Kalászi Adrián

Konformációs flexibilitás és 3D QSAR

Témavezető: dr. Farkas Ödön, docens

Kémia Doktori Iskola – vezetője: Dr. Inzelt György, egyetemi tanár

Szintetikus kémia, anyagtudomány, biomolekuláris kémia doktori program – vezetője: Dr. Horváth István Tamás, egyetemi tanár

Készült az

ELTE TTK, Kémiai Intézet, Kémiai Informatika Laboratóriumban

2007

„…És mit jelent az, hogy szeretettel munkálkodjatok?
Ezt jelenti: sződd a szövetet a szívedből vett szálakkal, mintha szerelmesed öltené majd magára.
Ezt jelenti: építsd a házat olyan szeretettel, mintha szerelmesed lakoznék majd benne.
Ezt jelenti: vesd a magot gyöngédséggel és arass örvendezve, mintha szerelmesed élvezné majd gyümölcsét.
Ezt jelenti: tölts meg mindent, amit megalkotsz, a benned lévő szellem leheletével…”

Kahlil Gibran: A próféta

Köszönetnyilvánítás

Mindenekelőtt szeretnék köszönetet mondani témavezetőmnek, Dr. Farkas Ödönnek az eredményes konzultációkért, Imre Gábornak a programozási gyakorlatom elmélyítéséhez nyújtott segítségéért, és Dr. Jákli Imrének az adatbáziskezelés megvalósításáért és a dolgozatom gondos átolvasásáért.

Köszönettel tartozom Dr. Hudecz Ferencnek és Dr. Mező Gábornak a téma felvetéséért, a dolgozatban szereplő peptidek és biológiai mérések elvégzéséért.

Köszönetet szeretnék mondani Sarkadi Áginak az angol szövegrészek gondos ellenőrzéséért.

A doktori munkám elkészítését a Gaussian Inc., a Richter Gedeon Centenáriumi Alapítvány szakmai továbbképési pályázata, az Infopark Alapítvány két nyertes pályázata, a GVOP-3.1.1.-2004-05-0451/3.0 és az HEF_06_2-IQSAR jelű NKTH-KPI pályázatok, valamint a ChemAxon Kft támogatta.

Köszönöm kedvesemnek, Szász Csillának a dolgozat gondos átolvasását, és mindazt a szeretet és türelmet, amely nélkül e munka nem születhetett volna meg.

Budapest, 2007 május 15.

1. Tartalomjegyzék

1. Tartalomjegyzék 1
2. Az értekezés alapjául szolgáló közlemények 3
3. Egyéb saját közlemények 4
4. Bevezetés 5
   4.1. In silico módszerek a hatóanyag tervezésben 6
   4.2. A farmakofor és a bioaktív konformáció 8
   4.3. QSAR 10
      4.3.1. 3D QSAR 11
      4.3.2. Konformációs flexibilitás és QSAR 13
   4.4. A QSAR modellek statisztikai ellenőrzése 14
      4.4.1. Korrelációs együttható és a determinációs együttható 15
      4.4.2. Belső ellenőrzés 18
      4.4.3. Külső ellenőrzés 22
      4.4.4. A Kubinyi paradoxon 22
      4.4.5. Összefoglalás 24
5. Célkitűzés 25
6. A 3D QSAR elméleti háttere flexibilis molekulákra 26
   6.1. A ligand konformációs flexibilitásának hatása a kötődésre 26
   6.2. Relatív koncentráció meghatározása kísérleti módszerekkel 28
      6.2.1. Telítési kötődési vizsgálat 28
      6.2.2. Dózis-hatás görbe felvételén alapuló módszerek 29
   6.3. A molekulatár speciális tulajdonságai 30
   6.4. A konformációs flexibilitás – hatás összefüggés 32
7. Számítási eljárások 34
   7.1. A konformációs távolság definíciói 35
      7.1.1. Összeforgatás után, atomi pozíciók alapján számított távolság 35
      7.1.2. A peptidgerinc ϕ és ψ torziószögei alapján számított távolság 36
      7.1.3. A torzióalapú és koordináta-alapú konformációs távolságmértékek összehasonlítása 37
   7.2. Az r és a q^2 összehasonlítása 39
   7.3. A konformerkereső algoritmus bemutatása 42

