

# **Vak vezet világtalant: hogyan lesz rendezetlen peptidekből rendezett komplex?**

## **doktori dolgozat**

Györffy Dániel

Témavezetők:

Dr. Závodszy Péter

Dr. Szilágyi András



Pázmány Péter Katolikus Egyetem  
Információs Technológiai és Bionikai Kar

Budapest, 2014.



Vajjon a vak vezetheti-é a világtalant? avagy nem  
mindketten a verembe esnek-é?

*Lukács 6, 39*



# Tartalomjegyzék

<b>Bevezetés</b>	<b>7</b>
Fehérjék felgombolyodásának statisztikus mechanikája . . . . .	8
Rendezetlen fehérjék . . . . .	10
A rendezetlenség jóslása . . . . .	11
Molekuláris felismerés biológiai makromolekuláknál . . . . .	12
Rendezetlen fehérjék kapcsolt felgombolyodás-kötődése . . . . .	14
Két- és háromállapotú dimerek . . . . .	17
Egyszerűsített fehérjemodellek . . . . .	18
Hálózatmodellek . . . . .	19
<b>Célkitűzések</b>	<b>22</b>
<b>Módszerek</b>	<b>23</b>
A HP modell . . . . .	23
A HPN modell . . . . .	24
A pull moves mozgáskészlet . . . . .	24
Az állapottér föltérképezése . . . . .	26
Enumerációs vizsgálatok . . . . .	26
Mintavételi eljárások . . . . .	26
Állapothálózat . . . . .	29
Perron-klaszter Klaszterelemzés . . . . .	29
Diszkonnectivitási gráf . . . . .	30
Átmenetiútvonal-elmélet . . . . .	30
A vizsgált szekvenciák . . . . .	32
A van't Hoff és a kalorimetrikus entalpia számítása . . . . .	32
A használt programok . . . . .	33
<b>Eredmények</b>	<b>34</b>
Alkonyzóna fehérjék rendezettsége és rendezetlensége között . . . . .	34
Monte Carlo-szimulációk . . . . .	37
A vizsgálatba bevont szekvenciák kiválasztása . . . . .	37
A megfelelő mozgáskészlet kiválasztása . . . . .	38
A legmegfelelőbb energiafüggvény kiválasztása . . . . .	42
Távolságfüggő energiafüggvény . . . . .	43
„Klaszter” energiafüggvény . . . . .	43
Négyzetes diagonális energiafüggvény . . . . .	44

Az energiafüggvények összevetése . . . . .	44
A négyzetrácsmodell szimmetriái . . . . .	45
Kétrétegű állapot-hálózat modell . . . . .	46
A kétrétegű modell elemzése előre definiált állapotok segítségével . . . . .	48
Kötődési fluxustájkép . . . . .	57
A két lánc viselkedésének szimmetriája . . . . .	58
A kétrétegű modell belső dinamikájának elemzése . . . . .	58
<b>Az eredmények megvitatása</b>	<b>65</b>
Sokállapotú rendszerek kinetikai leírásának két fő módja . . . . .	65
Kétrétegű állapot-hálózat modell . . . . .	65
A kötődés – felgombolyodás három fő mechanizmusa és azok aránya . . . . .	66
Steady-state kontra egyensúly . . . . .	67
Metastabilis állapotok . . . . .	68
Derék és nemderék szekvenciák dimerképzésének sebessége – az előképzett szerkezetek szerepe . . . . .	68
Aszimmetria . . . . .	69
Az alkalmazott megközelítés korlátai . . . . .	69
<b>Összefoglalás</b>	<b>70</b>

# Bevezetés

A fehérjék funkciójukat kölcsönhatásaikon keresztül fejtik ki. A fehérjék molekuláris felismerési folyamatainak megértése kulcsfontosságú a molekuláris rendszerek működésének feltárásában. Ma már jól ismert, hogy a fehérjék nem valamilyen statikus, merev testként viselkednek, hanem a szerkezetük dinamikusan változik; igaz ez azokban a folyamatokban is, amelyekben a fehérjék valamilyen molekuláris komplexet alakítanak ki.

A fehérjék egy nemrégén felfedezett csoportjában a szerkezeti dinamikának még a szerkezetnél magánál is lényegesebb szerepe van. Az ún. rendezetlen fehérjék szabad állapotban, tehát amikor nem kötődnek valamilyen biológiai partnerhez, nem rendelkeznek jól meghatározott háromdimenziós szerkezettel, hanem konformációs állapotok közötti gyors fluktuációt végeznek. Sok esetben azonban, amikor valamilyen célmolekulához kötődnek, rendezett szerkezetet vesznek fel. *A kapcsolt felgombolyodás és kötődés leírására két modell létezik, az indukált illeszkedés és a konformációkiválasztás.* Az indukált illeszkedés szerint a ligandum kötődése váltja ki azokat a konformációváltozásokat, amelyek során a rendezetlen lánc a komplexre jellemző rendezett szerkezetet veszi föl. A konformációkiválasztás esetében a rendezetlen fehérje szabad állapotban különböző konformációs állapotok között fluktuál, köztük a komplexbeli konformációt is fölveszi, és a ligandum mintegy kiválasztja ezt a konformációt azáltal, hogy ezzel képes a legkedvezőbb kötést kialakítani. A legújabb eredmények azt mutatják, hogy egy adott komplex képződésében mindkét mechanizmus szerepet játszhat, de különböző mértékben.

A kapcsolt felgombolyodás és kötődés elméletét olyan esetekre dolgozták ki, ahol az egyik résztvevő lényegében merevnek tekinthető – a kapcsolt felgombolyodás és kötődés tehát csak a flexibilis partner esetében értelmezhető. Rendezetlen peptidek homodimerképzése során azonban mindkét lánc átmegy egy rendezetlen-rendezett átmeneten. Nyilvánvaló, hogy ha a hagyományos fogalomrendszer szerint próbáljuk leírni a folyamatot, meg kell válaszolni azt a kérdést, hogy míg az egyik lánc rendezett szerkezetének kialakulása folyamatban van – függetlenül attól, hogy a kötődés előtt, közben, vagy után következik be –, mi történik a másik láncal, hiszen a dimer kialakulása során mindkét lánc rendeződése végbemegy. Tehát nem egyszerűen a kötődési folyamat és egyetlen lánc rendeződési folyamatának időbeli viszonyát kell vizsgálni, hanem a kötődésen túl két lánc rendeződését, és azoknak időbeli viszonyait is figyelemmel kell kísérni.

További kérdéseket vet föl a két lánc szekvenciájának azonosságából eredő inherens szimmetria. Vajon a szekvenciabeli szimmetriából az alapállapot szimmetriája is következik? Ha pedig az alapállapot szimmetrikus, vajon a dimerképződés során is szimmetrikusan viselkedik a két lánc?

Munkámban két rendezetlen peptid által képzett rendezett komplex kialakulásának fogalmi leírását kíséreltem meg a hagyományos fogalomrendszerből kiindulva, és választ keresek arra a kérdésre, hogy a szekvenciák szimmetriája megnyilvánul-e a dimerképződési folyamat során. Tudomásom szerint ezidáig senki nem vizsgálta egzakt modellek segítségével rendezetlen pepti-

dek homodimerképződését. Munkám során tehát először tesztek kísérletet egyszerűsített, egzakt modellek, és a modellek segítségével definiált dinamikus hálózatok alkalmazásával rendezetlen peptidek kölcsönös kapcsolt felgombolyodásának és kötődésének kvantitatív leírására.

## Fehérjék felgombolyodásának statisztikus mechanikája

Anfinsen 1961-es kísérletében kimutatta, hogy a ribonukleáz A aktív állapotának szerkezete termodinamikailag meghatározott [1]. Később ezt általánosította „termodinamikai hipotézis” néven. A termodinamikai hipotézis kimondja, hogy egy fehérje natív állapota normális fiziológiás körülmények (pH, oldószer, ionerősség stb.) között az az állapot, amelyben az egész rendszerre (fehérje és oldószer) nézve a szabadentalpia minimális [2].

Levinthal gondolatkísérletében megvizsgálta, hogy lehetséges-e az, hogy a fehérjék véletlen keresés során találjanak rá natív állapotokra. Levinthal gondolatmenetét követve, feltételezve, hogy egy 150 aminosav hosszú peptidláncban minden aminosavnak 3 lehetséges állapota van, a lehetséges konformációk száma  $3^{150} \approx 4 \cdot 10^{71}$ . Ha a lánc 1 ps-onként látogat meg egy konformációt, akkor  $10^{59}$  s  $\approx 3 \cdot 10^{51}$  év szükséges az összes konformáció végigpróbálásához, ami jóval több, mint az univerzum becsült kora.

A mérések azonban azt mutatják, és a biológiai kényszerek is megkövetelik, hogy egy fehérje felgombolyodása akár mikroszekundumos időskálán is végbemehet. Az ellentmondás Levinthal szerint úgy oldható fel, ha feltételezzük, hogy a felgombolyodás jól definiált útvonalakon zajlik, ami lehetővé teszi, hogy a csillagászati számú konformációnak csak egy töredékét kell végigjárni a felgombolyodás során [3]. Ennek egyik lehetséges módja, ha lokális kölcsönhatások lokális szerkezeteket hoznak létre, és ezek mintegy felgombolyodási magokként szolgálnak és megszabják a harmadlagos szerkezet kialakulásának menetét [4].

Sokáig az a nézet uralta a fehérjék felgombolyodásáról való gondolkodást, hogy a felgombolyodás során a fehérje jól definiált állapotok során keresztül jut el a natív állapotba. Egy útvonal lehet például

$$D \rightarrow I_1 \rightarrow I_2 \rightarrow I_3 \rightarrow N, \quad (1)$$

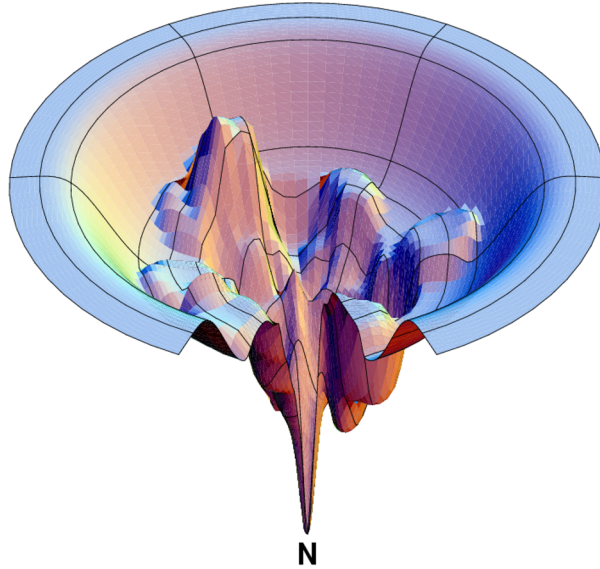
ahol  $D$  a denaturált,  $N$  pedig a natív állapot, az  $I$  állapotok pedig különböző köztes állapotokat jelölnek.

Több kísérletben is azt találták, hogy a felgombolyodást különböző állapotokból elindítva a fehérje ugyanabba a natív állapotba gombolyodott föl, tehát legalább néhány különböző útvonalnak léteznie kell [5].

A fehérjék energetikájára és felgombolyodására jelenleg is elfogadott nézetet a spinüvegek elmélete ihlette. A spinüvegek olyan rendszerek, amelyekben az egyes kölcsönhatások kapcsolt-sága miatt nagy a frusztráció – azaz az egyes kedvező kölcsönhatások nem elégíthetők ki egyszerre, és ennek következtében a szabadenergiafelszín durva. A frusztráció következtében az átalakulások lassúak, csak nagyon magas hőmérsékleten gyorsulnak föl, és ilyen magas hőmérsékleten, ha van is a rendszernek egy kitüntetett, a natív állapottal analóg állapota, az a magas hőmérséklet miatt csak kis valószínűséggel valósul meg.

Aminosavak véletlen heteropolimereit vizsgálva, azt találták, hogy ezeknek a szekvenciáknak a legtöbbször hasonlóan viselkedik a spinüvegekhez [6]. Az élő szervezetekben talált, rendezett natív állapottal rendelkező fehérjék azonban nem ilyenek. Ezek natív állapota a viszonylag gyors felgombolyodás hőmérsékletén is stabilis. Ezt kvantitatíven a  $T_f \gg T_g$  összefüggés írja le, ahol  $T_f$  a felgombolyodási hőmérséklet,  $T_g$  pedig az úgynevezett üvegátmeneti hőmérséklet [7]. Az





1. ábra. Fehérje hipotetikus energia-felšíne valamilyen felgombolyodási paraméter függvényében. A tölcsér alak teszi lehetővé egyes szekvenciák gyors felgombolyodását olyan hőmérsékleten, ahol a natív állapot stabilis [9].

üvegátmeneti hőmérsékletre többféle definíció is ismert, de ezek mind arra a jelenségre reflektálnak, hogy az olyan sok szabadsági fokú és frusztrált rendszerek dinamikája, mint amilyenek a fehérjék vagy a spinüvegek, egy bizonyos hőmérséklet alatt hirtelen nagyon lelassul, a rendszer lényegében néhány alacsony energiájú állapot valamelyikébe befagy [8].

A felgombolyodási hőmérsékleten a natív állapot még kellően stabilis és a hőmérséklet csökkentésével a stabilitása növekszik. A rendszer dinamikája lassabb lesz, és ezzel a natív állapot eléréséhez szükséges idő viszont növekszik a hőmérséklet csökkentésével. Ha tehát a felgombolyodási hőmérsékleten a rendszer még kellően dinamikus – tehát  $T_f > T_g$ , akkor a rendszer gyorsan képes elérni a natív állapotát, ami kellően stabilis is.

A szabadenergia-felšíne durvaságát a kölcsönhatások frusztráltsága okozza, ezért az élő szervezetben található rendezett fehérjék kölcsönhatásai minimálisan frusztráltak. Az evolúció során tehát a minimálisan frusztrált szekvenciák maradtak meg. Az ilyen minimálisan frusztrált szekvenciák energiafelšíne egy tölcsérre emlékeztet, azaz minél jobban hasonlít egy konformáció a natív konformációra, átlagosan annál alacsonyabb az energiája. Lehetnek ugyan a tölcsér falán lokális minimumok, de ezek a tölcsér központi üregének a mélységéhez viszonyítva sekélyek (1. ábra) [7].

A felgombolyodási képesség gyakorlati szempontú kritériuma szerint azok a fehérjék jól felgombolyodók, amelyekre fennáll az

$$\frac{\langle E \rangle_d - E_{\text{nat}}}{\sigma_d E} \gg 1 \quad (2)$$

összefüggés, ahol  $E_{\text{nat}}$  a natív állapot energiája,  $\langle E \rangle_d$  a denaturált állapotok energiájának átlaga és  $\sigma_d E$  a denaturált állapotok energiáinak szórása [10].

A felgombolyodási képesség kapcsolatba hozható a *kétállapotú kooperativitással* is; a kétállapotú kooperativitás egy gyakran használt kritériuma az, hogy a ( $\Delta H_{vH}$ ) van't Hoff entalpia és a ( $\Delta H_{cal}$ ) kalorimetrikus entalpia aránya közel legyen 1-hez, azaz

$$\Delta H_{vH}/\Delta H_{cal} \approx 1, \quad (3)$$

amit *kalorimetrikus kétállapotúságnak* nevezünk. Ha tehát meg akarjuk vizsgálni, hogy egy energiafüggvény mennyire teljesíti a kalorimetrikus kétállapotúság kritériumát, ki kell számítanunk ezt az arányt.

Mivel a fehérjék asszociációja során ugyanazok a kölcsönhatástípusok játszanak szerepet, mint a fehérjék felgombolyodásakor, ezért feltételezhető, hogy az asszociációs energiafelszín is tölcser alakú, akárcsak a felgombolyodás energiafelszíne [11, 12]. A kötődés során a különböző láncok asszociációja megfeleltethető a „hidrofób felgombolyodási egységek”, „mikrodomének”, „aldomének” asszociációjának, magasabb szinten a már felgombolyodott domének asszociációjának. Az egyetlen különbség a láncok folytonossága [12]. A komplexképződés mechanizmusát az határozza meg, hogy milyen a felgombolyodási és a kötődési tölcser egymáshoz viszonyított alakja [13].

## Rendezetlen fehérjék

Évtizedeken keresztül a fehérjék működési mechanizmusára az egyetlen magyarázat az volt, hogy a fehérjék valamilyen jól meghatározott szerkezettel bírnak, és ez a szerkezet határozza meg a funkciót, amit ellátnak. A kilencvenes évek végén, kétezres évek elején több kutatócsoportban is felfigyeltek arra, hogy van a fehérjéknek egy bizonyos csoportja, amelyeknek az aminosavösszetétele jellemzően eltér a korábban részletesen vizsgált és a natív állapotban jól meghatározott szerkezettel rendelkező fehérjék aminosavösszetételétől [14, 15, 16].

Ezek a fehérjék a különböző szerkezetmeghatározó módszerek szerint is (pl. NMR, röntgen-diffrakció, cirkuláris dikroizmus) sajátos, a rendezett fehérjéktől eltérő viselkedést mutatnak [17]. Általában a rendezetlenség indikátorának tartják, ha a fehérjelánc bizonyos szakaszai nem látszanak a röntgenszerkezetben [18, 19]. A rendezetlen fehérjék CD spektruma is jellegzetes, a módszer alkalmas a rendezetlen fehérjékre jellemző, reziduális másodlagos szerkezeti elemek kimutatására is [20, 21]. Habár jól definiált szerkezetük nincs, mégis jól meghatározott funkciót képesek ellátni, amihez nemhogy nem szükséges a jól meghatározott szerkezet, de bizonyos esetekben kimondottan annak hiánya az, ami a funkció ellátását lehetővé teszi [22, 23, 24].

Habár szabad állapotban a rendezetlen fehérjék nem rendelkeznek jól meghatározott szerkezettel, gyakran valamilyen partnerhez való kötődés során rendezett szerkezetűvé válnak. Ezt a jelenséget nevezik kapcsolt felgombolyodás-kötődésnek [25]. A kötőpartner lehet kisméretű részecske, pl. ion, vagy kisméretű szerves molekula, de lehet biológiai makromolekula, pl. DNS vagy egy másik fehérje. Előfordul az az eset is, hogy a rendezetlen fehérje rendezett homodimert képez, pl. Arc represszor [26].

Bár korábban szinte teljesen kiestek a fehérjekutatók látóköréből, kiderült róluk, hogy nagyon elterjedtek az élővilágban, különösen az eukarióta szervezetek között [27]. Például *Drosophila melanogaster* genomját vizsgálva azt találták, hogy a genomban található fehérjeszekvenciák 17%-a rendezetlen fehérjét kódol [16]. A proteomot vizsgálva azt találták, hogy míg az *E. coli* fehérjéinek 13,7%-a, addig a sörélesztő (*Saccharomyces cerevisiae*) fehérjéinek 49,1%-a tartalmaz legalább egy, legalább 30 aminosav hosszúságú rendezetlen szakaszt [28, 29].

A rendezetlen fehérjék különösen gyakoriak bizonyos molekuláris, ill. celluláris funkciók ellátásában szerepet játszó fehérjék között. Ilyen celluláris funkciók a

- transzkripció, transláció szabályozása
- jelátvitel, sejtciklus szabályozása
- mRNS feldolgozás, splicing
- citoszkeleton organizációja

[17, 30, 31, 32].

Megfigyelhető, hogy ezek mind olyan folyamatok, amelyekben központi szerepe van valamilyen makromolekuláris felismerési folyamatnak. A transzkripció és transláció, valamint az mRNS feldolgozás során elsősorban fehérje-nukleinsav, a jelátvitel és a sejtciklus szabályozásában pedig elsősorban fehérje-fehérje kölcsönhatások játszanak szerepet.

A rendezetlenségnek több előnye is van, amelyek megindokolják, hogy miért a fenti molekuláris folyamatok azok, ahol leggyakrabban találkozhatunk rendezetlen fehérjékkel. Az egyik ilyen, sokszor hangoztatott előny, hogy a rendezetlen fehérjék által létrehozott kötések nagy specificitásúak, de kis affinitásúak. A specificitás a nagyméretű interfésznek köszönhető, a kis affinitás pedig a nagy konformációsentrópia-csökkenésnek a kötődés – és az ezzel kapcsolt lánc-rendeződés – során. Ezt az általános nézetet kérdőjelezi meg azonban Dogan és munkatársai. Összehasonlítva a rendezett és a rendezetlen fehérjék komplexeinek kötési állandóit, nem találtak különbséget a két típus között [33].

Jellemző a rendezetlen fehérjékre, hogy több különböző partnerrel képesek komplex kialakítására, és ezekben a komplexekben más-más konformációt vehetnek fel [34, 35]. A rendezetlen fehérjék kötődése gyorsabb, mint a rendezetteké.

## A rendezetlenség jóslása

Mivel a rendezetlen fehérjék aminosavösszetétele jellegzetesen eltér a rendezett fehérjékétől [16, 36], sok rendezetlenségjósló eljárás az adott polipeptidlánc, vagy a polipeptidlánc egy szakaszának aminosavösszetétele alapján jóslja, hogy az adott szakasz rendezett vagy rendezetlen [37, 38, 39, 40, 41]. A teljes, húszdimenziós aminosavösszetétel-teret egy, a rendezetlenség mértékét meghatározó változóba való  $\mathbb{R}^{20} \rightarrow \mathbb{R}$  leképezés helyett a probléma kezelhetőségének javítása érdekében használható pl. az az  $\mathbb{R}^2 \rightarrow \mathbb{R}$  leképezés, ahol megelégszünk azzal, hogy a pontos aminosavösszetétel helyett az átlagos hidrofobicitással és az egy aminosavra eső nettó töltés abszolút értékével jellemezzünk egy adott szekvenciát. A hidrofobicitás–nettó töltés alapján végzett leképezés alkalmazásával a rendezetlen és a rendezett polipeptidláncok, ill. szakaszok elég jó elkülönülése érhető el [36].

Gyakori tapasztalat, hogy rövid láncokra, vagy rövid peptidszakaszokra a rendezetlenségjósló eljárások kevésbé hatékonyak, mint hosszabb láncokra [42, 43, 44]. A téves osztályozás mögött gyakran az a jelenség áll, hogy a hidrofobicitás–nettó töltés síkon a rendezett és a rendezetlen polipeptidláncokhoz tartozó pontokat tartalmazó részsíkok átfednek. Adódik tehát a feltételezés, hogy rövid láncokra az átfedő rész kiterjedtebb, mint hosszabbakra [45].

## Molekuláris felismerés biológiai makromolekuláknál

A hagyományos szerkezet-funkció paradigma szerint a fehérjék funkcióját a háromdimenziós szerkezetük határozza meg. Emil Fischer a XIX. század végén az enzimműködés magyarázatára állította fel híres hipotézisét, miszerint az enzim felszínének olyan a mintázata, hogy abba a szubsztrát mint kulcs a zárba illeszkedik bele [46]. Innen a közkeletű „kulcs-zár hipotézis” elnevezés. Nyilvánvaló, hogy Fischer a fehérjék szerkezetét merevnek tekintette.

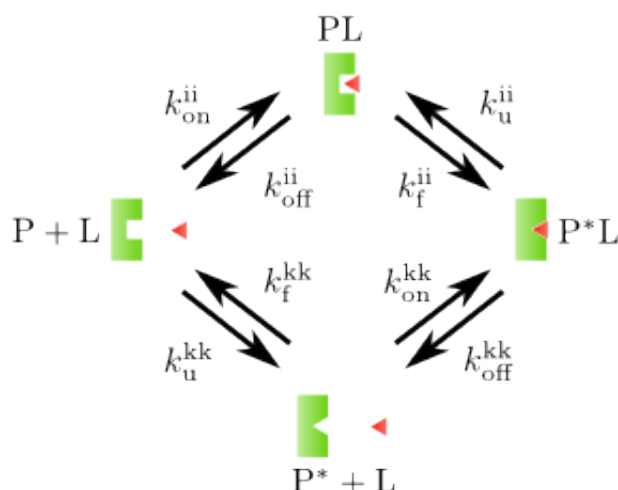
Később több olyan felfedezést is tettek az enzimműködést vizsgáló kutatók, amelyek arra mutattak, hogy a Fischer-féle magyarázat nem kielégítő. Fölfedeztek például olyan enzimeket, amelyek esetében a szubsztrát távollétében a kötőhely, ahova a szubsztrátnak kötnie kell, nem hozzáférhető. Ha azonban a szubsztrát valamilyen aspecifikus módon kötött az enzimhez, előidézte annak konformációs átalakulását, és ezáltal az aktív hely nyitottá vált, és a szubsztrát most már be tudott kötni a kötőhelyére [47]. Feltételezni kellett tehát, hogy a kötődés és a makromolekula konformációs átalakulása kapcsolt. Ennek a kapcsolt kötődés-konformációváltásnak speciális esete a rendezetlen fehérjék körében gyakori kapcsolt felgombolyodás-kötődés, ahol a szabad állapotban rendezetlen fehérje valamilyen partnerhez való kötődés során rendezett szerkezetűvé válik.

A kötődés és konformációs átalakulás kapcsoltágát leíró első modell a Koshland által bevezetett *indukált illeszkedés* („induced fit”) mechanizmusa [48]. Az indukált illeszkedés segítségével már magyarázni lehetett azokat az eseteket, ahol a szubsztrátkötőhely csak a szubsztrát jelenlétében válik elérhetővé. Az indukált illeszkedés hipotézise a kulcs-zár hipotézishez képest egyéb előnyökkel is rendelkezik. A kulcs-zár hipotézis ugyanis nem tudott kielégítő magyarázatot adni a szubsztrátspecifitásra. Bizonyos enzimek esetében ugyanis nagyon hasonló szerkezetű molekulák közül csak az egyik volt képes az enzimhez kötődni, míg esetleg teljesen eltérő szerkezetű molekulák átalakulását az enzim egyaránt katalizálni volt képes, ami csak úgy volt magyarázható, ha feltételezték, hogy a fehérjék nem teljesen merev szerkezetűek, és bizonyos konformációs átalakulásra képesek.

A XX. század hatvanas éveiben Straub F. Brúnó egy eltérő magyarázattal állt elő a szubsztrátkötés specifitásának magyarázatára, és ezt „fluctuation fitnek” nevezte el. A kiindulópont az volt, hogy mivel a fehérjék dinamikus szerkezetűek, ezért különböző konformációs állapotokban fordulhatnak elő, az adott konformációs állapotok szabadenergiáival arányos valószínűséggel. Ha pedig ez így van, akkor feltételezhető, hogy az enzim az enzim – szubsztrát komplexre jellemző konformációját felveszi a szubsztrát távollétében is. Ilyen esetben a szubsztrát mintegy kiválasztja a sok lehetséges konformáció közül azt, ami a kötődés szempontjából a legmegfelelőbb, ami éppen az enzim komplexbeli konformációja [49].

Kétezer-kilencben a hipotézis újra megjelent, és most már *konformációkiválasztás* („conformational selection”) néven [50] terjedt el. A konformációkiválasztás, bár általánosabb kontextusban tárgyalja a problémát, tehát a molekuláris felismerés egyik lehetséges mechanizmusaként írja le, az alapelvét tekintve megegyezik a Straub által bevezetett fluktuációs fit mechanizmusával [51]. Nevezik a jelenséget *populációeltolódásnak* is („population shift”), arra utalva, hogy a konformációs sokaságban az egyes konformációk egymáshoz viszonyított valószínűségei változnak a kötődés hatására [52].

Az indukált illeszkedés és a konformációkiválasztás „hívei” között ma is komoly viták folynak és máig nem eldöntött kérdés, hogy melyiknek mekkora a szerepe a biológiai makromolekulák molekuláris felismerési folyamataiban.



2. ábra. Ligandumkötés egyszerűsített kinetikai sémája. A  $P^*L$  komplex kétféle útvonalon alakulhat ki, attól függően, hogy az asszociáció ( $P + L \rightarrow PL$ ) vagy a konformációs átalakulás ( $P + L \rightarrow P^* + L$ ) megy végbe előbb. Az előbbi az indukált illeszkedésnek, az utóbbi a konformációkiválasztásnak felel meg. A nyilak mentén az egyes folyamatok sebességi együtthatói vannak feltüntetve.

A problémát több szempontból is megközelítették már és a megközelítésmódtól függően maguk a fogalmak is más-más jelentést kapnak. A legtöbb helyen a tárgyalást arra a speciális esetre szűkítették, ahol a két kötőpartner közül csak az egyik rendelkezik jelentős konformációs flexibilitással, a másik partner szerkezete merevnek tekinthető. Ilyen eset, ahol egy enzim kisméretű szubsztrát átalakulását katalizálja, ilyen általában a receptor-ligandum kötődés vagy az antigén-antitest kölcsönhatás.

Talán a legelterjedtebb tárgyalásmód szerint a két mechanizmust az egyes elemi lépések sorrendje határozza meg. Indukált illeszkedés esetén először a kisméretű ligandum kötődése következik be, és ezt követi a makromolekulában valamilyen konformációs átrendeződés (2. ábra felső folyamat). Ezzel szemben konformációkiválasztás során a flexibilis partner konformációs átalakulásai során kialakul a kötő konformáció, és ebben az állapotban történik meg a kisméretű molekula kötése (2. ábra alsó folyamat).

Azt, hogy a komplex kialakulása melyik mechanizmus szerint megy végbe, az egyes elemi lépések sebességi együtthatói határozzák meg. Már ebből a tárgyalásmódból is jól látszik, hogy egy adott kölcsönhatás esetén mindkét mechanizmus szerepet játszhat, de különböző mértékben. Pontosabb képet kapunk a dominanciaviszonyokról, ha a sebességi együtthatók helyett az egyes útvonalak fluxusait hasonlítjuk össze [53].

Azt, hogy melyik mechanizmus milyen mértékben vesz részt egy adott komplex létrejöttében, a kötőpartnereken kívül a külső körülmények is befolyásolhatják. Okazaki és munkatársai molekuladinamikai szimulációk segítségével vizsgálták, hogy mely paraméterek és hogyan befolyásolják, hogy melyik mechanizmus domináns az adott szituációban. Azt találták, hogy ha a partnerek közötti kötés erős, és elsősorban hosszútávú kölcsönhatások felelősek a kialakulásáért, akkor az indukált illeszkedés a preferált mechanizmus. Rövid távú és gyenge kölcsönhatások

pedig a konformációkiválasztásnak kedveznek. Ezen túl a szerzők szerint azokban az esetekben, ahol pl. fehérje kisméretű ligandumot köt, ott várhatóan konformációkiválasztás zajlik le, míg fehérje más makromolekulákhoz való kötődése során inkább indukált illeszkedés [54].

A ligandum anyagi minőségén túl annak koncentrációja is befolyással lehet a mechanizmusra. Ha feltételezzük, hogy a konformációs átalakulás lassú a kötődéshez viszonyítva, akkor az egyensúly megközelítésének sebessége, amit  $k_{\text{obs}}$  ad meg, indukált illeszkedés esetén nő a ligandumkoncentráció növekedésével, míg konformációkiválasztás esetén csökken. Ha ezt a feltételezést elhagyjuk, akkor csak abban az esetben tehetünk egyértelmű kijelentést a komplexképződés mechanizmusára, ha  $k_{\text{obs}}$  csökken a ligandumkoncentráció növekedésével, ekkor konformációkiválasztás történik. Ellenkező esetben mind a konformációkiválasztás, mind az indukált illeszkedés lehetséges [55]. Zhou receptorok ligandumkötését egy négyállapotú modell segítségével vizsgálta. A modellben a receptor aktív és inaktív állapota közötti átmenet sebessége függött attól, hogy köti-e a ligandumot, valamint a kötési energia is különbözött az aktív és az inaktív állapotban. A modellen végzett számítások alapján, ha a konformációs átalakulás gyors a relatív diffúzió sebességéhez mérten, akkor indukált illeszkedés a komplex kialakulásának fő mechanizmusa, míg ha a konformációs átalakulás lassú, akkor az elsődleges komplexképződési mechanizmus a konformációkiválasztás [56].

