szakaszainak nagyítását mutatják.



**18. ábra.** Az esszenciális dinamikai (ED) módusok szuperpozíciója meghatároz egy átmeneti irányt a holo konformációból az apo konformáció felé. Az esszenciális alteret meghatározó két kollektív módus szuperpozíciója látható az ábrákon. Az egyes  $C_{\alpha}$  atomok elmozdulásának irányát és relatív nagyságát a fekete vektorok jelzik. Az oldalnézet (A) mutatja az NBD-k egymástól való távolodását. Az elforgatott nézeten (B) látható a transzmembrán régió extracelluláris (felső) szakaszainak záródása. Alulnézetben (C) látható az NBD-k oldalirányú kitérése, amely a rendszert az MsbA fehérje esetén megfigyelt „alul zárt” apo konformáció (PDB ID: 3B5X) irányába mozgatja.

### 5.1.5. Az „alul nyitott” szerkezet instabilitása

A hMDR1 transzporter holo és „alul nyitott” apo konformációjának dinamikai összehasonlítására tett kísérlet végül nem járt sikerrel, mert az apo hMDR1 modell instabilnak mutatkozott a molekuláris dinamikai szimulációk során (19. ábra). Annak ellenőrzésére, hogy az instabilitás nem a homológiamodell pontatlanságainak köszönhető, elvégeztem egy hasonló szimulációt a mintaként használt egér MDR3 szerkezettel is (PDB ID: 3G5U). A szimuláció során azonban ez a szerkezet is hasonló módon, jelentősen eltávolodott a kezdeti állapottól, és instabilnak mutatkozott (20. ábra). A továbbiakban ezért ezt elemzem.

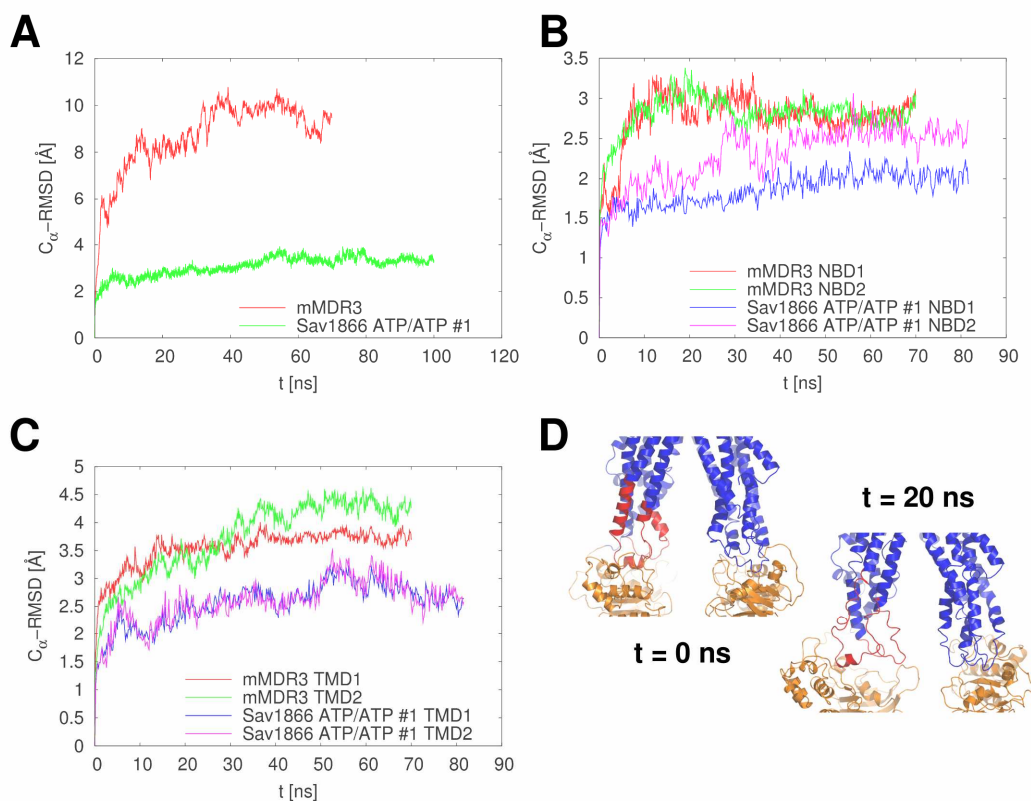


**19. ábra. A humán „alul nyitott” MDR1 szerkezet instabil molekuláris dinamikai szimulációkban.** (A) A  $C_{\alpha}$  atomok kezdeti szerkezettől számolt átlagos négyzetes gyök eltérése (RMSD) erősen megnövekedett a szimuláció során a teljes fehérjére. (B) Az RMSD értékekben csak mérsékelt növekedés látható az NBD-k esetén az apo rendszerben. (C) A TMD-k esetén az apo és holo formát jellemző RMSD értékek közötti különbség nagyobb, mint az NBD-k esetén. A teljes fehérje magas RMSD értékeit egyrészt ez okozza, másrészt az NBD-k merevtestszerű elmozdulása, amelynek az oka a transzmembrán hélixek citoplazmatikus szakaszainak megnövekedett flexibilitása lehet. (D) Az utóbbi mozgékonyágnövekedés forrása az, hogy a szimuláció során ezek a szakaszok nagyon hamar (20 ns) elvesztik másodlagos szerkezetüket. Az instabil szakaszokat pirossal jelöltük a kezdeti és a trajektória végén kapott szerkezeten, 100 ns szimulációs idő után.

Az „alul nyitott” konformáció instabilitásának jellemzésére a szimulációs idő függvényében kiszámoltam a kezdeti szerkezettől való eltávolodást a teljes fehérjére (20.A ábra), és külön a TMD és NBD régiókra is (20.B. és 20.C. ábra). A dupla ATP-kötött Sav1866 rendszer egyik trajektóriáját használtam stabil szimulációra vonatkozó referenciaként. Az egér MDR3 fehérje esetén kiszámolt eltávolodás (RMSD  $\sim 9$  Å) sokkal nagyobb, mint a stabil fehérje esetén (RMSD  $\sim 2,5$  Å; 20.A. ábra), viszont az egyes doménekre kiszámolt RMSD értékek esetén ezek a különbségek kisebbek. A fluktuációk az NBD-k esetében 2,5 és 3 Å között voltak, szemben a Sav1866 rendszerre kapott 2–2,5 Å körüli értékekkel (20.B. ábra), a TMD-k esetében az eltávolodás RMSD értéke  $\sim 4$  Å, míg a Sav1866 rendszerben  $\sim 3$  Å (20.C. ábra). Az RMSD értékek összehasonlítása azt mutatja, hogy a nagyobb szerkezeti változások a két NBD domén merevtestszerű elmozdulása miatt



történnek. Ugyanakkor mind az NBD-k, mind a TMD-k esetén az RMSD értékek az mMDR3 szimulációban általában magasabbak, mint a Sav1866 esetén, és a különbség a TMD-k esetén kifejezettebb. A trajektória fehérjeszerkezeteinek elemzése azt mutatja, hogy a legnagyobb instabilitást a transzmembrán domének citoplazmatikus hélicei mutatják (19.D. és 20.D. ábra). Ennek a régióknak a megnövekedett mozgékonyasága miatt az NBD-k merevtestszerűen elmozdulnak, amely a teljes fehérje megnövekedett RMSD értékeinek a forrása. A szerkezeti instabilitás jellemzésére kiszámoltam a transzmembrán domének hélixtartalmát a trajektóriából származó szerkezetek alapján. A TMD-kben azoknak a kezdetben helikális oldalláncoknak a hányada, amelyek a szimuláció során dominánsan helikálisak is maradtak, összesen 63,13%, szemben a Sav1866 trajektóriára kapott 90,04%-os értékkel. A humán MDR1 fehérje esetében a helikális állapotban megmaradt aminosavak hányada 64,30% az apo és 91,84% a holo rendszerben. A másodlagos szerkezet elvesztése az érintett régiókban, azaz a citoplazmatikus hélixek széttekeredése, az NBD-k transzmembrán doménekről történő lecsatolódását okozza, jelentős torzulásokat okozva a szerkezetben.