7.4. Bioaktív-konformáció kereső algorimus 45
7.4.1. Algoritmus 1 46
7.4.2. Algoritmus 2 50
7.5. IQsar 52
8. Alkalmazások és eredmények 54
8.1. Konjugálási hely hozzáférésnek becslése 54
8.1.1. Oligotuftsin: egy új gyógyszermolekula hordozó 54
8.1.2. A kiindulási konformerek előállítása 56
8.1.3. Az oldatbeli konformációs viszonyok feltérképezése 61
8.1.4. A molekulafelszín számítása 62
8.1.5. Kísérleti eredmények 63
8.1.6. Az eredmények összefoglalása 64
8.2. Mucin epitóp peptidek bioaktív konformációjának meghatározása 66
8.2.1. A kiindulási konformerek előállítása 67
8.2.2. Az oldatbeli konformációs viszonyok feltérképezése 68
8.2.3. Bioaktív-konformáció keresés 69
8.2.4. Az eredmények értelmezése 76
8.3. β-Amiloid epitóp peptid bioaktív konformációjának meghatározása 78
8.3.1. A kiindulási konformerek előállítása 79
8.3.2. Az oldatbeli konformációs viszonyok feltérképezése 81
8.3.3. Bioaktív-konformáció keresés 82
8.3.4. Az eredmények értelmezése 85
9. Következtetések, eredmények összefoglalása 86
10. Kitekintés 90
A. Függelék : r^2 személetes jelentésének levezetése 93
B. Függelék : Kiegészítés a relatív koncentráció számításához 95
C. Függelék : Atomi pozíciók alapján számított konformációs távolság 97
D. Függelék : Átemelt ábrák 99
11. Irodalomjegyzék 101

2. Az értekezés alapjául szolgáló közlemények

1. **Kalászi, A.**; Farkas, O. Lead conformer prediction based on a library of flexible molecules. *Journal of Molecular Structure-THEOCHEM* **2003,** *666,* 645-649.
2. Mezo, G.; **Kalaszi, A.**; Remenyi, J.; Majer, Z.; Hilbert, A.; Lang, O.; Kohidai, L.; Barna, K.; Gaal, D.; Hudecz, F. Synthesis, conformation, and immunoreactivity of new carrier molecules based on repeated tuftsin-like sequence. *Biopolymers* **2004,** *73,* 645-656.
3. Manea, M.; **Kalaszi, A.**; Mezo, G.; Horvati, K.; Horvath, A.; Farkas, O.; Przybylski, M.; Hudecz, F. Antibody recognition and conformational flexibility of a plaque-specific β-amyloid epitope modulated by non-native peptide flanking regions. *Journal of Medical Chemistry – kisebb javításokkal elfogadva* **2007**. *(megjegyzés: a publikáció első szerzője megosztott helyen* Manea, M. és **Kalaszi, A.***)*

3. Egyéb saját közlemények

4. Mező, G.; **Kalászi, A.;** Reményi, J.; Mihala, N.; Majer, Zs.; Hilbert, Á.; Láng, O.; Kőhidai, L.; Barna, K.; Gaál, D. Synthesis of new carrier molecules based on repeated tuftsin sequences. In *Peptides 2000.*; Martinez, J.; Fehrenz, J. A. Eds.; EDK: Paris, **2001**; pp 253-254.

5. **Kalaszi, A.**; Farkas, O. Predicting lead compounds using libraries of flexible molecules. *Biopolymers* **2003,** *71*, 331.

6. Bagyi, I.; Balogh, B.; Czajlik, A.; Elias, O.; Gaspari, Z.; Gergely, V.; Hudaky, I.; Hudaky, P.; **Kalaszi, A.**; Karolyhazy, L.; Keseru, K.; Kiss, R.; Krajsovszky, G.; Lang, B.; Nagy, T.; Racz, A.; Szentesi, A.; Tabi, T.; Tapolcsanyi, P.; Vaik, J.; Koo, J. C. P.; Chass, G. A.; Farkas, O.; Perczel, A.; Matyus, P. Generation and analysis of the conformational potential energy surfaces of N-acetyl-N-methyl-L-alanine-N '-methylamide. An exploratory ab initio study. *Journal of Molecular Structure-Theochem* **2003,** *625*, 121-136.