Bizonyítékok a konformációkiválasztásra elsősorban az NMR-es mérésekből származtak. Ezek a bizonyítékok azon alapulnak, hogy ha sikerül a ligandum távollétében is kimutatni azt a konformációt, amit a makromolekula a komplexben fölvesz, az arra utal, hogy a konformációkiválasztás működhet az adott esetben [57]. Még erősebb bizonyító erejűek azok az esetek, ahol egy adott makromolekula több különböző ligandummal is képes komplexet kialakítani, de más-más konformációban. Amikor megvizsgálták a konformációs állapotait szabad állapotban, akkor a különböző komplexekre jellemző konformációkat sikerült kimutatni [58].

Önmagában az, hogy a komplexbeli konformáció megfigyelhető a ligandum távollétében is, még nem bizonyítja azt, hogy az ilyen konformációjú molekulákhoz történik a kötődés. Nyilvánvaló az is, hogy abban az esetben, amikor a kötőhely hozzáférhetősége függ a ligandum jelenlététől, egyedül az indukált illeszkedés lehet működőképes [47].

Elméleti számítások is vannak azonban, amelyek a konformációkiválasztás jelentőségét támasztják alá. Egy négyállapotú modell segítségével kísérik meg kimutatni, hogy ha az indukált illeszkedés során az ütközési komplex konformációja nem kimagasló a kiindulási anyagok koncentrációihoz viszonyítva, akkor az e mechanizmus szerint végbemenő komplexképződési reakció nagyon lassú a konformációkiválasztás szerinti reakcióhoz viszonyítva, aholis a kötő konformáció a kötődést megelőzően is már jelen van valamilyen kis valószínűséggel, és ez a kis valószínűségű kötő konformáció vesz részt a tényleges kötődésben [59].

## Rendezetlen fehérjék kapcsolt felgombolyodás-kötődése

Gyakori jelenség, hogy a szabad állapotban rendezetlen fehérjék rendezett szerkezetűvé válnak a funkcionális célmolekulához való kötés során. Ez a kapcsolt felgombolyodás és kötődés jelensége [25]. Habár a rendezett fehérjék komplexeinek kialakulása során a kötődés mellett a lánc felgombolyodásának is végbe kell menni, mégis hasonló sebességgel alakulnak ki, mint a rendezetlen fehérjék komplexei, ahol már utólagos felgombolyodásra nincs szükség [60]. Gyakori megfigyelés, hogy a rendezetlen fehérjék nem teljesen rendezetlenek szabad állapotban sem, hanem bizonyos előképződött szerkezeteket tartalmaznak. Ezek a szerkezeti elemek gyakran hasonlítanak az adott fehérje komplexeiben megfigyelhető szerkezetekre [61, 62]. Ezekre a bi-

zonyítékokra alapozva többen is fölvetették, hogy ezek az előképződött szerkezeti elemek kötőhelyekként szolgálnak és elősegítik a gyors asszociációt [63, 64]. Ez tulajdonképpen megfelel a konformációkiválasztás mechanizmusának. Ezek a bizonyítékok azonban kizárólag a szabad állapot vizsgálatából származnak, és az önmagában még nem bizonyíték a konformációkiválasztásra, hogy a kötött állapotbeli konformációk szabad állapotban is megjelennek.

Mások éppen a rendezetlen fehérjék nagyobb flexibilitásában látják a gyors komplexképződés magyarázatát. Chen a p53 – S100B komplex kialakulásának molekuladinamikai szimulációját végezte atomi szintű reprezentációt használva. A kapott többdimenziós szabadenergiafelszín vizsgálata azt mutatta, hogy a natív komplex kialakulása aspecifikus ütközési komplexen keresztül zajlik, és nem az előformált szerkezeti elemek, mint inkább a fly-casting (lásd később) mechanizmus járul hozzá a natív állapot gyors eléréséhez [65].

A KIX – CREB pKID domén kötődéshez kapcsolt felgombolyodását vizsgálták Gō-modell [66] segítségével és azt találták, hogy a monomer stabilitásának növelése lassította a komplex kialakulását, míg aspecifikus kontaktusok növelték a sebességet, de a mechanizmust nem változtatták meg. Ezek az eredmények megfeleltek az NMR-el kapott eredményeknek [67], ahol aspecifikus hidrofób kontaktusok által összetartott ütközési komplex létét mutatták ki. Hasonló eredményeket kaptak a CBD – Cdc42 kölcsönhatás – durvaszemcsés Gō-szimulációval kiegészített – egymolekulás kísérletben történt vizsgálata során. Itt a natív állapot mellett egy lazán kötött állapotot is találtak. A korábbi Gō-szimulációkról kimutatták, hogy túlbecsülik a szabad állapotú rendezetlen fehérje rendezettszerkezet-tartalmát és alulbecsülik a láncok közötti kölcsönhatások erősségét, aminek következtében olyan esetekben is a konformációkiválasztást tartották a felismerés mechanizmusának, ahol valójában indukált felgombolyodás játszódott le [68].

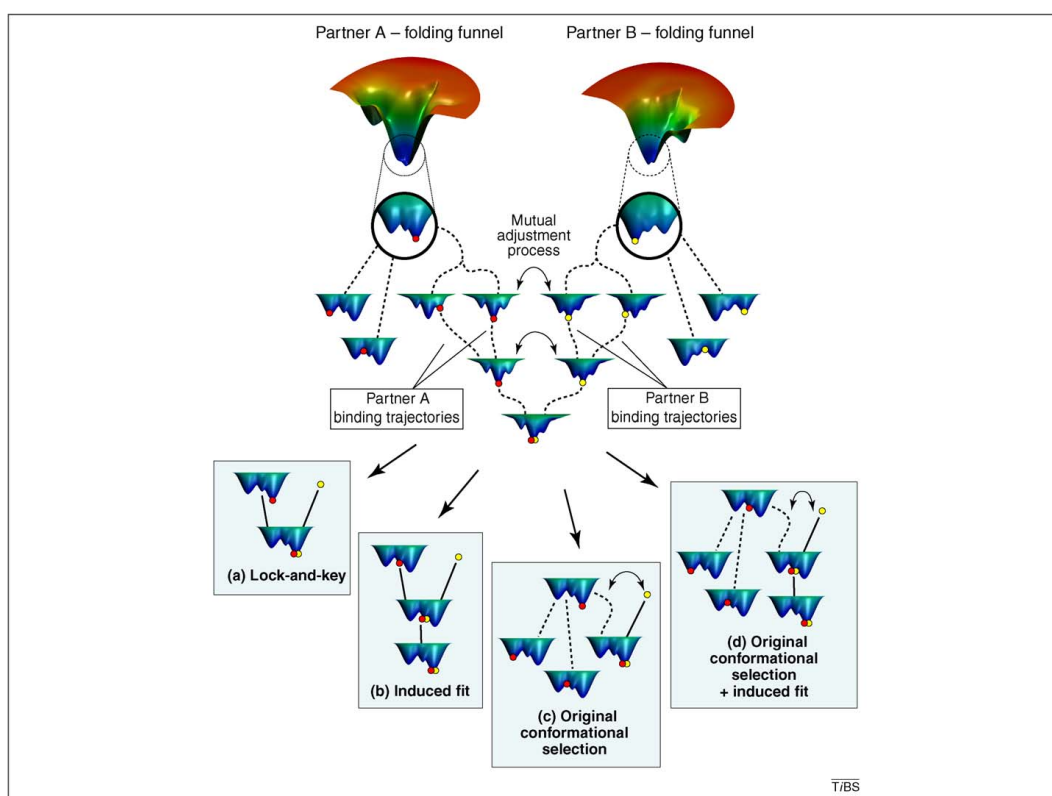
Monte Carlo szimulációk segítségével megmutatták, hogy a rendezetlen p27 Cdk2-höz történő kötődéséhez kapcsolt felgombolyodása során a felgombolyodás mechanizmusát és sebességét nem az előképződött  $\alpha$ -hélix határozza meg, hanem a specifikusan kialakított natív kötőfelület. Egyensúlyi és kinetikai mérésekben azt mutatták ki, hogy az RNáz S S-peptidjének, ami szabad állapotban rendezetlen, S-fehérjéhez való kötődése során a kötődés megelőzi az S-peptid felgombolyodását. Kezdetben a kötődésben néhány hidrofób régió vett részt. Ezek azonban jól meghatározott régiók voltak, ha más hidrofób régiókat vittek be a szekvenciába, az lassította a komplexképződést [69].

A fenti eredmények mind arra mutatnak, hogy a rendezetlen fehérjék kapcsolt felgombolyodás és kötődésének elsődleges mechanizmusa az indukált felgombolyodás lehet. Ennek magyarázatára több elmélet is napvilágot látott. Shoemaker szerint a rendezetlen fehérjék nagyobb befogási sugárral rendelkeznek, mint a rendezett fehérjék, és ezáltal nagyobb távolságból képesek aspecifikus kölcsönhatásokat kialakítani. Ezt a jelenséget a szerző „fly-casting” mechanizmusnak nevezte el [70]. Az  $IA_3$  inhibitor élesztő aszpartát proteázhoz (YPrA) való kötődése során, ahol az inhibitor N-terminálisa hélixszé rendeződik, a rendezetlen C-terminális rész részt vesz a kötésben, de a jelenléte a kötés erősségét nem befolyásolja, ami a fly casting mechanizmus kísérleti bizonyítékának tekinthető [71]. Huang és Liu 2009-ben publikált cikkükben azonban megmutatták, hogy a nagyobb befogási sugár lassabb diffúzióval párosul, és bár a rendezetlen fehérjék kötődése valóban gyorsabban megy végbe, mint a rendezett láncoké, de ez nem a „fly casting” mechanizmusnak tulajdonítható, hanem az alacsonyabb kötődési energiágnak [60]. Hosszú távú elektrosztatikus kontaktusok is elősegíthetik a rendezetlen fehérjék gyors kötődését [72, 73].

Habár úgy tűnik, a rendezetlen fehérjék kötődése valóban gyorsabb a globuláris fehérjék kötődéséhez viszonyítva, ez a különbség csekély [33]. Ez szükségessé teszi, hogy a rendezet-

len – rendezett átmenet is gyorsan lejátszódik. Bár nem teljesen tisztázott, hogy hogyan érik el a rendezetlen fehérjék a gyors felgombolyodást, az feltehetően hozzájárul ehhez, hogy azok a szegmensek, amelyek rendeződnek a kötődés során, viszonylag kis méretűek és a felvett szerkezetek is egyszerű topológiájúak, alacsony kontaktusrenddel [74, 73]. Szerepe lehet a kötődést követő gyors rendeződésben az elektrosztatikus kölcsönhatásoknak, amelyek nemcsak a megfelelő szerkezetű ütközési komplex létrejöttét segítik elő, hanem megakadályozzák az ütközési komplex felbomlását is, amivel elősegítik az ütközési komplex natív komplexszé való átalakulását [58].

Ezekben a vizsgálatokban azonban csak az egyik kötőpartner volt rendezetlen, a másik rendezett szerkezetű volt. Az általánosabb esetre, ahol mindkét lánc rendezetlen – rendezett átmeneten esik át a kötődés során, Csermely és munkatársai 2010-ben a konformációkiválasztás modelljének kiterjesztésével egy általános modell kidolgozására tettek kísérletet [75]. Modelljük szerint ahogy a kötőpartnerek közelednek egymás felé, úgy változik mindkét fél szabadenergiafelcsíné a partner hatására. Eközben ha az egyik lánc bizonyos konformációba jut, egy konformációkiválasztási lépés során egy szorosabb komplex jön létre, ahol a kötődés hatására újabb konformációs átrendeződések zajlanak le mindkét láncban (3. ábra).



3. ábra. A kiterjesztett konformációkiválasztás sémája. A mechanizmus magyarázatát lásd a szövegben [75]

Sokkal kevesebb vizsgálatot végeztek olyan komplexekre, ahol mindkét partner rendezetlen állapotból rendezett állapotba megy át. A CBP NCBD domén és az ACTR komplex kialakulását vizsgálták mind NMR-el [76] mind molekuladinamikai szimuláció segítségével [74, 58]. Míg az ACTR szabad állapotban szinte semmilyen másodlagos szerkezettel nem rendelkezik, addig az



NCBD magas másodlagos szerkezet-tartalmú olvadt gombóc állapotban van. Mindkét molekula három hélixet tartalmaz a komplexben. A komplex kialakulásában a kezdeti lépés az NCBD egy rendezetlen szakaszának aspecifikus kötődése, de a későbbiekben az előképzett másodlagos szerkezetek is szerepet kapnak a kötődés során [74], ami megfelel a Csermely által javasolt kiterjesztett konformációkiválasztásnak [75]. Hasonló mechanizmus szerint zajlik az NRSF – mSin3 PAH domén komplexének kialakulása is [77].

## Két- és háromállapotú dimerek

A fehérjék a funkciójukat leggyakrabban valamilyen komplex részeként fejtik ki. A komplexeket a fehérjéken kívül egyéb biológiai makromolekulák, pl. RNS, és egyéb szerves és szervetlen anyagok alkotják. Vizsgálatunk szempontjából most azok a komplexek érdekesek, amelyeknek fehérjék az alkotórészei. Ezekben belül is most a két fehérjeláncból kialakuló dimerekre fókuszálunk.

Korábban a vizsgálatok szinte kizárólag azokra a fehérjékre irányultak, amelyek jól meghatározott háromdimenziós térszerkezettel rendelkeznek abban az állapotban, amelyikben a funkciójukat ellátják. Nyilvánvaló volt tehát a feltételezés, hogy két ilyen molekula úgy képez komplexet, hogy a két, már felgombolyodott lánc kapcsolódik össze dimerré. A specificitást a felszín komplementer, ill. nem komplementer volta biztosítja. Az ilyen komplexek kialakulása során három termodinamikai állapot különíthető el: (i) mindkét lánc denaturált állapotban van, nincs köztük kontaktus (ii) mindkét lánc a natív konformációban van, nincs köztük kontaktus (iii) és a natív dimer. A dimerképződés kinetikája a



sémával írható le, ahol  $D$  és  $N$  rendre az egyes láncok denaturált és natív konformációi, míg  $N_2$  a natív dimer. Mivel itt három állapot különíthető el, ezért ezeket a dimereket *háromállapotú dimereknek* nevezik.

Később egyensúlyi denaturációs kísérletek [78] és NMR [76] segítségével során kimutatták, hogy bizonyos dimerek esetén a láncok felgombolyodása és asszociációja kapcsolt. Ezeknek a dimereknek a kialakulása a



séma szerint megy végbe. Az ilyen séma szerint kialakuló dimereket *kétállapotú dimereknek* nevezik [78]. A tanulmány szerzői szerint az dönti el, hogy egy adott dimer melyik séma szerint alakul ki, hogy viszonyul egymáshoz a kötési és a felgombolyodási szabadenergia.

A kétállapotú dimerek kialakulása lényeges hasonlóságot mutat az egyedi láncok kétállapotú felgombolyodásához, míg a háromállapotú dimereké nem. A kétállapotú felgombolyodás-kötődés hasonlít az instabil építőelemek asszociációjának és stabilizálódásának folyamatára a monomer felgombolyodása során [12].

A két- és háromállapotú dimerek jellegzetes különbségeket mutatnak a kötőfelszíneik tulajdonságaiban is. A kötőfelszín szerkezeti motívumait vizsgálva azt találták, hogy míg a kétállapotú dimerek kötőfelszíneinek szerkezeti motívumai hasonlítanak a rendezett monomerek belsejében található szerkezeti motívumokhoz, addig a háromállapotú dimerek esetében inkább a különbségek dominálnak [79]. A kötőfelszín méretében is különbség van: míg a háromállapotú dimerek kötőfelszíne relatíve kicsi a teljes felszínhez képest, addig a kétállapotú dimerek

nagyméretű interfésszel rendelkeznek [80]. A háromállapotú dimerek kötőfelszínei kevésbé hidrofób karakterűek, mint a kétállapotú dimerek kötőfelszínei, annak ellenére, hogy a rendezett fehérjék szekvenciája a rendezetlenekéhez viszonyítva kevésbé hidrophil jellegű [72, 81].

Xu és munkatársai kompaktságanalízissel vizsgáltak kísérletes adatok alapján két- és háromállapotúként, valamint doméncseréltként azonosított homodimereket. A vizsgálat lényege, hogy egy szimmetrikus homodimer egyik láncát egy bizonyos aminosavnál képzeletben elvágják, a másik láncot az ennek megfelelő aminosavnál szintén és a két szabad véget megfelelően – tehát az egyik lánc N-terminálisát a másik lánc C-terminálisával – összekötik. Az így kapott, szintén szimmetrikus dimerre kiszámítják a kompaktsági együttható értékét (minél alacsonyabb ez az érték, annál kompaktabb a dimer). Ezt minden pozícióban történő vágásra kiszámítják és így kapnak egy kompaktsági profilt. A doméncserélt és a kétállapotú homodimerek esetében négy, míg a háromállapotú dimerek esetén három típust különítettek el. Egy elméleti fejtegetést is adnak abból a feltevésből kiindulva, hogy a dimerkialakulás kinetikája az adott komplex evolúciós kialakulásának mechanizmusát tükrözi [82].

Papoian és Wolynes a "Random Energy Model" általánosítását alkalmazták a felgombolyodás és kötődés kapcsoltságának leírására. Egy felgombolyodási és egy kötődési paraméter függvényében vizsgálták a szabadenergiát. Mindkét paramétertől függő entrópiatag kapcsolja össze a két folyamatot (ti. a felgombolyodást és a kötődést). A szabadenergia-kifejezésben szereplő egyéb paraméterek értékeit (pl. energiáris a natív állapot és a denaturált sokaság átlagenergiája között, stb.) ötszáz fehérje-fehérje komplex statisztikus elemzése alapján számították. (A denaturált állapotsokaságot a szekvencia permutálásával állították elő.) Azt találták, hogy legalább a komplexek tizenöt százalékában az egyik kötőpartner szabad állapotban instabil. Egy kiválasztott fehérjeláncra, ill. a teljes adatbázisra ábrázolták a szabadenergiát a felgombolyodási és kötődési paraméter függvényében és azt találták, hogy a kettő jellegzetesen különbözik (de az ismereteknek megfelelően). A hidrofób és hidrophil kontaktusokat vizsgálva mind a láncon belül, mind az interfész régióban, azt találták, hogy a többivel ellentétben az interfész hidrophil kontaktusai annál kedvezőbb energiájúak, minél kevésbé alakult ki az interfész. Ezt a tapasztalatnak ellentmondó eredményt víz-közvetített kontaktusok bevezetésével remélik orvosolni [83].

## Egyszerűsített fehérjemodellek

A „termodinamikai hipotézis” kimondja, hogy a fehérjék natív állapota termodinamikailag meghatározott [2]. Ha egy fehérje natív állapotát akarjuk vizsgálni, akkor ismernünk kell a szerkezetét. A *termodinamikai hipotézis* következményeként a natív állapot megtalálása egy optimalizációs probléma a fehérje–oldószer rendszer állapotterében. Ez a probléma azonban *NP-nehéz* bármely két- vagy háromdimenziós modellben [84, 85], ami azt eredményezi, hogy a keresési idő nagyon gyorsan növekszik a rendszer méretének növekedésével.

Az összes atomot figyelembe vevő fehérjemodellek legalacsonyabb energiájú állapotának megtalálása a jelenlegi számítógépes kapacitással még kisméretű fehérjék esetén is megoldhatatlan feladat. Még rosszabb a helyzet, ha nem egyszerűen a legalacsonyabb energiájú állapot megtalálása, hanem a rendszer teljes termodinamikai leírása a cél. Különböző egyszerűsítések tehetők azonban, amelyek lehetővé teszik, hogy valamilyen hozzávetőleges képet kapjunk a natív állapot szerkezetéről, a felgombolyodás folyamatáról, vagy a rendszer egy hozzávetőleges termodinamikai leírását adjuk.

Csökkenthetjük a számítások során alkalmazott modell szerkezetének felbontását pl. azáltal, hogy nem vesszük külön-külön figyelembe a modell összes atomját, hanem bizonyos atomo-

kat egy nagyobb egységbe vonunk össze és a számításokat már ezeken a nagyobb egységeken végezzük. Ha a megválaszolni kívánt kérdés olyan jellegű, akár teljesen el is tekinthetünk az oldalláncok explicit modellezésétől [86, 87]. Más esetekben egy adott oldallánc összes atomját egy nagyobb méretű atom reprezentálja [88].

Egy másik lehetőség a keresés felgyorsítására az állapottér méretének csökkentése. Egy gyakran alkalmazott módszer az állapottér diszkretizálása, azaz annak a feltételezése, hogy a modell egyes elemei csak bizonyos előre meghatározott pozíciókat foglalhatnak el a topográfiai térben. Ezek az ún. *rácsmodellek* [86, 87].

A fehérjéket alkotó aminosavakat szokás a fizikai-kémiai tulajdonságaik alapján csoportosítani (pl. hidrofób, poláros, negatívan töltött, pozitívan töltött). Ha az azonos tulajdonságú aminosavakat azonosaknak tekintjük, akkor ezáltal a szekvenciatér csökkenését érhetjük el. Ez ugyan nem gyorsítja meg egy bizonyos szekvencia alapállapotának megtalálását, de ha valamilyen általános kérdés megválaszolása a cél, akkor akár a teljes *szekvenciatér* is feltérképezhetővé válik.

Az egyes állapotok energiáját leíró *energiafüggvény* is egyszerűsíthető. Lehetséges pl. az, hogy ne az összes kontaktust vegyük figyelembe az adott állapot energiájának kiszámítása során, hanem csak például azokat, ahol a kölcsönható partnerek egymástól legfeljebb bizonyos távolságra helyezkednek el, vagy csak bizonyos aminosav-típusok közötti kontaktusokat. Az ún. *Gō-modellekben* [66] csak azokat a kontaktusokat vesszük figyelembe, amelyek a natív állapotban is jelen vannak. Az egyes kölcsönhatások energiáját leíró függvény is egyszerűsíthető, pl. egy folytonos függvény helyett valamilyen lépcsős függvény bevezetésével; ezt teszi a *DMD diszkrét molekuladinamika* eljárás [89, 90].

## Hálózatmodellek

Komplex, sok szabadsági fokú rendszerek termodinamikai vizsgálatának gyakori eszköze a szabadenergia-felszín vizsgálata. A szabadenergia-felszín valamilyen előre definiált rendparaméter vagy reakciókoordináta függvényében ábrázolt szabadenergia. A szabadenergia-felszín vizsgálatán alapuló módszerek egyik hátránya, hogy a kapott eredmények nagymértékben függhetnek a reakciókoordináta megválasztásától. Még relatíve jól megválasztott reakciókoordináták esetén is az esetleges nagymértékű dimenziószám-csökkenés következtében előfordul, hogy azonos pozícióba kerülnek olyan állapotok, amelyek között jelentős különbségek vannak, és ami még fontosabb, hogy amelyek közötti átalakulás sokkal lassabb, mint esetleg a reakciókoordináta mentén távol eső állapotok között.

Egy másik lehetőség az állapothálózat vizsgálata. Ha a vizsgált rendszer állapottere diszkrét, akkor közvetlenül felépíthető egy állapothálózat úgy, hogy a hálózatot reprezentáló gráf (ezután állapothálózaton mindig ezt a gráfot értem) csúcsai a mikroállapotok, és a csúcsokat összekötő élek a lehetséges átmenetek az állapotok között. Folytonos állapotterű rendszerek vizsgálatakor, pl. hagyományos molekuladinamikai szimulációk során, az állapottér diszkretizálásával szintén felépíthető egy állapothálózat [91]. Azután ennek az állapothálózatnak a statisztikai vizsgálata segítségével kinetikailag jellemezhető a rendszer, azonosíthatóak a metastabilis állapotok [92], a fontos átmenetek, és meghatározhatóak ezeknek az átmeneteknek a kinetikai paraméterei, pl. a sebességi együttható [93, 94].

*Metastabilis állapoton* az állapottér olyan részhalmazát értjük, amelyre igaz, hogy a rendszer átlagosan jóval több időt tölt az adott állapotban, mint amennyi idő ahhoz szükséges, hogy az állapoton belül az állapotra vonatkozó valószínűségeloszlás megfeleljen az egyensúlyi eloszlásnak. Legyen  $X_t$  diszkrét idejű, diszkrét állapotterű homogén Markov-lánc,  $S$  véges halmaz

a rendszer állapottere és legyen

$$\tau_I^x \equiv \inf \{t > 0 : X_t \in I | X_0 = x\} \quad (6)$$

tetszőleges  $I \subset S$ -re, tehát tetszőleges  $x \in S$  állapotból vesszük tetszőleges  $I \subset S$  részhalmaz bármely elemébe való jutás idejét, és ezek közül a legkisebbet választjuk ki.

Legyen  $M \subset S$ . Ha  $M$ -re teljesül a következő két feltétel pozitív véges  $a$  és  $b$ ,  $a \ll b$  konstansokra:

(i) minden  $z \in S$ -re

$$p(\tau_M^z \leq \tau_z^z) \geq b \quad (7)$$

(ii) bármely  $x \neq y$ ,  $x \in M$ ,  $y \in M$ -re

$$p(\tau_y^x < \tau_x^x) \leq a, \quad (8)$$

akkor  $M$ -re azt mondjuk, hogy *metastabilis pontok* halmaza. Az (i) feltétel annak a valószínűségét fejezi ki, hogy a rendszernek legalább annyi ideig tart visszajutni a  $z$  állapotba, mint eljutni valamelyik metastabilis pontba. A (ii) feltétel pedig azt a valószínűséget fejezi ki, hogy a rendszer gyorsabban eljut valamelyik másik metastabilis pontba, minthogy visszajusson  $x$ -be. Nyilvánvalóan a (ii)-ben kifejezett valószínűség metastabilis pontokra nagyon kicsi.

Tetszőleges  $x \in M$  metastabilis pont *lokális medencéje*

$$A(x) \equiv \left\{ z \in S : p(\tau_x^z = \tau_M^z) = \sup_{y \in M} p(\tau_y^z = \tau_M^z) \right\}, \quad (9)$$

tehát azon mikroállapotok halmaza, amelyekből a legnagyobb valószínűséggel a metastabilis pontok közül  $x$ -be jut el leggyorsabban a rendszer. Egy metastabilis pont és a metastabilis pont lokális medencéje együttesen alkotnak egy metastabilis állapotot.

Ha ismerjük a teljes állapothálózatot, akkor módunkban áll kiszámítani az egyes átmenetek idejének várható értékét. Ekkor  $M \subset S$  metastabilis pontok halmaza, ha teljesíti az alábbi feltételt:

$$\frac{\inf_{x \in M} \mathbb{E}\tau_M^x}{\sup_{y \notin M} \mathbb{E}\tau_M^y} \gg 1, \quad (10)$$

ahol  $\mathbb{E}\tau$  a  $\tau$  valószínűségi változó várható értékét jelöli.

Az  $x \in M$  metastabilis pontok vonzási medencéje megadható az átmenetek idejének várható értékeivel is [95]:

$$A(x) = \left\{ z \in S : \mathbb{E}\tau_x^z = \inf_{y \in M} \mathbb{E}\tau_y^z \right\} \quad (11)$$

Ez tehát azt jelenti, hogy tetszőleges mikroállapotból a metastabilis pontok valamelyikébe átlagosan sokkal gyorsabban eljut a rendszer, mint egyik metastabilis pontból a másikba. A metastabilis állapotok megfelelnek az egyes metastabilis pontok lokális medencéinek. Minden metastabilis pont tekinthető egy metastabilis állapot reprezentatív mikroállapotának [96].

Átmenetek sebességei mikroállapotok tetszőleges sokaságai között számíthatóak az *átmenetiút-vonal-elmélet* [93] segítségével. Az átmenetiút-vonal-elmélet alkalmazása során az összes lehetséges tetszőleges hosszúságú trajektória közül kiválasztjuk azokat, amelyek egy tetszőlegesen definiált  $R \subset S$  halmazból egy  $P \subset S$ ,  $P \cap R = \emptyset$  halmazba vezetnek úgy, hogy a trajektória,

ha egyszer elhagyta  $R$ -t, akkor már nem tér vissza oda. Az ezeken a trajektóriákon átfolyó fluxusok összege adja a reakció sebességét. A módszer arra is lehetőséget nyújt, hogy megvizsgáljuk a reaktív útvonalak egymáshoz viszonyított fontosságát, azaz hogy az átmenetek mekkora része megy egy adott útvonalon keresztül [93, 94].

# Célkitűzések

A kapcsolt felgombolyodás és kötődés leírására hagyományosan alkalmazott fogalomrendszerből kiindulva kíséreltem meg két rendezetlen lánc kölcsönös kapcsolt felgombolyodás és kötődésének leírását. Meg kívántam vizsgálni egzakt modellek és dinamikus, hálózat alapú modellek alkalmazhatóságát a jelenség vizsgálatában. Célom volt olyan egzakt, hálózat alapú modell kidolgozása, amelynek segítségével már olyan hosszúságú láncok vizsgálata is lehetővé válik, ahol hidrofób mag kialakulására van mód.

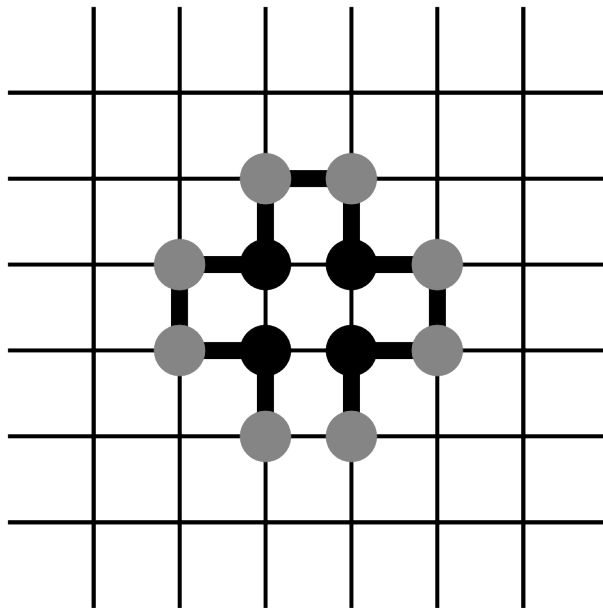
Munkám során az alábbi kérdésekre kerestem a választ:

- hogyan csökkenthető a két flexibilis láncból álló rendszer állapottere annyira, hogy teljes egészében fölépíthető legyen a rendszer állapotshálózata
- a kötődéshez kapcsolt konformációváltozás leírására használt fogalmak (indukált illeszkedés, konformációkiválasztás) hogyan alkalmazhatók olyan rendszerekben, ahol mindkét kötőpartner jelentős flexibilitással bír, ill. milyen mechanizmusok szerint mehet végbe rendezetlen fehérjék homodimerképzése
- az egyes mechanizmusoknak mekkora szerepe van a dimerképződés során
- az a szimmetria, amely a két lánc szekvenciaazonosságából származik, megjelenik-e magában a dimerképződési folyamatban
- mennyire megfelelőek a hagyományosan használt kinetikai sémák komplex rendszerek kinetikájának leírására .

# Módszerek

## A HP modell

Munkám során a fehérjeláncokat a Lau és Dill által bevezetett ún. *HP (hidrofób poláros) négyzet-rácsmodell* segítségével modelleztem [87]. A modell egy négyzetrácson elhelyezett önelkerülő bolyongás, ami úgy képzelhető el, mint egy madzagra felfűzött gyöngysor, ahol a gyöngyök (a továbbiakban *lánctmonomerek*) egymástól azonos távolságra helyezkednek el, amely távolság megfelel a négyzetrács rácspontjai közötti távolságnak. Minden ilyen lánctmonomer a négyzetrács egy rácspontján helyezkedik el; egy rácsponton egyszerre csak egy lánctmonomer lehet (4. ábra). A lánctmonomerek megfeleltethetők az egyes aminosavaknak, de akár aminosavnál nagyobb szerkezeti egységeknek is.



4. ábra. Egy *HP* lánc négyzetrácson elhelyezve.

A szekvenciátér kételemű: a láncokat hidrofób és poláros lánctmonomerek alkotják. Az egyes állapotok energiáinak meghatározására több energiafüggvényt is alkalmaztam, ezekben azonban

közös volt, hogy csak két hidrofób láncmonomer közötti kölcsönhatás járult hozzá effektív módon az állapot energiájához.

