

Fehérjék térszerkezetének és evolúciójának vizsgálata a mozgékonytű tükrében

- Doktori értekezés tézisei -



Ángyán Annamária Franciska

Kémia Doktori Iskola

vezetője: Dr. Inzelt György, D.Sc.

Szintetikus kémia, Anyagtudomány, Biomolekuláris Kémia Doktori Program

vezetője: Dr. Perczel András, D.Sc.

Témavezető: Dr. Gáspári Zoltán, Ph.D.

Eötvös Loránd Tudományegyetem

Kémiai Intézet, Szerves Kémia Tanszék

Szerkezeti Kémia és Biológia Laboratórium

Budapest

- 2012 -

Bevezetés

A molekuláris biológia területén az elmúlt években több szemléletváltásnak lehettünk tanúi. Egyrészt, a fehérjemolekulák belső, inherens dinamikus mivolta előtérbe kerül, másrészt a fehérjeszerkezet alatt már nem csupán globuláris fehérjéket értjük, hanem a polipeptidlánc további, strukturális és fizikokémiai jellemzői alapján elkülöníthető konformációit; ennek főleg az ún. rendezetlen fehérjék leírásában van szerepe.

Nem csak a térszerkezet, hanem a dinamika is megváltozik a kötődés során, aminek a kötődési szabadentalpia entropikus járulékanak megértésében van jelentősége. Az ún. rendezetlen fehérjék esetében ez látványos, például a kötés közbeni feltekeredés (folding and binding) mechanizmus során, de a globuláris fehérjéknél is fellép ez a jelenség. A konformációs szelekció fontosságát számos publikáció támasztja alá. Ennek ellenére kevés tanulmány van, ahol ezt atomi szinten is bemutatnák a kutatók. A fehérjék térszerkezetének és dinamikájának atomi szintű jellemzésére az NMR spektroszkópia a legalkalmasabb módszer. Nagy előnye, hogy a mérés a fiziológiához némileg hasonló körülmények között, oldatban történik. A NMR minta oldata nagyjából 10^{16} - 10^{17} fehérjemolekulát tartalmaz. Az oldatbeli konformerek mozgékonyosságát és szerkezeti heterogenitását egyetlen konformerrel nem érdemes jellemezni, ugyanis csak térben és időben átlagoltan mérhető minden kísérleti adat. Léteznek protokollok, amelyek különböző időskálákon a fehérjék belső dinamikáját figyelembe veszik. Ilyen típusú szerkezetszámolás eredménye egy ún. dinamikus szerkezeti sokaság, amely a vizsgált makromolekula belső dinamikáját kísérleti paramétereiből származtatva írja le. A fehérjeszerkezetek vizsgálódási köre kiszélesedni látszik: a statikus, egyetlen konformerrel szemben a dinamikus, egyre szélesebb időskálát leíró fehérjemodellek veszik át a terepet, és ez a biológiai folyamatok atomi szinten történő mélyebb megértéséhez elengedhetetlen. A kísérleti adatok értékelése viszont még számos kihívást rejt magában.

A dinamika megváltozhat a kötődés, mutáció vagy akár rendezett-rendezetlen átmenet hatására is. A globuláris szerkezettel szemben a rendezetlen fehérjék funkcionális előnnyel bírnak, a számos kórkép kapcsán leírt fehérjeaggregátumok pedig termodinamikai értelemben stabilabbak. Evolúciós skálán döntő szerepet játszik a fehérjék konformációs változatossága, dinamikája, ami kihat a biológiai rendszerekben betöltött funkciókra. Ha a fehérjék aggregációs hajlama eleve inherensen magas és nem számolunk védekezési mechanizmusokkal, akkor az új fehérjék keletkezése gyakorlatilag lehetetlenné és elhanyagolható jelenséggé válna. Az elmúlt években viszont egyre több kísérleti bizonyíték van, hogy fehérjék ténylegesen keletkeznek *de novo*, azaz spontán átíródás révén is. Ez a kísérleti tény új megvilágításba helyezi a fehérjék evolúcióját, mivel a nemkódoló DNS szakaszok spontán átíródása új fehérje keletkezése mellett sokkal gyakoribb jelenség lehet, mint eredetileg gondolták a kutatók. Felmerül a kérdés, hogy evolúciós

skálán miképpen jöttek létre a fehérjék, hogy milyen térszerkezetet vettek fel az újonnan kialakuló polipeptidláncok, valóban jelentős-e az aggregációra való hajlamuk, és hogy a fehérjék mozgékony-sága és dinamikája milyen szerepet játszott ebben a folyamatban.