7. **Kalaszi, A.**; Imre, G.; Jakli, I.; Farkas, O. Theoretical basis and a computational tool to identify binding conformations of flexible ligands. *European Journal of Medicinal Chemistry - beküldve* **2007**.

4. Rövidítések jegyzéke

QSAR	kvantitatív szerkezet-hatás összefüggés (quantitative structure-activity relationship)
VS	virtuális szűrővizsgálat (virtual screening)
HTS	nagy áteresztőképességű szűrővizsgálat (high throughput screening)
2D	két dimenzió
3D	három dimenzió
ADME	abszorpció, disztibúció, metabolizmus, exkréció
HIV	A szerzett immunhiány kórokozója (human immunodeficiency virus)
CDK2	ciklin függő kináz 2 (cyclin-dependent kinase 2)
PLS	parciális legkisebb négyzetek módszere (Partial Least Square)
CoMFA	összehasonlító molekulatér elemzés (Comparative Molecular Field Analysis, CoMFA)
MLR	többváltozós lineáris regresszió (Multiple Linear Regression)
LOO	egy érték kihagyásával történő keresztellenőrzés (leave one out crossvalidation)
LMO	több érték kihagyásával történő keresztellenőrzés (leave multiple out crossvalidation)
BC	bioaktív konformáció
V_{BC}	bioaktív konformációnak megfelelő konformációs térfogat
ρ_{BC}	V_{BC}-hez rendelhető konformációs sűrűségérték
PCORE	különböző molekulák azon farmakofor pontjai, amelyek a kötődés szempontjából ekvivalensek, és részt vesznek a fehérje – kismolekula kölcsönhatás kialakításában.
MD	molekuladinamika
rmsd	átlagos eltérés-négyzetösszegek gyöke (root mean square deviation)
AC	egyetlen, horgony a konformáció (anchor conformation). Az AC-kkel mintavételezem az egyes mulekulák konformációs terét.
uAC	azok a konformációk, amelyekkel a PCORE által leírt konformációs teret mintavételezem (unified anchor conformation)
CD	cirkuláris dikroizmus

5. Bevezetés

A gyógyszeripar egyik legfontosabb feladata, az új és minél biztonságosabb gyógyszerek minél gyorsabb előállítása, egyre jelentősebb kutatás-fejlesztési ráfordításokat kíván meg. Egyes becslések szerint egy új gyógyszer kifejlesztésére fordított kiadások már meghaladják a 800 millió USD-t[1]. A költségeket leginkább a késő klinikai fázisban kudarcot vallott gyógyszerek növelik meg, ezért minden olyan eljárás, amely minél korábban leállítja a sikertelen hatóanyagok kutatását, nagy érdeklődést vált ki[1]. Ilyen eljárások például a gyógyszerfejlesztés során egyre intenzívebben használt ún. *in silico*, modellező és szimulációs módszerek. Ezeket a módszereket a gyógyszerfejlesztés egész folyamatán keresztül fel lehet használni, a vezérmolekulák keresésétől kezdve a klinikai próbák szimulációjáig. Egyes jóslatok szerint az *in silico* megközelítés a kutatásfejlesztési ráfordítások akár 10-15%-át kiteheti a következő 5-10 évben[2].