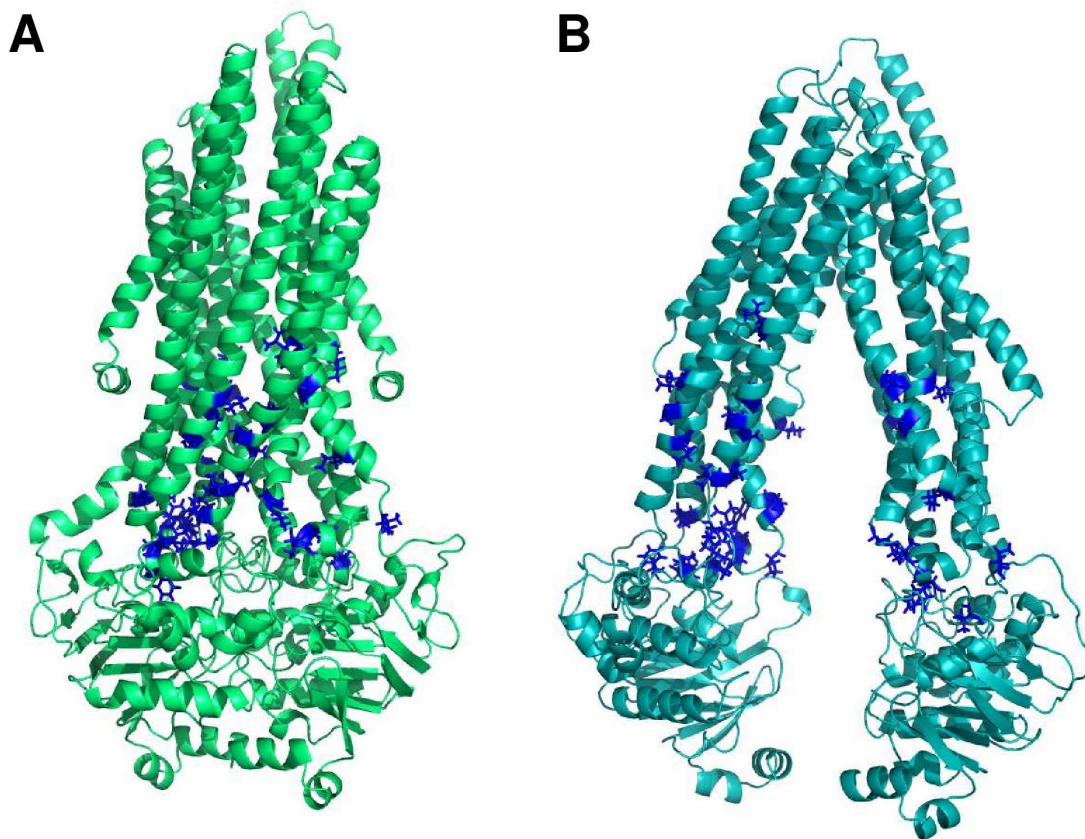


**20. ábra. Instabil szakaszok az mMDR3 „alul nyitott” apo szerkezetben.** (A) A teljes egér MDR3 fehérjére számolt RMSD értékek a hMDR1 fehérjével végzett szimulációkhoz hasonlóan nagyon magasak (8–10 Å), ami a kezdeti szerkezettől való nagymértékű eltávolodást jelez. Stabil rendszerként az egyik reprezentatív Sav1866 szimulációt tüntettem fel. (B) Az alacsony RMSD értékek az NBD-k stabilitására utalnak. (C) Az egér MDR3 szerkezet instabilitását egyrészt a TMD-kben bekövetkezett szerkezeti változások, másrészt az NBD-k merevtestszerű elmozdulása okozza. (D) Az mMDR3 szerkezetben a transzmembrán hélixek citoplazmatikus szakaszai széttekerednek. Az instabil régiókat piros szín jelöli.

Elképzelésem szerint az instabilitás egyik lehetséges oka azon hidrofób oldalláncoknak a megnövekedett száma, amelyek vizes fázisba kerülnek az „alul nyitott” szerkezetben. Ennek az alátámasztására kiszámoltam az oldószer által hozzáférhető felszín (ASA) a két hMDR1 modell intracelluláris hurkaiban található minden egyes aminosavra. A kapott eredményeket a 4. táblázat foglalja össze, ahol az egyszerűség kedvéért csak azok az oldalláncok vannak feltüntetve, amelyeknek a hozzáférhető felszíne az „alul nyitott” konformációban legalább  $20 \text{ \AA}^2$ , és legalább kétszer akkora, mint az „alul zárt” konformációban. Ezeket a többségében hidrofób, és az IL2-ben illetve a *coupling* hélixben található oldalláncokat a szerkezeteken is bejelöltem (21. ábra). Bizonyos oldalláncok esetében az apo konformációban a hozzáférhető felszín jelentősen megnövekedett (a Leu236 esetén ~23-szoros, az Ala238 esetén ~30-szoros emelkedés), más esetben a növekedése kisebb, de az abszolút értéke igen jelentős (kb. 3-szoros növekedés az 163, 804, 904 fenilalaninok esetében, ahol az értékek  $113,4 \text{ \AA}^2$ ,  $41,4 \text{ \AA}^2$ ,  $67,3 \text{ \AA}^2$  az „alul nyitott” konformációban). Ezek alapján az „alul nyitott” szerkezet instabilitásának a forrása valóban lehet a megnövekedett hidrofób felszín.

	aminosav	ASA zárt	ASA nyitott
<b>IL-1</b>	PHE 163	34.0	113.4
	VAL 165	30.8	112.5
	VAL 168	18.4	40.1
	GLY 169	9.8	55.0
	ILE 186	2.4	20.8
<b>IL-2</b>	LEU 236	0.9	22.9
	LEU 244	12.1	79.1
	ALA 246	15.1	31.2
	ALA 248	2.2	59.7
	ALA 252	8.2	67.8
	VAL 253	1.2	64.4
	VAL 257	2.7	51.5
	LEU 258	10.6	31.0
	ALA 260	3.3	39.0
	ILE 261	16.8	47.4
	VAL 264	10.8	32.0
	ILE 265	4.8	28.4
	PHE 267	14.6	34.2
	LEU 281	6.3	38.5
	ALA 284	9.0	60.9
	ILE 287	20.9	120.8
<b>LINKER-1</b>	ALA 354	16.1	33.2
	ALA 356	0.5	49.1
	GLY 360	0.8	43.6
	ILE 368	13.2	62.3
	ILE 375	41.2	144.8
<b>IL-3</b>	LEU 797	13.2	45.2
	TRP 803	28.6	64.7
	PHE 804	11.1	41.4
	PRO 807	41.8	118.7
	GLY 812	0.1	53.4
	ALA 813	1.1	45.2
<b>IL-4</b>	LEU 879	17.9	62.4
	ALA 883	10.3	42.4
	PHE 904	22.9	67.3
	VAL 908	13.7	39.2
<b>LINKER-2</b>	ILE 1003	6.6	104.2

**4. táblázat. Intracelluláris hurkokban található aminosavak hozzáférhető felszíne (ASA).** A táblázatban azok az apoláros vagy hidrofób aminosavak szerepelnek, amelyeknek a hozzáférhető felszíne az „alul nyitott” konformációban legalább 20 Å<sup>2</sup>, és legalább kétszer akkora, mint az „alul zárt” konformáció esetében. A bemutatott értékek mértékegysége Å<sup>2</sup>. IL: intracelluláris hurok, LINKER: a TMD és NBD közötti összekötő régiók.



**21. ábra.** Az intracelluláris hurok apoláros és hidrofób aminosavjainak oldószer által hozzáférhető felülete (ASA) megnövekedett az „alul nyitott” konformációban. A 4. táblázatban felsorolt, megnövekedett ASA értékkel rendelkező oldalláncok kék színű pálcika modellel vannak feltüntetve az „alul zárt” (A) és az „alul nyitott” (B) hMDR1 modellen. A legtöbb megnövekedett hozzáférhetőséggel rendelkező oldallánc az IL2 hurokban és az azon belüli *coupling* hélixben található. Molekuláris dinamikai szimulációkban ez a régió mutatja a legnagyobb arányú széttékeredést.

## 5.2. *Diszkusszió*

Munkám során ABC fehérjék különböző, lipid kettősrétegbe ágyazott szerkezetével végeztem molekuláris dinamikai szimulációkat. Ezek a szerkezetek azokat a transzportfolyamat során felvett különböző konformációkat reprezentálják, amelyek belső dinamikai tulajdonságainak részletes ismerete igen fontos. A kötött nukleotid hidrolízisének hatását a teljes fehérje dinamikai korrelációs hálózatára a Sav1866 **bakteriális** fehérje „alul zárt” holo szerkezetében vizsgáltam. Az egyes oldalláncok közötti kölcsönös információ kiszámításával azonosítottam a korrelált mozgást végző aminosavakat az ATP/ATP és az ATP/ADP rendszerben is. A két rendszerre kapott eredmények összehasonlításával megmutattam, hogy az ATP ADP-re való cseréje közvetlenül megváltoztatja a fehérjén belüli korrelációs hálózat mintázatát. A nukleotidcsere a hidrolízis utáni pillanatot modellezi, az elhasadt  $\gamma$ -foszfát kötés energiájának disszipálódását nem vettem figyelembe. A dinamikai

korrelációs hálózatban csomópontként szerepelnek a Q-hurok és az X-hurok egyes aminosavjainak oldalláncai. Ezek a hurkok a korábbi vizsgálatok alapján is fontos szerepet játszanak az NBD és a TMD közötti allosztérikus kommunikációban. Érdekes módon az ATP/ATP rendszerben megjelenő korrelációk nagy része az ATP/ADP rendszerből hiányzik, ami arra utal, hogy a hidrolízis még a konformáció jelentősebb megváltozása nélkül is képes átalakítani az oldalláncok dinamikai korrelációs hálózatát. A dinamikában bekövetkezett változások vezethetnek az NBD–NBD interfész és a kulcsfontosságú NBD–TMD kölcsönhatások destabilizálódásához, amely a fehérjét a transzportciklus következő lépése felé vezetheti.

Elsőként mutattam meg, hogy dinamikai csatoltság létezik a transzporter TM3 és TM6 hélicei és a nukleotidkötött NBD-k között. Az eredmények alapján azt feltételezem, hogy a TM3-ban is található a funkció szempontjából fontos aminosavak. Érdekes megjegyezni, hogy a citoplazmatikus régióval korrelált mozgást végző oldalláncok találhatóak a TM1 és TM2 közötti extracelluláris hurokban, amely a humán MDR1 fehérjében hosszabb, mint a Sav1866 fehérjében. Ez a megfigyelés is utalhat arra, hogy ennek az extracelluláris huroknak szerepe lehet a transzportfolyamat során bekövetkező aszimmetrikus konformációs változásokban.