Dolgozatomban a fehérjék dinamikáját és evolvabilitását járom körbe a mozgékony-ság tükrében. A dolgozat első részében a dinamikus szerkezeti sokaságokkal kapcsolatos fejlesztéseimet mutatom be. Ez a rész az I., II. és III. saját közleményekre támaszkodik. A dolgozat második részében evolúciós távlatokban vizsgálom a *de novo* fehérjék dinamikus tulajdonságait a rendezetlenségen keresztül. Arra a kérdésre kerestem választ, hogy az aggregációs hajlam befolyásolja-e az újonnan keletkező fehérjék kialakulását és stabilitását. Ez a rész a IV. saját közleményre támaszkodik.

I. rész: Az NMR kísérleti adatok és szerkezeti konformer-sokaságok jellemzése

Célkitűzés

Az első részben a fehérje térszerkezete és dinamikája kapcsolatát jártam körbe. Egyik célom egy NMR orientált fehérjetérszerkezet becslő algoritmus fejlesztése és megvalósítása volt. Másrészt, az NMR adatokból származtatott, a fehérje dinamikáját jellemző paraméterek és a fehérje térszerkezetének megfeleltetését és jellemzését elősegítő protokoll megvalósítása volt.

Az alkalmazott módszerek

- Az RCSB PDB, a RECOORD, és a SCOP ASTRAL adatbázisokat használtam. A PDB és a RECOORD adatbázisokból fehérjeszerkezeteket és kísérleti adatokat használtam fel. A SCOP térszerkezet-oszályozó adatbázist a módszerfejlesztés során használtam.
- A szükséges molekuladinamikai szimulációkat GROMACS programcsomaggal, OPLS-AA erőterrel számoltam. A SHIFTX és a PALES programokat implementáltuk. Az elemzéseket saját PERL programokkal végeztem.
- A fejlesztésekhez használt programokat PERL programnyelven írtam. A szervek PERL és C++ programnyelven vannak megírva. Az eloszlások összevetésére kontingencia-analízist használtam.

Új tudományos eredmények

1. Kifejlesztettem a PRIDE-NMR algoritmust (PRObability of IDentity - NMR), amely H-H távolsági eloszlások alapján képes megbecsülni a fehérjeszerkezetet. A módszer robusztusságát és hatékonyságát teszteltem több adatkészleten (40 fehérjés tesztkészlet, PDB adatbázisbeli adatok). A Novotny által alkalmazott fehérjeszerkezet-becslő szerverek teszteléséhez hasonló tesztet állítottam össze a PRIDE-NMR hatékonyságának becslésére. A PRIDE-NMR algoritmus teljesítménye megközelíti a legjobb fehérjeszerkezet-becslő algoritmusok teljesítményét.
2. A PRIDE-NMR módszer elérhető a világhálón. A háttáradatbázis küszöbértéke, a távolságeloszlás minimális távolsága, a lánchossz-súlyozás és a százalékos szűrés változtatható az eredmények minél szélesebb körű értékelése érdekében. A szerver használhatóságát egy SH3 domén példáján mutattam be.
3. Adatszonkítási teszttel megmutattam, hogy a NOE adatok specifikusan tartalmazzák a térszerkezetre jellemző információkat, szemben a térszerkezetekből visszszámolt H-H adatképpel, amelyek hiába jelentenek tízszeres adathalmazt, csonkítva már egyáltalán nem képesek jellemezni a térszerkezetet.
4. A DER, a MUMO és egy saját MUMO szimulációból származó, ubiquitinre számolt dinamikus szerkezeti sokaságokat vettem össze egy tisztított, távolság jellegű kényszerfeltétel listával. A kapott eredményeim alátámasztják, hogy a PRIDE-NMR módszer, a távolság jellegű kényszerfeltételeken alapuló térszerkezetbecslésen túl, a számolt szerkezeti sokaságok minőségellenőrzésére is alkalmas.
5. Kifejlesztettük a CoNSEnsX szervert (Compliance of NMR-derived Structural Ensembles with experimental data), amely a NOE adatokon túl további NMR dinamikai paramétereknek (S^2 , RDC-k, J-csatolás, kémiai eltolódások) felelteti meg a számolt szerkezeti sokaságot. A program a beadott szerkezetek atomi koordinátái alapján számolja vissza az adott paramétert (implementált egyéb programok), és a rendelkezésre álló kísérleti adatokkal veti össze. A szerver az egyes konformereket mind külön-külön, mind konformer-sokaságként is tudja kezelni. A kimenetben mind az egyes konformerekre, mind a teljes sokaságra átlagolva ki vannak értékelve az adatok. A módszert egy globuláris (ubiquitin) és egy belsőleg rendezetlen fehérje (PDE 5/6 γ -alegység) példáján mutattam be.