A HP modell bevezető cikkében [87] egy szomszédság alapú energiafüggvényt definiáltak. Monomer esetében egy adott  $\Gamma$  konformáció energiája az energiafüggvény szerint [87]:

$$E(\Gamma) = \varepsilon \sum_{i,j>i+1} \delta_{i,j} \Delta_{i,j}, \quad (12)$$

ahol  $\varepsilon < 0$  a kölcsönhatási energia,

$$\delta_{i,j} = \begin{cases} 1 & \text{ha } d_{i,j} = 1 \\ 0 & \text{egyébként} \end{cases}, \quad (13)$$

és

$$\Delta_{i,j} = \begin{cases} 1 & \text{ha } s_i = \text{H és } s_j = \text{H} \\ 0 & \text{egyébként} \end{cases}, \quad (14)$$

ahol  $d_{i,j}$  az  $i$ -dik és a  $j$ -dik *láncmonomer*<sup>1</sup> euklideszi távolsága és  $s = (s_1, s_2 \dots s_l)$  a szekvencia.

Tetszőleges,  $n$  számú monomerből álló rendszer adott  $\Gamma$  állapotának energiája pedig:

$$E(\Gamma) = \varepsilon \left( \sum_{k=1}^n \sum_{j_k > i_k + 1} \delta_{i_k, j_k} \Delta_{i_k, j_k} + \sum_{k, m \neq k} \sum_{j_m} \delta_{i_k, j_m} \Delta_{i_k, j_m} \right), \quad (15)$$

ahol  $i_k, j_k, j_m$  rendre a  $k$ -adik lánc  $i$ -edik, a  $k$ -adik lánc  $j$ -edik és az  $m$ -edik lánc  $j$ -edik láncmonomerét jelöli.

## A HPN modell

Ha szeretnénk figyelembe venni a töltött aminosavak hatását is, a HP modell már elégtelennek bizonyul. Azokban a vizsgálatokban, ahol a töltések szerepét is vizsgálni kívántuk, az ún. HPN modellt használtuk, ahol a H, a P, és az N rendre hidrofób, pozitívan töltött és negatívan töltött aminosavakat reprezentálnak [97]. Egy állapot energiáját úgy határozzuk meg, hogy az összes kontaktus energiáját összegezzük. Az egyes kontaktusokhoz pedig aszerint rendelünk energiát, hogy az milyen típusú láncmonomerek között jön létre,  $\varepsilon_{\text{HH}} = -1$ ,  $\varepsilon_{\text{PN}} = -0.75$ ,  $\varepsilon_{\text{PP}} = \varepsilon_{\text{NN}} = 0.75$  és  $\varepsilon_{\text{PH}} = \varepsilon_{\text{NH}} = 0$  szerint.

## A pull moves mozgáskészlet

Bizonyos mintavételezési eljárások (pl. Markov-lánc Monte Carlo módszerek (lásd később)) szükségessé teszik az állapotok közötti átmenetek definiálását. Egy adott modellben ezeknek az átmeneteknek a halmazát *mozgáskészletnek* nevezzük.

Lesh és munkatársai [98] 2003-ban publikálták az általuk *pull moves*nak nevezett mozgáskészletet [98, 99]. A mozgáskészlet mozgásait az 5. ábra mutatja.

Jelölje  $x_i$ ,  $i = 1, 2 \dots l$  azt a rácspontot a négyzet rácson, amit az  $i$ -edik láncmonomer foglal el. Jelöljön  $L$  egy olyan rácspontot, ami szomszédos  $x_{i+1}$ -gyel és szabad, azaz nem foglalja el

<sup>1</sup>*láncmonomernek* nevezem a modell lánc egy rácspontot elfoglaló egységét, megkülönböztetendő a *monomertől*, ami egyetlen láncot jelöl



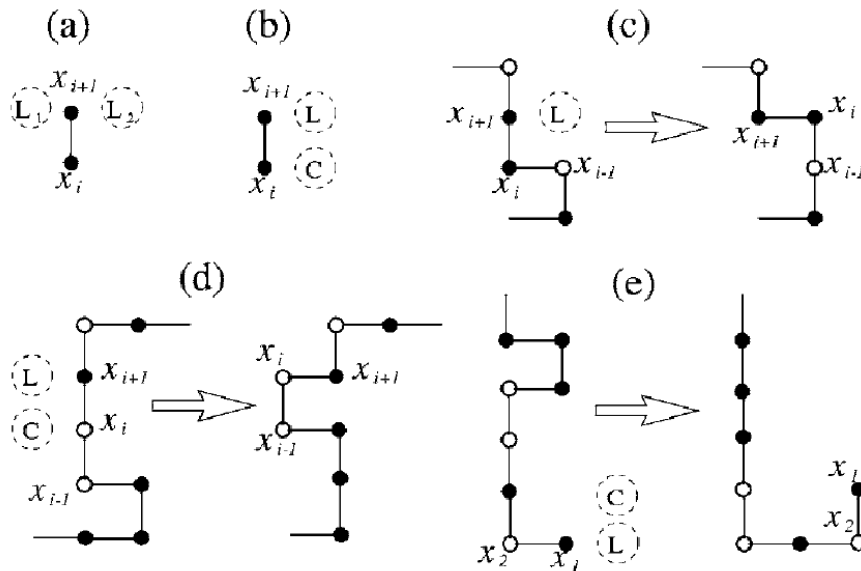
egyetlen láncmonomer sem. Jelölje  $C$  annak a négyzetnek a negyedik csúcsát, amelynek a másik három csúcsa  $x_i$ ,  $x_{i+1}$  és  $L$  (5. b) ábra). Ha  $C = x_{i-1}$ , akkor az  $i$ -edik monomert egyszerűen az  $L$  pozícióba helyezzük, azaz  $x'_i = L$  (5. c) ábra), ahol  $x'_i$  jelöli az  $i$ -edik monomer pozícióját a mozgás eredményeként létrejövő konformációban, és ezzel egy új érvényes konformációt kapunk. Ha  $C$  szabad, akkor  $x'_i = L$  és  $x'_{i-1} = C$ . Ha az így kapott konformáció a lánc egy érvényes konformációja, a mozgás befejeződött. Ha a kapott konformáció nem érvényes konformáció, akkor addig folytatjuk a  $x'_{i-k} = x_{i-k+2}$ ,  $k = 1, 2, \dots$  műveletet, amíg vagy egy érvényes konformációt kapunk, vagy elérjük a lánc végét (5. d) ábra).

A mozgáskészlet reverzibilitása érdekében a szerzők egy végmozgást is bevezettek. Itt most jelölje  $L$  a lánc valamelyik végén elhelyezkedő láncmonomer pozíciójával (ezt itt most jelölje az egyszerűség kedvéért  $x_1$ , de az elmondottak ugyanúgy érvényesek maradnak a lánc másik végére) szomszédos szabad rácspontot,  $C$  pedig egy – az  $L$  pozíciójával szomszédos – szintén szabad rácspontot úgy, hogy  $C \neq x_1$ . Végezzük el a  $x'_1 = L$  és  $x'_2 = C$  áthelyezéseket. Ha az így kapott konformáció érvényes konformáció, a mozgás befejeződött. Ha a kapott konformáció nem érvényes konformáció, akkor addig folytatjuk a  $x'_k = x_{k-2}$   $k = 3, 4, \dots$  áthelyezéseket, amíg vagy egy érvényes konformációt kapunk, vagy elérjük a lánc végét (5. e) ábra).

A mozgáskészlet teljes, tehát bármely konformációból bármely másikba el lehet jutni kizárólag a mozgáskészlet által definiált mozgások segítségével, tehát

$$\forall (\Gamma_i \in \Gamma, \Gamma_j \in \Gamma) : \exists (m = m_1 \circ m_2 \circ \dots \circ m_n : m_1, m_2, \dots, m_n \in \mathfrak{M}) : (m(\Gamma_i) = \Gamma_j) \quad (16)$$

ahol  $\mathfrak{M} = \{m_i\}$  a mozgáskészlet,  $\Gamma$  a lehetséges konformációk halmaza és  $\circ$  az elemi mozgások kompozícióját jelenti, tehát azt, hogy a következő mozgást a megelőző mozgás által létrehozott konformációra alkalmazzuk; lokális, azaz olyan mozgásokat tartalmaz, amelyek a lehető legkevesebb láncmonomer áthelyezésével járnak és azok sem kerülnek távolra az eredeti pozíciójuktól [98].



5. ábra. A pull moves mozgáskészlet [99].

Az eredeti cikk állításával szemben azonban a  $\mathfrak{M}$  mozgáskészlet nem teljesen reverzibilis, azaz nem teljesül a

$$\forall(\Gamma_i, \Gamma_j : \Gamma_i, \Gamma_j \in \Gamma) : \exists(m_k \in \mathfrak{M} : m_k(\Gamma_i) = \Gamma_j) \rightarrow \exists(m_l \in \mathfrak{M} : m_l(\Gamma_j) = \Gamma_i) \quad (17)$$

reverzibilitás. Egyszerűen reverzibilissé tehető azonban, ha a végmozgások közül elhagyjuk azokat a mozgásokat, ahol a  $C$  pozíció szomszédos az  $x_2$  pozícióval [100] (a magyarázatot és a bizonyítást lásd később).

## Az állapotér feltérképezése

Egy rendszer termodinamikai leírásához szükséges az *állapotsűrűség-függvény* ismerete. Az  $\Omega(E)$  állapot-sűrűség-függvény megadja, hogy folytonos energia esetén hány mikroállapot energiája esik egy kicsiny,  $E + \delta E$  intervallumba, diszkrét energiaszintek esetén pedig azt, hogy hány mikroállapot rendelkezik éppen  $E$  energiával.

Az energiákból és a hozzájuk tartozó állapot-sűrűségekből kiszámítható a  $Z$  kanonikus állapotösszeg, ami diszkrét esetre

$$Z = \sum_i e^{-E_i/k_B T} = \sum_{E_i} \Omega(E_i) e^{-E_i/k_B T}. \quad (18)$$

A kanonikus állapotösszegekből pedig már minden fontos termodinamikai függvény (pl. szabadenergia, entrópia, belső energia stb.) értéke kiszámítható.

Ahhoz, hogy meghatározzuk az állapot-sűrűség-függvényt, szükség van az állapotér feltérképezésére. Ha van rá lehetőség, akkor enumeráció segítségével a teljes állapotér feltérképezhető. Ha azonban az állapotér mérete nem teszi lehetővé az enumerációt, valamilyen mintavételi eljárás alkalmazása válik szükségessé.

## Enumerációs vizsgálatok

Az enumerációs vizsgálatnál az állapotteret teljes mértékben feltérképeztem, azaz kiszámítottam az összes lehetséges állapot energiáját. Enumeráció segítségével megtalálható a rendszer alapállapota, valamint egzaktul megkapható az állapot-sűrűség.

Az enumerációs vizsgálatoknál a fő nehézséget a modell tulajdonságaiból következő szimmetriák kezelése okozza. A szimmetriák megfelelő kezelése esetén azonban az állapotér csökkenése érhető el.

## Mintavételi eljárások

A biológiai makromolekulák állapotterének feltérképezése során alkalmazott mintavételi eljárások alapvetően két csoportba sorolhatók. Mindkét csoport eljárásainak közös jellemzője, hogy az egyes állapotok között átmeneteket definiál és ezeknek az átmeneteknek a segítségével igyekszik az állapotternek minél nagyobb részét bejárni, törekedve arra, hogy bizonyos paraméterek (pl. energia) mintabeli eloszlása minél jobban megközelítse a tényleges állapotok közötti eloszlást.

Ha a  $t$  időpontbeli állapot egyértelműen meghatározza a  $t + \delta t$  (illetve diszkrét idejű modellek esetében  $t + 1$ ) időpontbeli állapotot, akkor determinisztikus eljárásról beszélünk. Ebbe

a csoportba tartoznak a hagyományosan használt molekuladinamikai szimulációk [101], vagy a DMD [89, 90].

Ha a  $t$  időpontbeli állapot csak azt határozza meg, hogy a  $t + \delta t$  ( $t + 1$ ) időpontban milyen valószínűséggel találjuk a rendszert az egyes állapotokban, akkor sztochasztikus mintavételezési eljárásról beszélhetünk. A sztochasztikus mintavételezési eljárások legfontosabb (és általam egyedülként alkalmazott) csoportja a Monte Carlo-módszerek csoportja.

### Monte Carlo-módszerek

A Monte Carlo-módszerek sztochasztikus mintavételi és optimalizációs eljárások, amelyeknek célja egy meghatározott mennyiség megfelelő valószínűségi eloszlásával jellemezhető mintasokaság előállítása. A Monte Carlo-módszerek egy fontos osztálya, az ún. „rejection sampling”-módszerek a következő elv alapján működnek: legyen  $\xi$  és  $\eta$  két, egyazon  $H$  eseménytéren értelmezett, diszkrét (folytonos) valószínűségi változó. A  $\xi$  valószínűségi változó tömegfüggvénye (sűrűségfüggvénye) legyen  $f(x)$ , míg az  $\eta$  valószínűségi változóé  $g(x)$ , amelyekre egy megfelelően választott  $c$  konstansra érvényes a

$$f(x) \leq c \cdot g(x) \quad (19)$$

összefüggés.  $g(x)$  egy könnyen mintavételezhető eloszlás, nevezzük *segédeloszlásnak*,  $f(x)$  pedig az a nehezen mintavételezhető eloszlás, amelynek megfelelő mintasokaságot kívánunk előállítani; nevezzük ezt *báziseloszlásnak*.

Válasszuk ki véletlenszerűen a  $H$  egy tetszőleges  $y$  elemét, és generáljunk egy  $u$  véletlen számot a  $[0, c \cdot g(y)]$  intervallumon egyenletes eloszlásból. Ha  $u \leq f(y)$ , akkor  $y$ -t elfogadjuk, és belekerül a mintába, ellenkező esetben elutasítjuk [102].

Ez az eljárás azt eredményezi, hogy a mintavételezés során kapott minta eloszlása megfelel az  $f(x)$  eloszlásnak.

**Markov-lánc Monte Carlo-módszerek** Legyen  $S$  véges halmaz egy eseménytér. Definiáljunk ekkor egy  $G(V, E)$  súlyozott gráfot, amelyre  $V = S$  és  $E \subseteq (S \times S)$ . Tételezzük fel, hogy a gráf összefüggő. Végezzünk véletlen bolyongást a  $G$  gráfon és legyen az  $(i, j) \in E$  él  $w_{ij}$  súlya annak a valószínűsége, hogy az  $i$ -edik csúcsból a következő lépésben a  $j$ -edik csúcsba jutunk. Az egy csúcsból kiinduló élek súlyainak összege értelemszerűen 1.

Ekkor a  $G$  súlyozott gráfot egy véges állapotterű, diszkrét idejű *Markov-modellnek* nevezzük, a véletlen bolyongást magát pedig véges állapotterű, diszkrét idejű *Markov-láncnak*.

Most tekintsünk egy olyan Monte Carlo mintavételezést, ahol a segédeloszlás mintavételezése a Markov-modellen végrehajtott véletlen bolyongás eredménye. Ekkor, ha minden  $i, j \in S$ ,  $i \rightarrow j$ ,  $(i, j) \in E$  átmenet valószínűségére érvényes a

$$p_{ij} = \min \left( 1, \frac{f(\xi(i))}{f(\xi(j))} \right) \quad (20)$$

összefüggés, akkor a mintasokaság eloszlása éppen az  $f(x)$  báziseloszlás lesz [103].

**Metropolis – Hastings Markov-lánc Monte Carlo-módszer** A Metropolis Markov-lánc Monte Carlo módszer feltételezi, hogy az *a priori* valószínűségek (azaz, hogy a rendszer milyen valószínűséggel kísérli meg az adott átmenetet) azonosak az ellentétes irányú átmenetekre.

Azonban ez a feltétel nem minden Markov-láncrea érvényes. A Metropolis-algoritmus általánosítása arra az esetre, amikor az *a priori* valószínűségek nem feltétlenül egyenlőek, a Metropolis–Hastings-algoritmus [104].

Állandó hőmérsékletű és nyomású fizikai rendszerek állapotainak eloszlása a

$$p_i = e^{-E_i/k_B T} \quad (21)$$

kanonikus vagy más néven Boltzmann-eloszlást követi, ahol  $T$  az abszolút hőmérséklet,  $k_B$  pedig a Boltzmann-állandó. Ha tehát állandó hőmérsékletű és nyomású fizikai rendszer állapotainak mintavételezése a cél, akkor a kanonikus eloszlás lesz tehát a Monte Carlo mintavételezés báziseloszlása.

Ha tehát a megfelelő átmenetek valószínűségeit a

$$p_{ij} = \min \left( 1, \frac{P_{ji}^{ap}}{P_{ij}^{ap}} \cdot e^{-(E_j - E_i)/k_B T} \right). \quad (22)$$

*Metropolis–Hastings-kritérium* adja meg [103, 104], a kapott mintában az állapotok eloszlása meg fog felelni a Boltzmann-eloszlásnak.

**Wang–Landau-mintavételezés** Míg a Metropolis–Hastings Markov-lánc Monte Carlo-módszer kanonikus sokaság előállítására alkalmas, addig a Wang–Landau mintavételezés [105, 106] célja közvetlenül az állapotsűrűség-függvény meghatározása.

Jelölje  $\Omega(E)$  az állapotsűrűség-függvényt. Ha egy tetszőleges  $i \rightarrow j$  átmenetre az átmenet valószínűsége arányos az  $\Omega(E_i)/\Omega(E_j)$  hányadossal, akkor az egyes energiaszintek eloszlása egyenletes lesz. A módszer az egyes energiákhoz tartozó állapotsűrűségek iteratív megváltoztatásával, az egymást követő ciklusokban egyre pontosabban közelíti a tényleges értékeket. Az egyes ciklusok akkor érnek véget, ha a  $H(E)$  energia-hisztogram kellően sima.

Az eljárás kezdetekor  $\Omega(E) = 1$  és  $H(E) = 0$  minden  $E$ -re. Definiálunk egy  $f > 1$  módosító faktort, aminek a kezdeti értékét leggyakrabban  $f = e$ -nek választják. Ezután egy Markov-lánc Monte Carlo szimulációt végzünk

$$p_{ij} = \min \left( 1, \frac{\Omega(E_i)}{\Omega(E_j)} \right) \quad (23)$$

átmeneti valószínűségekkel. Minden lépésben elvégezzük az

$$\Omega(E_i) \rightarrow \Omega(E_i) \cdot f \text{ és a} \quad (24)$$

$$H(E_i) \rightarrow H(E_i) + 1 \quad (25)$$

műveleteket. A szimulációt addig végezzük, amíg a

$$H(E) \geq w \cdot \langle H(E) \rangle \text{ minden } E\text{-re} \quad (26)$$

feltétel nem teljesül, ahol  $w$  adja meg, hogy mennyire kell simának lenni a kapott hisztogramnak a továbblépéshez.  $w$  értéke esetünkben 0,8 volt.

Ha a (26) feltétel teljesül, a szimulációt leállítjuk, visszaállítjuk a hisztogramot  $H(E) = 0$ -ra és a módosító faktor értékét egy olyan függvény szerint változtatjuk meg, ami csökkenti

a módosító faktor értékét és a függvény ismételt alkalmazásával értéke egyre jobban megközelíti egyet. A gyakorlatban általában a négyzetgyökfüggvény használatos, azaz

$$f \rightarrow \sqrt{f}. \quad (27)$$

A szimulációkat addig ismétljük, amíg a módosító faktor értéke kisebb nem lesz egy előre definiált értéknél pl.  $f_{vgs} \leq e^{10^{-8}}$ . Minden ciklusban az állapotsűrűség-függvény értékeit  $\ln f$ -el arányos pontossággal tudjuk meghatározni. Tehát  $f_{vgs}$  megfelelő megválasztásával elvileg tetszőleges pontosság érhető el [105].

## Állapothálózat

A Markov-lánc Monte Carlo-szimulációk során az átmeneteket és azok valószínűségeit definiáló Markov-modellt nem ismerjük teljes egészében, arról csak egy mintavételi folyamat során szerzünk információt. Rövidebb láncokra azonban előállítható a teljes gráf, ami azután részleteiben – pl. hálózatelméleti módszerekkel – vizsgálható. Ezt a gráfot nevezzük állapotahálózatnak. Az állapotahálózat – mint súlyozott gráf – csúcsai a rendszer egyes állapotai, a köztük lévő élek pedig a közöttük lehetséges átmenetek, amit itt valamelyik Monte Carlo mozgáskészlet mozgásai definiálnak. Egy adott irányított élhez rendelt súly az adott átmenet valószínűsége.

Az állapotahálózat mátrix formában is felírható. Ha az állapotok száma  $n$ , akkor felírható egy  $\mathbf{T}^{n \times n}$  mátrix, amelynek  $T_{ij}$  eleme az átmeneti hálózat  $i$ -edik állapotából a  $j$ -edik állapotába mutató él súlya, azaz annak a valószínűsége, hogy a következő lépésben a rendszer a  $j$ -edik állapotba jut, feltéve, hogy jelen pillanatban az  $i$ -edik állapotban van. A  $\mathbf{T}$  mátrixot az állapotahálózat *átmeneti mátrixának* nevezzük.

## Perron-klaszter Klaszterelemzés

Egy sok szabadságfokkal rendelkező rendszernek, amilyenek a biológiai makromolekulák is, rengeteg mikroállapota valósulhat meg. Tekintsük most a mikroállapotok halmazának két diszjunkt részhalmazát. Ha a két részhalmazon belül a mikroállapotok közötti átmenetek lényegesen gyorsabbak, mint a részhalmazok közötti átmenetek, akkor a két részhalmazt a rendszer metastabilis állapotainak nevezzük. Ez azt jelenti, hogy ha a rendszer egy metastabilis állapotba kerül, akkor az ott egyhuzamban eltöltött idő várható értéke nagy. Általában a metastabilis állapotok alacsony szabadenergiájú állapotok és bizonyos feltételek mellett megfeleltethetők a hagyományos szabadenergia-felszínek szabadenergia-gödreinek.

Matematikai értelemben a metastabilitás megfelel az ún. Markov-tulajdonságnak, azaz annak a valószínűsége, hogy bizonyos idő elteltével a rendszer egy bizonyos metastabilis állapotban lesz, csak a jelenlegi állapottól függ, és nem függ attól, hogy a jelenlegi állapotába honnan jutott a rendszer.

Egy Markov-modell metastabilis állapotainak elkülönítésére szolgál a *Perron-klaszter Klaszterelemzés* („Perron-Cluster Cluster Analysis”) [92]. Tekintsünk most egy ergodikus rendszert leíró, véges állapotterű Markov-modellt. Ekkor a  $\mathbf{T}^{n \times n}$  átmeneti mátrix, ahol  $n$  a Markov-modell állapotainak száma,  $\lambda_i$  sajátértékei mind valóságos lesznek és  $\lambda_1 = 1$ , valamint  $|\lambda_i| < 1$  minden  $i > 1$ -re. Minden egyes sajátérték megfelel egy  $t_i = -\frac{\tau}{\ln \lambda_i}$  időskálájú mozgásnak [107]. Azt, hogy a megfelelő mozgás milyen állapotok között zajlik, a sajátértékhez tartozó jobboldali sajátvektor komponenseinek előjelelemzésével határozhatjuk meg [91].

Ha az első  $k$  számú sajátérték nagyságban jól elkülönül a többitől, akkor ezt a  $k$  sajátértéket Perron-klaszternek nevezzük, és ez azt jelzi, hogy az állapotok nagy része  $k$  számú metastabilis állapotba sorolható.

## Diszkonnektivitási gráf

Energiafelszínek topográfiájának feltárására alkalmas eszköz a diszkonnektivitási gráf. Osszuk föl a vizsgált rendszer állapotterét diszkrét, át nem fedő részterekre (részhalmazokra) úgy, hogy minden ilyen részter egy  $\alpha$  lokális minimumhoz tartozó  $R^{E_0}(\alpha)$  vonzási medencét (attraction basin) reprezentáljon, ahol minden  $s \in R^{E_0}(\alpha)$  állapot energiája kisebb egy bizonyos  $E_0$ -nál. Ezek az  $R^{E_0}(\alpha)$  vonzási medencék nagyobb méretű  $R^{E_i}(\alpha')$  szupermedencékbe egyesíthetők úgy, hogy minden  $s \in R^{E_i}(\alpha')$  állapot  $E(s)$  energiája kisebb legyen  $E_i > E_0$ -nál és  $R^{E_i}(\alpha')$  bármely két állapota legyen összekötve olyan útvonallal, ami magában a részterben vezet.  $\alpha'$  itt a legalacsonyabb energiájú lokális minimum azok között, melyek  $R^{E_i}(\alpha)$ -ben vannak, azaz

$$\alpha' = \min \{ \alpha : \alpha \in R^{E_0}(\alpha) \subset R^{E_i}(\alpha') \} \quad (28)$$

A vonzási medencéknek a hierarchikus összevonása addig ismételhető, amíg a teljes állapotteret egyetlen szupermedencébe nem egyesítettük.

Ábrázoljuk a vonzási medencék hierarchiáját gráf formájában úgy, hogy a gráf csúcsai legyenek a különböző energiaszinten definiált vonzási (szuper)medencék, és két ilyen csúcsot akkor kössön össze él, ha az egyik csúcs által reprezentált (szuper)medence a másik csúcs által reprezentált szupermedencének részhalmaza. Gyakorlati szempontból érdemes a gráfot úgy ábrázolni a kétdimenziós térben, hogy a gráf csúcsainak valamelyik dimenzió szerinti koordinátái megfeleljenek  $E_i$ -nek. Ekkor a gráfról leolvasható, hogy az egyes vonzási medencék közötti energiagátak milyen energiájúak és ezáltal következtetni lehet a rendszer dinamikájára [108].

## Átmenetiútvonal-elmélet

A reakciókinetikában a sebességi együttható elméleti származtatásának legszélesebb körben alkalmazott elmélete az átmenetiállapot-elmélet. Eszerint a reaktáns állapotból kiindulva a rendszernek keresztül kell mennie egy magas energiájú állapotban, amit átmeneti állapotnak nevezünk. Ha egyetlen reakciókoordináta mentén ábrázoljuk a rendszer szabadenergiáját, akkor két alacsony szabadenergiájú állapot között – ezek a reaktáns és a termék – a szabadenergia-görbének maximuma van, és az ehhez a maximumhelyhez tartozó állapot az átmeneti állapot. A reakció sebességét egyrészt a reakciókoordináta menti diffúzió sebessége, másrészt az átmeneti állapot és a reaktáns állapot szabadenergiájának különbsége – az aktiválási szabadenergia – határozza meg. Az elmélet feltételezése szerint ha a rendszer a reakciókoordináta mentén túljutott az átmeneti állapotban, akkor a reakció mindenképpen lezajlik.

Az elmélet alapvető feltételezése az, hogy a reaktáns és a termék állapot ún. metastabilis állapotok. A metastabilitás azt jelenti, hogy a mikroállapotok között a metastabilis állapotban belüli átmenetek jóval kisebb időskálán zajlanak le, mint a metastabilis állapotok közötti átmenetek. Matematikailag ez azt jelenti, hogy a rendszer egy Markov-modellként írható le, ahol a Markov-modell csúcsai a metastabilis állapotok, a köztük lévő átmenetek az élek, és az élek súlya arányos a sebességi együttható értékével.

Ha kettőnél több állapotot és a köztük lévő átmeneteket akarjuk figyelembe venni, akkor felírható a sebességi együtthatók segítségével egy differenciálegyenlet-rendszer, amelynek megoldásával megkaphatók az egyes metastabilis állapotok mennyiségeinek időgörbéi. Hangsúlyozni kell azonban, hogy ez a vizsgálat csak akkor írja le pontosan a rendszer dinamikáját, ha a modell állapotai valóban metastabilis állapotok.

Mit lehet tenni, ha a rendszernek nincsenek metastabilis állapotai illetve ha általunk önkényesen definiált állapotból induló átmeneteket vizsgálunk? Erre jelenthet megoldást az *Átmeneti útvonalelmélet* [93].

Az Átmeneti Útvonal Elmélet (a továbbiakban az angol elnevezésből származó TPT (Transition Path Theory) rövidítést fogom használni) segítségével kiszámítható, hogy az átmeneti hálózat élein átmenő fluxus mekkora hányada reaktív, azaz mekkora hányada éri el a termék állapotot.

A módszer első lépéseként definiáljuk tetszés szerint az állapotok halmazának két át nem fedő részhalmazát, ezek egyike lesz a reaktánshalmaz ( $A$ ), a másik a termékshalmaz ( $B$ ). Ezután az átmeneti hálózat élein a reaktánshalmazból a termékshalmazba irányuló reaktív fluxust vizsgáljuk. Ehhez a ( $q^+$ ), ún. *elköteleződési függvény* („committor function”) értékeinek kiszámítása szükséges. Az ún. *előrefelé elköteleződési függvény* („forward committor function”) értéke egy adott állapotban azt adja meg, hogy a rendszer milyen valószínűséggel jut el a termékshalmazba az adott állapotból anélkül, hogy meglátogatná a reaktánshalmazt. A ( $q^-$ ) *hátrafelé elköteleződési függvény* („backward committor function”) értéke pedig azt adja meg, hogy mi a valószínűsége, hogy a rendszer a reaktánshalmaz valamelyik állapotából érkezett a termékshalmaz érintése nélkül.

Ha  $T_{ij}$  az átmeneti mátrix  $i$ -edik sorának  $j$ -edik eleme, akkor az előrefelé elköteleződési függvény értéke a

$$q^+_i = \sum_j T_{ij} q^+_j \quad (29)$$

egyenletrendszer megoldásával határozható meg [94], feltéve, hogy

$$\begin{aligned} q^+_i &= 0, \text{ minden } i \in A\text{-ra} \\ q^+_i &= 1, \text{ minden } i \in B\text{-re.} \end{aligned}$$

Reverzibilis Markov-modellekre  $q^-_i = 1 - q^+_i$ , tehát a hátrafelé elköteleződési függvény értékeinek kiszámítására nincs szükség.

Az egyensúlyi valószínűségek és az előrefelé elköteleződési függvény értékeinek segítségével kiszámítható az egy élen átfolyó reaktív fluxus:

$$f_{ij}^{AB} = \begin{cases} \pi_i (1 - q^+_i) T_{ij} q^+_j & \text{ha } i \neq j \\ 0 & \text{egyébként,} \end{cases} \quad (30)$$

ahol  $\pi_i$  az  $i$  állapot egyensúlyi valószínűsége.

A reakció sebessége ( $v_{AB}$ ) ezek alapján

$$v_{AB} = \sum_{i \in A, j \in S} f_{ij}^{AB}, \quad (31)$$

ahol  $S$  a rendszer állapotainak halmaza.

## A vizsgált szekvenciák

Vizsgálataimba azokat a HP szekvenciákat vontam be, ahol a dimer alapállapota egyedi, azaz létezik egy olyan állapot, amelynek mindegyik másik állapotnál alacsonyabb az energiája. A monomer alapállapotára nem tettem ilyen kikötést.

Jelölje  $\{N_M\}$  a monomer alapállapotú konformációinak halmazát és  $N_D$  pedig azt a konformációt, amit a monomer a dimer alapállapotában fölvesz (itt most feltételeztem, hogy a dimer alapállapota szimmetrikus, azaz a két lánc konformációja megegyezik). Ha egy szekvenciára  $N_D \in \{N_M\}$ , az adott szekvenciát *derék* szekvenciának nevezem, ha  $N_D \notin \{N_M\}$ , a szekvencia *nem derék*.

A számítási kapacitás korlátozottsága miatt a legnagyobb lánchossz, amelyre még tudtam enumerációt végezni,  $l = 10$ . Vizsgálataimat azokon a  $4 \leq l \leq 10$  szekvenciákon végeztem, amelyekre a dimer alapállapota egyedi.

## A van't Hoff és a kalorimetrikus entalpia számítása

A kalorimetrikus kétállapotúság kritériuma szerint

$$\Delta H_{\text{vH}}/\Delta H_{\text{cal}} \approx 1, \quad (32)$$

ahol  $\Delta H_{\text{vH}}$  és  $\Delta H_{\text{cal}}$  rendre a van't Hoff és a kalorimetrikus entalpia. Mind a van't Hoff, mind a kalorimetrikus értéke kiszámítható az állapotosság-függvény ismeretében.