5.1. In silico módszerek a hatóanyag tervezésben

A gyógyszerkutatás egyik fontos lépése a vezérmolekula (lead compound) keresése. A vezérmolekulának rendelkeznie kell a szükséges biológiai aktivitással, bár ez lehet gyenge, vagy nem szelektív. A vezérmolekulának nem szabad túl polárosnak vagy túl lipofilnek lennie, nem tartalmazhat toxikus csoportokat, és metabolitjai sem lehetnek toxikusak. A vezérmolekula általában nem reagálhat irreverzibilisen a célfehérjével (bár ez alól vannak kivételek). Kezdetben a hatóanyag- és vezérmolekulák inkább természetes eredetűek voltak (növényi eredetű, endogén neurotranszmitterek, szteroid hormonok). Később hatalmas, kombinatorikus kémiai úton előállított molekulakönyvtárakban kerestek nagy áteresztőképességű szűrővizsgálat (high throughput screening, HTS) segítségével optimálható vezérmolekulákat (hit). Annak ellenére, hogy ez a megközelítés számos olyan vezérmolekulát eredményezett, amelyikből gyógyszer lett, általános megoldássá mégsem vált[3]. A klasszikus kutatás során is az egy gyógyszermolekulára jutó kipróbált molekulák száma gyakran elérte a 10000-et. A kombinatorikus kémia és a HTS alkalmazása során ez a szám elérheti a százezres, akár a milliós nagyságrendet is.

A következő lépés, a vezérmolekulák optimálása, mely sikeres esetben aktív, szelektív, lehetőleg orálisan adható, nem toxikus, bizonyos molekulatömeg- és lipofilitási

tartományba eső hatóanyag jelölt molekulához vezet. A sikeres jelöltek további ellenőrzésére klinikai vizsgálatokat indítanak.

Napjainkban mind a vezérmolekulák keresése során, mind pedig az optimálás szakaszában az (automatizált) párhuzamos szintézisek játszanak döntő szerepet. A szintézis során számítási úton megtervezett, fókuszált vegyületkönyvtárakat állítanak elő. A tervezéshez legtöbbször virtuális szűrővizsgálatot (virtual screening, VS) alkalmaznak[4]. A VS olyan számítási eljárások gyűjtőneve, amelyek egy nagyszámú molekulát tartalmazó (virtuális) molekulakönyvtár méretét lecsökkentik, és fókuszált molekulakönyvtárat hoznak létre (**1. táblázat**).

Szubstruktúra keresés	
Alapkérdés:	Az adatbázis melyik molekulái tartalmazzák a keresett (query) szubstruktúrát?
Modell:	A molekulák amelyek tartalmazzák keresett szubstruktúrát valószínűsíthetően aktívak.
Szükséges hozzá:	2D szubstruktúra vagy 3D farmakofor, amely közös az aktív molekulákban
Hasonlóság alapú keresés	
Alapkérdés:	Az adatbázis melyik molekulái „hasonlóak” a keresendő (query) molekulához?
Modell:	Teljes molekulák, amelyek a keresendőhöz hasonlítanak, valószínűsíthetően aktívak.
Szükséges hozzá:	Egy vagy több aktív molekula.
Dokkolás	
Alapkérdés:	Az adatbázis melyik molekulái illeszthetők be a fehérje vagy receptor ismert kötőhelyére?
Modell:	Azok a molekulák, amelyeknek a kölcsönhatását a kötőhellyel magas pontszámmal jellemezhetjük, valószínűleg aktívak.
Szükséges hozzá:	A célfehérje 3D struktúrája (röntgen szerkezet, vagy homológia modellezés)
QSAR	
Alapkérdés:	Melyik molekuláknak van a legnagyobb jósolt aktivitásuk?
Modell:	Magas jósolt aktivitással rendelkező molekulák, valószínűleg aktívak.
Szükséges hozzá:	Elegendő molekulára mért aktivitás adat.

***1. táblázat**: A virtuális szűrővizsgálat fajtái*[5].

A vezérmolekula keresésének első lépésekor (hit discovery) a VS olyan, minél diverzebb molekulakönyvtárak elkészítésére használható, ahol a különböző molekuláktól elvárhatunk bizonyos alakbeli, vagy kémiai összetételbeli tulajdonságokat[4]. Ezen a szinten az aktivitás általában nem szigorú feltétel, egy molekula, amelynek már μmolos tartományba esik az aktivitása, már találatnak (hit) minősül. Miután a találatokat azonosították és ellenőrizték, elkezdődik a találat-optimalizálási folyamat. Ekkor VS-t már meglévő, vagy virtuális molekulák felhasználásával az egyes találatokra fókuszált molekulakönyvtárak előállítására használják. Ha az eredeti találat aktivitását fel tudtuk javítani μmol alatti tartományba, akkor a találatból vezérmolekula lesz. A hit optimálása során felmerült nehézségek előre vetíthetik a vezérmolekula további optimálása során várható nehézségeket.