A Becker és munkatársai által közölt tanulmányban az ATP-kötött és a nukleotid nélküli Sav1866 szerkezetek vizsgálatakor azt figyelték meg, hogy nukleotid hiányában a transzporter szerkezete az extracelluláris oldalon spontán záródásnak indul [132]. A nukleotid nélküli szerkezetben megfigyelt kollektív mozgások a TM3/TM4 és a TM6 citoplazmatikus szakaszainak összehangolt mozgását mutatták. A szerkezetben lényegesen több megmaradó kontaktust találtak a *coupling* hélixek és az X-hurok illetve a Q-hurok között, mint az ATP-kötött szerkezetben. Érdekes módon az általam vizsgált ATP/ADP rendszer meglepően sok dinamikai hasonlóságot mutat a Becker-féle nukleotidmentes rendszerrel, mint például a Q-hurok és a *coupling* hélixek közötti csatoltság. Emellett a kezdeti szerkezettől legnagyobb eltávolodást mutató ATP/ADP rendszer (#2, 13. ábra) a Becker és munkatársai által is megfigyelt extracelluláris záródást mutatja a transzmembrán doméneknél. A két rendszer közötti hasonlóság alapján elképzelhető, hogy a transzporterben már egyetlen ATP hidrolízise is elindíthat egy olyan folyamatot, amely a transzporter záródását okozza az extracelluláris oldalon. Ez a spontán extracelluláris záródás hozzájárulhat az ABC transzporterekben megfigyelt magas alap ATPáz aktivitáshoz azzal, hogy a nukleotidkötő doménekhez csatolva a *coupling* hélixeken keresztül elősegíti az NBD-k disszociációját. Mivel nincsenek teljesen megbízható adatok arra vonatkozóan, hogy hol és

hogyan történik a transzportálandó szubsztrátok kötése, ezért az is elképzelhető, hogy az „alul zárt” ATP-kötött állapotban az extracelluláris régiók záródása a szubsztrát bekötődését imitálja, és ezzel a hidrolízist elindítva járul hozzá a megfigyelt magas alap ATPáz aktivitáshoz.

A munkám és a Becker és munkatársai által megfigyelt eredmények alapján azt feltételezem, hogy az ATP-kötött „alul zárt” konformációt az X-hurok és a tetrahélix-köteg közötti kapcsoltság stabilizálja. Egyetlen ATP molekula hidrolízise esetén ezeknek a régióknak és az NBD-nek a dinamikus csatoltsági hálózata nagyrészt megszűnik, és megjelenik az NBD-k és a *coupling* hélixek közötti csatoltság. A transzporter belső mozgásainak efféle megváltozása fontos esemény, egy jel az NBD-ktől a TMD-k felé, amely az extracelluláris régiók spontán záródásával együtt elősegítheti a rendszer továbblépését a transzportciklusban. A számítógépes szimulációk által elérhető korlátozott időskála miatt azonban sajnos a konformációs átmenet teljes útvonala számításos módszerekkel, atomi részletességgel nem követhető, de rá lehet mutatni a folyamatban várhatóan résztvevő oldalláncokra.

Az izolált NBD dimerekkel ATP hiányában végzett szimulációk egy részében a domének mutattak szétnyílást, más részében nem [65, 74-78]. A Sav1866 ATP/ADP rendszerben nem figyeltem meg az NBD-k spontán disszociációját (16. ábra), ezért azt feltételezem, hogy a TMD-k jelenléte megtartja az NBD-k viszonylag szoros asszociációját. Az NBD-k és a TMD-k közötti erős csatoltság szempontjából az apo formára előnyösebb modellt adna az „alul zárt” apo konformáció, és ebből arra is következtethetünk, milyen mértékben nyílik fel a citoplazmatikus oldal a transzportciklus során. A kérdéshez kapcsolódó jelentős kísérleti munkát végzett a humán MDR1 fehérjén Loo és munkatársai [64], akik a transzmembrán hélixek citoplazmatikus régióinak keresztkötésével megmutatták, hogy a közöttük levő távolság rögzítése nem akadályozza az ATPáz aktivitást. Ezek alapján elképzelhető, hogy az „alul nyitott”, az NBD-k nagymértékű eltávolodását mutató szerkezet előfordulása nem szükséges feltétele a transzport végbemenetelésnek. Hasonló, keresztkötéses kísérleteket végeztek van Veen és munkatársai [133], akik az MsbA fehérje konformációjában ATP hatására bekövetkezett változásokat próbálták megragadni. Az ő eredményeik alapján azonban nem lehet egyértelműen megmondani, hogy az NBD-k eltávolodnak-e egymástól, vagy sem. A réz-fenantrolin által mediált, 208-as pozícióba beültetett ciszteinek közötti keresztkötés hiánya nem feltétlenül nagy konformációs változásokból ered. Az „alul zárt” apo MsbA szerkezetben, ahol az NBD-k egymáshoz viszonylag közel helyezkednek el, a Glu208 oldalláncokat két hélix választja el egymástól, ami ennek a konformációnak az esetén

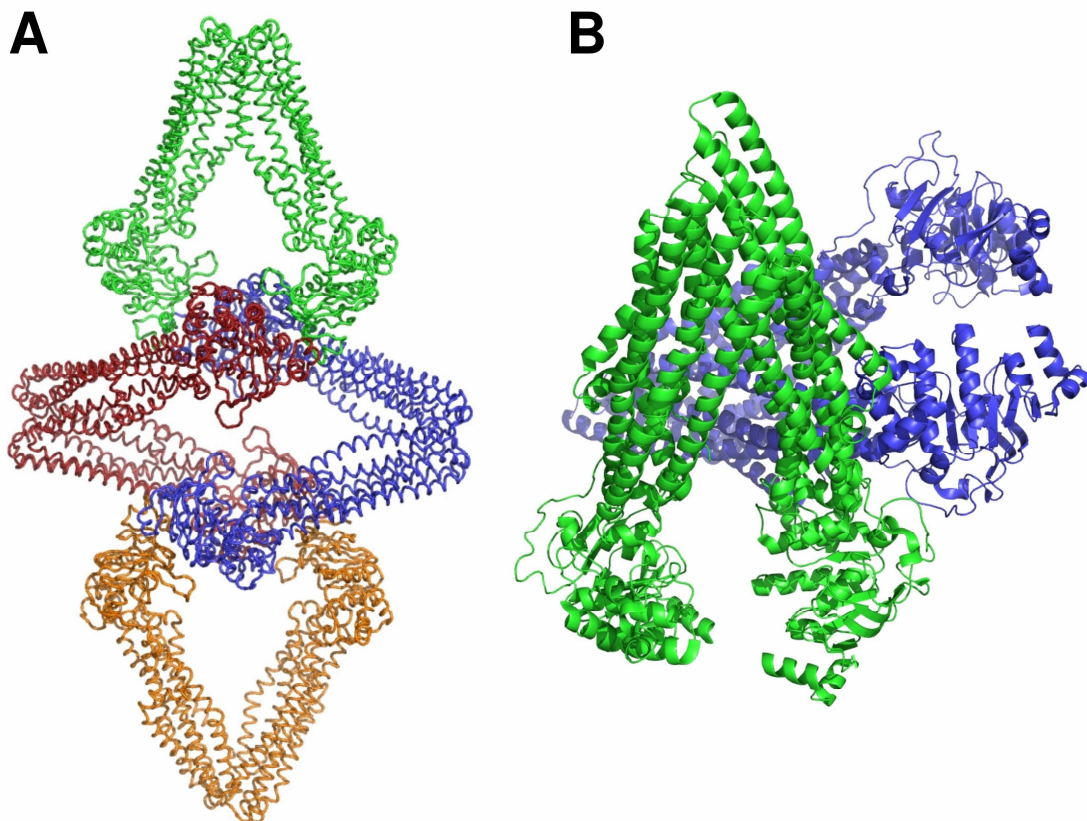
kizárja a két 208-as pozíció közötti keresztkötés lehetőségét. Az EPR kísérletekből származtatott távolságadatok az irányra vonatkozó információ hiányában szintén nem tudják megválaszolni, hogy az „alul zárt” vagy az „alul nyitott” apo konformáció valósul meg a nukleotid nélkül végzett kísérletekben [54].

Az esszenciális dinamikai számítások alapján a holo hMDR1 transzporter képes az „alul zárt” apo konformáció irányába történő elmozdulásra, az NBD-k egymáshoz képesti oldalirányú elmozdulásával. Ezt az elmozdulást csak három szimulációból egy esetben sikerült megfigyelni, a másik két esetben a kollektív mozgásokban nem találtam értelmezhető összefüggést a TMD-k és az NBD-k mozgása között. A megfigyelt esetek rossz arányának oka lehet a konformációs mintavételezés elégtelensége, amely egy ilyen nagyméretű rendszer esetén várható. Ez az eredmény ezért nem tekinthető robusztusnak, és mindenképpen óvatossággal kell kezelni, ennek ellenére felveti annak a lehetőségét, hogy az „alul zárt” apo konformáció egy köztes vagy akár egy végleges állapotot képviseljen a transzportciklusban, az „alul nyitott” szerkezet MD szimulációkban mutatott instabilitása miatt. Érdekes még megjegyezni, hogy a megfigyelt kollektív mozgás mindkét doménre nézve aszimmetrikus, de ezt csak egyetlen szimuláció esetében tudtam megfigyelni, ezért nem tekinthető szignifikáns eredménynek.