A van't Hoff entalpiát a

$$\Delta H_{\text{vH}} = \langle \Delta H \rangle_{\text{D}}^{T_f} = \langle H \rangle_{\text{D}}^{T_f} - H_{\text{N}} \quad (33)$$

egyenlet adja meg, ahol  $\langle \dots \rangle_{\text{D}}^T$  a denaturált állapothoz tartozó mikroállapotokra vett Boltzmann-átlag a  $T$  hőmérsékleten,  $H_{\text{N}}$  a natív állapot entalpiája és  $T_f$  a felgombolyodási hőmérséklet, amit a

$$p_{\text{N}}(T_f) = p_{\text{D}}(T_f) \quad (34)$$

definiál, ahol  $p_{\text{N}}(T_f)$  és  $p_{\text{D}}(T_f)$  rendre a natív és a denaturált állapot valószínűsége a  $T_f$  hőmérsékleten.

A felgombolyodási hőmérsékletet úgy számítottam, hogy a *szelőmódszer* segítségével megoldottam  $T$ -re a

$$q_{\text{N}} e^{-E_{\text{N}}/T} = \frac{1}{2} \sum_i q_i e^{-E_i/T} \quad (35)$$

egyenletet, ahol  $q_i$  és  $E_i$  rendre az  $i$ -edik állapot multiplicitása és energiája.

A kalorimetrikus entalpia értéke

$$\Delta H_{\text{cal}} = \langle \Delta H \rangle^{\infty} = \langle H \rangle^{\infty} - H_{\text{N}}, \quad (36)$$

ahol  $\langle \dots \rangle^T$  az összes mikroállapotra vett Boltzmann-átlag a  $T$  hőmérsékleten.

Kanonikus sokaság esetén végtelen hőmérsékleten minden mikroállapot egyformán valószínű. Az általam vizsgált modellben minden állapot egy mikroállapot-sokaságnak felel meg. A modellben definiált állapotokra számított átlag tehát végtelen hőmérsékleten

$$\langle H \rangle^{\infty} = \frac{\sum_i H_i q_i}{\sum_j q_j}, \quad (37)$$



ahol  $q_i$  az  $i$ -edik állapot multiplicitása. Mivel a modellben nincs nyomás definiálva, ezért egy mikroállapot entalpiáját egyenlőnek tekintem az energiájával.

## A használt programok

A legtöbb feladatot saját magam által írt programokkal oldottam meg. Az enumerációs számításokat végző, a Monte Carlo szimulációkat végző és az állapothálózatokat előállító programokat C nyelven írtam. Az adatok kiértékeléséhez használt egyszerűbb programokat Python nyelven. Kiterjedten használtam a Python numerikus számítások végzésére készült modulját, a Numpyt.

Az átmeneti hálózatok átmeneti mátrixai a legtöbb szekvencia esetében ritka mátrixok voltak. A ritka mátrixokon végzett számításokat (pl. mátrixok spektrális elemzése) a Python Scipy moduljának Scipy.sparse és Scipy.sparse.linalg moduljában definált osztályok és függvények segítségével végeztem.

A hálózatokon végzett számításokhoz a Networkx modult használtam, a grafikonok készítéséhez pedig a Matplotlib modul.

# Eredmények

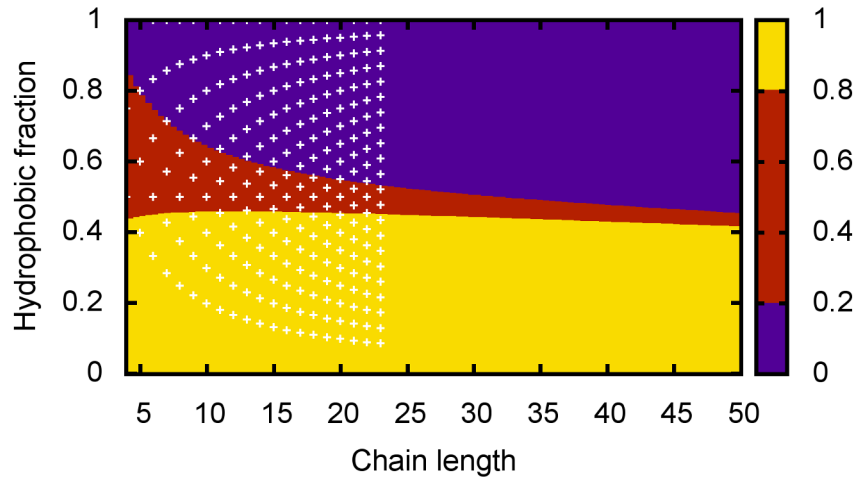
## Alkonyzóna fehérjék rendezettsége és rendezetlensége között

A rendezetlen fehérjék az aminosavösszetételük alapján jól megkülönböztethetők a rendezett fehérjéktől. Jó elkülönülést kapunk, ha a szekvenciák aminosavösszetételét az átlagos hidrofobicitással és az abszolút átlagos nettó töltéssel jellemezzük. Ha a vizsgált szekvenciákat az átlagos hidrofobicitás-abszolút átlagos nettó töltés síkon helyezzük el, a rendezett és a rendezetlen szekvenciák által elfoglalt részsíkok átfednek. Az átfedő zónát *alkonyzónának* nevezzük.

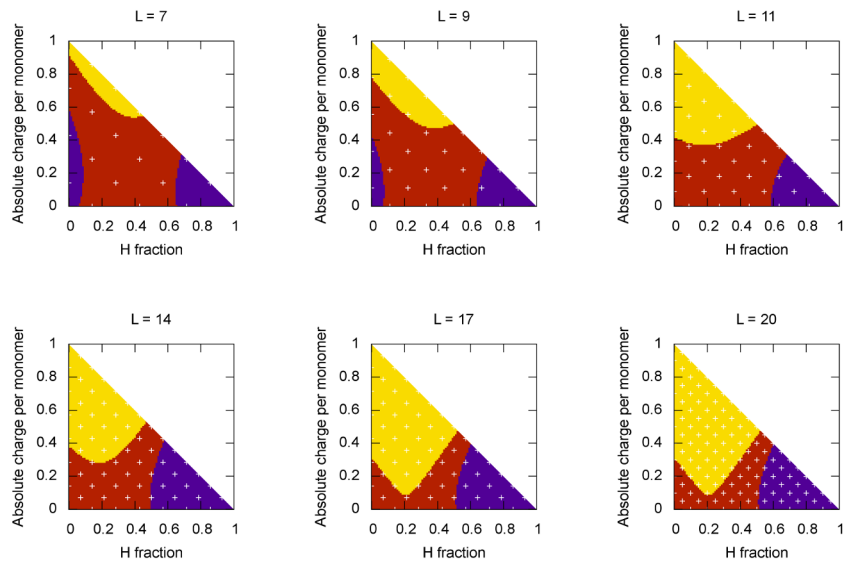
HP és HPN szekvenciákat vizsgáltunk abból a célból, hogy megvizsgáljuk, hogyan függ a lánchosszúságtól az alkonyzóna szélessége. HP szekvenciák esetében 4–23 lánconomer hosszúságig az összes lehetséges szekvenciát megvizsgáltuk, és minden szekvencia esetén enumerációval meghatároztuk az alapállapot energiáját. Harminc, negyven és ötven lánconomer hosszúságú lánccok esetén pedig mintavételeztük a szekvenciateret, és minden szekvencia esetén az alapállapot energiát Metropolis–Hastings Markov-lánc Monte Carlo módszer segítségével becsültük. HPN szekvenciák esetén 4–20 lánconomer hosszúságig enumerációval határoztuk meg az alapállapot energiákat, 14 lánconomer hosszúságig az összes lehetséges szekvenciára, hosszabb lánccokra pedig a szekvenciatér mintavételezésével kapott szekvenciákra. Szintén mintavételeztük a 30, 40 és 50 lánconomer hosszúságú lánccok szekvenciaterét, és a kapott szekvenciák alapállapot energiáit Monte Carlo szimuláció segítségével becsültük.

Azt a láncot tekintettük rendezetlennek, amelynek az egy lánconomerre eső alapállapot energiája nagyobb volt egy előre meghatározott értéknél, ami a mi esetünkben 0,3 volt. Mind a HP, mind a HPN modellek esetén egy lánchosszúságfüggő kontaktusenergiá alapján számítottuk az alapállapot energiáit. HP modellek esetében az átlagos hidrofobicitás, míg HPN modellek esetében az átlagos hidrofobicitás és az abszolút átlagos nettó töltés függvényében vizsgáltuk a rendezetlen szekvenciák részarányát, és megvizsgáltuk, hogy milyen széles az alkonyzóna, azaz az a tartomány, ahol a rendezetlen szekvenciák aránya 0,2 és 0,8 között van. A 6. és a 7. ábrák mutatják az eredményeket. Jól látszik, hogy az alkonyzóna szélessége a lánchosszúság növekedésével csökken. Lánchosszúságtól függő kontaktusenergia segítségével az is elérhető, hogy az alkonyzóna pozíciója ne változzon a HP és a HPN modellek esetében, ahogy azt valódi fehérjék esetében láthatjuk a 8. ábrán [45].

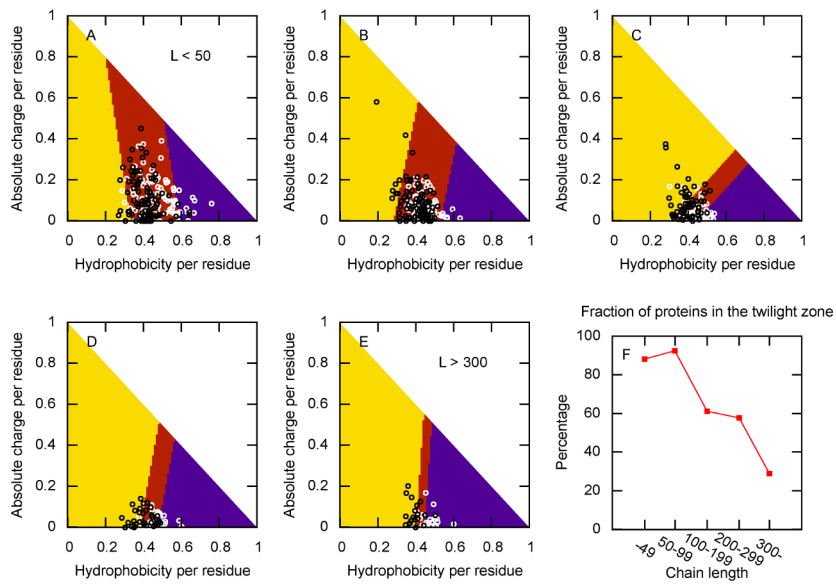
A dimerképződés szempontjából az alkonyzóna szélességének azért van jelentősége, mert nyilvánvaló, hogy rendezetlen peptidek rendezett homodimerjeinek képződésekor azonos aminosavösszetétel mellett a monomer rendezetlen, míg a dimer rendezett, tehát a rendezettséget a lánchosszúság határozza meg.



6. ábra. A rendezetlen szekvenciák aránya az összes HP szekvencia között az átlagos hidrofobicitás és a lánc hosszúság függvényében, lánc hosszfüggő kontaktusenergiával számolva. A pontokra az  $1/(1 + \exp(ax^2 + by^2 + cxy + dx + ey + f))$  alakú kétváltozós négyzetes logisztikus függvényt illesztettük (amit a színskála jelöl). Az adatpontok helyét fehér keresztek jelzik.



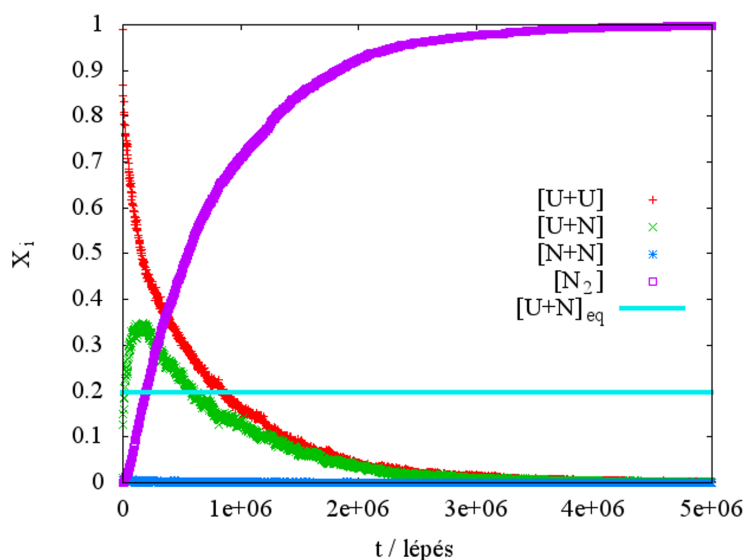
7. ábra. A rendezetlen szekvenciák aránya HPN szekvenciák között az átlagos hidrofobicitás és az abszolút átlagos nettó töltés függvényében, lánc hosszfüggő kontaktusenergiával számolva, különböző lánc hosszúságokra. A pontokra az  $1/(1 + \exp(ax^2 + by^2 + cxy + dx + ey + f))$  alakú függvényt illesztettük. A színskála megfelel a 6. ábrán láthatóknak.



8. ábra. (A-E) A rendezetlen szekvenciák részaránya valódi fehérjék között az átlagos hidrofobicitás és az abszolút átlagos nettó töltés függvényében különböző lánc hosszúság-tartományokban (-49, 50-99, 100-199, 200-299, 300-). A lánc hosszúság-tartományok rendre 194, 184, 126, 64 és 38 láncot tartalmaztak; fele-fele arányban rendezett és rendezetlen láncokat. Az adatokra  $1/(1 + \exp(ax + by + c))$  alakú kétváltozós logisztikus függvényt illesztettünk. A színskála megegyezik a 6. ábrán láthatóval. (F) Az alkonyzóna szélessége a különböző lánc hosszúság-tartományba eső láncok esetén.

## Monte Carlo-szimulációk

A kérdések megválaszolására először Markov-lánc Monte Carlo szimulációkat végeztem  $l = 4-10$  hosszúságú, egyedi dimer alapállapottal rendelkező láncokra. A szimulációkat a felgombolyodási hőmérsékleten végeztem, amit úgy definiáltam, mint az a hőmérséklet, ahol a natív állapot valószínűsége 0,5. A dimerek lehetséges állapotait a monomer láncok felgombolyodottsága alapján csoportosítottam. Figyelemmel kísértem a szimuláció során az egyes csoportok koncentrációit és azt összevettem az egyensúlyi értékekkel. Megfigyeltem, hogy derék szekvenciák esetén, tehát amikor a lánc dimer alapállapotbeli konformációja a monomernek is alapállapota, az az állapot, amelyben egy natív és egy nemnatív konformációjú lánc van jelen, az egyensúlyi értékéhez képest átmenetileg felhalmozódik (lásd 9. ábra). Ez arra utalt, hogy egy aszimmetrikus köztes állapot alakul ki a dimerképződés során, amelyben a két lánc közül az egyik már a natív konformációban van. Más köztes állapotok esetén ilyen feldúsulást nem tapasztaltam.



9. ábra. A PHPPHPPPH szekvenciájú lánc dimerképződésének kinetikai görbéi. Az  $U + U$  állapotban mindkét lánc denaturált állapotban van, az  $U + N$  állapotban az egyik lánc natív, a másik denaturált állapotban van, míg az  $N + N$  állapotban mindkét lánc a natív konformációban. Az  $N_2$  állapot a natív dimert jelöli. [...] a pillanatnyi koncentrációkat jelöli, míg [...]eq az egyensúlyi koncentrációt.

A szimulációk segítségével egy átfogó képet kaptam a dimerképződés folyamatáról, de nem voltak alkalmasak arra, hogy a folyamatok részletes leírását adjam a segítségükkel. Abból a célból, hogy a dimerképződés részletes mechanizmusát föltárjam, fölépítettem a rendszer teljes állapothálózatát, amit a szimuláció során csak mintavételeztem. A teljes hálózat felépítéséhez az összes állapotot és a köztük lévő átmeneteket is egyszerre kell tárolni a számítógép memóriájában. Ez azonban csak nagyon rövid láncokra lehetséges.

## A vizsgálatba bevont szekvenciák kiválasztása

A vizsgálat célja fehérje homodimerek képződési mechanizmusának vizsgálata volt, különös tekintettel arra, hogyan képeznek rendezetlen peptidok rendezett komplexet. Megkerestem tehát

azokat a HP szekvenciákat, amelyek dimer alapállapota egyedi, azaz – a modell tulajdonságai-  
ból eredő szimmetriáktól eltekintve – a legalacsonyabb energiaszinten egyetlen állapot található.  
Enumerációval  $l = 4-8$  lánchosszúságig meghatároztam minden egyes HP szekvencia esetén  
a dimer alapállapotát és annak energiáját, és ezek közül kiválogattam azokat, amelyek egye-  
di alapállapottal rendelkeznek. Ezek között a szekvenciák között vannak olyanok, amelyeknek  
monomer alapállapotban is egyedi az alapállapotuk, a többségre azonban a monomer alapállá-  
potának degeneráltsága jellemző. A szekvenciákat csoportosítottam aszerint is, hogy a láncok  
natív dimerbeli konformációja a monomernek is alapállapoti konformációja-e. Az 1. táblázat  
az így kiválasztott szekvenciákat mutatja, a monomer alapállapotok számával együtt. A táblá-  
zatban feltüntettem, hogy az adott szekvencia derék vagy nem derék, ill. azt is, ha az alapállapot  
aszimmetrikus. Aszimmetrikus alapállapotú szekvenciák esetében az egyik konformáció minden  
esetben a monomernek is alapállapota. A 10. ábra a megfelelő dimerek alapállapotait mutatja.

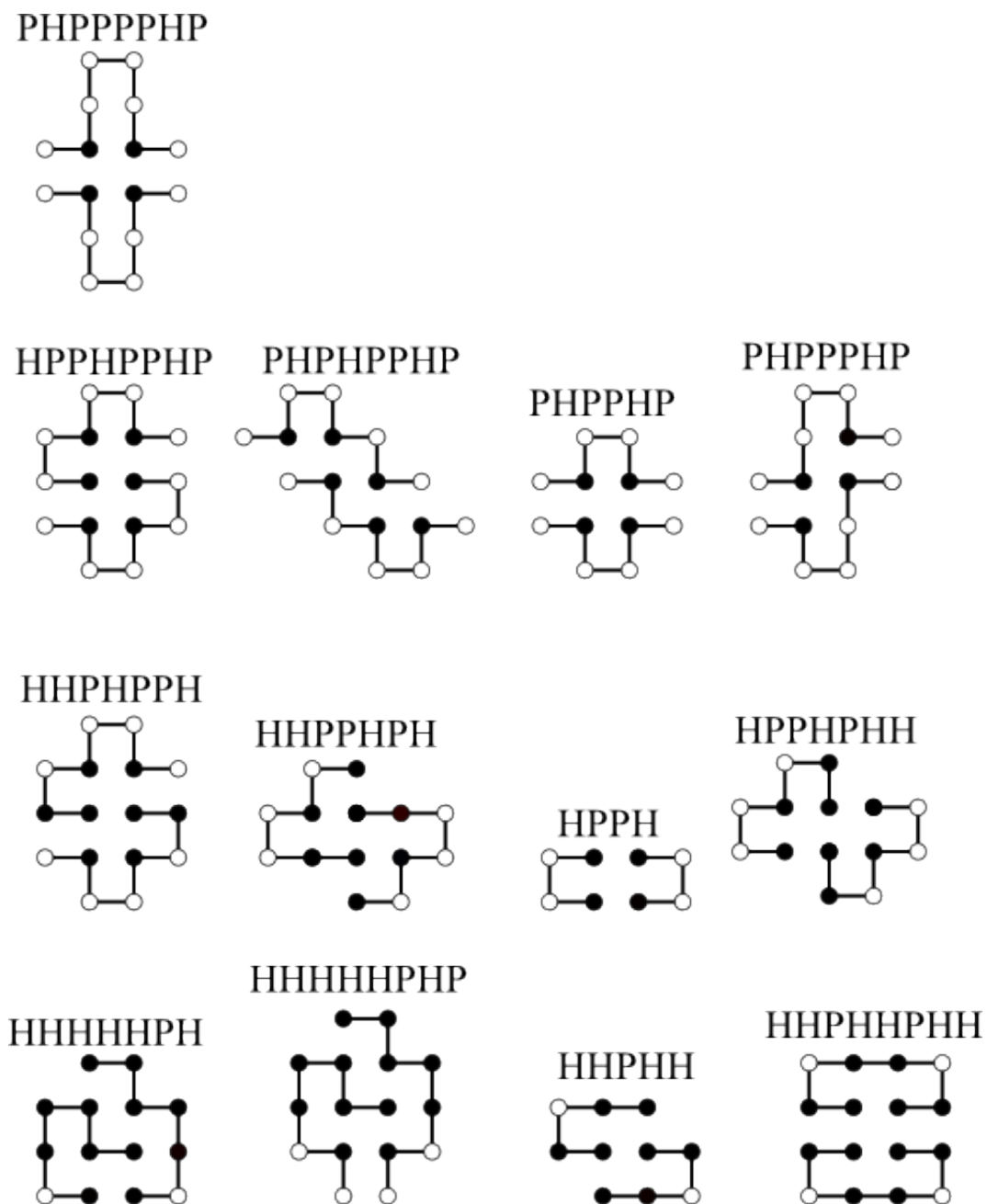
Szekvencia	A monomer alapállapotainak száma	Derék–nem derék–aszimmetrikus
HPPH	1	derék
HHPHH	1	derék
PHPPHP	3	derék
HPPHPPH	2	derék
HHHHHPH	3	aszimmetrikus
PHPPHP	5	derék
HHPHPPH	1	nem derék
PHPHPPH	2	nem derék
HHPHPPHP	3	derék
PHPPHP	5	derék
HPPHPPHP	2	derék
HHHHHPH	6	aszimmetrikus
HHPHPPH	3	derék

1. táblázat. Az egyedi dimer alapállapottal rendelkező,  $l = 4-10$  hosszúságú szekvenciák és a monomereik  
alapállapotainak száma.

## A megfelelő mozgáskészlet kiválasztása

A Markov-lánc Monte Carlo módszerekben szükséges definiálni az állapotok közötti átmenete-  
teket. A lehetséges átmeneteknek ezt a halmazát mozgáskészletnek nevezzük. Négyzetrácsmo-  
dellekre többféle mozgáskészletet definiáltak. Ilyen mozgáskészlet pl. az angol terminológiában  
„ $k$ -bead move”-nak nevezett mozgáskészlet, ahol egy előre meghatározott  $k$  számú lánconomer  
pozíciója változik meg egyetlen mozgás során [109], azonban ez a mozgáskészlet hosszú láncok-  
ra nem *ergodikus* [110]. Az ún. *könyökmozgás* („pivot move”) során egyetlen kiválasztott lán-  
conomer körül végzünk elforgatást [111, 112]. A könyökmozgás ergodikus, de hátránya, hogy  
kompakt konformációk esetén az elfogadási arány nagyon alacsony [111, 113]. Alapvetően el-  
térő módszer a kötés-átrendezés („bond-rebridging”) [114], amelyre ugyan magas az elfogadási  
arány, viszont nem ergodikus [113].

Lesh és munkatársai 2003-ban publikálták „pull moves” nevű Monte Carlo mozgáskészle-  
tüket [98]. A mozgáskészletet bevezető cikkben adnak egy bizonyítást a mozgáskészlet rever-  
zibilitására. A bizonyítás a következő. Ha a mozgás egy vagy két lánconomer pozíciójának



10. ábra. Az egyedi dimer alapállapottal rendelkező szekvenciák alapállapotai.

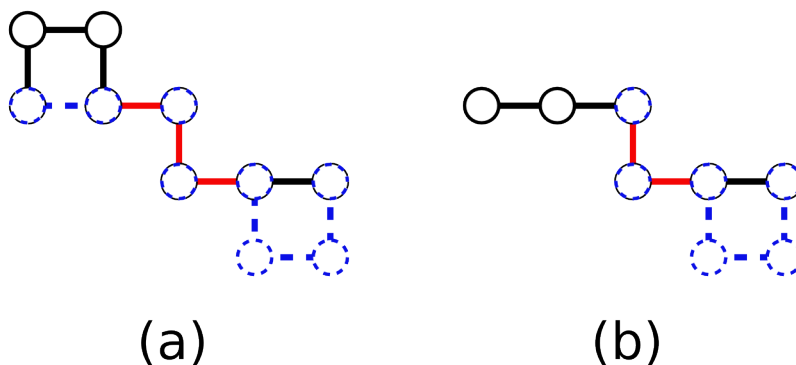
megváltozásával jár, a reverzibilitás triviális (11. ábra), ezért vizsgáljuk azokat az eseteket, ahol legalább három láncmonomer kerül új pozícióba.

Minden, legalább három láncmonomert érintő mozgás egy olyan huroknak a létrejöttével jár, ahol a lánc a négyzetrács egy celláját három oldalról körbeveszi. Vegyük most először azt az esetet, amikor a mozgás a láncvég elérése előtt véget ér. Ekkor egy ugyanolyan hurok megszűnik, mint ami a mozgás kezdetekor a lánc egy másik részén létrejött. A megszűnő és a képződő hurok közötti részen a négyzetrács ugyanazon pozíciói lesznek foglaltak a kiindulási és a végkonfor-



11. ábra. Egy és két láncmonomert érintő mozgások. Könnyen belátható, hogy mindkét mozgástípus reverzibilis.

mációban. Ha most azt a mozgást tekintjük, ami létrehozza azt a hurkot, amit a fenti mozgás megszüntetett, akkor, miközben a közbeeső részeket szintén változatlanul hagyja, a fenti mozgás által létrehozott hurkot szünteti meg és ezzel a mozgás azon a ponton be is fejeződik (12. (a) ábra). Ha az eredeti mozgás elérte a lánc végét (legyen most ez az  $n$ -dik láncmonomer), akkor az  $n$ -dik és az  $n - 1$ -dik láncmonomerekkel kezdődő végmozgás visszaállítja az eredeti konformációt (12. (b) ábra) [98].

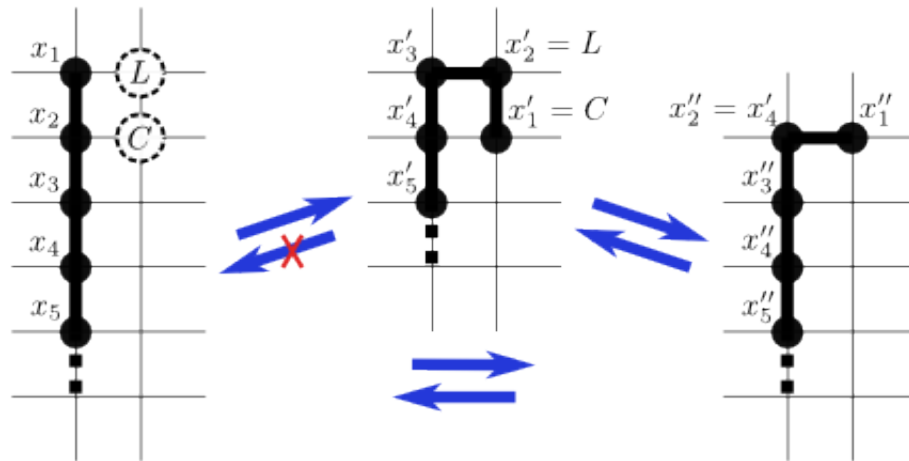


12. ábra. Legalább három láncmonomer pozíciójának megváltozásával járó mozgások. Fekete színnel van ábrázolva a kiindulási konformáció, kék szaggatottal a végkonformáció. Azok a pozíciók, amiket a piros élek kötnék össze, mind a kiindulási, mind a végkonformációban foglaltak. (a) Mozgás, ami nem változtatja meg a láncvégi láncmonomer pozícióját. (b) Mozgás, ami a láncvégi láncmonomer pozíciójának megváltoztatásával jár.

A cikkben adott bizonyítás azonban nem tér ki a végmozgások reverzibilitására. Tekintsük most az 1 és 2 láncmonomereket érintő olyan végmozgást, ahol a  $C$  pozíció szomszédos az  $x_2$  pozícióval, tehát az  $x_1, x_2, L$  és  $C$  pozíciók egy négyzet négy sarkán helyezkednek el. Ekkor kezdjük el az ellentétes irányú mozgást, úgy, hogy az eredeti mozgás lépéseinek fordítottját hajtjuk végre fordított sorrendben. Mivel az  $x'_4$  pozíció szomszédos az  $x'_1$  pozícióval, mikor elérjük a mozgás során az  $x'_4$  pozíciót, egy érvényes konformációt kapunk, tehát a mozgás leáll. Ebből kifolyólag azonban az 1 sorszámú láncmonomer nem kerül vissza az eredeti pozíciójába, tehát a végmozgás nem volt reverzibilis (13. ábra) [100].

Most vizsgáljuk meg milyen következménye van ennek az irreverzibilitásnak. Markov-lánc Monte Carlo szimulációkban tetszőleges  $i \rightarrow j$  átmenet valószínűsége arányos az *a priori* való-





13. ábra. Az irreverzibilis végmozgás [100].

színűségek hányadosával, azaz

$$p_{ij} \propto \frac{p_{ji}^{ap}}{p_{ij}^{ap}}. \quad (38)$$

Az átmenet *a priori* valószínűsége:

$$p_{ij}^{ap} = \frac{n_{ij}}{n_i}, \quad (39)$$

ahol  $n_{ij}$  az  $i$  állapotból kiinduló,  $j$  állapotot létrehozó lehetséges mozgások száma,  $n_i$  pedig az összes, az  $i$  csúcsból kiinduló lehetséges mozgás száma. Ha egy mozgás irreverzibilis úgy, hogy  $n_{ji} = 0$  (ez a helyzet az irreverzibilis végmozgások esetében), akkor  $p_{ij}$  nulla lesz és a mozgást soha nem fogadjuk el a szimuláció során. Ekkor a számítás ugyan korrekt eredményre vezet, de a szükségesnél több CPU időt kell fölhasználni.

Az algoritmus implementációja során azonban a (38) egyenlet egy egyszerűsített alakja is használható. Ha feltételezzük, hogy a mozgáskészlet reverzibilis, ahogy azt a [115, 116, 114] tanulmányokban teszik, akkor

$$n_{ij} = n_{ji}, \quad (40)$$

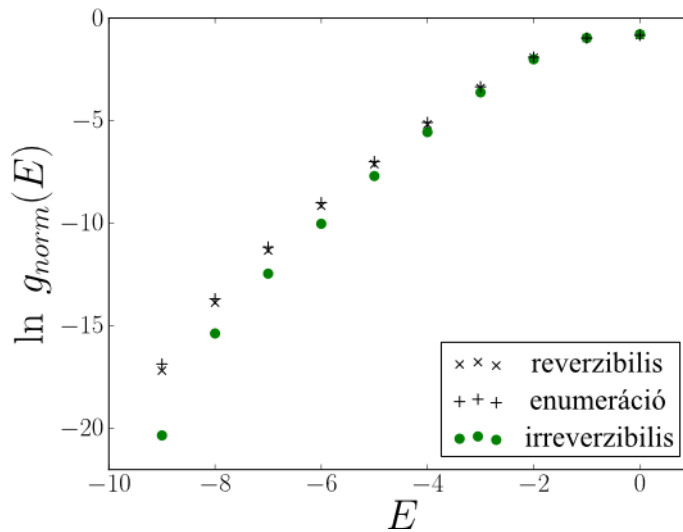
és (38)-ből

$$p_{ij} \propto \frac{n_i}{n_j} \quad (41)$$

lesz.

Ha azonban a mozgáskészlet irreverzibilis, akkor (40) nem lesz érvényes és az irreverzibilis mozgások valószínűsége nullától eltérő lesz, ami nem megfelelő mintavételezéshez vezet.