Ezért fontos, hogy ezen a szinten több találatból válogathassanak. A vezérmolekula optimálása során a VS további, célzott vegyületkönyvtárak előállítására használható, amelyek már nemcsak az aktivitás növekedését célozzák, hanem a szelektivitás, toxicitás és az ADME tulajdonságok javítását is.

A folyamat előrehaladtával a VS hatékonysága egyre inkább függ attól, hogy van-e valamilyen, figyelembe vehető szerkezeti információ. Amikor a célfehérje 3D szerkezete ismert, dokkolási eljárásokat előszeretettel használnak VS-re. Sok esetben azonban nem ismert a kötőhely 3D szerkezete, ekkor a molekulák 2D (topológiai) vagy 3D (koordináta) reprezentációján alapuló, ún. ligand alapú VS eljárásokat alkalmaznak.

5.2. A farmakofor és a bioaktív konformáció

A 3D molekuláris hasonlóság alapú módszerek és a 3D QSAR eljárások esetében kulcsfontosságú egy ún. 3D farmakofor[6] ismerete. A farmakofor (pharmacophore) kifejezést Ehrlich vezette be az 1900-as évek elején[7]. Napjainkban farmakofornak nevezik olyan sztérikus és elektronikus tulajdonságok összességét, amelyek szükségesek ahhoz, hogy teljesüljenek a biológiai célszerkezettel való optimális kölcsönhatás és a megfelelő válasz kiváltásának ill. blokkolásának feltételei[8]. A modell szerint szerkezetileg különböző molekulák közös farmakoforja azonos módon lép a kötőhellyel kölcsönhatásba. Az egyes molekulák bioaktív konformációja az a konformáció, amely a kötődés során megvalósul, és magában foglalja a farmakofort. A farmakoforokat használhatják VS-re úgy, hogy ujjlenyomatokat (fingerprint) készítenek belőlük, és ezt hasonlítják össze a könyvtár molekuláinak ujjlenyomataival (például[9-11]), vagy az adatbázis szűrését közvetlen molekuláris összeforgatás[12, 13] (alignment, overlay) alapján végzik.

A farmakofor meghatározásának első lépése a farmakofor pontok előállítása legtöbbször úgy történik, hogy az atomokhoz (vagy egyes csoportokhoz) megfelelő tulajdonságokat rendelnek (hidrogénkötés akceptor, donor, aromás, hidrofób, pH=7-en pozitív vagy negatív töltés). Ezután történik az egyes molekulák 3D szerkezetének (automatikus) előállítása és a konformációs analízis. Ezek általában számításigényes feladatok, mivel a molekulák nagyszámú releváns konformerrel és sokféle összeforgatási lehetőséggel rendelkezhetnek. Számos módszert kidolgoztak e feladat elvégzésére, például az AAA[14]-t, DISCO[15]-t, de legnépszerűbb talán a Catalyst[16] programcsomagban található HypoGen[17, 18] nevű eljárás.

Chen és munkatársai ilyen eljárások teljesítményét hasonlították össze[19], 145 ismert kristályszerkezetű fehérje-ligand komplex esetén. Többek között azt vizsgálták, hogy a különféle módszerekkel megkapható-e a kísérletileg meghatározott, a számítás során ismeretlennek föltételezett bioaktív konformáció, pusztán a ligand molekulák alapján. Azt találták, hogy az eredmények kevéssé függnek attól, hogy milyen eljárást használnak, viszont jelentősen függnek az alkalmazott „templát" (amire a többi molekulát forgatják, azaz farmakofor) molekula szerkezetétől és konformációjától. A templátoknak elég pontosan (1.5Å rmsd) meg kell közelíteni (a számítás során ismeretlennek tekintett) a bioaktív konformációt ahhoz, hogy a rájuk illesztett molekulák is helyes konformációban és orientációban helyezkedjenek el. Megállapították, hogy ha a templát bioaktív konformációja nem ismert, a molekuláris összeillesztés nagyon pontatlan is lehet. Példát is említenek, a különösen flexibilis HIV inhibitorok esetén a helyes összeillesztések aránya (a teljes összeillesztések számához képest) kisebb volt, mint 1%, ugyanakkor a kevéssé flexibilis ligandok esetén ez az arány némileg javult (4% elasztáz, 9% CDK2, 53% tripszin ligandok esetén). Tehát a molekulák helyes bioaktív konformációjának és a közös farmakofornak az azonosítása lényegesen függ a konformációs flexibilitástól.