Az „alul nyitott” hMDR1 modell és az egér MDR3 röntgendiffrakciós szerkezet is instabilnak bizonyult a szimulációkban. Ivetac és Sansom munkája megmutatta, hogy molekuláris dinamikai szimulációkkal ki lehet szűrni hibás röntgendiffrakciós szerkezeteket, amelyek rövid időtartam alatt is jelentős instabilitást mutatnak szimulációkban [83]. Erre az azóta visszavont MsbA szerkezet szolgáltatta a példát. Az általam vizsgálat „alul nyitott” apo szerkezet ugyan nem mutatott olyan mértékű instabilitást, mint a hibás MsbA szerkezet, az eredmények alapján mégis elképzelhető, hogy a transzmembrán domének citoplazmatikus szakaszai közötti és az NBD-k közötti távolság a valóságban nem olyan nagy, mint a kristályokban megfigyelt „alul nyitott” apo szerkezetben. Az MD szimulációink alapján az NBD-k ilyen mértékű eltávolodása az NBD-k leválását okozza a TMD-kről, ami elrontja a két domén dinamikai csatoltságát, és ezzel lehetetlenné teheti a transzporter működését.

A megfigyelt instabilitásnak többféle oka is lehet. Elképzelhető, hogy a röntgendiffrakciós szerkezet alacsony felbontása miatt (3,8 Å) az oldalláncok elrendezése az egér MDR3 szerkezetben nem optimális, elsősorban az intracelluláris hurkok azon szakaszaiban, amelyek a legnagyobb instabilitást mutatják. Ez arra figyelmeztethet bennünket, hogy nagy körültekintéssel járjunk el olyankor, amikor az oldalláncok irányultságára építünk, pl. droggötési vizsgálatok tervezésekor és értelmezésekor. Másrészt

több, az „alul zárt” holo szerkezetben eltemetett apoláros és hidrofób oldallánc oldószer általi hozzáférhetősége az „alul nyitott” szerkezetben nagymértékű, ami elképzelhető, hogy csak a kristályosítási körülmények (alacsony hőmérséklet, lipid kettősréteg hiánya) mellett számít kedvezőnek. Szintén elképzelhető, hogy az „alul nyitott” konformációt az elemi cellában levő molekulák kölcsönhatásai stabilizálják. Ezt a konformációt eddig az MsbA (PDB ID: 3B5W) bakteriális fehérje és az egér MDR3 (PDB ID: 3G5U) kristályaiban figyelték meg, ahol az elemi cella egynél több teljes transzportert tartalmazott. Bár az azonos transzporterben belüli NBD-k egymástól távol helyezkednek el, kölcsönhatást létesítenek a szomszédos molekulák NBD-jeivel (22. ábra). Ezek az intermolekuláris NBD–NBD kölcsönhatások nem a kanonikus „dimerizációs” interfészen jelentkeznek, és elképzelhető, hogy stabilizálják a kristályban előforduló konformációt, az „alul nyitott” szerkezetet. Ezzel ellentétben azokban a kristályokban, amelyek más konformációban tartalmazzák a transzportert (Sav1866 „alul zárt” holo, MsbA holo és „alul zárt” apo), a különböző fehérjékhez tartozó NBD-k egymástól távol helyezkednek el, nem lépnek kölcsönhatásba egymással.



**22. ábra.** A kristálykontaktusok stabilizálhatják az „alul nyitott” konformációt. Az ismert „alul nyitott” konformációt mutató röntgendiffrakciós szerkezetben (MsbA, mMDR3) található intermolekuláris kontaktzfelszín az NBD-k között. Ezek a kontaktusok nem a fiziológiás „dimer” esetében megfigyeltek, és elképzelhető, hogy egy nem-fiziológiás kristálybeli konformációban stabilizálják a transzportereket. (A) MsbA (PDB ID: 3B5W), (B) mMDR3 (PDB ID: 3G5U).



Az apo konformációról az imént bemutatott kérdések miatt elképzelhetőnek tartom, hogy a fiziológias körülmények között jelen lévő apo forma leírására az „alul zárt” apo szerkezet jobb modellt ad. Az MRP1 fehérje esetén megfigyelt 2D kristálybeli apo szerkezet is egy félig nyitott konformációt mutat, az „alul zárt” konformációhoz hasonlóan [134]. A sejtbeli ATP koncentrációt (2–5 mM) összehasonlítva az ABC exporterek tipikus  $K_M$  értékével (200–500  $\mu\text{M}$ ) az is elképzelhető, hogy az „alul nyitott” apo konformáció egy igen rövid életidejű, átmeneti állapot, amelyet a kristályosítási körülmények mellett sikerült stabilizálni és megfigyelni.

## 6. Összefoglalás

Munkám során kifejlesztettem egy új, konformációs entrópia becslésére használatos módszert, amely Gauss-keverék függvények illesztésén alapul. A módszert öt, kis méretű peptidből álló molekuláris rendszeren teszteltem. A kapott entrópiaértékeket összehasonlítottam egyrészt más, ismert módszerek által adott eredményekkel, másrészt a konformációs sokaság egzakt entrópiaértékével is, amelyet minden rendszerre a teljes konformációs tér feletti integrálással kiszámoltam. Az egzakt entrópiával való átlagos legjobb egyezést az általam kifejlesztett módszer adta. A Gauss-keverék módszer elvileg közvetlenül alkalmazható nagyobb rendszerek, akár fehérjék konformációs entrópiájának számítására, de további munkát igényel a szükséges Gauss-csúcsok számának automatikus megállapítása. Az algoritmus polinomiálisan skálázódik a dimenziószám és a mintaméret függvényében is, ezért nagyobb rendszerek esetén szükséges lehet olyan társmódszerek kidolgozása, amellyel a rendszer effektív dimenziószámát csökkenthetjük. A módszer ezen korlátok ellenére egy koncepcionálisan egyszerű és könnyen átlátható elméleti háttérrel ad a további fejlesztésekhez.

Az ABC fehérjéken végzett munkában különböző transzporterszerkezetek dinamikai jellemzőit vizsgáltam. A nukleotidkötött, „alul zárt” holo konformációban az egymással korrelált mozgást végző oldalláncok feltérképezésével megállapítottam, hogy akár egyetlen ATP hidrolízise is okozhat olyan változást a dinamikailag korrelált aminosavak hálózatában, amely elősegíti a transzporter citoplazmatikus oldalának felnyílását. Az „alul nyitott” szerkezet MD szimulációinkban instabilitását mutatott. Elképzeléseim szerint az instabilitás oka a nagymértékű exponált hidrofób felület az intracelluláris hurkokban, vagy az oldalláncok nem-optimális elrendezése alacsony felbontásból eredő bizonytalanság miatt. Az „alul nyitott” apo szerkezetet a kristályban megfigyelt intermolekuláris NBD–NBD kontaktusok és a kristályosítás körülményei is stabilizálhatják, fiziológias körülmények között azonban elképzelhető, hogy egy rövid életidejű, átmeneti konformációról van szó. Az „alul zárt” holo szerkezet esszenciális dinamikai analízise alapján azt feltételezem, hogy az ATP hidrolízise után a transzporter képes az „alul zárt” apo konformáció irányába elmozdulni. Ezen lehetséges holo–apo útvonal és az „alul nyitott” szerkezet instabilitása alapján azt javaslom, hogy az „alul zárt” apo szerkezet relevánsabb leírását adja a transzporter szerkezetének, mint az „alul nyitott” apo konformáció. Ezek a szimulációk segítenek megérteni az ABC fehérjék működéséhez szükséges szerkezeti és dinamikai jellemzőiket, ezáltal új megközelítést ajánlva

az ABC transzporterek felgombolyodásának, stabilitásának és működésének befolyásolását célzó racionális gyógyszertervezés számára.

## 7. Függelék

### 7.1. A Gibbs-eloszlás származtatása

Termosztált rendszer esetén a rendszer számára elérhető mikroállapotok számának megbecsléséhez tekintsük a következő gondolatmenetet. Írjuk a rendszer teljes energiáját  $E = E_1 + E_2$  formába, ahol  $E_1$  az általunk vizsgált részrendszer,  $E_2$  pedig a termosztát energiája. Annak a valószínűsége, hogy a részrendszer egy adott  $E_1$  energiájú mikroállapotban található, arányos a teljes rendszer azon állapotainak számával, amelyben a részrendszer energiája  $E_1$ . A teljes rendszer azon állapotainak száma, amelyben a részrendszer egy rögzített  $E_1$  energiájú állapotban van, megegyezik a termosztát  $E - E_1$  energiájú állapotainak számával, amit jelöljön  $\Gamma_2(E - E_1)$ . Ez alapján a részrendszer egy  $E_1$  energiájú állapotának valószínűsége

$$w(E_1) = c \cdot \Gamma_2(E - E_1) \quad (73)$$

ahol  $c$  egy arányossági tényező. A részrendszer energiaeloszlására így felírható az

$$1 = \sum_{E_1} \Gamma_1(E_1) w(E_1) = c \sum_{E_1} \Gamma_1(E_1) \Gamma_2(E - E_1) \quad (74)$$

normálási feltétel. Az összegzés mögött levő kifejezésről azonban megmutatható, hogy egy átlagos,  $E_1 = \langle E_1 \rangle$  érték körül éles csúcsot mutat [135]. Ezek alapján az összegzés helyettesíthető a

$$\sum_{E_1} \Gamma_1(E_1) w(E_1) = \Delta\Gamma \cdot w(\langle E_1 \rangle) = 1 \quad (75)$$

kifejezéssel, ahol  $w(\langle E_1 \rangle)$  az átlagos energiaérték előfordulási valószínűsége,  $\Delta\Gamma$  pedig a részrendszer által bejárt állapotok átlagos száma, amit az (75) összefüggés definiál. Ezek alapján a részrendszer entrópiája felírható