Wang–Landau szimuláció segítségével becsültük egy  $l = 20$  hosszúságú kétdimenziós HP monomer állapotsűrűség-függvényét abból a célból, hogy megvizsgáljuk mi a hatása, ha figyelmen kívül hagyjuk az irreverzibilitást. Elvégeztük a szimulációt úgy, hogy megengedtük, ill. nem engedték meg az irreverzibilis mozgásokat, és ezeket összevetettük az enumerációval kapott egzakt értékekkel. Az eredményeket a 14. ábra mutatja. Ahogy az ábra mutatja, hogy ha



14. ábra. A  $HPHPHHPHPHPHHPHHPH$  kétdimenziós 20mer [118] normált állapotsűrűség-függvény ( $g_{norm}(E)$ ) logaritmus 50 Wang–Landau szimuláció átlagából számolva reverzibilis ill. irreverzibilis mozgáskészletet alkalmazva, valamint egzakt enumerációból.

megengedjük az irreverzibilis mozgásokat, akkor a módszer szignifikánsan alulbecsüli a legtöbb állapotsűrűséget, különösen az alacsony energiák esetében. Hasonló, bár kevésbé kifejezett eredményt kaptunk három korábban már vizsgált szekvencia [116, 117, 118]: egy kétdimenziós 64mer, egy háromdimenziós 103mer és egy 18mer homopolimer vizsgálatokor.

## A legmegfelelőbb energiafüggvény kiválasztása

Ha egy rendszert statisztikus termodinamikai szempontból kívánunk jellemezni, elengedhetetlen ismerni a rendszer állapotainak energiáját. Egy olyan nagyon leegyszerűsített modellnél, mint a kétdimenziós HP négyzetrácsmodell, egyáltalán nem triviális, hogy milyen energiafüggvény segítségével jellemezzük az egyes állapotokat. A jó energiafüggvénynek meg kell felelnie néhány gyakorlati kritériumnak:

- a rendszer kellően nagyszámú energiaszinttel rendelkezzen, de lehetőleg kevesebb, mint a mikroállapotok száma
- az egyedi alapállapottal rendelkező szekvenciák száma minél nagyobb legyen
- legalább a szekvenciák egy része könnyen felgombolyodó legyen, ami kapcsolatba hozható az adott szekvencia (kalorimetrikus) kétállapotúságával. A kalorimetrikus kétállapotúság gyakran megfigyelhető kisméretű fehérjéknél [119].

Célunk az volt, hogy egy olyan energiafüggvényt definiáljunk, ami lehetőleg teljesíti a fenti összes kritériumot.

Minél nagyobb az energiaszintek száma, az energiafelszín annál simább lesz, és fordítva, kevés energiaszint kevésbé sima energiafelszínre hoz létre. Minél simább az energiafelszín, a felgombolyodás annál gyorsabban megy végbe, ami a valós fehérjék esetében biológiai kényszer.

Az egyedi alapállapot megfelel az Anfinsen-dogmának, azaz hogy egy fehérje denaturáció után mindig ugyanabba az állapotába tér vissza, mégpedig az állapotok közül a legalacsonyabb szabadenergiájúba. Rendezetlen fehérjék esetén is gyakran megfigyelték, hogy valamilyen partnerhez való kötődés során a szabad állapotban rendezetlen lánc a kötődés során rendezett szerkezetet vesz fel [25], tehát a komplex alapállapota egyedinek tekinthető akkor is, ha a komplexet alkotó láncok alapállapotai önmagukban nem is azok.

A [87]-ben definiált, szomszédság alapú energiafüggvénynek van néhány hátránya. Először is az energiaszintek száma kicsi. Gyakori, hogy akár egy tíz láncmonomer hosszú, egyedi alapállapotú szekvencia is mindössze négy energiaszinttel rendelkezik, miközben a monomer lehetséges konformációinak a száma 2034 (pl. PHPPHPPHHH). Ha a 4-8 láncmonomer közötti szekvenciákat vizsgáljuk, azoknak a szekvenciáknak a száma, ahol a dimer alapállapota egyedi, mindössze tizenegy (ha azokat a szekvenciákat egynek számítjuk, ahol a két lánc szekvenciája azonos, csak ellentétes irányban). A van't Hoff és a kalorimetrikus entalpia aránya még a Gőpotenciál [66] alkalmazása és a HP+ [120] energiafüggvény használata esetén is jóval 1 alatt marad [119].

Különböző energiafüggvények összehasonlító elemzését végeztem el monomer láncokon, hogy olyan energiafüggvényt találjak, amely lehetőleg kielégíti a viszonylag sok energiaszint, a sok egyedi alapállapot és a gyors felgombolyodás kritériumait.

### Távolságfüggő energiafüggvény

Első kísérletként a szomszédság alapú energiafüggvény helyett egy olyan távolságfüggő energiafüggvényt vezetünk be, ahol minden olyan H-H láncmonomerpár hozzájárul az összenergiához, amelyek a szekvencia mentén nem szomszédosak. Egy ilyen adott H-H láncmonomerpár energiahozzájárulása távolságuk valamilyen hatványával arányos. Monomer esetén tehát egy  $\Gamma$  konformáció energiája:

$$E(\Gamma) = \varepsilon \sum_{i,j>i+1} \frac{1}{(d_{i,j})^a} \Delta_{i,j}, \quad (42)$$

ahol  $\varepsilon < 0$  a kölcsönhatási energia, és

$$\Delta_{i,j} = \begin{cases} 1 & \text{ha } s_i = \text{H és } s_j = \text{H} \\ 0 & \text{egyébként} \end{cases}, \quad (43)$$

ahol  $d_{i,j}$  az  $i$ -edik és a  $j$ -edik láncmonomer euklideszi távolsága,  $a$  a távolságfüggés hatványkitevője, és  $s = (s_1, s_2, \dots, s_l)$  a szekvencia.

Tetszőleges  $n$  számú monomerből álló rendszer adott  $\Gamma$  állapotának energiája pedig:

$$E(\Gamma) = \varepsilon \left( \sum_{k=1}^n \sum_{i_k, j_k > i_k + 1} \frac{1}{(d_{i_k, j_k})^a} \Delta_{i_k, j_k} + \sum_{k,m} \sum_{i_k, j_m} \frac{1}{(d_{i_k, j_m})^a} \Delta_{i_k, j_m} \right), \quad (44)$$

ahol  $i_k, j_k, j_m$  rendre a  $k$ -adik lánc  $i$ -edik, a  $k$ -adik lánc  $j$ -edik és az  $m$ -edik lánc  $j$ -edik láncmonomerét jelöli.

### „Klaszter” energiafüggvény

Nevezzük H-H klaszternek azon hidrofób monomerek halmazát a négyzetrácson, amelynek minden eleme elérhető minden másiktól szomszédsági kontaktusokon keresztül (beleértve most

a lánckontaktust is). Ezután a klaszter – a láncc mentén nem szomszédos – tagjai között kiszámítjuk a kölcsönhatási energia értékét a távolságfüggő energiafüggvény szerint.

Ha elemezzük az energiafüggvényt, láthatjuk, hogy kéttagú klaszter esetén a klaszterhez tartozó energia megegyezik a szomszédság alapú energiafüggvénnyel kapott energiával. Ha a klaszter valamelyik tagjának szomszédságába egy harmadik hidrofób lánccmonomer kerül, akkor a szomszédság alapú energiafüggvénnyel kapott energiához hozzáadódik a nem szomszédos lánccmonomerek közötti távolságfüggő energiát. A távolságfüggő energiafüggvény esetében, ahogy két hidrofób lánccmonomer közeledik egymáshoz, a kölcsönhatási energia fokozatosan csökken. Az új energiafüggvényben ezek a tagok csak akkor adódnak a teljes állapot energiájához, ha az adott lánccmonomerek azonos klaszterbe kerülnek.

### Négyzetes diagonális energiafüggvény

Induljunk ki a távolságfüggő energiafüggvényből, de csak azokhoz a H-H párokhoz rendeljük nullától különböző energiát a fent definiált távolságfüggő energiafüggvény szerint, amelyeknek  $d$  euklideszi távolságára érvényes a

$$d^2 \leq 2 \quad (45)$$

összefüggés.

Miután az összes releváns H-H párra kapott energiákat összegezem, az összeg abszolút értékét négyzetre emelem, és a négyzet ellentettjét veszem.

### Az energiafüggvények összevetése

A távolságfüggő energiafüggvény nyilvánvaló előnye a szomszédság alapúval szemben az, hogy mind az energiaszintek, mind az egyedi alapállapottal rendelkező szekvenciák száma jóval nagyobb. Az energiafüggvény azonban additív, ezért nem várható tőle kooperatív viselkedés [119]. A vizsgálatok azt mutatták, hogy meglepő módon még a szomszédság alapú energiafüggvénynél is kevésbé mutatja a kétállapotú kooperativitást.

A szomszédság alapú és a távolságfüggő energiafüggvényt kombinálva definiáltuk a klaszter energiafüggvényt. A klaszter energiafüggvény sajátja, hogy minél nagyobb méretű egy klaszter, annál alacsonyabb (nagyobb negatív érték) az egy kontaktusra jutó kölcsönhatási energia. Ha például nyolc hidrofób lánccmonomer két, egyenként négy hidrofób lánccmonomerből álló klaszterbe csoportosul, a kölcsönhatási energia magasabb lesz, mintha ezt a két klasztert egyetlen, nyolc lánccmonomerből álló klaszterbe egyesítenénk. A várakozásainkkal ellentétben a klaszter energiafüggvény, bár bizonyos szekvenciák esetén növelte a  $\Delta H_{\text{vH}}/\Delta H_{\text{cal}}$  arányt, nem javította lényegesen a termodinamikai kétállapotúságot, annak ellenére, hogy jelentősen nőtt az állapotok energiái és a natív állapottól való távolságaik közötti korreláció (az eredményeket nem mutatjuk).

Igazán kétállapotú kooperativitást a négyzetes diagonális energiafüggvény bevezetésével értünk el. Ismert, hogy azok a heteropolimer-modellek mutatnak kétállapotú kooperativitást, ahol az alapállapot energiája és a denaturált állapot átlagos energiája közötti energiárés nagy a denaturált állapotok energiáinak szórásához viszonyítva, ami nem érhető el additív energiafüggvény használatával [119]. A négyzetes diagonális energiafüggvény szerint egy állapot energiájának abszolút értéke a lineárisnál gyorsabban növekszik a kontaktusok számának növekedésével. Az energiafüggvénnyel minden szekvenciára 1-hez közeli (vagy 1-et meghaladó)  $\Delta H_{\text{vH}}/\Delta H_{\text{cal}}$  arányt kaptunk. Mivel a vizsgált energiafüggvények közül egyedül a négyzetes diagonális energiafüggvényre jellemző a kalorimetrikus kétállapotúság, és mivel az energiafüggvény továbbra

Szekvencia	Klaszter	$\Delta H_{vH}/\Delta H_{cal}$		
		szomszédság alapú	Távolságfüggő	Négyzetes diagonális
HHPHH	0,71		0,82	1,01
HPPPH	0,68		0,74	1,01
HPPHH	0,68		0,74	1,01
HRPHHH	0,82	0,76	0,71	0,90
HPPPHH	0,62	0,82	0,76	1,00
HHPPPH	0,62	0,78	0,74	0,96
HHPHPH	0,82	0,78	0,71	0,90
HPPHHH	0,62	0,78	0,74	0,96
HRPPPH	0,88	0,78	0,76	0,96
HPPHPH	0,88	0,78	0,76	0,96
HHPHPHH	0,22		0,35	
HPPHPPH	0,49		0,51	0,89
HHHHHHH	0,20		0,20	
HRPHHHH	0,78	0,71	0,70	0,91
HHHPHPH	0,54	0,71	0,67	0,94
PHPHHH	0,54	0,71	0,67	0,94
PHPHPH	0,77	0,71	0,64	0,86
HHHPHHH	0,28		0,39	0,63
HRPPPHH	0,81	0,71	0,64	0,90
PHPHPH	0,84	0,71	0,70	0,93
HRPHHPH	0,77	0,71	0,64	0,86
HRPPHPH	0,84	0,71	0,70	0,93
HHHPHPH	0,78	0,71	0,70	0,91
HHPHPH	0,81	0,71	0,64	0,90

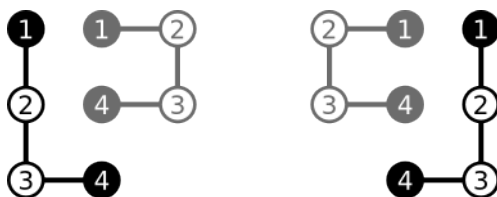
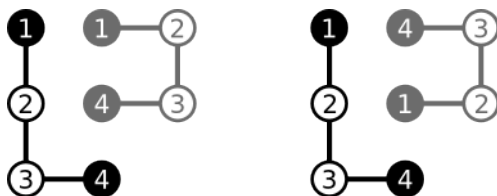
2. táblázat. *Van't Hoff*  $\Delta H_{vH}$  és kalorimetrikus  $\Delta H_{cal}$  entalpia hányadosa különböző energiafüggvények használata esetén egyedi alapállapotú,  $l = 5-8$  hosszúságú szekvenciák monomerjeire.

is rendelkezik azzal a tulajdonsággal, hogy nem tartalmaz az alapállapot szerkezetére vonatkozó információt, a további számolásokhoz a négyzetes diagonális energiafüggvényt használtam. A különböző energiafüggvényekre számított  $\Delta H_{vH}/\Delta H_{cal}$  értékeket a 2. táblázat mutatja.

## A négyzetrácsmodell szimmetriái

A HP modellben, annak jellegéből fakadóan bizonyos konformációk szimmetriaviszonyban állnak egymással. A legegyszerűbb ilyen viszony, a tükörképi viszony (15. ábra), azaz az egyik konformáció a másiknak tükörképi párja. Már ez az egy szimmetria is azt okozza, hogy az alapállapot degenerált lesz, hiszen az egymással tükörképi viszonyban álló konformációk azonos energiájúak lesznek. Ha az adott lánc szekvenciája palindrom, akkor a két különböző irányú lánclefutás szintén megkülönböztethetetlen egymástól (16. ábra). Dimerek esetében a láncok felcserélhetősége újabb egymással egyenértékű konformációpárokat definiál (17. ábra).

A szimmetriák kezelését – egy erre a célra írt program segítségével – a következőképpen oldottam meg: generáltam a monomer láncre az összes lehetséges önelkerülő konformációt. Mindkét lánc kezdőpontját a  $2 \cdot l$  oldalhosszúságú, négyzet alakú doboz összes lehetséges pozíciójába helyeztem, valamint mindkét lánc kezdővektorát mind a négy lehetséges irányba forgattam. Az eljárással megkaptam a rendszer összes lehetséges állapotát. Az összes állapotot a 15–17. áb-

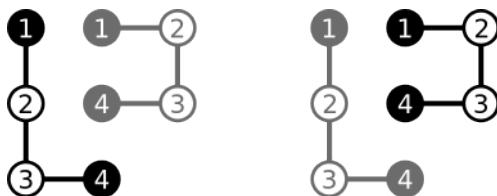
15. ábra. *HP állapotok tükörképi szimmetriája.*16. ábra. *A szekvencia palindromitásából származó szimmetria.*

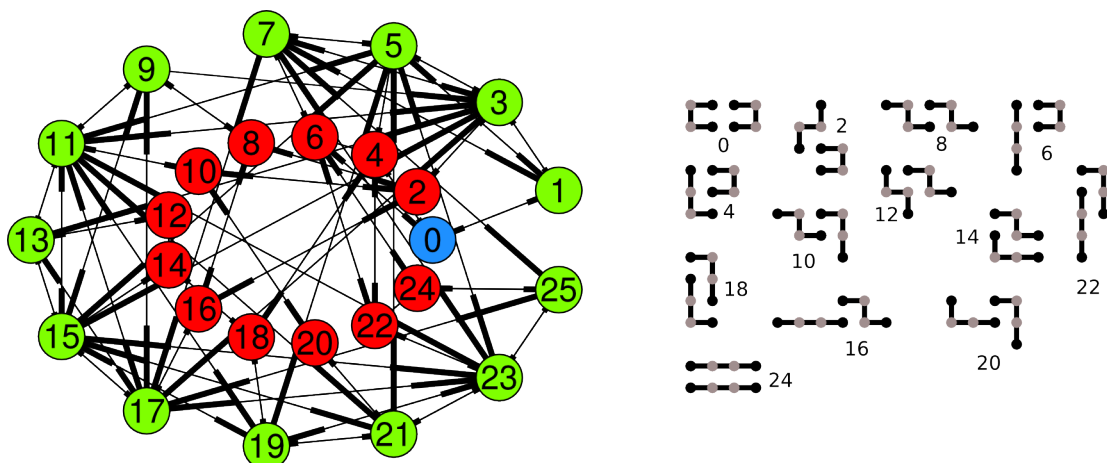
rán bemutatott szimetriatranszformációk segítségével egy olyan egységesített állapotba alakítottam át, hogy az első lánc kezdőpontja a  $(0, 0)$  pozícióban legyen, a kezdővektora az  $(1, 0)$  vektor legyen, és az első olyan belső láncmonomernél, ahol a lánc nem egyenesen folytatódik, jobbra forduljon. A kapott egységesített állapotok mindegyike egy konformációpárral jellemezhető. Meghatároztam azoknak az állapotoknak a számát, amelyeket az adott konformációpárral jellemezhető egységesített állapotba alakítottam át, és – mint a konformációpár multiplicitását – fölhasználtam egy állapot (szabad)energiájának kiszámításában.

## Kétrétegű állapothálózat modell

Az alapmodellben a két láncot egy  $2 \cdot l$  oldalhosszúságú, négyzet alakú dobozban helyeztem el. Ez azt jelenti, hogy egy  $l = 10$  hosszúságú lánc esetén minden konformációpárra  $\sim 100$  különböző elhelyezkedés lehetséges (ha az egyik lánc kezdőpontját az origóba helyezzük és a másik kezdőpontját az összes lehetséges rácspontra viszem – a szimmetriák miatt ezt megtehetem). Minden ilyen elhelyezkedés esetén a második lánc 4 különböző orientációban állhat a másik lánchoz képest. Ez nagymértékben megnöveli az állapotok számát és így az állapothálózat méretét is.

Minden állapotnak a két lánc konformációján kívül az egyik legfontosabb – és mivel a vizsgálatunk az asszociáltság és a felgombolyodás kapcsolatára fókuszál, ebből a szempontból legfontosabb – sajátága, hogy a két lánc asszociált állapotban van-e. Asszociált állapotnak tekintjük

17. ábra. *A láncok felcserélhetőségéből adódó szimmetria.*



18. ábra. A HPPH szekvencia kétrétegű modellje (balra) és az asszociált állapotok (jobbra). A bal oldali ábrán piros színnel vannak jelölve az asszociált állapotok, zölddel a disszociált állapotok, a natív állapot pedig kék színnel.

a két láncot, ha van legalább egy olyan láncmonomerpár, amely pár tagjai különböző lánc részei és egymás mellett helyezkednek el a rácson.

Olyan hálózatmodellt vezetünk be tehát (18. ábra), ami ezeket a sajátságokat veszi elsődlegesen figyelembe. Az ún. *kétrétegű modellben* minden konformációpárhoz két állapot tartozik: egy asszociált és egy disszociált.

Tegyük fel most, hogy a láncok diffúziós mozgásához szükséges idő elhanyagolható a konformációs mozgásokhoz szükséges időhöz képest, és úgy tekintjük, mintha a rendszer az asszociált állapotok közül csak a legalacsonyabb energiájú állapotban tartózkodna.

Két asszociált állapot közé akkor húzunk élt, ha létezik olyan mozgás a mozgáskészletben, ami az egyik asszociált állapotot átviszi egy másik asszociált állapotba. Ha csak olyan mozgás létezik, amely az asszociált állapotot egy másik konformációpár olyan állapotába viszi át, amely annak nem asszociált állapota, akkor a két konformációpárnak megfelelő disszociált állapotokat kötjük össze, ami megfelel annak, hogy a két lánc először disszociál, bekövetkezik a konformációváltozás, majd újra kötődnek. Minden konformációpárhoz tartozó asszociált és disszociált állapotok össze vannak kötve egymással.

A számításokat a kétrétegű állapothálózatot reprezentáló  $\mathbf{T}$  átmeneti mátrix segítségével végeztük. Az átmeneti mátrix  $T_{ij}$ -edik elemét a

$$T_{ij} = \begin{cases} \min\left(1, \frac{n_i}{n_j} \cdot e^{-(E_j - E_i)/k_B T}\right) / n_i, & \text{ha } i \neq j \\ 1 - \sum_{k \neq i} T_{ik}, & \text{ha } i = j \end{cases} \quad (46)$$

egyenlet adja meg, ahol  $n_i$  és  $n_j$  rendre az  $i$ -edik és  $j$ -edik állapotból kiinduló élek száma,  $E_i$  és  $E_j$  az  $i$ -edik és  $j$ -edik állapot energiája,  $k_B$  a Boltzmann-állandó és  $T$  a hőmérséklet, amit a Monte Carlo-szimulációkhoz hasonlóan úgy állítottuk be, hogy a natív állapot valószínűsége 0,5 legyen.

Vizsgálatainkat alapvetően kétféle koncepció szerint végeztük. Az egyik koncepció az volt, hogy valamilyen – a kérdésfeltevés szempontjából releváns – *mezoállapotokat*<sup>2</sup> (a mezoáll-

<sup>2</sup>Mezoállapotnak nevezem mikroállapotok bármilyen halmazát, amely a rendszernek nem makroállapota.

potoknak ezt a típusát *előre definiált mezoállapotoknak* nevezem) definiáltunk, megfigyelhető paraméterek segítségével, a modell felépítésétől függetlenül, és a mikroállapotok mindegyikét ezen előre definiált mezoállapotok valamelyikébe soroltuk. Ezek után vizsgáltuk az egyes előre definiált mezoállapotok közötti átmeneteket, azok sebességi együtthatóit és fluxusait. A másik koncepció szerinti vizsgálat során a mezoállapotok elkülönítésére nem vittünk be önkényes információt, hanem magának a modellnek a belső kinetikai struktúráját tártuk fel.

### A kétrétegű modell elemzése előre definiált állapotok segítségével

Mivel a kutatás a rendezetlen fehérjék kapcsolt felgombolyodás–kötődésére irányult, célszerű volt valamilyen, a láncok felgombolyodottságát és asszociáltságot kifejező paramétert definiálni. A mikroállapotokat a láncok felgombolyodottsági és asszociációs paraméterei szerint különböző mikroállapotokba soroltuk. A legtöbb számításban az egyszerűség kedvéért kétállapotú paramétereket használtunk. Ebben az esetben a mikroállapotokat a következő hat mezoállapot valamelyikébe soroltuk (ahol az „*u*” és „*f*” rendre a denaturált állapotot és a natív állapotbeli konformációs állapotot jelöli, míg a „*d*” és „*a*” rendre a disszociált és az asszociált állapotot):

- *udu*: egyik lánc sincs a natív konformációban és nincs közöttük kontaktus
- *uau*: egyik lánc sincs a natív konformációban és van közöttük kontaktus
- *fd*: pontosan az egyik lánc a natív konformációban van és nincs a láncok között kontaktus
- *fau*: pontosan az egyik lánc a natív konformációban van és van kontaktus a láncok között
- *fdf*: mindkét lánc a natív konformációban van és nincs közöttük kontaktus
- *faf*: mindkét lánc a natív konformációban van és van közöttük kontaktus, ami a natív dimer .

Az *uau* és a *fau* állapotokat további két-két részre osztottam aszerint, hogy az adott asszociált mikroállapotból elérhető-e a natív állapot kizárólag asszociált állapotokon keresztül, vagy szükség van legalább egy disszociációs lépésre. Ezek alapján tehát az *uau* és a *fau* helyett a következő négy állapot valamelyikébe soroltam a megfelelő mikroállapotokat

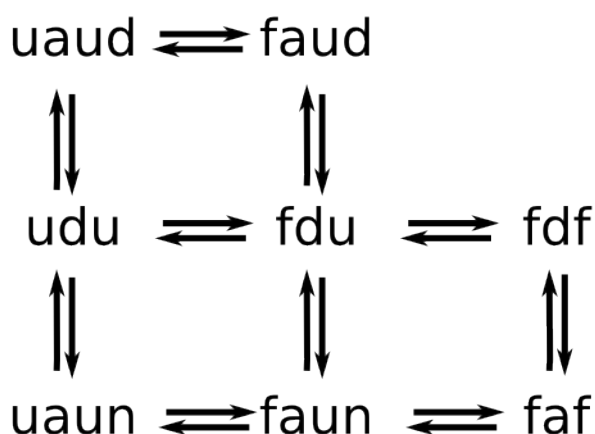
- *uaun* és *faun*: a natív állapot elérhető kizárólag asszociált állapotokon keresztül
- *uau* és *fau*: a natív állapot csak disszociált állapot(ok)on keresztül érhető el .

A fenti nyolc állapot segítségével felrajzolható a komplex kialakulásának kinetikai sémája. Az *uau* és a *fau* mezoállapothoz nem minden szekvencia esetében tartoznak mikroállapotok. A nyolc állapot segítségével definiált kinetikai sémát a 19. ábra mutatja.

### Merev dokkolás, indukált felgombolyodás, konformációkiválasztás

Aszerint, hogy a kötődéskor az egyes kötőpartnerek milyen felgombolyodottsági állapotban vannak, három dimerképződési mechanizmus különíthető el. Indukált felgombolyodásnak, tehát ahol mindkét lánc a kötődés hatására veszi fel a natív konformációját, tekinthető az az eset, amikor két denaturált lánc kapcsolódik össze. Abban az esetben, amikor a kötődéskor az egyik lánc már a natív dimerbeli konformációban van, a másik viszont még denaturált állapotban, habár





19. ábra. A kétrétegű állapothálózatra az általunk előre definiált mezoállapotok segítségével felírt kinetikai séma.

a későbbiek során indukált felgombolyodás is végbemehet, a kezdeti lépés egy konformációki-választási lépés. Amikor két, a natív dimerbeli konformációjában lévő lánc kapcsolódik, merev dokkolásról beszélhetünk.

Azt, hogy az egyes mechanizmusok mekkora szerepet játszanak a dimer kialakulásában, az egyes mechanizmusok szerinti útvonalakon átfolyó fluxusok kiszámításával lehet meghatározni. Mivel a mechanizmusok a monomereknek a kötődés pillanatában megfigyelhető felgombolyodottsági állapotától függenek, a két réteget összekötő, tehát a kötődési lépéshez tartozó éleken átmenő fluxust kell kiszámítani, és a fluxusokat összegezni a mezoállapotra vonatkozóan. Nem minden, a két réteget összekötő élen átfolyó fluxus reaktív, azaz olyan, hogy a rendszer újabb disszociáció nélkül jut el a natív állapotba.

Annak a kiderítésére, hogy a fluxus mekkora hányada reaktív, az átmenetiútvonal-elméletből ismert elköteleződési függvényt hívtuk segítségül. Az átmenetiútvonal-elméletben egy tetszőlegesen definiált  $R \subset S$  reaktánshalmaz, és egy, a reaktánshalmazzal át nem fedő  $P \subset S$ ,  $R \cap P = \emptyset$  termékalmazt definiálunk, és minden  $s \in S$ ,  $s \notin R \cup P$  állapotra kiszámítjuk, hogy mekkora annak a valószínűsége, hogy  $s$ -ből kiindulva  $P$ -be jutunk anélkül, hogy közben  $R$ -t meglátogatnánk. Az  $s$  állapotokhoz ezt a valószínűséget hozzárendelő függvényt nevezzük elköteleződési függvénynek. Ha a reaktánshalmazt úgy definiáljuk, hogy az legyen a disszociált állapotok halmaza, a termékalmazba pedig a natív állapot tartozzon, akkor az elköteleződési függvény azt adja meg, hogy milyen valószínűséggel jut el a rendszer egy asszociált, nem natív állapotból a natív állapotba anélkül, hogy közben disszociálna a két lánc. Nevezzük az így definiált elköteleződési függvényt *asszociációs elköteleződési függvénynek*. Ha tehát a két réteget összekötő éleken az asszociált réteg felé folyó fluxusokat megszorozzuk az asszociációs elköteleződési függvény értékével abban az állapotban, ahova az él fut, akkor megkapjuk az adott élen átfolyó *reaktív fluxust*.

A merev dokkolás fluxusát a

$$f_{\text{md}}^{\text{rea}} = \sum_{(i,j) \in fdf \times faf} p_i T_{ij} q_j^{\text{assz}}, \quad (47)$$

ahol  $p_i$  az  $i$  állapot pillanatnyi valószínűsége,  $T_{ij}$  az átmeneti mátrix  $i$ -edik sorának  $j$ -edik eleme és  $q_j^{\text{assz}}$  pedig az asszociációs elköteleződési függvény értéke a  $j$  állapotban.

Ugyanígy az indukált felgombolyodás fluxusát

$$f_{\text{if}}^{\text{rea}} = \sum_{(i,j) \in udu \times uau} p_i T_{ij} q_j^{\text{assz}}, \quad (48)$$

míg a konformációkiválasztás fluxusát

$$f_{\text{kk}}^{\text{rea}} = \sum_{(i,j) \in fdu \times faun} p_i T_{ij} q_j^{\text{assz}} \quad (49)$$

adja meg.

A fent definiált fluxusok pillanatnyi fluxusok. Azt, hogy melyik mechanizmusnak mekkora a szerepe a dimerképződésben, valójában az összfluxusokkal lehet jellemezni, amit a pillanatnyi fluxusok időbeli összegzésével kapunk. Ha a rendszer egyensúlyban, vagy valamilyen steady-state állapotban van, akkor a pillanatnyi fluxusok függetlenek az időtől, tehát arányuk jól jellemzi az egyes mechanizmusok relatív jelentőségét.

#### Fluxusok időgörbéi

Induljunk ki a  $T = \infty$  hőmérsékletnek megfelelő valószínűségi eloszlásból. Legyen az  $n$  natív állapot nyelő állapot, tehát

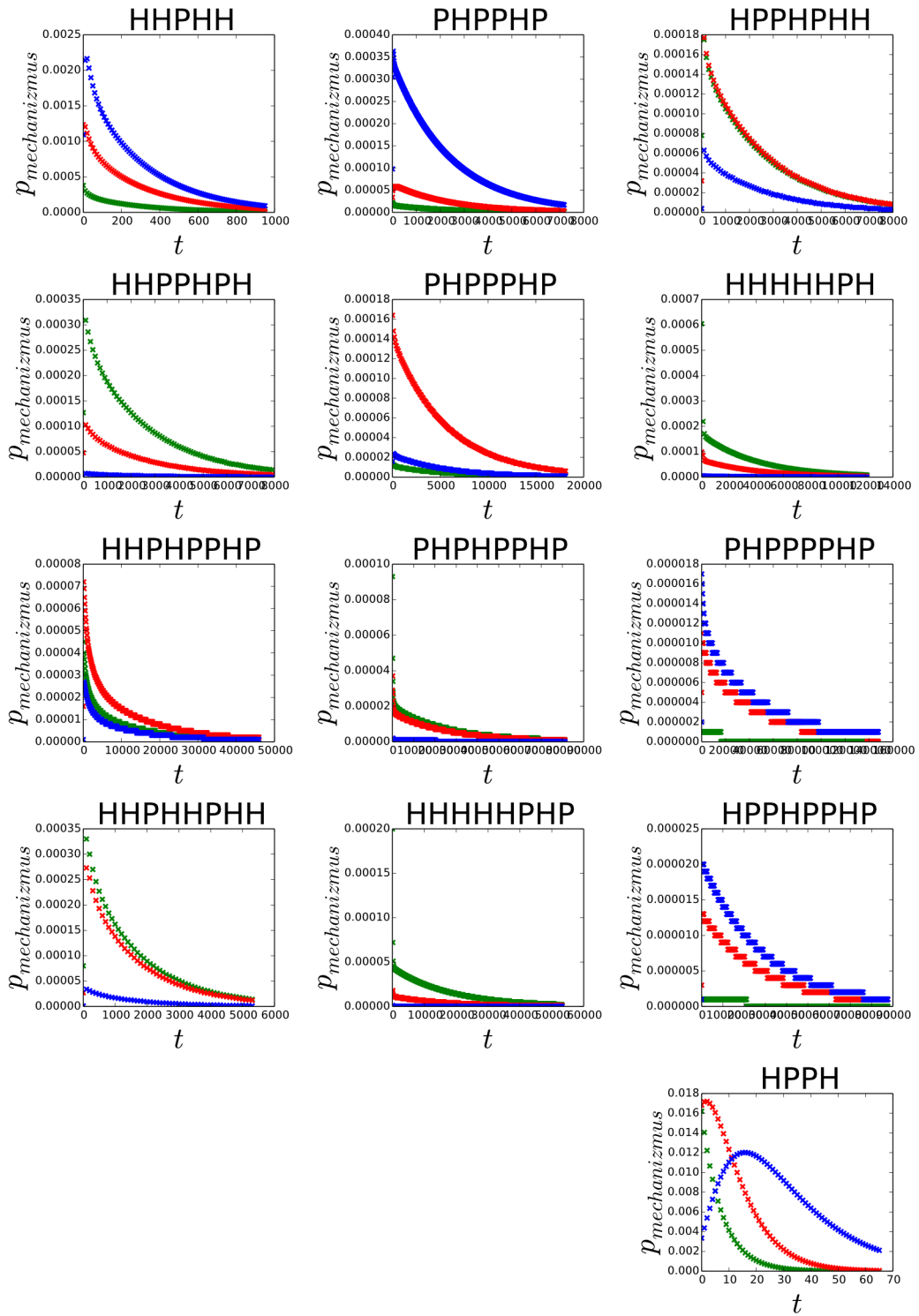
$$T_{ni} = \begin{cases} 0 & \text{ha } i \neq n \\ 1 & \text{ha } i = n. \end{cases} \quad (50)$$

és az eloszlásnak megfelelő  $\mathbf{p}$  valószínűségi vektort minden  $t$  időpontban szorozzuk meg jobbról az átmeneti mátrixszal, tehát

$$\mathbf{p}(t+1) = \mathbf{p}(t) \mathbf{T}. \quad (51)$$

Számítsuk ki minden lépésben (47), (48) és (49) alapján a reaktív fluxusokat. A három mechanizmusnak megfelelő reaktív fluxusokat az idő függvényében a 20. ábra mutatja. Az ábrán látható, hogy az, hogy melyik mechanizmus a domináns, függ a szekvenciától, és bizonyos esetekben időben is változhat. Megfigyelhető az is, hogy mind a három mechanizmus lehet domináns.