Következtetések

- A PRIDE-NMR szervert a távolság jellegű kényszerfeltételek alapján becsüli a fehérje térszerkezetét. Megmutattam, hogy a NOE adatok elegendő specifikus információt tartalmaznak a fehérje várható térszerkezetéről.
- Az NMR spektroszkópiai adatokból nyerhető paraméterek a vizsgált makromolekula térszerkeze mellett a belső dinamikájáról, a mozgékonyaságáról is szolgáltatnak információkat. Az a tény, hogy a molekulát oldatban vizsgáljuk, közelebb viszi az eredményeket a valós biológiai rendszerekben tapasztalható körülményekhez.
- A különböző időskálákon leírt molekuláris mozgások dinamikájának figyelembevételével a szimuláció során lehetővé teszi, hogy a kapott szerkezeti sokaságok strukturális heterogenitása ne a kényszerfeltételek bizonytalanságából eredjen, hanem a molekula tényleges belső mozgékonyaságát tükrözze adott időskálán. A fehérjék térszerkeze a belső mozgékonyaság figyelembevételével kiegészül a negyedik dimenzióval. Fontos a figyelembevett paraméterek időskáláját szem előtt tartani, hiszen a fehérjék dinamikája több nagyságrendet ölel fel, és egyes paraméterek csak bizonyos időskálára vonatkoznak.

II. rész: Véletlenszerű fehérjeszekvenciák *in silico* szerkezetvizsgálata

Céltűzés

A dolgozat második felében a fehérjék dinamikáját evolúciós távlatba kívántam helyezni a rendezetlenségen keresztül. Munkámban a *de novo* fehérjék keletkezését modellezve, *in silico* strukturális elemzést végeztem annak eldöntése végett, hogy az újonnan keletkező fehérjékre nézve az aggregáció tényleg jelentős veszély-e.

Az alkalmazott módszerek

- Random DNS szekvenciák segítségével modelleztem a *de novo* fehérjéket. Adott GC-tartalom mellett véletlenszerűen generáltam nukleotidszekvenciát, a STOP kodonok kizárása mellett, majd a standard genetikai kód alapján lefordítottam őket fehérjére. 10%-os lépésközzel, 10%-tól 90%-ig minden GC-tartalom mellett 10.000 random szekvenciát generáltam. 160 aminosavas szekvenciákat állítottam elő, az általánosan elfogadott átlagos globuláris doménméretnek megfelelően. A BLAST elemzést az "nr" (nem-redundáns, alapértelmezett) adatbázisra futtattam.

- További random adatkészletet állítottam elő a humán és egér proteomból, a szekvenciák aminosav-összetételét és a lánchosszat megtartva. Minden szekvenciára N véletlenszerű aminosavak közötti permutációt végeztem, ahol N a kérdéses fehérje lánchossza.
- A valós fehérjék adatkészletei homológia-szűrésen átesett adatbázisokból származnak (ASTRAL40, AmyPDB, DISPROT, PDBTM, valamint a teljes humán és egér proteom).
- Az irodalomban leírt, három humán, *de novo*, azaz nemkódoló DNS szakasz spon-tán átíródása révén keletkezett fehérje mRNS és fehérje szekvenciáit az UniProt adatbázisból töltöttem le. Kiszámoltam a GC-tartalmat mind a teljes, mind csak a fehérjét kódoló mRNS szakaszra vonatkozóan.
- Kizárólag szekvencia alapú prediktorokat használva, konszenzus predikcióval be-csültem a random szekvenciák rendezetlenségi, transzmembrán hélixképző és agg-regációs hajlamát. A nettó töltést és a hidrofobicitást az Uversky által definiált módon számoltam. A programok futtatását és az elemzéseket saját PERL progra-mokkal végeztem.
- Az adatokat Pearson korrelációval, valamint egy- és kétdimenziós Kolomorov-Smirnov analízissel elemeztem. Az eloszlásokat 2D-ban ábrázoltam, és a különböző adat-készletek által lefedett területeket összevettem.