$$S_1 = k_B \ln \Delta\Gamma = -k_B \ln w(\langle E_1 \rangle) \quad (76)$$

alakban. A (76) képlet érdekessége, hogy azt mutatja, hogy egy rendszer entrópiája kiszámítható pusztán egyetlen, az átlagos energiának megfelelő állapot valószínűségéből. A  $w(E_1)$  eloszlást azonban a legtöbbször nem ismerjük egzaktul, ugyanis az eloszlás normálási faktorának kiszámításához az összes energiaállapot valószínűségét ismernünk kell. A  $w(E_1)$  eloszlás megbecsléséhez azt a feltételezést használhatjuk, hogy a termosztát sokkal nagyobb, mint a vizsgált részrendszer ( $E_1 \ll E_2$ ), és így energiájának fluktuációi gyakorlatilag elhanyagolhatók. A termosztát ezzel gyakorlatilag mikrokanonikus rendszernek tekinthető, vagyis entrópiáját az adott energiájú állapotainak száma határozza meg. Mivel a részrendszer

energiája nagyon kicsi a rendszer teljes energiájához képest ( $E_1 \ll E$ ), a termostát entrópiája sorfejtéssel helyettesíthető:

$$k_B \ln \Gamma_2(E - E_1) = S_2(E - E_1) \approx S_2(E) - \left. \frac{\partial S_2(E')}{\partial E'} \right|_{E'=E} E_1 = S_2(E) - \frac{E_1}{k_B T_2} \quad (77)$$

ahol  $T_2 = \frac{1}{k_B} \left( \frac{\partial S_2(E)}{\partial E} \right)^{-1}$  a termostát hőmérséklete. A (77) egyenlet jobb oldalának első tagja nem függ  $E_1$ -től, ezért a  $w(E_1)$  eloszlásfüggvényhez csak konstans szorzófaktorral járul hozzá, ami így felírható

$$w(E_1) = Z^{-1} e^{-\frac{E_1}{k_B T}} = Z^{-1} e^{-\beta E_1} \quad (78)$$

ahol  $\beta = 1/(k_B T)$  az ún. hőmérsékleti faktor,  $Z$  pedig az eloszlás normálási faktora, amit állapotösszegnek vagy más néven partíciós függvénynek neveznek.

## 7.2. A konfigurációs és kinetikus entrópiájárulékok szétválasztása

Klasszikus rendszerek esetén a (16) összefüggést gyakran érdemes további tagokra bontani. Descartes-koordinátarendszer használata esetén a mozgási energia az impulzusok négyzetes függvénye lesz, és ha a rendszerben nincsenek disszipatív erők, akkor az energia és az entrópia kinetikus és konfigurációs tagokra bontható. Ekkor a rendszer energiafüggvénye  $E(p, q) = E_{kin}(p) + E_{pot}(q)$  alakú, és a rendszer entrópiája (16) alapján

$$\begin{aligned} S &= k_B (\ln(h^{-3N} Z) + \beta \langle E \rangle) = \\ &= k_B (\ln(h^{-3N} Z_{kin} Z_{conf}) + \beta \langle E_{kin} \rangle + \beta \langle E_{pot} \rangle) = \\ &= k_B (\ln(h^{-3N} Z_{kin}) + \beta \langle E_{kin} \rangle) + k_B (\ln Z_{conf} + \beta \langle E_{pot} \rangle) = S_{kin} + S_{conf} \end{aligned} \quad (79)$$

Az entrópia kifejezése tehát két tagra bontható, ahol  $S_{kin}$  a kinetikus járulék

$$S_{kin} = k_B (\ln(h^{-3N} Z_{kin}) + \beta \langle E_{kin} \rangle) \quad (80)$$

amelyben szereplő tagok

$$Z_{kin} = \int e^{-\beta E_{kin}(p)} dp \quad (81)$$

$$\langle E_{kin} \rangle = Z_{kin}^{-1} \int E_{kin}(p) e^{-\beta E_{kin}(p)} dp \quad (82)$$

a kinetikus állapotösszeg ( $Z_{kin}$ ) és az átlagos kinetikus energia ( $\langle E_{kin} \rangle$ ). A (79) képlet jobb oldalán szereplő másik tag az ún. konfigurációs entrópia ( $S_{conf}$ ), amely

$$S_{conf} = k_B (\ln(Z_{conf}) + \beta \langle E_{pot} \rangle) \quad (83)$$

formában a konfigurációs állapotösszegeből ( $Z_{conf}$ ) és az átlagos potenciális energiából ( $\langle E_{pot} \rangle$ ) származtatható, amelyeket a

$$Z_{conf} = \int e^{-\beta E_{pot}(q)} dq \quad (84)$$

$$\langle E_{pot} \rangle = Z_{conf}^{-1} \int E_{pot}(q) e^{-\beta E_{pot}(q)} dq \quad (85)$$

kifejezések definiálnak.

A kinetikus állapotok  $Z_{kin}$  integrálja ebben az esetben analitikusan is kiszámolható,

$$Z_{kin} = \int e^{-\beta E_{kin}(p)} dp = \prod_k \int e^{-\beta \frac{p_k^2}{2m_k}} dp_k = \prod_k \left( \frac{2\pi m_k}{\beta} \right)^{3/2} \quad (86)$$

ahol  $m_k$  és  $p_k$  a  $k$ . atom tömege és impulzusa. A rendszer átlagos kinetikus energiája szintén kiszámítható analitikusan,

$$\begin{aligned} \langle E_{kin} \rangle &= Z_{kin}^{-1} \int \left( \sum_k \frac{p_k^2}{2m_k} \right) e^{-\beta \sum_k \frac{p_k^2}{2m_k}} dp = \\ &= \sum_k \left( \frac{2\pi m_k}{\beta} \right)^{-3/2} \int \frac{p_k^2}{2m_k} e^{-\beta \frac{p_k^2}{2m_k}} dp_k = N \frac{3}{2} \beta^{-1} \end{aligned} \quad (87)$$

és ez által a kinetikus entrópiajárulék

$$\begin{aligned} S_{kin} &= k_B (\ln(h^{-3N} Z_{kin}) + \beta \langle E_{kin} \rangle) = k_B \ln \prod_k \left( \frac{2\pi m_k}{\beta h^2} \right)^{3/2} + k_B \frac{3N}{2} = \\ &= k_B \frac{3}{2} \sum_k \left( 1 + \ln \frac{2\pi m_k}{\beta h^2} \right) \end{aligned} \quad (88)$$

formában megkapható. Látható, hogy az entrópia kinetikus járuléka a koordinátáktól teljesen független, és pusztán az atomi tömegek segítségével kiszámítható. Ezért a legtöbb módszer az entrópia kiszámításakor csak a konfigurációs entrópiára fókuszál.

## 8. Publikációk jegyzéke

1. Gyimesi G., Závodszy P., Hegedűs T., and Szilágyi A., „Calculation of configurational entropy from molecular dynamics trajectories using Gaussian mixtures.”, beküldés alatt
2. Gyimesi G., Ramachandran S., Kota P., Dokholyan N.V., Sarkadi B., Hegedűs T., „ATP hydrolysis at one of the two sites in ABC transporters initiates transport related conformational transitions.” *Biochimica et Biophysica Acta – Biomembranes*, **2011**. 1808(12): p. 2954-2964.
3. Papp A., Szommer T., Barna L., Gyimesi G., Ferdinándy P., Spadoni C., Darvas F., Fujita T., Ürge L., Dormán G., „Enhanced hit-to-lead process using bioanalogous lead evaluation and chemogenomics: application in designing selective matrix metalloprotease inhibitors.” *Expert Opinion on Drug Delivery*, **2007**. 2(5): p. 707-723.
4. Flachner B., Varga A., Szabó J., Barna L., Hajdú I., Gyimesi G., Závodszy P., Vas M., „Substrate-assisted movement of the catalytic Lys 215 during domain closure: site-directed mutagenesis studies of human 3-phosphoglycerate kinase.” *Biochemistry*, **2005**. 44(51): p. 16853-65.

## 9. Irodalomjegyzék

- [1] Karplus M., McCammon J.A., "Molecular dynamics simulations of biomolecules". *Nat Struct Biol*, **2002**. 9(9): p. 646-52.
- [2] Baldwin R.L., "Temperature dependence of the hydrophobic interaction in protein folding". *Proc Natl Acad Sci U S A*, **1986**. 83(21): p. 8069-72.
- [3] Grunberg R., Nilges M., Leckner J., "Flexibility and conformational entropy in protein-protein binding". *Structure*, **2006**. 14(4): p. 683-93.
- [4] D'Aquino J.A., Freire E., Amzel L.M., "Binding of small organic molecules to macromolecular targets: evaluation of conformational entropy changes". *Proteins*, **2000**. Suppl 4: p. 93-107.
- [5] Chang C.E., Chen W., Gilson M.K., "Ligand configurational entropy and protein binding". *Proc Natl Acad Sci U S A*, **2007**. 104(5): p. 1534-9.
- [6] Amzel L.M., Siebert X., Armstrong A., Pabon G., "Thermodynamic calculations in biological systems". *Biophys Chem*, **2005**. 117(3): p. 239-54.
- [7] Tusnady G.E., Dosztanyi Z., Simon I., "PDB\_TM: selection and membrane localization of transmembrane proteins in the protein data bank". *Nucleic Acids Res*, **2005**. 33(Database issue): p. D275-8.
- [8] Marreddy R.K.R., Geertsma E.R., Poolman B., "Recombinant Membrane Protein Production: Past, Present and Future", in *Supramolecular Structure and Function 10*. **2011**. p. 41-74.
- [9] Rousseau F., Schymkowitz J., "A systems biology perspective on protein structural dynamics and signal transduction". *Curr Opin Struct Biol*, **2005**. 15(1): p. 23-30.
- [10] Landau L.D., Lifshic E.M., "Elméleti fizika V. (Statisztikus fizika I.)". **1981**, Budapest: Tankönyvkiadó.
- [11] Tribus M., "Thermostatcs and Thermodynamics". **1961**, New York: Van Nostrand.
- [12] Herschbach D.R., Johnston H.S., Rapp D., "Molecular Partition Functions in Terms of Local Properties". *Journal of Chemical Physics*, **1959**. 31(6): p. 1652-1661.
- [13] Beveridge D.L., DiCapua F.M., "Free Energy via Molecular Simulation: Applications to Chemical and Biomolecular Systems". *Annu Rev Biophys Biophys Chem*, **1989**. 18: p. 431-92.
- [14] Zwanzig R.W., "High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases". *Journal of Chemical Physics*, **1954**. 22(8): p. 1420-26.
- [15] Kumar S., Bouzida D., Swendsen R.H., Kollman P.A., Rosenberg J.M., "The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method". *Journal of Computational Chemistry*, **1992**. 13(8): p. 1011-1021.