Azt, hogy egy adott szekvencia esetén melyik mechanizmus dominál, nagyrészt meghatározza az, hogy az adott szekvencia derék-e vagy sem. Derék szekvenciák esetén, ahol a dimer alapállapotbeli konformáció monomer állapotban is stabil, azt várjuk, hogy vagy a merev dokkolás vagy a konformációkiválasztás a domináns mechanizmus, míg nem derék szekvenciák esetén éppen azt várjuk, hogy az indukált felgombolyodás a legjelentősebb. A 20. ábrán látható eredmények, habár nem tökéletesen, de igazolták a várakozásainkat. A legrövidebb, HPPH szekvencia esetén időben is változik, hogy melyik mechanizmus dominál. A jelenség magyarázata az lehet, hogy az állapotok kis száma miatt az egyes állapotok relatív pillanatnyi valószínűségei rövid idő alatt nagymértékben képesek megváltozni, és az átmeneti valószínűségek időbeli változatlanságai miatt a fluxusok időbeli változásait a disszociált állapotok relatív valószínűségeinek időbeli változásai fogják meghatározni.



20. ábra. Merev dokkolás ( $x$ ), indukált felgombolyodás ( $x$ ) és konformációkiválasztás ( $x$ ) szerinti pillanatnyi fluxusok az idő függvényében.

### Átmenetiútvonal-elmélet

A reakciókinetikai elméletben a reakciók sebességét leggyakrabban az *átmenetiállapot-elmélet* segítségével számították. Az átmenetiállapot-elmélet azonban csak metastabilis állapotok közötti átmenetek sebességeinek meghatározására alkalmas. Komplex rendszerekben, amelyek viselkedése nem közelíthető jól egy kétállapotú rendszer viselkedésével, vagy olyan reakciókban, ahol a reaktáns és a termék állapot a rendszernek nem metastabilis állapotai, az átmenetiútvonal-elmélet alkalmas a reakciósebességek meghatározására. Az átmenetiútvonal-elmélet bizonyos körülmények (pl. egyensúly, steady-state állapot) között a sebességi együttható számítására is felhasználható.

Az átmenetiútvonal-elméletben tetszőleges – át nem fedő – reaktáns és termék állapot-halmazokat definiálhatunk, és az így definiált reakció sebességét számíthatjuk ki. Mivel a cél a felgombolyodási és kötődési folyamat időbeli viszonyának feltárása volt, ezért a reaktáns halmazt úgy definiáltuk, hogy az legyen egyenlő az *udu* halmazzal, míg a termék-halmaz a natív állapotot tartalmazta. Az *udu* halmazból indított reakciókra kiszámítottuk a három dimerkialakulási mechanizmus szerinti reaktív fluxusokat:

$$f_{\text{md}}^{\text{rea}} = \sum_{(i,j) \in \text{fdf} \times \text{faf}} \pi_i T_{ij} q^+_j (1 - q^+_i) q_k^{\text{assz}} \quad (52)$$

$$f_{\text{if}}^{\text{rea}} = \sum_{(i,j) \in \text{fdu} \times \text{faun}} \pi_i T_{ij} q^+_j (1 - q^+_i) q_k^{\text{assz}} \quad (53)$$

$$f_{\text{kk}}^{\text{rea}} = \sum_{(i,j) \in \text{udu} \times \text{uaun}} \pi_i T_{ij} q^+_j (1 - q^+_i) q_k^{\text{assz}}, \quad (54)$$

ahol  $\pi_i$  az  $i$  állapot egyensúlyi valószínűsége,  $T_{ij}$  az átmeneti mátrix  $i$ -edik sorának  $j$ -edik eleme,  $q^+_i$  az elköteleződési függvény értéke az  $i$  állapotban,  $q_k^{\text{assz}}$  pedig az asszociációs elköteleződési függvény értéke a  $j$  állapotban. A három mechanizmus szerinti fluxusok különböző szekvenciákra számított értékeit a 3. táblázat mutatja.

Szekvencia	Merev dokkolás	Konformációkiválasztás	Indukált felgombolyodás
HPPH	<b>0,811</b>	0,178	0,011
HHPHH	0,097	<b>0,587</b>	0,316
PHPPHP	<b>0,998</b>	0,001	$5,52 \cdot 10^{-4}$
HPPHPPH	0,169	<b>0,419</b>	0,412
HHPHPPH	$7,514 \cdot 10^{-3}$	0,137	<b>0,856</b>
PHPPHP	<b>0,913</b>	0,084	$3,155 \cdot 10^{-3}$
HHHHHPH	$2,514 \cdot 10^{-3}$	0,104	<b>0,893</b>
HHPHPPHP	<b>0,500</b>	0,398	0,102
PHPHPPHP	0,105	<b>0,544</b>	0,351
PHPPPPHP	<b>0,998</b>	$2,023 \cdot 10^{-3}$	$1,352 \cdot 10^{-4}$
HHPHHPHH	0,060	0,389	<b>0,552</b>
HHHHHPHP	0,027	<b>0,513</b>	0,460
HPPHPPHP	<b>0,778</b>	0,220	$2,01 \cdot 10^{-3}$

3. táblázat. *Merev dokkolás, indukált felgombolyodás és konformációkiválasztás szerinti egyensúlyi fluxusok az átmenetiútvonal-elmélet segítségével számolva. A reaktánshalmaz az udu állapotok halmaza. Félkövérrel vannak szedve a domináns mechanizmusok fluxusai.*

Kiszámítottuk az  $udu \rightarrow faf$  átmenet reakciósebességét is a

$$v(udu \rightarrow faf) = \sum_{(i,j) \in udu \times \overline{udu}} f_{ij} \quad (55)$$

egyenlet segítségével, ahol  $\overline{udu}$  az  $udu$  halmaz komplementer halmaza és

$$f_{ij} = \pi_i T_{ij} q^+_j (1 - q^+_i) \quad (56)$$

reaktív fluxus. Egyensúlyi állapotban, illetve ha az  $udu$  halmaz metastabilis állapot, a reakciósebességből számítható a sebességi együttható

$$k(udu \rightarrow faf) = \frac{v(udu \rightarrow faf)}{\sum_{i \in udu} p_i} \quad (57)$$

szerint, ahol egyensúlyban  $p_i = \pi_i$ . Az általunk számított egyensúlyi sebességi együtthatókat a 4. táblázat tartalmazza.

Szekvencia	$k^{eq}(udu \rightarrow faf)$
HPPH	0,01577
HHPHH	0,045128
PHPPHP	0,000311
HPPHPHH	0,00528
HHPHPHP	0,002849
PHPPHP	$6,3 \cdot 10^{-5}$
HHHHHPH	0,018113
HHPHPHP	0,00042
PHHPHPHP	$5,9 \cdot 10^{-5}$
PHPPHPHP	$10^{-5}$
HHPHHPHH	0,003519
HHHHHPHP	0,000133
HPPHPHP	$9,7 \cdot 10^{-5}$

4. táblázat. Az átmenetiútvonal-elmélet segítségével számított egyensúlyi sebességi együtthatók.

#### Steady-state reaktív fluxusok

Az élő rendszerek nemegyensúlyi rendszerek, sok esetben azonban steady-state körülményeket feltételezhetünk; azáltal, hogy kívülről energiát viszünk be, a belső környezet viszonylag állandó állapotban tartható. Az élő sejtben is, egy adott fehérje esetében, amelyre a sejtnek folyamatosan szüksége van, feltételezhetjük, hogy folyamatosan termelődik, miközben folyamatosan el is bomlik a natív állapotú fehérje, komplexek esetében pedig a komplex és az őt alkotó molekulák.

Ha tehát a sejtben lévő viszonyokat kívánjuk modellezni, akkor kézenfekvő steady-state viszonyokat feltételezni. Steady-state körülmények között, ha az általunk definiált reaktánshalmaz a rendszernek nem metastabilis állapota, az átmenetiútvonal-elmélet segítségével tudjuk kiszámítani a reakció sebességi együtthatóját.

Többféle steady-state állapot definiálható, mi azt az esetet vizsgáltuk, ahol az egyedi láncok denaturált állapotban képződnek, és mivel az egyes láncok külön-külön képződnek, a kezdő

állapot mindig az  $udu$ , és a keletkezés folyamatossága miatt az  $udu$  állapot koncentrációja állandó. Tekintsünk tehát egy olyan ergodikus rendszert, ahol nyelő és forrásállapotokat definiálunk. Jelölje  $P$  a nyelő,  $R$  pedig a forrásállapotok halmazát. *Kvázistacionárius* vagy más néven *steady-state* állapotban a rendszer minden  $i \notin (P \cup R)$  állapotára érvényes az anyagmegmaradás, azaz

$$\sum_j m_i^{ss} T_{ij} - \sum_j m_j^{ss} T_{ji} = 0, \quad (58)$$

ahol  $m_i^{ss}$  a steady-state anyagmennyiség az  $i$  állapotban,  $T_{ij}$  pedig az átmeneti mátrix  $i$ -edik sorának  $j$ -edik eleme.

Mivel

$$T_{ii} = 1 - \sum_j T_{ij}, \text{ minden } i\text{-re}, \quad (59)$$

(58)

$$\sum_j m_j^{ss} - m_i^{ss} = 0 \quad (60)$$

alakban írható.

A nyelő állapotokban legyen az anyagmennyiség nulla, azaz

$$m_k = 0, \text{ minden } i \in P\text{-re}, \quad (61)$$

a forrásállapotokban pedig legyen az anyagmennyiség állandó, azaz

$$m_l = c = \text{konstans}. \quad (62)$$

Megoldva az egyenletrendszert megkapjuk az  $m_i^{ss}$  értékeket az összes állapotra. A steady-state valószínűségek értékei most már a

$$P_i^{ss} = \frac{m_i^{ss}}{\sum_j m_j^{ss}} \quad (63)$$

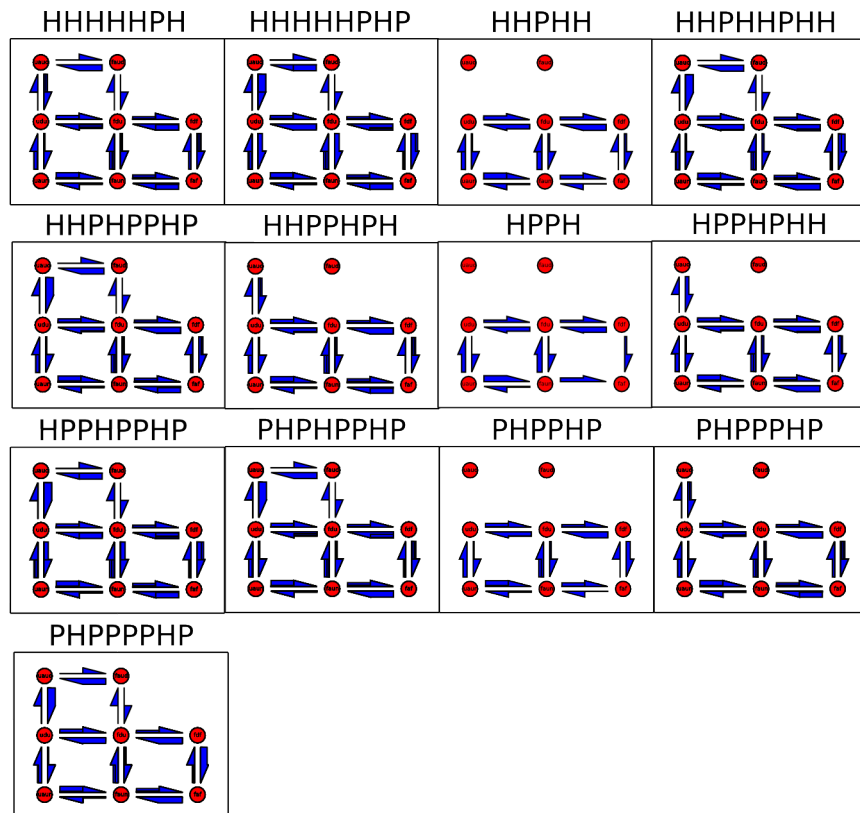
egyenlet segítségével könnyen meghatározhatók.

Az egyensúlyi állapothoz hasonlóan a steady-state állapotban is kiszámítottuk az egyes mechanizmusok szerinti reaktív fluxusokat, és a kapcsolt felgombolyodás és kötődés sebességi együtthatóit a különböző szekvenciákra. Az adatokat az 5. és a 6. táblázat tartalmazza.

#### *A homodimerképződés kinetikai sémái*

A fent definiált nyolc előre definiált mezoállapotra az átmenetiútvonal-elmélet segítségével kiszámítottuk az egyensúlyi és a steady-state fluxusokat, ill. a sebességi együtthatókat. A steady-state sebességi együtthatók segítségével felírt kinetikai sémák a 21. ábrán láthatók. Az ábrákból látszik, hogy nem minden szekvencia esetén található mikroállapot az  $uau$  és a  $fau$  állapotban.

A 22. ábrán a HPPHPH szekvencia előre definiált mezoállapotok segítségével felírt kinetikai sémája látható, a sebességi együtthatók helyett az átmenetek reaktív fluxusait feltüntetve. Ha összehasonlítjuk az  $udu$  állapotból az  $fdu$ , ill. az  $uau$  állapotokba irányuló fluxusok nagyságát, láthatjuk, hogy az  $fdu$  állapotba irányuló fluxus lényegesen nagyobb. Az 5. táblázat adataiból azonban az látszik, hogy a domináns mechanizmus az indukált felgombolyodás, ami az  $udu \rightarrow uau$  átmenetnek felel meg. Ez a látszólagos ellentmondás föloldható, ha az előremutató



21. ábra. Az átmenetiútvonal-elmélet segítségével az előre definiált mezoállapotok közötti átmenetek steady-state sebességi együtthatói. Az állapotok elhelyezkedése megfelel az állapotok elhelyezkedésének a 19. ábra sémájában. A nyilak vastagsága arányos a sebességi együttható értékével.

Szekvencia	Merev dokkolás	Konformációkiválasztás	Indukált felgombolyodás
HPPH	<b>0,752</b>	0,231	0,017
HHPHH	0,132	<b>0,596</b>	0,273
PHPPHP	<b>0,988</b>	$8,172 \cdot 10^{-3}$	$4,062 \cdot 10^{-3}$
HPPHPHH	0,098	0,369	<b>0,533</b>
HHPHPHP	0,017	0,251	<b>0,732</b>
PHPPHP	<b>0,833</b>	0,162	$4,757 \cdot 10^{-3}$
HHHHHPH	$9,772 \cdot 10^{-3}$	0,347	<b>0,643</b>
HHPHPHP	0,250	<b>0,535</b>	0,215
PHHPHPHP	0,057	<b>0,490</b>	0,453
PHPPHPHP	<b>0,996</b>	$3,906 \cdot 10^{-3}$	$2,776 \cdot 10^{-4}$
HHPHHPHH	0,106	<b>0,623</b>	0,272
HHHHHPHP	0,045	<b>0,672</b>	0,283
HPPHPHP	<b>0,561</b>	0,434	$5,172 \cdot 10^{-3}$

5. táblázat. Merev dokkolás, indukált felgombolyodás és konformációkiválasztás szerinti steady-state fluxusok az átmenetiútvonalemélet segítségével számolva. A reaktánshalmaz az  $udu$  állapotok halmaza. Félkövérrel vannak szedve a domináns mechanizmusok fluxusai.

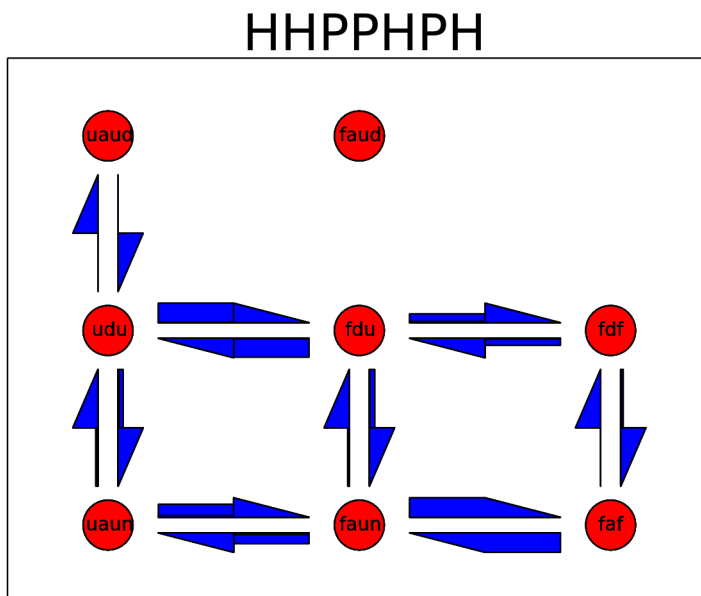
Szekvencia	$k^{ss}$ ( $udu \rightarrow faf$ )
HPPH	0,010771
HHPHH	0,034073
PHPPHP	0,000273
HPPHPHH	0,00109
HHPHPHP	0,000782
PHPPHP	$6,3 \cdot 10^{-5}$
HHHHHPH	0,00431
HHPHPHP	0,000142
PHHPHPHP	0,000143
PHPPHPHP	$10^{-5}$
HHPHHPHH	0,000377
HHHHHPHP	0,000282
HPPHPHP	$5,9 \cdot 10^{-5}$

6. táblázat. Az átmenetiútvonalemélet segítségével számított steady-state sebességi együtthatók.

fluxusok mellett a fordított irányú reakciók fluxusait is tekintetbe vesszük. Látható, hogy míg az  $udu \rightarrow fdu$  és az  $fdu \rightarrow udu$  átmenetek reaktív fluxusai nagyjából megegyeznek, addig az  $udu \rightarrow uau$  átmenet fluxusa többszöröse az  $uau \rightarrow udu$  átmenet fluxusáénak.

A 23. ábrán a HHHHPH szekvencia steady-state és egyensúlyi kinetikai sémái láthatók; a bal oldali felső ábrán a nyilak vastagsága a steady-state reaktív fluxusokkal, a jobb oldali felső ábrán a steady-state sebességi együtthatókkal, míg az alsó ábrán az egyensúlyi reaktív fluxusokkal arányos. Összhangban azzal, hogy az 5. táblázat szerint a domináns mechanizmus az indukált felgombolyodás, a HHPHPHP szekvenciához hasonlóan itt is az  $udu \rightarrow uau$  átmenet reaktív fluxusa többszöröse az ellentétes irányú átmenet reaktív fluxusának, ugyanez a sebességi együtthatók esetében éppen fordított, tehát az  $uau \rightarrow udu$  átmenet sebességi együtthatója nagyobb a fordított reakció sebességi együtthatójánál.





22. ábra. Az átmenetiútvonal-elmélet segítségével az előre definiált mezoállapotok közötti átmenetek steady state reaktív fluxusai a HHPHPH szekvencia esetében. A nyilak vastagságai arányosak a reaktív fluxusokkal.

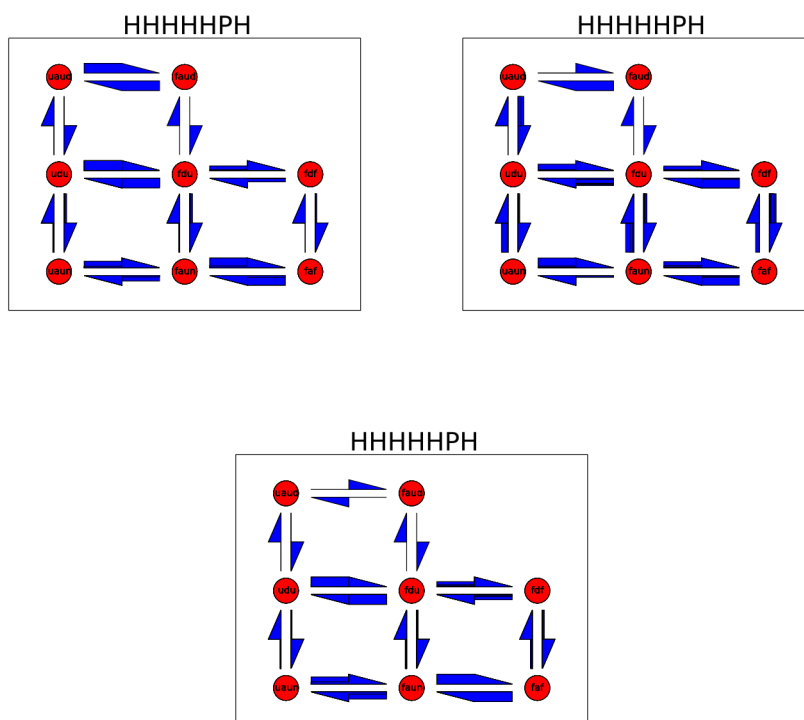
### Kötődési fluxustájkép

Célunk volt részletes képet kapni arról, hogy a kötődés a két lánc milyen felgombolyodottsági állapotában megy végbe. Ehhez definiáltunk egy felgombolyodottsági mérőszámot és a két lánc felgombolyodottságának függvényében ábrázoltuk a kötődési fluxust.

Egy adott állapot felgombolyodottsági mérőszámát úgy definiáltam, hogy az legyen az adott állapotból a natív állapotba történő eljutás idejének várható értéke („mean first passage time”) egy olyan hálózaton, aminek a topológiája megegyezik az eredeti átmeneti hálózat topológiájával, de az élek súlyai megfelelnek annak, mintha minden állapot azonos energiájú lenne.

A konformációkat a felgombolyodásiparaméter-értékük szerint  $n + 1$  ( $n = 4$  az  $l = 4$  és  $l = 5$  hosszúságú láncokra,  $n = 9$  az  $l > 5$  hosszúságú láncokra) csoportba soroltam, amiket az általuk tartalmazott konformációk felgombolyodásiparaméter-értékeik szerint sorba rendeztem. Jelölje most  $f_1$  és  $f_2$  rendre annak a csoportnak a sorszámát, amelyikbe egy adott disszociált állapotban az egyik illetve a másik lánc tartozik, úgy, hogy  $f_1 \leq f_2$ . Ekkor minden  $f_1, f_2$  párra összegzem a hozzájuk rendelt állapotokhoz tartozó kötődési fluxust, azaz az azon az élen átmenő fluxust, ami az asszociált rétegbe vezet. Az eredmények egyensúlyi állapotban lévő rendszerre a 24. ábrán, míg steady-state állapotban lévő rendszerre a 25. ábrán láthatók.

Látható, hogy a bináris felgombolyodási paramétert használó leírás szerinti  $udu \rightarrow uau$  átmenet többféle különböző folyamatot takar. Az is látszik, hogy a legtöbb esetben az asszociáció nem szimmetrikus abban az értelemben, hogy a két lánc nem azonos állapotban van a kötődés pillanatában.



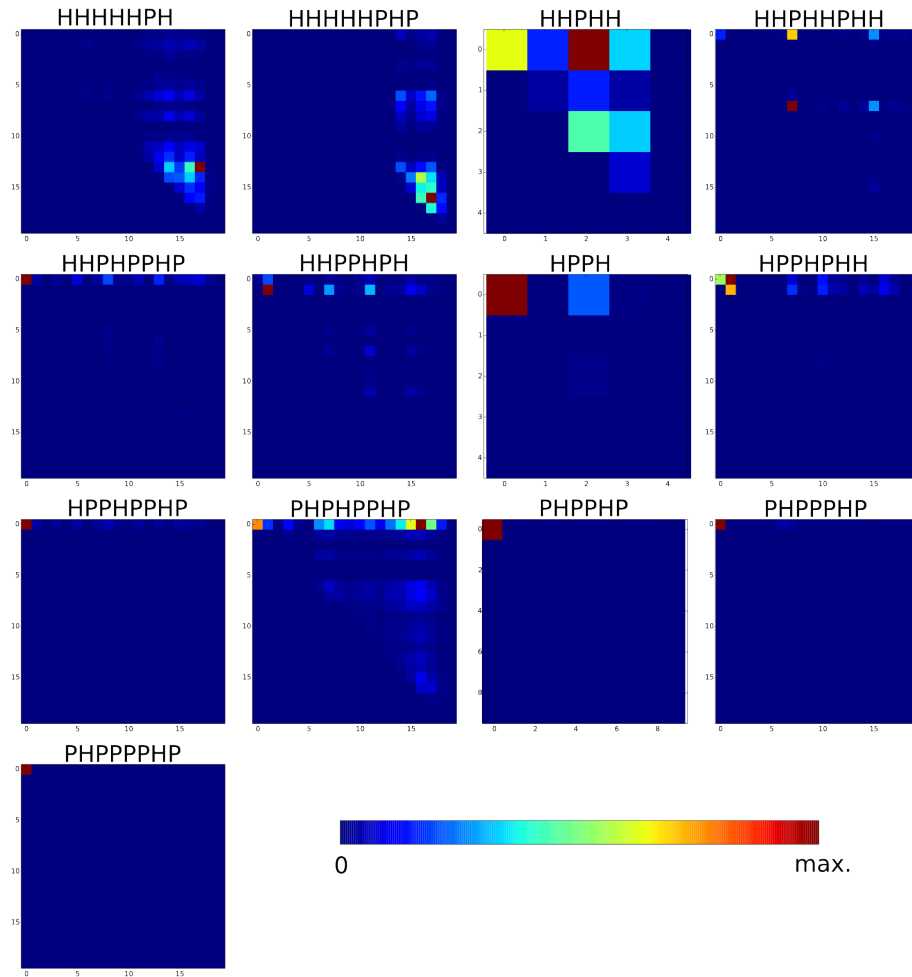
23. ábra. A HHHHHPH szekvencia egyensúlyi és steady-state kinetikai sémái láthatók. A bal oldali felső ábrán a nyilak vastagsága a steady-state reaktív fluxusokkal, a jobb oldali felső ábrán a steady-state sebességi együtthatókkal, míg az alsó ábrán az egyensúlyi reaktív fluxusokkal arányos.

### A két lánc viselkedésének szimmetriája

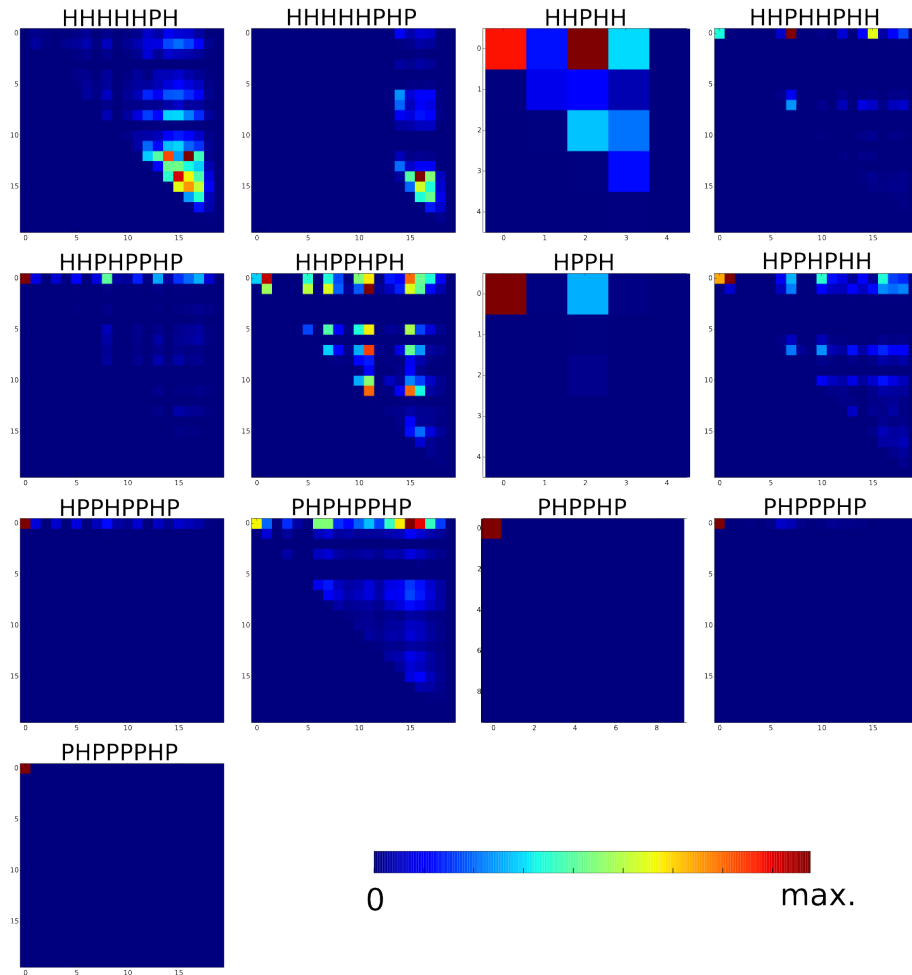
Már a kötődési fluxustájkép is nyújtott arra vonatkozó információt, hogy a két lánc azonosan viselkedik-e a dimerképződési folyamat során. Emellett definiáltam egy aszimmetria paramétert, ami minden állapatra megadja, hogy a két lánc konformációja milyen mértékben tér el egymástól. Két lánc távolságát úgy határoztam meg, hogy a monomerre fölépíttem az állapothálózatot, és minden konformációpár között kiszámítottam a konformációs átmenet idejének várható értékét. Az oda-vissza irányú várható időket átlagoltam, és az átlagolt érték adta meg két konformáció távolságát. A  $T = \infty$  hőmérsékletnek megfelelő állapotból kiindulva az átlagos aszimmetria időfüggését vizsgáltam és azt találtam, hogy bizonyos mértékű aszimmetria minden szekvencia esetében föllép. Van azonban olyan szekvencia is, ahol, legalábbis a folyamat kezdetén az átlagos aszimmetria értéke jelentősen meghaladta azt az értéket, amit egyensúly esetében várnánk (a 26. ábra).