Új tudományos eredmények

1. Megmutattam, hogy a véletlenszerű szekvenciák szerkezeti preferenciáit a megfe-lelő DNS szekvencia GC-tartalma irányítja. Míg random fehérjeszekvenciák rende-zetlenségi hajlama nő, a transzmembrán hélixképző és az aggregációs hajlam csök-ken a GC-tartalommal. A random fehérjeszekvenciákra 1% alatti találati arányban voltak hasonló szekvenciák, tehát teljesen *de novo*-ank, egyedinek tekinthetők.
2. Elemeztem a 2D eloszlásokat. A random fehérjék rendezetlenségi és aggregációs hajlamának mértéke megegyezik a valós fehérjék esetében tapasztaltakkal. A valós fehérjék nagyobb százalékban tartalmazzak transzmembrán hélixképző részeket, mint a GC-tartalom alapján generált random fehérjeszekvenciák.
3. Megmutattam, hogy a random szekvenciák esetében tapasztalt tendenciák nem lán-c-hossz függők. Ugyanis a humán proteom és a randomizált humán proteom össze-vetése ugyanazokat a tendenciákat eredményezte, mint a humán proteom összeve-tése a GC-tartalom szerint randomizált, 160 aminosavas fehérjékkel szemben.

4. A három humán *de novo* fehérje konszenzus elemzése során tapasztalt szerkezeti preferenciák és GC-tartalmak közötti összefüggések teljes összhangban vannak a random szekvenciákra leírt tendenciákkal.

Következtetések

- Az eredményeim olyan *de novo* szintetizálódó fehérjékre vonatkoznak, amelyek még semmilyen kölcsönhatásba nem léptek sem egymással, sem más sejtkomponenssel.
- A humán proteom és a biológiailag releváns 40 és 60% közötti GC-tartalmú random szekvenciák predikcióinak összevetése egészen új megvilágításba helyezi a térszerkezeti jellemzők közötti összefüggéseket.
- Az újonnan szintetizálódó fehérjék, még akkor is ha először nagyon kis mennyiségben jelennek meg, feltételezhetően kisebb veszélyt jelentenek a sejtre komoly aggregációs hajlam esetén is, mint a nagy mennyiségben folyamatosan szintetizálódó fehérjék.

Összefoglalás

A doktori munkámban bemutatott két témakör közös pontja a mozgékonyág kérdése. A fehérjék belső dinamikájának leírásával lehetőség nyílik a biológiai folyamatok új szempontból történő vizsgálatára, az újonnan keletkező fehérjék szerkezeti preferenciáinak vizsgálata pedig a fehérjeevolúciót helyezi új dimenzióba.

Megmutattam, hogy az NMR spektroszkópiai adatokból származtatott távolsági jellegű kényszerfeltételek alapján lehet becsülni a fehérje várható térszerkezetét távolságeloszlások segítségével. A fehérjeszerkezetek és az NMR kísérleti adatok megfeleltetése a CoNSEnsX integrált szervert állítottam össze. A biológiai folyamatokban tehát a mozgékonyág és a különböző időskálájú dinamikus viselkedés ismerete elengedhetetlen előfeltétele annak, hogy a megfigyelt jelenségekre kielégítő szerkezeti vagy funkcionális modelleket tudjunk felállítani.