Nicklaus és munkatársai 1995-ben megvizsgálták[20], hogy 33 ligand bioaktív konformációja (amely minden esetben itt is ismert a fehérje-ligand komplex röntgenszerkezetből) és a konformációs analízis során kapott legkisebb energiájú konformer miként viszonyul egymáshoz. Megállapították, hogy a flexibilis ligandok esetén a bioaktív konformációt a globális energiaminimumhoz képest viszonylag széles energiatartományban találhatjuk meg. Magas korrelációs értéket találtak a forgatható torziók száma és az energiakülönbség (r^2=0.633) között, és a forgatható torziók száma és a molekuláris szerkezetek különbözősége (rmsd) között is (r^2=0.827) (D. Függelék **41. ábra**). Megállapították, hogy a ligand molekulák kötődés során fellépő deformációja az elforgatható kötések számán túl függ még a hidrogénkötésben résztvevő atomok számától. A deformáció hajtóereje tehát az, hogy a molekula hidrogénkötés kialakítására alkalmas pontokat keres a fehérjén.

2003-ban Perola és munkatársai[21] ugyancsak a ligandok kötődése során fellépő konformációs átrendeződésből származó energiát vizsgálták. Ők 150 fehérje-ligand komplex kristályszerkezetéből indultak ki. A vizsgálatokhoz két erőtér (MMFF és OPLS-AA) és három programcsomag konformációs analízis eljárásait (Catalyst, OMEGA, ICM) használták. Megerősítették azt az általánosan elfogadott megállapítást, hogy a ligand molekulák bioaktív konformációja ritkán esik egybe a globális energiaminmumnak megfelelő konformációval. Az

általuk vizsgált 150 ligand 60%-ának bioaktív konformációja még lokális energiaminimumnak sem felelt meg. A kötődés során bekövetkező konformációs energiaváltozást a bioaktív konformáció és a talált legalacsonyabb energiájú konformer energiakülönbségével jellemezték. Ők is találtak korrelációt a ligand flexibilitása és a kötődés során bekövetkező konformációs energiaváltozás között, viszont a flexibilitás és a kötődési erősség között nem. Megfigyelték azt is, hogy a ligandok általában nyújtott (extended) konformációban kötődnek, amit azzal magyaráztak, hogy a hidrofób csoportok, amelyek vízben lehetőség szerint eltemetettek a molekulában, a fehérjéhez kötődve feltárulkoznak.

Megvizsgálták, hogy az egyes konformációs analízis eljárások milyen arányban találják meg a bioaktív konformációkat. A fentiekben is említett 1.5Å-ös távolságküszöb mellett a módszerek 85-95%-ban megtalálják a bioaktív konformációt. Meg kell azonban említeni, hogy a vizsgált ligandok zöme kevés forgatható torziót tartalmazott. A globális energiaminimum keresésre a három módszer közül az ICM tűnt a legjobbnak, ugyanakkor a legdiverzebb konformációs sokaságot a Catalyst-tal állították elő.

Aszódi és munkatársai öt, a farmakoforok építéséhez használható konformációkereső program (DGEOM, QXP, ROTATE, LMOD, OMEGA) teljesítményét hasonlították össze[22] egy 604 ligand molekulából álló könyvtáron. Az eredmények kiértékelése során energia geometria és gyorsaság szerinti összehasonlításokat végeztek. A talált konformerek energiája alapján az LMOD volt a legsikeresebb, a többi módszer azonban geometriailag diverzebb konformereket eredményezett, emellett a leggyorsabbnak az OMEGA nevű eljárás bizonyult.