- [16] Meirovitch H., Cheluvaraja S., White R.P., "Methods for calculating the entropy and free energy and their application to problems involving protein flexibility and ligand binding". *Curr Protein Pept Sci*, **2009**. 10(3): p. 229-43.
- [17] Kollman P.A., Massova I., Reyes C., Kuhn B., Huo S., Chong L., Lee M., Lee T., Duan Y., Wang W., Donini O., Cieplak P., Srinivasan J., Case D.A., Cheatham T.E., 3rd, "Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models". *Acc Chem Res*, **2000**. 33(12): p. 889-97.
- [18] Karplus M., Kushick J.N., "Method for estimating the configurational entropy of macromolecules". *Macromolecules*, **1981**. 14(2): p. 325-332.
- [19] Schlitter J., "Estimation of absolute and relative entropies of macromolecules using the covariance matrix". *Chem Phys Lett*, **1993**. 215(6): p. 617-621.
- [20] Andricioaei I., Karplus M., "On the calculation of entropy from covariance matrices of the atomic fluctuations". *Journal of Chemical Physics*, **2001**. 115(14): p. 6289-92.
- [21] Baron R., van Gunsteren W.F., Hünenberger P.H., "Estimating the configurational entropy from molecular dynamics simulations: anharmonicity and correlation corrections to the quasi-harmonic approximation". *Trends Phys Chem*, **2006**. 11: p. 87-122.
- [22] Matsuda H., "Physical nature of higher-order mutual information: Intrinsic correlations and frustration". *Phys Rev E*, **2000**. 62(3): p. 3096-102.
- [23] Attard P., "Information content of signals using correlation function expansions of the entropy". *Phys Rev E*, **1997**. 56(4): p. 4052-67.
- [24] Reiss H., "Superposition Approximations from a Variation Principle". *Journal of Statistical Physics*, **1972**. 6(1): p. 39-47.
- [25] Numata J., Wan M., Knapp E.W., "Conformational entropy of biomolecules: beyond the quasi-harmonic approximation". *Genome Inform*, **2007**. 18: p. 192-205.
- [26] Loftsgaarden D.O., Quesenberry C.P., "A Nonparametric Estimate of a Multivariate Density Function". *Annals of Mathematical Statistics*, **1965**. 36(3): p. 1049-51.
- [27] Hnizdo V., Fedorowicz A., Singh H., Demchuk E., "Statistical thermodynamics of internal rotation in a hindering potential of mean force obtained from computer simulations". *J Comput Chem*, **2003**. 24(10): p. 1172-83.
- [28] Singh H., Misra N., Hnizdo V., Fedorowicz A., Demchuk E., "Nearest neighbor estimates of entropy". *Am J Math Manag Sci*, **2003**. 23: p. 301.
- [29] Hnizdo V., Darian E., Fedorowicz A., Demchuk E., Li S., Singh H., "Nearest-neighbor nonparametric method for estimating the configurational entropy of complex molecules". *J Comput Chem*, **2007**. 28(3): p. 655-68.
- [30] Hensen U., Grubmüller H., Lange O.F., "Adaptive anisotropic kernels for nonparametric estimation of absolute configurational entropies in high-dimensional

- configuration spaces". *Phys Rev E Stat Nonlin Soft Matter Phys*, **2009**. 80(1 Pt 1): p. 011913.
- [31] White R.P., Meirovitch H., "Lower and upper bounds for the absolute free energy by the hypothetical scanning Monte Carlo method: application to liquid argon and water". *J Chem Phys*, **2004**. 121(22): p. 10889-904.
- [32] Chelvaraja S., Meirovitch H., "Simulation method for calculating the entropy and free energy of peptides and proteins". *Proc Natl Acad Sci U S A*, **2004**. 101(25): p. 9241-6.
- [33] Chelvaraja S., Meirovitch H., "Calculation of the entropy and free energy of peptides by molecular dynamics simulations using the hypothetical scanning molecular dynamics method". *J Chem Phys*, **2006**. 125(2): p. 024905.
- [34] Wang J., Brüschweiler R., "2D Entropy of Discrete Molecular Ensembles". *J Chem Theory Comput*, **2006**. 2(1): p. 18-24.
- [35] Li D.-W., Khanlarzadeh M., Wang J., Huo S., Brüschweiler R., "Evaluation of Configurational Entropy Methods from Peptide Folding–Unfolding Simulation". *J Phys Chem B*, **2007**. 111(49): p. 13807-13813.
- [36] Eckford P.D., Sharom F.J., "ABC efflux pump-based resistance to chemotherapy drugs". *Chem Rev*, **2009**. 109(7): p. 2989-3011.
- [37] Dean M., Rzhetsky A., Allikmets R., "The human ATP-binding cassette (ABC) transporter superfamily". *Genome Res*, **2001**. 11(7): p. 1156-66.
- [38] Riordan J.R., Rommens J.M., Kerem B., Alon N., Rozmahel R., Grzelczak Z., Zielenski J., Lok S., Plavsic N., Chou J.L., et al., "Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA". *Science*, **1989**. 245(4922): p. 1066-73.
- [39] Senior A.E., Gadsby D.C., "ATP hydrolysis cycles and mechanism in P-glycoprotein and CFTR". *Semin Cancer Biol*, **1997**. 8(3): p. 143-50.
- [40] Wada M., Toh S., Taniguchi K., Nakamura T., Uchiumi T., Kohno K., Yoshida I., Kimura A., Sakisaka S., Adachi Y., Kuwano M., "Mutations in the canalicular multispecific organic anion transporter (cMOAT) gene, a novel ABC transporter, in patients with hyperbilirubinemia II/Dubin-Johnson syndrome". *Hum Mol Genet*, **1998**. 7(2): p. 203-7.
- [41] Reis A.F., Ye W.Z., Dubois-Laforgue D., Bellanne-Chantelot C., Timsit J., Velho G., "Association of a variant in exon 31 of the sulfonylurea receptor 1 (SUR1) gene with type 2 diabetes mellitus in French Caucasians". *Hum Genet*, **2000**. 107(2): p. 138-44.
- [42] Mosser J., Douar A.M., Sarde C.O., Kioschis P., Feil R., Moser H., Poustka A.M., Mandel J.L., Aubourg P., "Putative X-linked adrenoleukodystrophy gene shares unexpected homology with ABC transporters". *Nature*, **1993**. 361(6414): p. 726-30.
- [43] Gottesman M.M., Ling V., "The molecular basis of multidrug resistance in cancer: the early years of P-glycoprotein research". *FEBS Lett*, **2006**. 580(4): p. 998-1009.

- [44] Szakacs G., Varadi A., Ozvegy-Laczka C., Sarkadi B., "The role of ABC transporters in drug absorption, distribution, metabolism, excretion and toxicity (ADME-Tox)". *Drug Discov Today*, **2008**. 13(9-10): p. 379-93.
- [45] Hyde S.C., Emsley P., Hartshorn M.J., Mimmack M.M., Gileadi U., Pearce S.R., Gallagher M.P., Gill D.R., Hubbard R.E., Higgins C.F., "Structural model of ATP-binding proteins associated with cystic fibrosis, multidrug resistance and bacterial transport". *Nature*, **1990**. 346(6282): p. 362-5.
- [46] Higgins C.F., "ABC transporters: from microorganisms to man". *Annu Rev Cell Biol*, **1992**. 8: p. 67-113.
- [47] Ye J., Osborne A.R., Groll M., Rapoport T.A., "RecA-like motor ATPases--lessons from structures". *Biochim Biophys Acta*, **2004**. 1659(1): p. 1-18.
- [48] Walker J.E., Saraste M., Runswick M.J., Gay N.J., "Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold". *Embo J*, **1982**. 1(8): p. 945-51.
- [49] Schneider E., Hunke S., "ATP-binding-cassette (ABC) transport systems: functional and structural aspects of the ATP-hydrolyzing subunits/domains". *FEMS Microbiol Rev*, **1998**. 22(1): p. 1-20.
- [50] Smith P.C., Karpowich N., Millen L., Moody J.E., Rosen J., Thomas P.J., Hunt J.F., "ATP binding to the motor domain from an ABC transporter drives formation of a nucleotide sandwich dimer". *Mol Cell*, **2002**. 10(1): p. 139-49.
- [51] Chen J., Lu G., Lin J., Davidson A.L., Quioco F.A., "A tweezers-like motion of the ATP-binding cassette dimer in an ABC transport cycle". *Mol Cell*, **2003**. 12(3): p. 651-61.
- [52] Zaitseva J., Jenewein S., Jumpertz T., Holland I.B., Schmitt L., "H662 is the linchpin of ATP hydrolysis in the nucleotide-binding domain of the ABC transporter HlyB". *Embo J*, **2005**. 24(11): p. 1901-10.
- [53] Dawson R.J., Locher K.P., "Structure of a bacterial multidrug ABC transporter". *Nature*, **2006**. 443(7108): p. 180-5.
- [54] Kerr I.D., Jones P.M., George A.M., "Multidrug efflux pumps: the structures of prokaryotic ATP-binding cassette transporter efflux pumps and implications for our understanding of eukaryotic P-glycoproteins and homologues". *Febs J*, **2010**. 277(3): p. 550-63.
- [55] Ward A., Reyes C.L., Yu J., Roth C.B., Chang G., "Flexibility in the ABC transporter MsbA: Alternating access with a twist". *Proc Natl Acad Sci U S A*, **2007**.
- [56] Aller S.G., Yu J., Ward A., Weng Y., Chittaboina S., Zhuo R., Harrell P.M., Trinh Y.T., Zhang Q., Urbatsch I.L., Chang G., "Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding". *Science*, **2009**. 323(5922): p. 1718-22.
- [57] Bakos E., Evers R., Szakacs G., Tusnady G.E., Welker E., Szabo K., de Haas M., van Deemter L., Borst P., Varadi A., Sarkadi B., "Functional multidrug resistance protein