A megállapítás, hogy a *de novo* fehérjék nem különösebben hajlamosak aggregációra, ellentmondani látszik az irodalomban leírt elvvel, hogy az aggregáció elleni védekezés optimalizációs lépés a fehérjeevolúció során. Ugyanakkor az alkalmazott predikciós módszer, ami csak három globális szerkezeti tulajdonságra ad becslést, nem mond semmit a részletes szerkezeti vagy funkcionális elemekről. Feltételezzük, hogy a *de novo* fehérje keletkezése után a szerkezet szelekció útján optimalizálódik a funkció betöltésére, és a

szerkezet, stabilitás és a mozgékonyág együtt lesz optimálva a funkcióval. E folyamat alatt az aggregációs hajlam megtartása vagy akár csökkentése, ami az újonnan keletkező *de novo* fehérje belső tulajdonsága, egy erős szelekciós nyomást gyakorol az evolúció során. Viszont ki kell hangsúlyozni, hogy az új kódoló szekvencia megjelenését nem gátolja feltétlenül az a tény, hogy a lefordított fehérjeszekvencia majd szokatlanul nagy aggregációs hajlammal rendelkezik.

Összefoglalva, doktori munkám során két szervert fejlesztettem ki és egy új szemléletű *in silico* szerkezeti elemzést végeztem el, a fehérjék térszerkezetét és evolúcióját vizsgálva a mozgékonyág tükrében.

Az értekezés alapjául szolgáló közlemények

- I. **Ángyán AF**, Perczel A, Pongor S, Gáspári Z:
Fast protein fold estimation from NMR-derived distance restraints.
Bioinformatics 24:272-275 (2008)
impakt faktor: 4.328

- II. Gáspári Z, **Ángyán AF**, Dhir S, Franklin D, Perczel A, Pinter A,
Pongor S:
Probing dynamic protein ensembles with atomic proximity measures.
Current Protein & Peptide Science 11:515-522 (2010)
impakt faktor: 3.830

- III. **Ángyán AF**, Szappanos B, Perczel A, Gáspári Z:
CoNSEnsX: an ensemble view of protein structures and NMR-derived
experimental data.
BMC Structural Biology 10:39 (2010)
impakt faktor: 2.258

- IV. **Ángyán AF**, Perczel A, Gáspári Z:
Estimating intrinsic structural preferences of *de novo* emerging random-
sequence proteins: is aggregation the main bottleneck?
FEBS Letters, feltételesen elfogadva (minor revision)

Bemutatott poszterek és előadások

1. **2006.** 3rd Central European Conference on Chemistry towards Biology (CTB3), Krakko, Lengyelország, poszter címe: *Fast protein fold estimation based on NMR distance restraints.* **Ángyán AF**, Perczel A, Gáspári Z.
2. **2006.** XII. Vegyészkonferencia, Csíkszereda, Románia, előadás címe: *Fehérje-térszerkezetek gyors becslése NMR kényszerfeltételek alapján.* **Ángyán AF**, Perczel A, Gáspári Z.
3. **2008.** 4th Central European Conference on Chemistry towards Biology (CTB4), Dobogókő, Magyarország, poszter címe: *Fast protein fold estimation based on NMR distance restraints derived from NMR spectra (Identifying protein fold from assigned NMR spectra)* **Ángyán AF**, Perczel A, Pongor S, Gáspári Z.
4. **2009.** ChemAxon User Group Meeting Training Day, Budapest, poszter címe: *Evaluation of small molecular libraries using molecular docking and binding profile analysis.* **Ángyán AF**, Iván G, Grolmusz V.
5. **2009.** 17th Annual International Conference on Intelligent Systems for Molecular Biology (ISMB) 8th European Conference on Computational Biology (ECCB), Stockholm, Svédország, poszter címe: *Evaluating small molecule libraries using molecular docking and binding profile analysis.* **Ángyán AF**, Iván G, Grolmusz V.
6. **2009.** Foldamers: building blocks, structure and function, Szeged, poszter címe: *Generation and analysis of realistic protein structural ensembles (What do dynamic protein structural ensembles tell us?)* Gáspári Z, **Ángyán AF**, Szappanos B, Várnai P, Pongor S, Perczel A.
7. **2010.** FEBS Congress, Göteborg, Svédország, előadás és poszter címe: *CoNSEnsX: an ensemble view of protein and peptide structures and NMR-derived experimental data.* **Ángyán AF**, Szappanos B, Perczel A, Gáspári Z.