### A kétrétegű modell belső dinamikájának elemzése

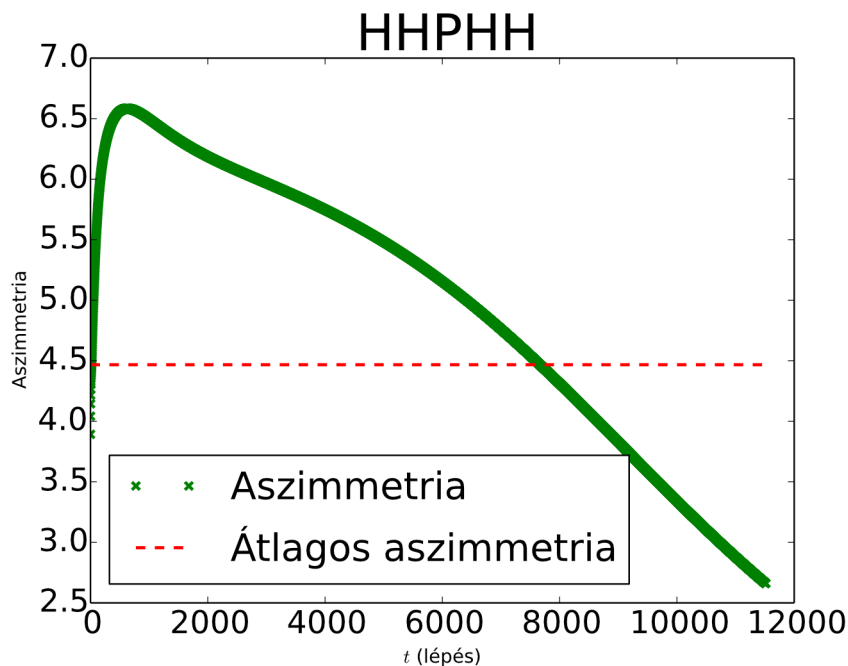
Állapothálózatok vizsgálhatók úgy is, hogy nem használunk föl önkényesen definiált mezoállapotokat a kinetika feltárásakor, hanem valamilyen hálózatelemző módszer segítségével a hálózat belső kinetikai szerkezetét tárjuk fel. A belső kinetikai szerkezet feltárása lehet például a rendszer metastabilis állapotainak azonosítása. A metastabilis állapotokra jellemző, hogy egy belső, részleges egyensúlyi állapot jóval rövidebb idő alatt kialakul, mint amennyi időre átlagosan szük-



24. ábra. Az asszociációs fluxusok az egyes láncok felgombolyodottságiparaméter-értékeinek függvényében egyensúlyi állapotban.



25. ábra. Az asszociációs fluxusok az egyes láncok felgombolyodottságiparaméter-értékeinek függvényében steady-state állapotban.



26. ábra. Az átlagos aszimmetria az idő függvényében a HHPHH szekvenciára végzett számítások alapján. Az ábrán föl van tüntetve az aszimmetria egyensúlyi várható értéke is.

ség van ahhoz, hogy a rendszer elhagyja a metastabilis állapotot, ha egyszer az adott állapotba került.

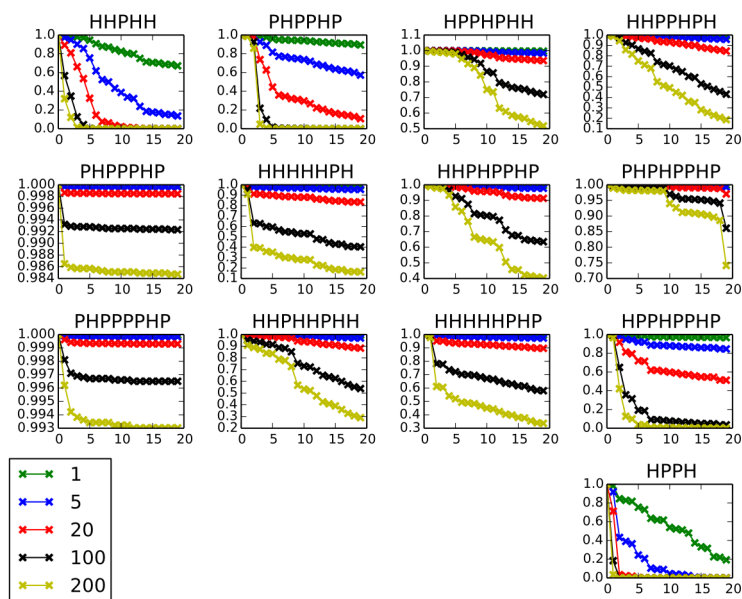
A metastabilis állapotok keresése állapothálózaton egy gráfklaszterezési probléma. A metastabilis állapotok azonosításának klaszterezési problémáját egyszerűen és elegánsan oldja meg a Perron-klaszter Klaszterelemzés, vagy az angol elnevezés alapján röviden PCCA algoritmus [92]. Az átmeneti mátrix spektrális elemzésén alapuló PCCA algoritmus egyszerűen a Markov-modellként definiált állapothálózat átmeneti mátrixának spektrális elemzése.

#### *Az állapothálózat spektrális elemzése*

Elvégeztük a vizsgált szekvenciák kétrétegű modelljeinek spektrális elemzését, hogy azonosítsuk a metastabilis állapotokat. Az, hogy egy állapot metastabilis állapotnak minősül-e, függ a vizsgált időskálától is. Ha a vizsgált időskála túlságosan rövid, akkor a metastabilis állapoton belüli egyensúly kialakulása sem megy végbe. Ha viszont a vizsgált időskála túl hosszú, a metastabilis állapotok között is kialakulhat az egyensúly és ekkor egyetlen metastabilis állapotot fogunk látni, ez pedig az állapottér összes állapotából fog állni. A vizsgált időskálát egyszerűen módosíthatjuk azáltal, hogy az átmeneti mátrixot különböző hatványaira emeljük, és ezeknek a hatványmátrixoknak végezzük el a spektrális elemzését.

Az egyes szekvenciák sajátértékspektrumait a 27. ábra mutatja. Látható, hogy a spektrumban nem különül el Perron-klaszter [92] az átmeneti mátrix egyetlen hatványánál sem, a sajátértékek nagyjából egyenletesen csökkennek. Ha a rendszer kétállapotú volna, akkor két sajátértékből álló Perron-klasztert kaptunk volna, ha háromállapotú, akkor három sajátértékből állót, stb. A mi esetünkben azonban a sajátértékspektrumok azt mutatják, hogy a rendszer nem kisszámú metasta-

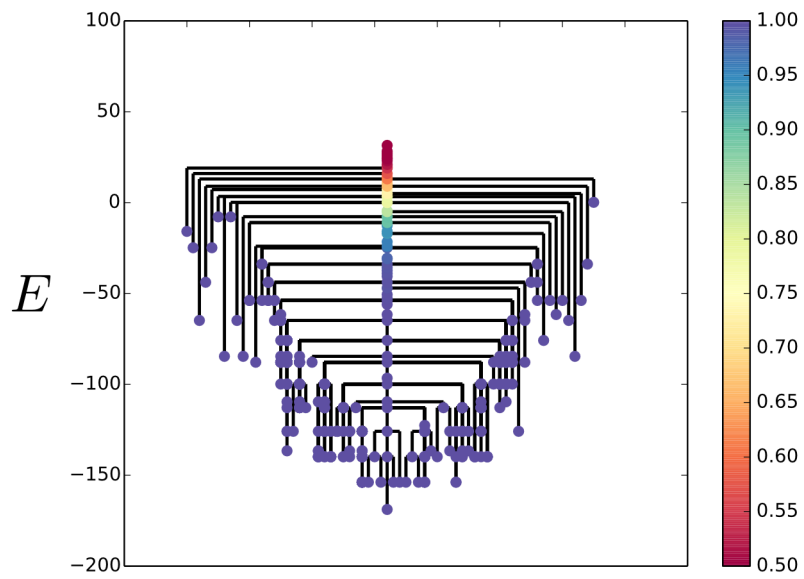
bilis állapottal rendelkezik, hanem sok ilyen metastabilis állapota van, amelyekből való szökési idők folytonosak.



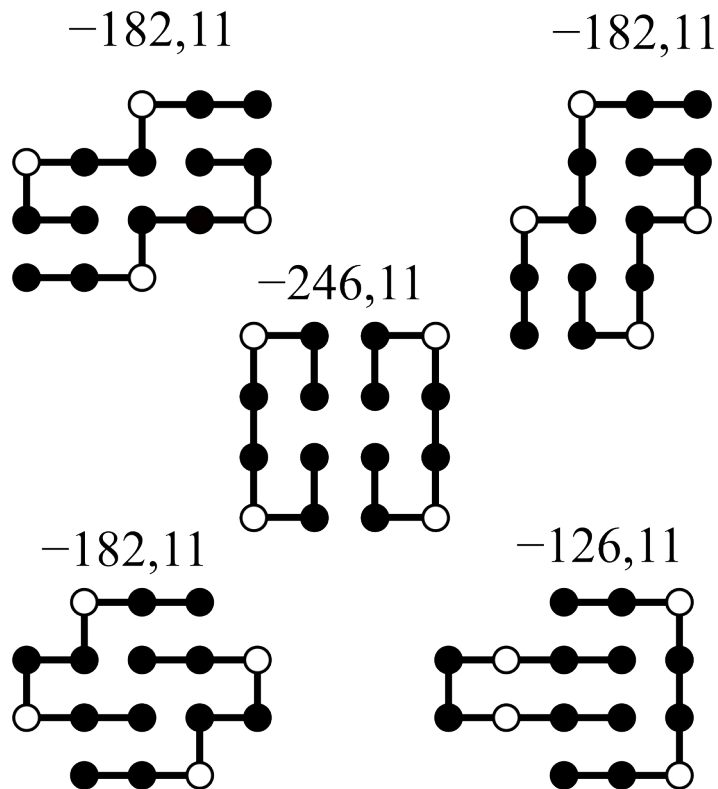
27. ábra. Átmeneti mátrixok különböző hatványainak (1., 5., 20., 100., 200.) sajátértékspektrumai. Az ábrán a húsz legnagyobb sajátérték van feltüntetve minden szekvenciára.

A sok metastabilis állapot létét, ill. az azokból való szökési idők folytonosságát támasztja alá a 28. ábrán látható diszkonnectivitási gráf is. A diszkonnectivitási gráf az energiaminimumok konnectivitását tükrözi, és ezzel jellemzi a sokdimenziós energiafelület alakját [108].

A legnagyobb sajátértékekhez tartozó jobboldali sajátvektorok elemzése alapján azonosíthatók a rendszer dinamikájának leglassabb folyamatai. Az általam vizsgált szekvenciák esetében azt találtam, hogy a leglassabb folyamatok bizonyos alacsony energiájú mikroállapotokból való szökési folyamatoknak feleltek meg. Ezek az alacsony energiájú mikroállapotok leggyakrabban olyan asszociált állapotoknak feleltek meg, ahol a láncok közötti kontaktusok száma nagy, amelyek egy részének fel kell bomlania ahhoz, hogy a rendszer elérhesse a natív állapotát. Egy példát a 29. ábra mutat.



28. ábra. A HHHHHPH szekvencia diszkonnectivitási gráfja. A csúcsok színe mutatja, hogy az adott részgráf milyen arányban tartalmaz disszociált állapotokat.



29. ábra. A leglassabb folyamatok bizonyos alacsony energiájú mikroállapotokból való szökéseknek felelnek meg. Az ábra a négy leglassabb folyamathoz tartozó mikroállapot mutatja a HHPHHPHH szekvencia esetében. Középen a natív állapot látható. Az állapotok fölött az energiájuk van feltüntetve.



## Az eredmények megvitatása

Rendezetlen fehérjék homodimereinek képződését vizsgáltam egyszerűsített fehérjemodellek és az ezek segítségével definiált egyszerűsített hálózati modell segítségével. Korábban is vizsgálták homodimerek kialakulását számítógépes modellezés segítségével, de ezek mindegyike vagy szerkezet alapú modellt alkalmazott, azaz az energiafüggvénybe beépített a natív állapotra vonatkozó információt (pl. Gō-modell) [121, 122, 123, 66, 68], vagy legalább az egyik lánc merev szerkezetű volt [69, 66, 68]. HP modell segítségével is vizsgálták rendezetlen lánc kötődés hatására bekövetkező rendeződését, de a kötőpartner itt is rendezett szerkezetű volt [124]. Munkám során tudomásom szerint én alkalmaztam először nem szerkezet alapú, egyszerűsített modellt és ezekből fölépített állapothálózatot két rendezetlen lánc dimerképződésének vizsgálatára.

### Sokállapotú rendszerek kinetikai leírásának két fő módja

A gépi tanulásban megkülönböztetnek *felügyelt* (supervised learning) és *felügyelet nélküli* (unsupervised learning) tanulást [102]. A felügyelt tanulás során a tanulóhalmaz elemei mellett a helyes megoldások is meg vannak adva, pl. osztályozás előre megadott osztályok valamelyikébe. A felügyelet nélküli tanulás során a tanuló algoritmus feladata a belső struktúrák feltárása, pl. bizonyos hasonlóságok alapján bizonyos bemeneti elemeket azonos kategóriába sorol. Az átmeneti hálózat elemzése során alapvetően kétféle megközelítést alkalmaztam. Az egyik szerint előre meghatározott mezoállapotokat definiáltam valamilyen általam definiált vizsgálati szempont szerint. Mivel a fehérjék felgombolyodásának és kötődésének viszonyát vizsgáltam, értelemszerű volt az állapotteret ilyen szempontok szerint felosztani. Ezt a felosztást tükrözi a redukált állapothálózat (19. ábra). Ez a megközelítés megfeleltethető a felügyelt tanulásnak, hiszen az állapotokat előre meghatározott osztályok valamelyikébe soroltam.

A másik megközelítésmód szerint nem mondtam előre semmit a rendszer lehetséges állapotairól, hanem megvizsgáltam, hogy hogyan osztható fel természetes módon az átmeneti hálózat. Az állapotternek metastabilis állapotokba való felosztása egy ilyen természetes felosztás, ami az állapotok közötti átmenetek sebességein alapul.

### Kétrétegű állapothálózat modell

A két láncból álló rendszerek állapottere az ugyanazon lánchosszúságú láncokból álló monomer rendszerekhez viszonyítva óriási, nem pusztán amiatt, mert a két lánc egymástól függetlenül bármilyen konformációt felvehet, hanem a koncentrációtól függő számú relatív pozícióban foglalhatnak helyet. A kiindulópont az volt, hogy ha sikerülne az állapotternek a láncok relatív elhelyezkedéséhez tartozó dimenzióját csökkenteni, az hosszabb láncok egzakt kezelését tenné lehetővé. Két feltételezést tettem, amelyek segítségével sikerült jelentősen csökkentenem a relatív pozíciókból származó állapotok számát. Az egyik feltevés az volt, hogy a translációs és

rotációs mozgások, ennek következtében a kötési és disszociációs lépések gyorsak a konformációs átalakulásokhoz viszonyítva [55, 125, 126]. A másik föltevés szerint az egy konformációpárhoz tartozó asszociált állapotok közül csak a legalacsonyabb energiájúban tartózkodik a rendszer figyelembeveendő ideig. Ez tulajdonképpen megfelel annak, hogy minden asszociált állapot egy szabadenergiagödörnek felel meg, ami a felgombolyodási szabadenergiafelszínnekhez hasonlóan tölcser alakú [7, 11, 12]. Ezenkívül a modell felépítése olyan, hogy ha kizárólag konformációs átalakulással el lehet jutni az egyik ilyen minimumból egy másikba, akkor azt feltételezem, hogy nem történik disszociációs/asszociációs folyamat. Minden más esetben a konformációs átalakulás a disszociált állapotok rétegében megy végbe.

Ezeknek az egyszerű feltételezéseknek a segítségével sikerült  $l = 8$  hosszúságú láncokra is a teljes dimer-állapothálózatot felépíteni, és a feltételezések lehetővé tették, hogy  $l = 8$  lánc-hosszúságig egzakt számításokat végezzek.

Markov-modell segítségével vizsgálták már fehérje-ligandum kötődés kinetikáját és mechanizmusát, de úgy, hogy a Markov-modell egy folytonos állapotterű modell alapján lett felépítve [127]. Egzakt számításokat végeztek HP modelleken fehérje felgombolyodást vizsgáló tanulmányokban [128, 118, 129], sőt fehérjék asszociációját is modellezték HP láncok segítségével [124], de ezekben az esetekben az eredményeket a szabadenergiafelszín koncepció segítségével értelmezték. A kétrétegű modell segítségével lehetővé vált, hogy egzakt számításokon alapuló modell alapján hozzak létre Markov-modellt, aminek segítségével a teljes rendszer dinamikája vizsgálható.

Kétrétegű modell nem csak HP modell segítségével építhető. Realisztikusabb fehérjemodellekre éppúgy érvényesek lehetnek a kiinduló feltevések, mint a HP modellre. Több láncból álló rendszerre végzett molekuladinamikai szimulációk eredményeinek értékelésében is használhatók állapothálózatok [130], és ezen állapothálózatok méretének csökkentése lehetővé teszi a konformációs állapottér, és ezen keresztül a kötődéshez kapcsolt konformációváltozások részletesebb felderítését.

## A kötődés – felgombolyodás három fő mechanizmusa és azok aránya

Hagyományosan a fehérje homodimereket aszerint, hogy a natív dimeren és a denaturált és disszociált állapoton kívül van-e valamilyen köztes állapotuk, két- és háromállapotú homodimerek közé sorolták. Ez a termodinamikai felosztás kapcsolatba hozható jellegzetes kinetikai mechanizmusokkal is. A kétállapotú dimerek esetében feltételezték, hogy a kötődés és a láncok felgombolyodása kapcsoltan megy végbe, míg a háromállapotú dimerek esetében a láncok először felgombolyodnak, és a felgombolyodott láncok kötődnek egymáshoz.

Munkám során azonban megmutattam, hogy a dimerképződés legalább háromféle különböző mechanizmus szerint mehet végbe, aszerint, hogy a kötődés pillanatában az egyik vagy mindkét lánc a natív konformációban van-e, vagy mindkét lánc denaturált állapotban van. Az egyes mechanizmusok szerinti fluxusokat a két réteget összekötő éleken átfolyó reaktív fluxusok összegei adják. A fluxusokat különböző körülmények között számítottam. Mivel az általam előre definiált állapotok a rendszernek nem metastabilis állapotai, ezért a különböző mechanizmusok szerinti fluxusok az időben változhattak. Ez okból kiszámítottam a fluxusokat egyensúlyi állapotban, steady-state állapotban, ahol úgy biztosítottam a steady-state körülményeket, hogy az  $udu$  (a két lánc disszociált állapotban van és mindkét lánc denaturált) állapot mennyiségét konstans értéken tartottam. Ekkor ugyan pillanatnyi fluxusokat számoltam, de mivel a különböző mechanizmusok szerinti fluxusok arányára voltam kíváncsi, ezért ez az információ elegendő. Ha nem egy

speciális állapotban (pl. egyensúly, steady-state) vagyok kíváncsi a rendszer viselkedésére, akkor ábrázolhatom a pillanatnyi fluxusokat az idő függvényében és a görbe alatti területek adják meg az összfluxusokat. Ebben az esetben azonban az eredmények jelentős mértékben függenek a kiindulási állapottól.

A különböző mechanizmusok relatív fluxusainak időgörbéi eltérőek a különböző szekvenciáknál, de mind a három mechanizmusra létezik olyan szekvencia, amelyben az adott mechanizmus a domináns. A derék és a nem derék szekvenciák pillanatnyi fluxus-görbéit összehasonlítva a várakozásoknak megfelelően a derék szekvenciáknál inkább a dokkolás és a konformációkiválasztás a domináns, míg a nem derék szekvenciáknál inkább az indukált folding (20. ábra). Azt az összefüggést, ami a láncon belüli és a láncok közötti kontaktusok aránya, valamint a dimerképződés mechanizmusa között fennáll [121], a kétrétegű modellben nem sikerült kimutatnom. Az ábrán látható eredményeket olyan számításokból kaptam, ahol a natív állapot nyelő volt. Ha a rendszer az egyensúly felé közeledett, kvalitatíve megegyező görbéket kaptam, tehát a dominanciaviszonyok nem függték attól, hogy a natív állapot nyelő vagy nem.

Mind a fluxus – idő-görbe, mint az átmenetiútvonalelmélet segítségével számított egyensúlyi és steady-state fluxusok azt mutatják, hogy általában egy adott szekvencia esetében mind a három mechanizmus jelen van, de az hogy milyen mértékben, nemcsak a szekvenciától függ, hanem változhat az időben is, és jelentősen függ az alkalmazott határfeltételektől, pl. a kiindulási állapottól. Habár az egyes dimerképződési mechanizmusoknak a fluxus – idő-görbék alapján kapott és az egyensúlyi, valamint a steady-state reaktív fluxusok alapján számított relatív fontosságai kvalitatíve nagyjából megegyeznek, kvantitatíve bizonyos különbségek figyelhetők meg közöttük. A kvantitatív eltérések felhívják a figyelmet arra, hogy az eredmények értelmezésénél nem szabad figyelmen kívül hagyni azt, hogy pontosan milyen feltételek között végeztük a számításokat, ill. a méréseket.

## Steady-state kontra egyensúly

A hagyományos kinetikai leírások azon a feltételezésen alapulnak, hogy az elemi reakciókban részt vevő reaktáns és termék állapotok a rendszernek metastabilis állapotai. Ebben az esetben a sebességi együtthatók függetlenek a rendszer makroállapotától. A metastabilis állapotokra az jellemző, hogy az egyes metastabilis állapotok közötti átmenetekhez szükséges átlagos idő jóval nagyobb, mint ami ahhoz szükséges, hogy a metastabilis állapoton belül egy részleges egyensúly alakuljon ki, tehát a metastabilis állapotot alkotó mikroállapotok egymáshoz viszonyított valószínűségei (koncentrációi) megfeleljenek az egyensúlyi helyzetben megfigyelhetőnek. Tehát egy metastabilis állapot a releváns időskálán úgy viselkedik, mintha egyensúly volna.

A fehérjék azonban komplex rendszerek, bonyolult szabadenergia-tájképpel, sok lokális minimummal [131]. Gyakran ezek a lokális minimumok – legalábbis látszólag – kevés számú metastabilis állapotot alkotnak ugyan, ez alapján lehet beszélni pl. két- és háromállapotú dimerekről [78], azonban nem mindig egyértelmű, hogy pontosan mik is a metastabilis állapotok. A metastabilis állapotok azonosítására hagyományosan a szabadenergia-tájkép elemzése volt a bevett módszer. Dimerképződés vizsgálatokor valamilyen felgombolyodási és valamilyen kötődési reakciókoordináta értékének függvényében ábrázolták a szabadenergiát. A szabadenergia-tájkép elemzésnek azonban hiányossága, hogy a dimenziócsökkentés miatt bizonyos, egyébként kinetikailag szeparált állapotok a reakciókoordináta-tér azonos pontjaira kerülhetnek. Ennek következtében előfordulhat, hogy a megfigyelt szabadenergia-gödrök valójában nem metastabilis állapotoknak felelnek meg, vagy több metastabilis állapotot is magukban foglalnak [132, 133].

Ha a kinetikai sémában feltüntetett állapotok a rendszernek nem metastabilis állapotai, vagy nem ismert, hogy metastabilis állapotok-e, akkor is van lehetőség a rendszer kinetikai elemzésére. Az átmenetiútvonal-elmélet [93] segítségével tetszőleges előre definiált reaktánshalmazból vizsgálható az átmenet sebessége tetszőleges, a reaktánshalmazzal át nem fedő termékhalmozba. Az így kapott reaktív fluxusból csak akkor lehet sebességi együtthatót számolni, ha a reaktánshalmaz metastabilis, ill. ha a rendszer egyensúlyban vagy steady-state állapotban van.

Az élő szervezetek nem egyensúlyi rendszerek, de az élő sejtek sem azok. Sok folyamat azonban megközelítőleg steady-state állapotúnak tekinthető. Feltételeztem, hogy a fehérje, aminek a dimerképződését vizsgáltam, folyamatosan képződik a sejtben, tehát a nyers, a fehérjeszintetizáló apparátusról lekerülő láncok mennyisége állandó. Ekkor, ha ezt az állapotot tekintem reaktánshalmaznak, akkor ki tudom számolni a dimerképződés sebességi együtthatóját.

A 23. ábrán látható egy olyan kinetikai séma, amelyen az egyensúlyi, és egy olyan, amelyen a steady-state reaktív fluxusok vannak feltüntetve. Ha összehasonlítjuk a két sémát, láthatjuk, hogy az  $uaun \rightarrow faun$ , ill. az  $faun \rightarrow uaun$  átmenethez tartozó reaktív fluxusok jelentősen különböznek egyensúlyi és steady-state állapotban. Ilyen különbségek más szekvenciáknál is megfigyelhetők. Ezek alapján, habár az  $udu \rightarrow faf$  átmenet sebességi együtthatóit nem befolyásolta jelentősen, hogy egyensúlyi vagy steady-state állapotban vizsgáltuk a rendszert, nincs arra garancia, hogy az *in vitro* végzett egyensúlyi kísérletek jól leírják a sejtben steady-state körülmények között lezajló folyamatokat. A jövőben tehát kísérletek végzése során érdemes tekintettel lenni erre, és megkísérelni a sejtben belüli steady-state körülményekhez hasonló körülményeket létrehozni a kémcsőben is.

## Metastabilis állapotok

Az átmeneti mátrix spektrális elemzésén alapuló Perron-klaszter Klaszterelemzés módszere segítségével megkíséréltem azonosítani a rendszer metastabilis állapotait. A kapott eredmények azt mutatják (27. ábra), hogy a vizsgált rendszerek nem rendelkeznek kisszámú, valós metastabilis állapottal, hanem sok lokális szabadenergia-minimummal rendelkeznek. A szökési idők az egyes lokális minimumokból nagyjából folytonos eloszlásúak. A rendszerek szabadenergia-felszíneit egy durva falú tölcsérhez lehet hasonlítani.

## Derék és nemderék szekvenciák dimerképzésének sebessége – az előképzett szerkezetek szerepe

A jelátviteli folyamatokban fehérje-fehérje komplexek gyors képződésére és elbomlására van szükség. Ezekben a folyamatokban a rendezetlen fehérjék nagy arányban képviseltetik magukat [17, 30], tehát szükséges, hogy képesek legyenek gyorsan kialakítani a komplexeiket. Gyakori jelenség, hogy a szabad állapotban rendezetlen fehérje komplexben rendezett szerkezetet vesz fel [25], tehát a kötődésen felül a lánc rendeződésének is le kell játszódnia a rendelkezésre álló rövid idő alatt. A fehérjék rendezetlenségének egyik előnye a rendezetlen fehérjéknek a rendezett fehérjékéhez viszonyítva gyors kötődése. Habár úgy tűnik, valóban gyorsabb kötődésre képesek a globuláris fehérjéknél [60], de a különbség csekély, mindössze 1,6-szoros [33]. Ez tehát azt jelenti, hogy a lánc rendeződésének is gyorsnak kell lennie. A gyors rendeződés egyik lehetséges magyarázata abból a tényből táplálkozik, hogy rendezetlen fehérjék még szabad állapotban sem mindig teljesen rendezettek, hanem bizonyos reziduális szerkezettel rendelkeznek. Felmerült tehát, hogy az előképződött szerkezeti elemek kötőhelyekként szolgálnak [64], és mivel már

szabad állapotban is jelen vannak, nem kell azzal időt tölteni, hogy rendeződjenek a kötődés során.

A derék szekvenciák esetében, a natív dimerbeli konformáció monomer állapotban is viszonylag stabil. Ez tehát megfeleltethető előképzett szerkezeti elemek [64] kialakulásának a kötőpartner távollétében. Összehasonlítva a derék és a nem derék szekvenciák egyensúlyi és steady-state fluxusait, azt tapasztaljuk, hogy a stabil monomer, tehát az előképzett szerkezeti elemek jelenléte nem növelte jelentősen a dimerkialakulás sebességét. Sőt, azok a szekvenciák, amelyek feltűnően lassan érik el a natív állapotukat, derék szekvenciák (6., 4. és 1. táblázat).

### **Aszimmetria**

A homodimerképződés elméleti leírásai egy-két kivételtől eltekintve nem vizsgálták a homodimerképződés szimmetriáját; kivételként említhető pl. [121]. A homodimereknek két- és három-állapotú dimerekre történő felosztása, és az ezeknek megfeleltetett dimerképződési kinetikák impliciten feltételezték, hogy a folyamat során a két lánc szimmetrikusan viselkedik. A kötődési fluxustájkép, és az átlagos aszimmetria változásának időbeli vizsgálata azt mutatták, hogy a két lánc az esetek nagy részében nem teljesen szimmetrikusan viselkedik. Bizonyos szekvenciák esetében az is előfordulhat, hogy a folyamat kezdetén az átlagos aszimmetria jelentősen meghaladja az egyensúlyi állapotban várható értéket (26. ábra).

### **Az alkalmazott megközelítés korlátai**

Annak ellenére, hogy az állapottér kis mérete nagyszámú szekvencia vizsgálatát teszi lehetővé a HP négyzetrácsmodell esetében, kevés az olyan szekvencia, amely egyedi alapállapottal rendelkezik. További nehézség, hogy egzakt számítások csak nagyon limitált lánchosszúságra végezhetők két láncból álló rendszerek esetén. Az egyszerűségéből kifolyólag az alkalmazott modell kevésbé reális. A modellrendszerek energiefelszínei rendkívül durvák, ami szintén eltér a valós fehérjék és realisztikusabb modellek viselkedésétől.

A felsorolt hátrányai ellenére érdemes ilyen egyszerűsített modelleket alkalmazni, mert általában, hogy egzakt számítások végezhetők, a kapott eredmények statisztikusan fizikailag korrektek, és általános következtetések levonására alkalmasak. Az alkalmazott modell további előnye, hogy nem alkalmaz olyan egyszerűsítéseket, mint pl. a Gō-modell, és lehetővé teszi két rendezetlen lánc kötődés során végbemenő kölcsönös felgombolyodásának vizsgálatát

# Összefoglalás

Ismert jelenség, hogy bizonyos rendezetlen peptidláncok biológiai partnerükhöz való kötődés során rendezett szerkezetet vesznek föl. Az is ismert, hogy bizonyos rendezetlen peptidek rendezett homodimereket képeznek. A folyamat elméleti leírása az állapottér nagy mérete miatt nehéz. Történtek kísérletek rendezetlen peptidek homodimerképzésének számítógépes vizsgálatára, de a vizsgálatokat ún. szerkezet alapú modelleken végezték, amelyek a natív állapotra vonatkozó külső információt vittek be a mintavételezési folyamat befolyásolására. Bizonyos egyszerűsítéseket tettem abból a célból, hogy a folyamat egzakt leírását adhassam legalább rövid láncokra, ügyelve arra, hogy az általam alkalmazott egyszerűsített modellek viselkedését semmilyen, a natív állapotra vonatkozó, kívülről bevitt információ ne befolyásolja.

- Mivel az állapottér nagy méretéért tekintélyes részben a két lánc lehetséges relatív pozícióinak nagy száma felelős, ezért számuk csökkentésével az állapottér méretének nagymértékű redukciója érhető el. Feltételeztem egyrészt, hogy a diffúziós és rotációs mozgások sebessége lényegesen meghaladja a láncok konformációs átalakulásainak sebességét, másrészt, hogy az egy konformációpárhoz tartozó asszociált állapotok közül csak a legalacsonyabb energiájú állapot populált szignifikánsan. A két feltételezés alapján egy kétrétegű hálózatmodellt definiáltam, amely lehetővé tette  $l = 8$  lánc hosszúságig egzakt számítások végzését.
- A kötődéshez kapcsolt konformációváltás leírására alkalmazott modellek (indukált illeszkedés, konformációkiválasztás) egyetlen lánc rendeződésének, ill. konformációváltásának leírására alkalmasak. Két flexibilis lánc esetén a folyamat során mindkét lánc konformációs állapotát tekintetbe kell venni. Úgy bővítettem ki tehát a hagyományos fogalomkészletet, hogy az mindkét lánc felgombolyodása és a köztük végbemenő kötődés időbeli viszonyainak leírására alkalmas legyen. A kibővített fogalomkészlet szerinti leírásmód alapja a két lánc felgombolyodottsági állapota az asszociáció pillanatában. Ha mindkét lánc denaturált állapotban van a kötődés pillanatában, indukált felgombolyodásról beszélünk, ha csak az egyik, a dimerképződés mechanizmusa konformációkiválasztás, míg ha mindkét lánc natív állapotban van, akkor a dimerképződés merev dokkolással történik.
- A kétrétegű hálózaton végzett szimulációk és egyensúlyi valamint steady-state állapotban, az átmenetiútval-elmélet segítségével megvizsgáltam, hogy az egyes mechanizmusok milyen szerepet játszanak a különböző szekvenciájú láncok dimerképződésében. Mind a szimulációk, mind az átmenetiútval-elmélet alapján számított fluxusok azt mutatták, hogy minden szekvenciánál mindhárom mechanizmus szerepet kap, de különböző mértékben. A domináns mechanizmus időben is változhat. Általánosságban elmondható, hogy azoknál a szekvenciáknál, ahol a monomer szabad állapotban stabilis, a merev dokkolás

vagy a konformációkiválasztás dominál, míg azoknál a szekvenciáknál, ahol a monomer nem stabilis, inkább az indukált felgombolyodás a jellemző mechanizmus.