- (MRP1) lacking the N-terminal transmembrane domain". *J Biol Chem*, **1998**. 273(48): p. 32167-75.
- [58] Loo T.W., Bartlett M.C., Clarke D.M., "Processing mutations disrupt interactions between the nucleotide binding and transmembrane domains of P-glycoprotein and the cystic fibrosis transmembrane conductance regulator (CFTR)". *J Biol Chem*, **2008**. 283(42): p. 28190-7.
- [59] Grote M., Polyhach Y., Jeschke G., Steinhoff H.J., Schneider E., Bordignon E., "Transmembrane signaling in the maltose ABC transporter MalFGK2-E: periplasmic MalF-P2 loop communicates substrate availability to the ATP-bound MalK dimer". *J Biol Chem*, **2009**. 284(26): p. 17521-6.
- [60] Goetz B.A., Perozo E., Locher K.P., "Distinct gate conformations of the ABC transporter BtuCD revealed by electron spin resonance spectroscopy and chemical cross-linking". *FEBS Lett*, **2009**. 583(2): p. 266-70.
- [61] Do Cao M.A., Crouzy S., Kim M., Becchi M., Cafiso D.S., Di Pietro A., Jault J.M., "Probing the conformation of the resting state of a bacterial multidrug ABC transporter, BmrA, by a site-directed spin labeling approach". *Protein Sci*, **2009**. 18(7): p. 1507-20.
- [62] Westfahl K.M., Merten J.A., Buchaklian A.H., Klug C.S., "Functionally important ATP binding and hydrolysis sites in Escherichia coli MsbA". *Biochemistry*, **2008**. 47(52): p. 13878-86.
- [63] Zou P., Bortolus M., McHaourab H.S., "Conformational cycle of the ABC transporter MsbA in liposomes: detailed analysis using double electron-electron resonance spectroscopy". *J Mol Biol*, **2009**. 393(3): p. 586-97.
- [64] Loo T.W., Bartlett M.C., Clarke D.M., "Human P-glycoprotein is active when the two halves are clamped together in the closed conformation". *Biochem Biophys Res Commun*, **2010**.
- [65] Jones P.M., O'Mara M.L., George A.M., "ABC transporters: a riddle wrapped in a mystery inside an enigma". *Trends Biochem Sci*, **2009**. 34(10): p. 520-31.
- [66] Senior A.E., al-Shawi M.K., Urbatsch I.L., "The catalytic cycle of P-glycoprotein". *FEBS Lett*, **1995**. 377(3): p. 285-9.
- [67] Jones P.M., George A.M., "Opening of the ADP-bound active site in the ABC transporter ATPase dimer: evidence for a constant contact, alternating sites model for the catalytic cycle". *Proteins*, **2009**. 75(2): p. 387-96.
- [68] Janas E., Hofacker M., Chen M., Gompf S., van der Does C., Tampe R., "The ATP hydrolysis cycle of the nucleotide-binding domain of the mitochondrial ATP-binding cassette transporter Mdl1p". *J Biol Chem*, **2003**. 278(29): p. 26862-9.
- [69] van der Does C., Tampe R., "How do ABC transporters drive transport?" *Biol Chem*, **2004**. 385(10): p. 927-33.

- [70] Higgins C.F., Linton K.J., "The ATP switch model for ABC transporters". *Nat Struct Mol Biol*, **2004**. 11(10): p. 918-26.
- [71] Aittoniemi J., de Wet H., Ashcroft F.M., Sansom M.S., "Asymmetric switching in a homodimeric ABC transporter: a simulation study". *PLoS Comput Biol*, **2010**. 6(4): p. e1000762.
- [72] Oancea G., O'Mara M.L., Bennett W.F., Tieleman D.P., Abele R., Tampe R., "Structural arrangement of the transmission interface in the antigen ABC transport complex TAP". *Proc Natl Acad Sci U S A*, **2009**. 106(14): p. 5551-6.
- [73] Weng J.W., Fan K.N., Wang W.N., "The conformational transition pathway of ATP binding cassette transporter MsbA revealed by atomistic simulations". *J Biol Chem*, **2010**. 285(5): p. 3053-63.
- [74] Jones P.M., George A.M., "Nucleotide-dependent allostery within the ABC transporter ATP-binding cassette: A computational study of the MJ0796 dimer". *J Biol Chem*, **2007**. 282(31): p. 22793-803.
- [75] Jones P.M., George A.M., "Mechanism of ABC transporters: a molecular dynamics simulation of a well characterized nucleotide-binding subunit". *Proc Natl Acad Sci U S A*, **2002**. 99(20): p. 12639-44.
- [76] Newstead S., Fowler P.W., Bilton P., Carpenter E.P., Sadler P.J., Campopiano D.J., Sansom M.S., Iwata S., "Insights into how nucleotide-binding domains power ABC transport". *Structure*, **2009**. 17(9): p. 1213-22.
- [77] Wen P.C., Tajkhorshid E., "Dimer opening of the nucleotide binding domains of ABC transporters after ATP hydrolysis". *Biophys J*, **2008**. 95(11): p. 5100-10.
- [78] Campbell J.D., Sansom M.S., "Nucleotide binding to the homodimeric MJ0796 protein: a computational study of a prokaryotic ABC transporter NBD dimer". *FEBS Lett*, **2005**. 579(19): p. 4193-9.
- [79] He L., Aleksandrov A.A., Serohijos A.W., Hegedus T., Aleksandrov L.A., Cui L., Dokholyan N.V., Riordan J.R., "Multiple membrane-cytoplasmic domain contacts in cfr mediate regulation of channel gating". *J Biol Chem*, **2008**.
- [80] Oloo E.O., Tieleman D.P., "Conformational transitions induced by the binding of MgATP to the vitamin B12 ATP-binding cassette (ABC) transporter BtuCD". *J Biol Chem*, **2004**. 279(43): p. 45013-9.
- [81] Sonne J., Kandt C., Peters G.H., Y F., Jensen M.O., Tieleman D.P., "Simulation of the coupling between nucleotide binding and transmembrane domains in the ABC transporter BtuCD". *Biophys J*, **2007**. 92(8): p. 2727-34.
- [82] Ivetac A., Campbell J.D., Sansom M.S., "Dynamics and Function in a Bacterial ABC Transporter: Simulation Studies of the BtuCDF System and Its Components". *Biochemistry*, **2007**. 46(10): p. 2767-78.
- [83] Ivetac A., Sansom M.S., "Molecular dynamics simulations and membrane protein structure quality". *Eur Biophys J*, **2008**. 37(4): p. 403-9.

- [84] Yoshida H., "Construction of higher order symplectic integrators". *Physics Letters A*, **1990**. 150(5-7): p. 262-268.
- [85] Verlet L., "Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules". *Phys Rev*, **1967**. 159(1): p. 98-103.
- [86] van Gunsteren W.F., Berendsen H.J.C., "A Leap-frog Algorithm for Stochastic Dynamics". *Molecular Simulation*, **1988**. 1(3): p. 173-185.
- [87] Van Der Spoel D., Lindahl E., Hess B., Groenhof G., Mark A.E., Berendsen H.J., "GROMACS: fast, flexible, and free". *J Comput Chem*, **2005**. 26(16): p. 1701-18.
- [88] Van der Spoel D., Lindahl E., Hess B., Kutzner C., van Buuren A.R., Apol E., Meulenhoff P.J., Tieleman D.P., Sijbers A.L.T.M., Feenstra K.A., van Drunen R., Berendsen H.J.C., "GROMACS User Manual". **2006**: <http://www.gromacs.org/>.
- [89] Hairer E., Lubich C., Wanner G., "Geometric numerical integration illustrated by the Störmer-Verlet method". *Acta Numerica*, **2003**. 12: p. 399-450.
- [90] Ryckaert J.P., Ciccotti G., Berendsen H.J.C., "Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes". *Journal of Computational Physics*, **1977**. 23(3): p. 327-341.
- [91] Miyamoto S., Kollman P.A., "SETTLE: an analytical version of the SHAKE and RATTLE algorithm for rigid water models". *Journal of Computational Chemistry*, **1992**. 13(8): p. 952.
- [92] Hess B., Bekker H., Berendsen H.J.C., Fraaije J.G.E.M., "LINCS: A linear constraint solver for molecular simulations". *Journal of Computational Chemistry*, **1997**. 18(12): p. 1463-1472.
- [93] Byrd R.H., Lu P., Nocedal J., Zhu C., "A Limited Memory Algorithm for Bound Constrained Optimization". *SIAM J. on Scientific Computing*, **1995**. 16(5): p. 1190-1208.
- [94] Berendsen H.J.C., Postma J.P.M., Van Gunsteren W.F., DiNola A., Haak J.R., "Molecular dynamics with coupling to an external bath". *J Chem Phys*, **1984**. 81(8): p. 3684.
- [95] Bussi G., Donadio D., Parrinello M., "Canonical sampling through velocity rescaling". *J Chem Phys*, **2007**. 126(1): p. 014101.
- [96] Metropolis N., Rosenbluth A.W., Rosenbluth M.N., Teller A.H., Teller E., "Equation of State Calculations by Fast Computing Machines". *J Chem Phys*, **1953**. 21(6): p. 1087.
- [97] Jones M.C., Marron J.S., Sheather S.J., "Progress in data-based bandwidth selection for kernel density estimation". *Computational Statistics*, **1996**. 11(3): p. 337-381.
- [98] Blundell T.L., Sibanda B.L., Sternberg M.J., Thornton J.M., "Knowledge-based prediction of protein structures and the design of novel molecules". *Nature*, **1987**. 326(6111): p. 347-52.