5.3. QSAR

A kvantitatív szerkezet-hatás összefüggés (Quantitative Structure Activity Relationship, QSAR) felállítását végző eljárások arra az általános alapelvre épülnek, hogy a vegyületek biológiai aktivitásának különbsége szerkezeti tulajdonságaik különbözőségével magyarázható. Sokféle, lényegesen különböző módon működő QSAR eljárást fejlesztettek ki Hansch 1963-ban elvégzett úttörő munkája[23] nyomán.

A 2D QSAR számítások—a 3D eljárásokkal szemben— térszerkezeti információt nem vesznek figyelembe, hanem a molekula topológiája alapján különféle deszkriptorokat készítenek, majd a deszkriptorok alapján többváltozós korrelációs módszerek, leggyakrabban többváltozós lineáris regresszió (multiple linear regression, MLR) alkalmazásával építik föl a szerkezet-hatás összefüggést[24]. A 2D módszerek amellett, hogy gyorsabbak, mint a három

dimenzióban működő eljárások, könnyebben is automatizálhatók, így alkalmazhatók nagyobb molekula adatbázisokon is.

A 2D QSAR és a 3D QSAR közötti átmenetnek tekinthető a hologram QSAR (HQSAR)[25] eljárás, ahol az egyes molekulákra előállítja az összes, 2D fragmens alapú deszkriptort, de emellett még figyelembe vesz sztereokémiai, és hibridizációs állapotra vonatkozó információkat is. Az eljárás a deszkriptorok alapján –a 3D QSAR eljárásoknál előszeretettel alkalmazott– a parciális legkisebb négyzetek módszerével (Partial Least Square, PLS)[26, 27] készíti el a szerkezet-hatás összefüggést.

5.3.1. 3D QSAR

A pontos 3D molekulaszerkezeteket előállító módszerek fejlődésével a QSAR eljárásokban lehetőség nyílt új, 3D deszkriptorok alkalmazására. A 3D QSAR módszereket főleg a vezérmolekula optimálása során használják, amikor már elegendő számú molekula rendelkezik mért kötődési aktivitás értékkel.

Talán az egyik legnépszerűbb* 3D QSAR eljárás a Cramer és munkatársai által kifejlesztett „összehasonlító molekulatér elemzés" (Comparative Molecular Field Analysis, CoMFA)[28]. A CoMFA módszer használata esetén először egy hasonló molekulákból előállított (congeneric) molekulasorozatot kell kiválasztani. Ezek a molekulák ugyanazzal a hatásmechanizmussal kell, hogy működjenek, és közös farmakoforral kell, hogy rendelkezzenek, de azonos molekuláris váz megléte nem szükséges. A módszer egy molekula esetén csak egy konformációt vesz figyelembe. Leggyakrabban a molekula számos forgatható torziószögét kell módosítani úgy, hogy egy olyan alacsony energiájú konformert kapjunk, amelyre a közös farmakofor jól illeszthető. A rigid, vagy különböző konformációs megkötéssel (constraint) rendelkező analógok vizsgálata segítséget nyújthat a flexibilisebb molekulák bioaktív konformációjának megállapításához, majd a molekulákat a közös farmakofor szerint összeforgatjuk. A CoMFA módszer a megfelelően egymásra forgatott molekulák körül egy közös, háromdimenziós rácsot készít, és a rácspontokban kiszámolja a sztérikus és elektrosztatikus kölcsönhatási energiákat. A kiszámolt értékeket egyfajta molekuláris deszkriptornak tekinti és PLS alkalmazásával 3D modellt épít fel. Az atomokhoz közeli rácspontokban a kiszámolt kölcsönhatási értékek numerikusan nagyon nagyok is lehetnek, ezért az ebből adódó problémák kezelése nagy körültekintést igényel.

* A dolgozat készítésének időpontjáig (2007 április) az idézett publikációra 1567 hivatkozás érkezett.