- Homodimerek esetén a dimert alkotó két lánc szekvenciája azonos. Azt várjuk tehát, hogy a két lánc a dimerképződés során is szimmetrikusan fog viselkedni. Az általam vizsgált szekvenciák legnagyobb részének dimer alapállapota szimmetrikus volt, azaz a két lánc konformációja megegyezett a natív állapotban. A szimmetrikus alapállapotú szekvenciák esetében is a két lánc viselkedésében a dimerképződés során különbség volt. A monomerek konformációi között definiált távolság segítségével a kétrétegű modellen végzett szimulációk során nyomon követtem az adott időpontra jellemző asszimetriát. A várakozásokkal ellentétben bizonyos mértékű aszimmetria mindegyik szekvenciára jellemző volt, de egyes szekvenciák jelentős aszimmetriát mutattak a dimerképződési folyamat kezdetén.
- A munka fontos tanulsága volt, hogy a hagyományos kinetikai leírásmód csak akkor alkalmazható a vizsgált rendszerre, ha a kinetikai leírásban definiált állapotok a rendszernek valódi metastabilis állapotai. Ha az állapotok a rendszernek nem valódi metastabilis állapotai, akkor a sebességi együtthatók nem függetlenek a pillanatnyi állapottól, amit a hagyományos leírásmód egyébként feltételez. Nem valódi metastabilis állapotok esetén a sebességi együtthatók időbeli változatlansága úgy biztosítható, ha a rendszert valamilyen steady-state állapotban vizsgáljuk. Az átmeneti mátrix spektrális elemzésének segítségével feltártam az állapothálózat belső kinetikai szerkezetét, és azt találtam, hogy az általam vizsgált rendszerek állapottere nem osztható kisszámú metastabilis állapotra, hanem nagyszámú lokális szabadenergiaminimummal rendelkezik, amelyek nem feleltethetők meg az általam előre definiált mezoállapotoknak. Az előre definiált mezoállapotokra felírt kinetikai séma tehát csak egyensúlyi vagy steady-state állapotban érvényes.

# Köszönetnyilvánítás

Szeretném megköszönni mindazon emberek támogatását, akik nélkül e munka nem jöhetett volna létre. Mindenekelőtt munkámat befogadó intézmények vezetőinek, Dr. Keserű György Miklósnak, az MTA TTK főigazgatójának, Dr. Buday Lászlónak, az MTA TTK Enzimológiai Intézet igazgatójának, hogy a munkámnak háttérrel biztosítottak. Köszönettel tartozom a PPKE ITK doktori iskola vezetőségének, hogy lehetővé tették számomra, hogy a doktori képzésben részt vegyek. Ezen kívül külön köszönettel tartozom Dr. Vida Tivadarnénak, aki a PPKE ITK adminisztrációs ügyei közötti eligazodásomat segítette. Mindenekelőtt azonban köszönettel tartozom témavezetőimnek, Dr. Szilágyi Andrásnak és Dr. Závodszy Péternek a hozzáértő és türelmes szakmai segítségért és támogatásért. Köszönet illeti meg a Szerkezeti Biofizika Kutatócsoport azon tagjait is, akik bármilyen segítséggel hozzájárultak a munkám sikeréhez. Végül köszönettel tartozom szüleimnek, hogy a tanulmányaimat támogatták és a hölgyeknek, akikkel töltött időből energiát meríthettem a munkámhoz.



# Az értekezés alapjául szolgáló publikációk

1. Szilágyi A, Györfy D és Závodszy P. „The twilight zone between protein order and disorder.” *Biophys J*, **2008**;95:1612–26.
2. Györfy D, Závodszy P és Szilágyi A. „„Pull moves” for rectangular lattice polymer models are not fully reversible.” *IEEE/ACM Trans Comput Biol Bioinform*, **2012**;9:1847–9.

# **Nemzetközi konferencián bemutatott poszter**

Györfly D, Závodszy P és Szilágyi A.: „The blind leading the blind: how disordered peptides form an ordered complex”. 3rd Prague Protein Spring meeting, Prága, 2014

# Irodalomjegyzék

- [1] Anfinsen CB és Haber E. „Studies on the reduction and re-formation of protein disulfide bonds.” *J Biol Chem*, **1961**;236:1361–3.
- [2] Anfinsen CB. „Principles that govern the folding of protein chains.” *Science*, **1973**; 181:223–30.
- [3] Levinthal C. „Are there pathways for protein folding?” *J Chim Phys*, **1968**;65:44–45.
- [4] Levinthal C. „How to Fold Graciously”. In JTP Debrunner és E Munck (Szerk.) „Mossbauer Spectroscopy in Biological Systems: Proceedings of a meeting held at Allerton House, Monticello, Illinois”, University of Illinois Press, 22–24.
- [5] Harrison SC és Durbin R. „Is there a single pathway for the folding of a polypeptide chain?” *Proceedings of the National Academy of Sciences*, **1985**;82(12):4028–4030.
- [6] Bryngelson JD és Wolynes PG. „Intermediates and barrier crossing in a random energy model (with applications to protein folding)”. *J Phys Chem*, **1989**;93(19):6902–6915.
- [7] Bryngelson JD, Onuchic JN, Socci ND és Wolynes PG. „Funnels, pathways, and the energy landscape of protein folding: a synthesis.” *Proteins*, **1995**;21:167–95.
- [8] Bryngelson JD és Wolynes PG. „Spin glasses and the statistical mechanics of protein folding.” *Proc Natl Acad Sci U S A*, **1987**;84:7524–8.
- [9] Dill KA és Chan HS. „From Levinthal to pathways to funnels.” *Nat Struct Biol*, **1997**; 4:10–9.
- [10] Buchler NE és Goldstein RA. „Universal correlation between energy gap and foldability for the random energy model and lattice proteins”. *The Journal of chemical physics*, **1999**; 111(14):6599–6609.
- [11] Tsai CJ, Kumar S, Ma B és Nussinov R. „Folding funnels, binding funnels, and protein function.” *Protein Sci*, **1999**;8:1181–90.
- [12] Tsai CJ, Xu D és Nussinov R. „Protein folding via binding and vice versa.” *Fold Des*, **1998**;3:R71–80.
- [13] Chu X, Gan L, Wang E és Wang J. „Quantifying the topography of the intrinsic energy landscape of flexible biomolecular recognition.” *Proc Natl Acad Sci U S A*, **2013**; 110:E2342–51.

- [14] Tompa P. „Intrinsically unstructured proteins.” *Trends Biochem Sci*, **2002**;27:527–33.
- [15] Wright PE és Dyson HJ. „Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm.” *J Mol Biol*, **1999**;293:321–31.
- [16] Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, Oldfield CJ, Campen AM, Ratliff CM, Hipps KW, Ausio J, Nissen MS, Reeves R, Kang C, Kissinger CR, Bailey RW, Griswold MD, Chiu W, Garner EC és Obradovic Z. „Intrinsically disordered protein.” *J Mol Graph Model*, **2001**;19:26–59.
- [17] Tompa P. *Structure and function of intrinsically disordered proteins*. Chapman & Hall/CRC, 1st kiadás, **2010**.
- [18] Redinbo MR, Stewart L, Kuhn P, Champoux JJ és Hol WGJ. „Crystal structures of human topoisomerase I in covalent and noncovalent complexes with DNA”. *Science*, **1998**;279(5356):1504–1513.
- [19] Kissinger CR, Parge HE, Knighton DR, Lewis CT, Pelletier LA, Tempczyk A, Kalish VJ, Tucker KD, Showalter RE, Moomaw EW, Gastinel LN, Habuka N, Chen X, Maldonado F, Barker JE, Bacquet R és Villafranca JE. „Crystal structures of human calcineurin and the human fkbp12-fk506-calcineurin complex”. *Nature*, **1995**;378(6557):641–644.
- [20] Hernández MA, Avila J és Andreu JM. „Physicochemical characterization of the heat-stable microtubule-associated protein map2”. *European Journal of Biochemistry*, **1986**;154(1):41–48.
- [21] Schweers O, Schönbrunn-Hanebeck E, Marx A és Mandelkow E. „Structural studies of tau protein and Alzheimer paired helical filaments show no evidence for  $\beta$ -structure.” *Journal of Biological Chemistry*, **1994**;269(39):24290–24297.
- [22] Pometun MS, Chekmenev EY és Wittebort RJ. „Quantitative observation of backbone disorder in native elastin”. *Journal of Biological Chemistry*, **2004**;279(9):7982–7987.
- [23] Tompa P és Csermely P. „The role of structural disorder in the function of RNA and protein chaperones”. *The FASEB Journal*, **2004**;18(11):1169–1175.
- [24] Dunker AK, Cortese MS, Romero P, Iakoucheva LM és Uversky VN. „Flexible nets”. *FEBS Journal*, **2005**;272(20):5129–5148.
- [25] Dyson HJ és Wright PE. „Coupling of folding and binding for unstructured proteins.” *Curr Opin Struct Biol*, **2002**;12:54–60.
- [26] Bowie JU és Sauer RT. „Equilibrium dissociation and unfolding of the Arc repressor dimer”. *Biochemistry*, **1989**;28:7139–43.
- [27] Romero R, Obradovic Z, Kissinger CR, Villafranca JE, Garner E, Guillot S és Dunker AK. „Thousands of proteins likely to have long disordered regions”. *Pacific Symp Bio-computing*, **1998**;3:437–448.
- [28] Hashimoto M, Ichimura T, Mizoguchi H, Tanaka K, Fujimitsu K, Keyamura K, Ote T, Yamakawa T, Yamazaki Y, Mori H, Katayama T és Kato Ji. „Cell size and nucleoid organization of engineered *Escherichia coli* cells with a reduced genome”. *Molecular Microbiology*, **2005**;55(1):137–149.

- [29] Giaever G, Chu AM, Ni L, Connelly C, Riles L, Veronneau S, Dow S, Lucau-Danila A, Anderson K, Andre B, Arkin AP, Astromoff A, El Bakkoury M, Bangham R, Benito R, Brachat S, Campanaro S, Curtiss M, Davis K, Deutschbauer A, Entian KD, Flaherty P, Foury F, Garfinkel DJ, Gerstein M, Gotte D, Guldener U, Hegemann JH, Hempel S, Herman Z, Jaramillo DF, Kelly DE, Kelly SL, Kotter P, LaBonte D, Lamb DC, Lan N, Liang H, Liao H, Liu L, Luo C, Lussier M, Mao R, Menard P, Ooi SL, Revuelta JL, Roberts CJ, Rose M, Ross-Macdonald P, Scherens B, Schimmack G, Shafer B, Shoemaker DD, Sookhai-Mahadeo S, Storms RK, Strathern JN, Valle G, Voet M, Volckaert G, Wang Cy, Ward TR, Wilhelmy J, Winzeler EA, Yang Y, Yen G, Youngman E, Yu K, Bussey H, Boeke JD, Snyder M, Philippsen P, Davis RW és Johnston M. „Functional profiling of the *Saccharomyces cerevisiae* genome”. *Nature*, **2002**;418(6896):387–391.
- [30] Xie H, Vucetic S, Iakoucheva LM, Oldfield CJ, Dunker AK, Uversky VN és Obradovic Z. „Functional anthology of intrinsic disorder. 1. biological processes and functions of proteins with long disordered regions.” *J Proteome Res*, **2007**;6:1882–98.
- [31] Iakoucheva LM, Brown CJ, Lawson JD, Obradović Z és Dunker AK. „Intrinsic disorder in cell-signaling and cancer-associated proteins.” *J Mol Biol*, **2002**;323:573–84.
- [32] Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, Obradović Z és Dunker AK. „Intrinsic disorder and protein function”. *Biochemistry*, **2002**;41:6573–6582.
- [33] Dogan J, Gianni S és Jemth P. „The binding mechanisms of intrinsically disordered proteins.” *Phys Chem Chem Phys*, **2014**;16:6323–31.
- [34] Kriwacki RW, Hengst L, Tennant L, Reed SI és Wright PE. „Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity”. *Proceedings of the National Academy of Sciences*, **1996**; 93(21):11504–11509.
- [35] Oldfield CJ, Meng J, Yang JY, Yang MQQ, Uversky VN és Dunker AK. „Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners.” *BMC genomics*, **2008**;9 Suppl 1:S1.
- [36] Uversky VN, Gillespie JR és Fink AL. „Why are "natively unfolded" proteins unstructured under physiologic conditions?” *Proteins*, **2000**;41:415–27.
- [37] Weathers EA, Paulaitis ME, Woolf TB és Hoh JH. „Reduced amino acid alphabet is sufficient to accurately recognize intrinsically disordered protein.” *FEBS Lett*, **2004**;576:348–52.
- [38] Coeytaux K és Poupon A. „Prediction of unfolded segments in a protein sequence based on amino acid composition.” *Bioinformatics*, **2005**;21:1891–900.
- [39] Romero P, Obradovic Z, Li X, Garner EC, Brown CJ és Dunker AK. „Sequence complexity of disordered protein.” *Proteins*, **2001**;42:38–48.
- [40] Linding R, Russell RB, Neduva V és Gibson TJ. „Globplot: Exploring protein sequences for globularity and disorder.” *Nucleic Acids Res*, **2003**;31:3701–8.

- [41] Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF és Jones DT. „Prediction and functional analysis of native disorder in proteins from the three kingdoms of life.” *J Mol Biol*, **2004**; 337:635–45.
- [42] Li X, Romero P, Rani M, Dunker AK és Obradovic Z. „Predicting protein disorder for n-, c-, and internal regions.” *Genome Inform Ser Workshop Genome Inform*, **1999**;10:30–40.
- [43] Obradovic Z, Peng K, Vucetic S, Radivojac P, Brown CJ és Dunker AK. „Predicting intrinsic disorder from amino acid sequence.” *Proteins*, **2003**;53 Suppl 6:566–72.
- [44] Melamud E és Moutl J. „Evaluation of disorder predictions in casp5.” *Proteins*, **2003**;53 Suppl 6:561–5.
- [45] Szilágyi A, Györfly D és Závodszy P. „The twilight zone between protein order and disorder.” *Biophys J*, **2008**;95:1612–26.
- [46] Fischer E. „Einfluss der configuration auf die Wirkung der Enzyme”. *Ber Dtsch Chem Ges*, **1894**;27:2985–93.
- [47] Sullivan SM és Holyoak T. „Enzymes with lid-gated active sites must operate by an induced fit mechanism instead of conformational selection”. *Proceedings of the National Academy of Sciences*, **2008**;105(37):13829–13834.
- [48] Koshland DE. „Application of a theory of enzyme specificity to protein synthesis.” *Proc Natl Acad Sci U S A*, **1958**;44:98–104.
- [49] Straub FB. „Formation of the secondary and tertiary structure of enzymes.” *Adv Enzymol Relat Areas Mol Biol*, **1964**;26:89–114.
- [50] Boehr DD, Nussinov R és Wright PE. „The role of dynamic conformational ensembles in biomolecular recognition.” *Nat Chem Biol*, **2009**;5:789–96.
- [51] Vértessy BG és Orosz F. „From "fluctuation fit" to "conformational selection": evolution, rediscovery, and integration of a concept.” *Bioessays*, **2011**;33:30–4.
- [52] Tsai CJ, Ma B és Nussinov R. „Folding and binding cascades: shifts in energy landscapes.” *Proc Natl Acad Sci U S A*, **1999**;96:9970–2.
- [53] Hammes GG, Chang YC és Oas TG. „Conformational selection or induced fit: a flux description of reaction mechanism.” *Proc Natl Acad Sci U S A*, **2009**;106:13737–41.
- [54] Okazaki K és Takada S. „Dynamic energy landscape view of coupled binding and protein conformational change: induced-fit versus population-shift mechanisms.” *Proc Natl Acad Sci U S A*, **2008**;105:11182–7.
- [55] Vogt AD és Di Cera E. „Conformational selection or induced fit? a critical appraisal of the kinetic mechanism.” *Biochemistry*, **2012**;51:5894–902.
- [56] Zhou HX. „From induced fit to conformational selection: a continuum of binding mechanism controlled by the timescale of conformational transitions.” *Biophys J*, **2010**; 98:L15–7.

- [57] Lange OF, Lakomek NA, Farès C, Schröder GF, Walter KF, Becker S, Meiler J, Grubmüller H, Griesinger C és de Groot BL. „Recognition dynamics up to microseconds revealed from an rdc-derived ubiquitin ensemble in solution.” *Science*, **2008**;320:1471–5.
- [58] Zhang W, Ganguly D és Chen J. „Residual structures, conformational fluctuations, and electrostatic interactions in the synergistic folding of two intrinsically disordered proteins.” *PLoS Comput Biol*, **2012**;8:e1002353.
- [59] Bosshard HR. „Molecular recognition by induced fit: how fit is the concept?” *News Physiol Sci*, **2001**;16:171–3.
- [60] Huang Y és Liu Z. „Kinetic advantage of intrinsically disordered proteins in coupled folding-binding process: a critical assessment of the "fly-casting" mechanism.” *J Mol Biol*, **2009**;393:1143–59.
- [61] Sivakolundu SG, Bashford D és Kriwacki RW. „Disordered p27kip1 exhibits intrinsic structure resembling the cdk2/cyclin a-bound conformation”. *Journal of Molecular Biology*, **2005**;353(5):1118 – 1128.
- [62] Song J, Guo LW, Muradov H, Artemyev NO, Ruoho AE és Markley JL. „Intrinsically disordered  $\gamma$ -subunit of cGMP phosphodiesterase encodes functionally relevant transient secondary and tertiary structure.” *Proc Natl Acad Sci U S A*, **2008**;105:1505–10.
- [63] Tsai CJ, Ma B, Sham YY, Kumar S és Nussinov R. „Structured disorder and conformational selection.” *Proteins*, **2001**;44:418–27.
- [64] Fuxreiter M, Simon I, Friedrich P és Tompa P. „Preformed structural elements feature in partner recognition by intrinsically unstructured proteins.” *J Mol Biol*, **2004**;338:1015–26.
- [65] Chen J. „Intrinsically disordered p53 extreme C-terminus binds to S100B ( $\beta\beta$ ) through "fly-casting"”. *J Am Chem Soc*, **2009**;131(6):2088–9.
- [66] Taketomi H, Ueda Y és Gō N. „Studies on protein folding, unfolding and fluctuations by computer simulation. i. the effect of specific amino acid sequence represented by specific inter-unit interactions.” *Int J Pept Protein Res*, **1975**;7:445–59.
- [67] Sugase K, Dyson HJ és Wright PE. „Mechanism of coupled folding and binding of an intrinsically disordered protein.” *Nature*, **2007**;447:1021–5.
- [68] Ganguly D és Chen J. „Topology-based modeling of intrinsically disordered proteins: balancing intrinsic folding and intermolecular interactions.” *Proteins*, **2011**;79:1251–66.
- [69] Verkhivker GM, Bouzida D, Gehlhaar DK, Rejto PA, Freer ST és Rose PW. „Simulating disorder-order transitions in molecular recognition of unstructured proteins: where folding meets binding.” *Proc Natl Acad Sci U S A*, **2003**;100:5148–53.
- [70] Shoemaker BA, Portman JJ és Wolynes PG. „Speeding molecular recognition by using the folding funnel: the fly-casting mechanism.” *Proc Natl Acad Sci U S A*, **2000**;97:8868–73.
- [71] Narayanan R, Ganesh OK, Edison AS és Hagen SJ. „Kinetics of folding and binding of an intrinsically disordered protein: the inhibitor of yeast aspartic proteinase YPrA.” *J Am Chem Soc*, **2008**;130:11477–85.

- [72] Ganguly D, Otieno S, Waddell B, Iconaru L, Kriwacki RW és Chen J. „Electrostatically accelerated coupled binding and folding of intrinsically disordered proteins.” *J Mol Biol*, **2012**;422:674–84.
- [73] Ganguly D, Zhang W és Chen J. „Electrostatically accelerated encounter and folding for facile recognition of intrinsically disordered proteins.” *PLoS Comput Biol*, **2013**; 9:e1003363.
- [74] Ganguly D, Zhang W és Chen J. „Synergistic folding of two intrinsically disordered proteins: searching for conformational selection.” *Mol Biosyst*, **2012**;8:198–209.
- [75] Csermely P, Palotai R és Nussinov R. „Induced fit, conformational selection and independent dynamic segments: an extended view of binding events.” *Trends Biochem Sci*, **2010**; 35:539–46.
- [76] Demarest SJ, Martinez-Yamout M, Chung J, Chen H, Xu W, Dyson HJ, Evans RM és Wright PE. „Mutual synergistic folding in recruitment of cbp/p300 by p160 nuclear receptor coactivators.” *Nature*, **2002**;415:549–53.
- [77] Higo J, Nishimura Y és Nakamura H. „A free-energy landscape for coupled folding and binding of an intrinsically disordered protein in explicit solvent from detailed all-atom computations.” *J Am Chem Soc*, **2011**;133:10448–58.
- [78] Neet KE és Timm DE. „Conformational stability of dimeric proteins: quantitative studies by equilibrium denaturation.” *Protein Sci*, **1994**;3:2167–74.
- [79] Tsai CJ, Xu D és Nussinov R. „Structural motifs at protein-protein interfaces: protein cores versus two-state and three-state model complexes.” *Protein Sci*, **1997**;6:1793–805.
- [80] Gunasekaran K, Tsai CJ és Nussinov R. „Analysis of ordered and disordered protein complexes reveals structural features discriminating between stable and unstable monomers.” *J Mol Biol*, **2004**;341:1327–41.
- [81] Mészáros B, Tompa P, Simon I és Dosztányi Z. „Molecular principles of the interactions of disordered proteins.” *J Mol Biol*, **2007**;372:549–61.
- [82] Xu D, Tsai CJ és Nussinov R. „Mechanism and evolution of protein dimerization.” *Protein Sci*, **1998**;7:533–44.
- [83] Papoian GA és Wolynes PG. „The physics and bioinformatics of binding and folding—an energy landscape perspective.” *Biopolymers*, **2003**;68:333–49.
- [84] Unger R és Moult J. „Finding the lowest free energy conformation of a protein is an np-hard problem: proof and implications.” *Bull Math Biol*, **1993**;55:1183–98.
- [85] Fraenkel AS. „Complexity of protein folding.” *Bull Math Biol*, **1993**;55:1199–210.
- [86] Skolnick J, Kolinski A és Yaris R. „Monte Carlo simulations of the folding of  $\beta$ -barrel globular proteins.” *Proc Natl Acad Sci U S A*, **1988**;85:5057–61.
- [87] Lau KF és Dill KA. „A lattice statistical mechanics model of the conformational and sequence spaces of proteins”. *Macromolecules*, **1989**;22:3986–3997.



- [88] Clementi C, Nymeyer H és Onuchic JN. „Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? an investigation for small globular proteins.” *J Mol Biol*, **2000**;298:937–53.
- [89] Krantz AT. „Analysis of an efficient algorithm for the hard-sphere problem”. *ACM Trans Model Comput Simul*, **1996**;6(3):185–209.
- [90] Dokholyan NV, Buldyrev SV, Stanley HE és Shakhnovich EI. „Discrete molecular dynamics studies of the folding of a protein-like model.” *Fold Des*, **1998**;3:577–87.
- [91] Noé F és Fischer S. „Transition networks for modeling the kinetics of conformational change in macromolecules.” *Curr Opin Struct Biol*, **2008**;18:154–62.
- [92] Deuffhard P. „Identification of almost invariant aggregates in reversible nearly uncoupled markov chains”. *Linear Algebra and its Applications*, **2000**;315(1-3):39–59.
- [93] E W és Vanden-Eijnden E. „Transition-path theory and path-finding algorithms for the study of rare events.” *Annu Rev Phys Chem*, **2010**;61:391–420.
- [94] Metzner P, Schütte C és Vanden-Eijnden E. „Transition path theory for markov jump processes”. *Multiscale Model Simul*, **2009**;7:1192–1219.
- [95] Bovier A. „Metastability”. In „Methods of contemporary mathematical statistical physics”, Springer-Verlag, Berlin, **2009**;177–221.
- [96] Bovier A, Eckhoff M, Gayrard V és Klein M. „Metastability and low lying spectra in reversible markov chains”. *Commun Math Phys*, **2000**;228:219–255.
- [97] Sindelar CV, Hendsch ZS és Tidor B. „Effects of salt bridges on protein structure and design.” *Protein Sci*, **1998**;7:1898–914.
- [98] Lesh N, Mitzenmacher M és Whitesides S. „A complete and effective move set for simplified protein folding”. In „Proceedings of the seventh annual international conference on Research in computational molecular biology”, RECOMB '03. ACM, New York, NY, USA, 188–195. URL <http://doi.acm.org/10.1145/640075.640099>.
- [99] Kou SC, Oh J és Wong WH. „A study of density of states and ground states in hydrophobic-hydrophilic protein folding models by equi-energy sampling.” *J Chem Phys*, **2006**;124:244903.
- [100] Györfy D, Závodszy P és Szilágyi A. „"Pull moves" for rectangular lattice polymer models are not fully reversible.” *IEEE/ACM Trans Comput Biol Bioinform*, **2012**;9:1847–9.
- [101] Binder K. *Monte Carlo and molecular dynamics simulations in polymer science*. Oxford University Press, New York, Oxford, **1995**.
- [102] Marsland S. *Machine Learning: An Algorithmic Perspective*. Chapman & Hall/CRC, 1st kiadás, **2009**.

- [103] Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH és Teller E. „Equation of state calculations by fast computing machines”. *Journal of Chemical Physics*, **1953**;21:1087–1092.
- [104] Hastings WK. „Monte Carlo sampling methods using Markov chains and their applications”. *Biometrika*, **1970**;57(1):97–109.
- [105] Wang F és Landau DP. „Efficient, multiple-range random walk algorithm to calculate the density of states.” *Phys Rev Lett*, **2001**;86:2050–3.
- [106] Wang F és Landau DP. „Determining the density of states for classical statistical models: a random walk algorithm to produce a flat histogram.” *Phys Rev E Stat Nonlin Soft Matter Phys*, **2001**;64:056101.
- [107] Prinz JH, Wu H, Sarich M, Keller B, Senne M, Held M, Chodera JD, Schütte C és Noé F. „Markov models of molecular kinetics: generation and validation.” *J Chem Phys*, **2011**; 134:174105.
- [108] Becker OM és Karplus M. „The topology of multidimensional potential energy surfaces: Theory and application to peptide structure and kinetics”. *The Journal of Chemical Physics*, **1997**;106(4):1495–1517.
- [109] Heilmann OJ és Rotne J. „Exact and Monte Carlo computations on a lattice model for change of conformation of a polymer”. *J Stat Phys*, **1982**;27:19–35.
- [110] Madras N és Sokal AD. „Nonergodicity of local, length-conserving Monte Carlo algorithms for the self-avoiding walk”. *J Stat Phys*, **1987**;47:573–595.
- [111] Madras N és Sokal AD. „The pivot algorithm: a highly efficient Monte Carlo method for the self-avoiding walk”. *J Stat Phys*, **1988**;50:109–186.
- [112] Lal M. „'Monte Carlo' computer simulation of chain molecules I.” *Molecular Physics*, **1969**;17(1):57–64.
- [113] Wüst T, Li YW és Landau DP. „Unraveling the beautiful complexity of simple lattice model polymers and proteins using Wang–Landau sampling”. *J Stat Phys*, **2011**; 144:638–51.
- [114] Deutsch JM. „Long range moves for high density polymer simulations”. *J Chem Phys*, **1997**;106:8849–8854.
- [115] Wüst T és Landau DP. „Versatile approach to access the low temperature thermodynamics of lattice polymers and proteins”. *Phys Rev Lett*, **2009**;102(17):178101.
- [116] Swetnam AD és Allen MP. „Improved simulations of lattice peptide adsorption.” *Phys Chem Chem Phys*, **2009**;11:2046–55.
- [117] Swetnam AD és Allen MP. „Improving the Wang–Landau algorithm for polymers and proteins.” *J Comput Chem*, **2011**;32:816–21.
- [118] Unger R és Moulton J. „Genetic algorithms for protein folding simulations.” *J Mol Biol*, **1993**;231:75–81.

- [119] Chan HS. „Modeling protein density of states: additive hydrophobic effects are insufficient for calorimetric two-state cooperativity.” *Proteins*, **2000**;40:543–71.
- [120] Chan HS és Dill KA. „Protein folding in the landscape perspective: chevron plots and non-arrhenius kinetics.” *Proteins*, **1998**;30:2–33.
- [121] Levy Y, Cho SS, Onuchic JN és Wolynes PG. „A survey of flexible protein binding mechanisms and their transition states using native topology based energy landscapes.” *J Mol Biol*, **2005**;346:1121–45.
- [122] Levy Y, Wolynes PG és Onuchic JN. „Protein topology determines binding mechanism.” *Proc Natl Acad Sci U S A*, **2004**;101:511–6.
- [123] Levy Y, Papoian GA, Onuchic JN és Wolynes PG. „Energy landscape analysis of protein dimers”. *Israel Journal of Chemistry*, **2004**;44(1-3):281–297.
- [124] Gupta N és Irbäck A. „Coupled folding-binding versus docking: a lattice model study.” *J Chem Phys*, **2004**;120:3983–9.
- [125] Tummino PJ és Copeland RA. „Residence time of receptor–ligand complexes and its effect on biological function”. *Biochemistry*, **2008**;47(20):5481–5492.
- [126] Van Aller GS, Nandigama R, Petit CM, DeWolf WE, Quinn CJ, Aubart KM, Zalacain M, Christensen SB, Copeland RA és Lai Z. „Mechanism of time-dependent inhibition of polypeptide deformylase by actinonin”. *Biochemistry*, **2005**;44(1):253–260.
- [127] Held M és Noé F. „Calculating kinetics and pathways of protein-ligand association.” *Eur J Cell Biol*, **2012**;91:357–64.
- [128] Liu J, Li G és Yu J. „Protein-folding simulations of the hydrophobic-hydrophilic model by combining pull moves with energy landscape paving.” *Phys Rev E Stat Nonlin Soft Matter Phys*, **2011**;84:031934.
- [129] Li YW, Wüst T és Landau DP. „Monte Carlo simulations of the HP model (the "Ising model" of protein folding).” *Comput Phys Commun*, **2011**;182:1896–1899.
- [130] Piana S, Lindorff-Larsen K és Shaw DE. „Atomistic description of the folding of a dimeric protein”. *The Journal of Physical Chemistry B*, **2013**;117(42):12935–12942.
- [131] Plotkin SS, Wang J és Wolynes PG. „Correlated energy landscape model for finite, random heteropolymers”. *Phys Rev E*, **1996**;53:6271–6296.
- [132] Krivov SV és Karplus M. „Hidden complexity of free energy surfaces for peptide (protein) folding.” *Proc Natl Acad Sci U S A*, **2004**;101:14766–70.
- [133] Muff S és Caffisch A. „Kinetic analysis of molecular dynamics simulations reveals changes in the denatured state and switch of folding pathways upon single-point mutation of a  $\beta$ -sheet miniprotein”. *Proteins: Structure, Function, and Bioinformatics*, **2008**;70(4):1185–1195.