- [99] Bajorath J., Stenkamp R., Aruffo A., "Knowledge-based model building of proteins: concepts and examples". *Protein Sci*, **1993**. 2(11): p. 1798-810.
- [100] Johnson M.S., Srinivasan N., Sowdhamini R., Blundell T.L., "Knowledge-based protein modeling". *Crit Rev Biochem Mol Biol*, **1994**. 29(1): p. 1-68.
- [101] Chenna R., Sugawara H., Koike T., Lopez R., Gibson T.J., Higgins D.G., Thompson J.D., "Multiple sequence alignment with the Clustal series of programs". *Nucleic Acids Res*, **2003**. 31(13): p. 3497-500.
- [102] Notredame C., Higgins D.G., Heringa J., "T-Coffee: A novel method for fast and accurate multiple sequence alignment". *J Mol Biol*, **2000**. 302(1): p. 205-17.
- [103] Subramanian A.R., Kaufmann M., Morgenstern B., "DIALIGN-TX: greedy and progressive approaches for segment-based multiple sequence alignment". *Algorithms Mol Biol*, **2008**. 3: p. 6.
- [104] Edgar R.C., "MUSCLE: multiple sequence alignment with high accuracy and high throughput". *Nucleic Acids Res*, **2004**. 32(5): p. 1792-7.
- [105] Eswar N., Webb B., Marti-Renom M.A., Madhusudhan M.S., Eramian D., Shen M.Y., Pieper U., Sali A., "Comparative protein structure modeling using Modeller". *Curr Protoc Bioinformatics*, **2006**. Chapter 5: p. Unit 5 6.
- [106] Marti-Renom M.A., Stuart A.C., Fiser A., Sanchez R., Melo F., Sali A., "Comparative protein structure modeling of genes and genomes". *Annu Rev Biophys Biomol Struct*, **2000**. 29: p. 291-325.
- [107] Sali A., Blundell T.L., "Comparative protein modelling by satisfaction of spatial restraints". *J Mol Biol*, **1993**. 234(3): p. 779-815.
- [108] Globisch C., Pajeva I.K., Wiese M., "Identification of Putative Binding Sites of P-glycoprotein Based on its Homology Model". *ChemMedChem*, **2008**. 3(2): p. 280-95.
- [109] Kukol A., "Lipid Models for United-Atom Molecular Dynamics Simulations of Proteins". *J. Chem. Theory Comput.*, **2009**. 5(3): p. 615-626.
- [110] Tusnady G.E., Dosztanyi Z., Simon I., "TMDET: web server for detecting transmembrane regions of proteins by using their 3D coordinates". *Bioinformatics*, **2005**. 21(7): p. 1276-7.
- [111] Tusnady G.E., Simon I., "The HMMTOP transmembrane topology prediction server". *Bioinformatics*, **2001**. 17(9): p. 849-50.
- [112] Amadei A., Linssen A.B., Berendsen H.J., "Essential dynamics of proteins". *Proteins*, **1993**. 17(4): p. 412-25.
- [113] Moro G., Bonati L., Bruschi M., Cosentino U., De Gioia L., Fantucci P.C., Pandini A., Papaleo E., Pitea D., Saracino G.A., Zampella G., "Computational approaches to shed light on molecular mechanisms in biological processes". *Theor Chem Acc*, **2007**. 117(5-6): p. 723-741.

- [114] Tama F., Sanejouand Y.H., "Conformational change of proteins arising from normal mode calculations". *Protein Eng*, **2001**. 14(1): p. 1-6.
- [115] Kabsch W., Sander C., "Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features". *Biopolymers*, **1983**. 22(12): p. 2577-637.
- [116] Lee B., Richards F.M., "The interpretation of protein structures: estimation of static accessibility". *J Mol Biol*, **1971**. 55(3): p. 379-400.
- [117] "The CCP4 suite: programs for protein crystallography". *Acta Crystallogr D Biol Crystallogr*, **1994**. 50(Pt 5): p. 760-3.
- [118] McClendon C.L., Friedland G., Mobley D.L., Amirkhani H., Jacobson M.P., "Quantifying Correlations Between Allosteric Sites in Thermodynamic Ensembles". *J Chem Theory Comput*, **2009**. 5(9): p. 2486-2502.
- [119] Maz'ya V., Schmidt G., "On approximate approximations using Gaussian kernels". *IMA J Numer Anal*, **1996**. 16(1): p. 13-29.
- [120] Dempster A.P., Laird N.M., Rubin D.B., "Maximum Likelihood from Incomplete Data via the EM Algorithm". *Journal of the Royal Statistical Society, Series B (Methodological)*, **1977**. 39(1): p. 1-38.
- [121] Arthur D., Vassilvitskii S., "k-means++: The Advantages of Careful Seeding", in *ACM-SIAM Symposium on Discrete Algorithms*. **2007**: New Orleans, Louisiana. p. 1027-1035.
- [122] Huber M.F., Bailey T., Durrant-Whyte H., Hanebeck U.D., "On entropy approximation for Gaussian mixture random vectors", in *IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*. **2008**. p. 181-188.
- [123] Lange O.F., Grubmuller H., "Full correlation analysis of conformational protein dynamics". *Proteins*, **2008**. 70(4): p. 1294-312.
- [124] Killian B.J., Yudenfreund Kravitz J., Gilson M.K., "Extraction of configurational entropy from molecular simulations via an expansion approximation". *J Chem Phys*, **2007**. 127(2): p. 024107.
- [125] Lugo M.R., Sharom F.J., "Interaction of LDS-751 with P-glycoprotein and mapping of the location of the R drug binding site". *Biochemistry*, **2005**. 44(2): p. 643-55.
- [126] Loo T.W., Clarke D.M., "Identification of residues in the drug-binding site of human P-glycoprotein using a thiol-reactive substrate". *J Biol Chem*, **1997**. 272(51): p. 31945-8.
- [127] Loo T.W., Bartlett M.C., Clarke D.M., "Transmembrane segment 7 of human P-glycoprotein forms part of the drug-binding pocket". *Biochem J*, **2006**. 399(2): p. 351-9.



- [128] Oliveira A.S., Baptista A.M., Soares C.M., "Conformational changes induced by ATP-hydrolysis in an ABC transporter: A molecular dynamics study of the Sav1866 exporter". *Proteins*, **2011**. 79(6): p. 1977-90.
- [129] Loo T.W., Clarke D.M., "Inhibition of oxidative cross-linking between engineered cysteine residues at positions 332 in predicted transmembrane segments (TM) 6 and 975 in predicted TM12 of human P-glycoprotein by drug substrates". *J Biol Chem*, **1996**. 271(44): p. 27482-7.
- [130] Loo T.W., Bartlett M.C., Clarke D.M., "Substrate-induced conformational changes in the transmembrane segments of human P-glycoprotein. Direct evidence for the substrate-induced fit mechanism for drug binding". *J Biol Chem*, **2003**. 278(16): p. 13603-6.
- [131] Storm J., O'Mara M L., Crowley E.H., Peall J., Tieleman D.P., Kerr I.D., Callaghan R., "Residue G346 in Transmembrane Segment Six is Involved in Inter-Domain Communication in P-Glycoprotein". *Biochemistry*, **2007**. 46(35): p. 9899-9910.
- [132] Becker J.P., Van Bambeke F., Tulkens P.M., Prevost M., "Dynamics and Structural Changes Induced by ATP Binding in SAV1866, a Bacterial ABC Exporter". *J Phys Chem B*, **2010**. 114(48): p. 15948-15957.
- [133] Doshi R., Woebking B., van Veen H.W., "Dissection of the conformational cycle of the multidrug/lipidA ABC exporter MsbA". *Proteins*, **2010**.
- [134] Rosenberg M.F., Oleschuk C.J., Wu P., Mao Q., Deeley R.G., Cole S.P., Ford R.C., "Structure of a human multidrug transporter in an inward-facing conformation". *J Struct Biol*, **2010**. 170(3): p. 540-7.
- [135] Huang K., "Lectures on Statistical Physics and Protein Folding". **2005**: *World Scientific Publishing Company*.