Az „összehasonlító molekula-hasonlósági vizsgálat” (comparative molecular similarity analysis, CoMSIA[29]) ezt a problémát úgy oldja meg, hogy Gauss alapú hasonlósági függvényeket alkalmaz.

Az úgynevezett „kvantitatív hasonlóság-aktivitás összefüggést” (Quantitative similarity-activity relationship QSiAR)[30] felállító eljárás esetén a 3D hasonlósági együtthatók mátrixából indulnak ki. Ez a mátrix n sorból és n oszlopból áll, ahol n a sorozat molekuláinak száma. A mátrix előállításához szükséges összeforgatásokat csak molekulapáronként végzik. Ha az eljárásban a SEAL[31] hasonlósági indexeket alkalmazzák, molekuláris rács előállítására nincs is szükség.

A 2D és 3D QSAR módszerfejlesztések legújabb trendjébe tartozik olyan eljárások kidolgozása, amelyek kiválasztanak bizonyos változókat a rendelkezésre álló deszkriptorok sokaságából úgy, hogy azokkal értelmes és statisztikailag szignifikáns modellt lehessen építeni. Ezt a leggyakrabban úgy valósítják meg, hogy sztohasztikus keresési módszert (pl.: szimulált megeresztést[32] (simulated annealing), genetikus algoritmust) és egy korrelációs módszert (MLR, PLS, vagy mesterséges ideghálózatok[33]) illesztenek össze. Változókiválasztási módszereket a 3D QSAR eljárások során arra alkalmaztak, hogy megkeressék az optimális régiót, amely alapján a molekulákat összeforgatva jobb QSAR modellek készíthetők, mint az eredeti, CoMFA technikával. Ilyen például a GOLPE[34], amelyet kemometriai alapelvek alapján fejlesztettek ki, és a q^2-GRS[35], amely a közeli molekulatér kis régióinak független elemzésére épül. A legújabb próbálkozás e téren a 2006-ban megjelent úgynevezett kNN-MFA módszer[36]. Ez az eljárás a k-legközelebbi szomszéd (k-nearest neighbour, kNN) algoritmusa alapján választja ki azokat a farmakoforokat, amelynek összeforgatásával készített molekulatér vizsgálat a legjobb minőségű modellt adja. A módszert ugyanakkor a szerzők is lassúnak találták, bár a kapott modell jobb minőségű volt (q^2), és jobb predikciót is adott, mint az összehasonlításhoz alkalmazott CoMFA.

Úgy tartják, hogy a biologiai aktivitás jóslására általában a 3D QSAR technikák használhatók a leghatékonyabban. Bár legtöbbször a szűk keresztmetszet a kémiai struktúrák pontos egymásra illesztése [12, 37-39] és a bioaktív konformáció előzetes ismerete. A molekulák egymásra illesztése (alignment, overlay) gyakran felhasználói beavatkozást igényel, ezért ez a leginkább időigényes részfolyamat. Számos félautomata technikát dolgoztak ki[12, 13]. Bizonyos 3D QSAR módszerek[40-45] úgy kerülték meg a molekuláris összeillesztés problémáját, hogy ún. „két és fél” dimenziós deszkriptorokat alkalmaztak. Ezek a deszkriptorok ugyan érzékenyek a molekula 3D struktúrájára, de nem függnek a molekula térbeli pozíciójától. Tehát a bioaktív konformáció ismerete ezeknél a típusú deszkriptoroknál is elengedhetetlen.

Az eredményül kapott modelleket általában nehezebb interpretálni, mint a tradicionális 3D QSAR technikák esetén.

5.3.2. Konformációs flexibilitás és QSAR

A 3D QSAR módszerek egyik hátránya, hogy a molekulákat egyetlen konformerrel írják le, ezért kidolgoztak olyan eljárásokat is, ahol több konformációt felhasználva a molekuláris flexibilitást is figyelembe veszik. Az egyik legismertebb[†] talán a Hopfinger és munkatársai által kifejlesztett 4D QSAR eljárás[46]. A negyedik dimenziót ekkor a – minden kiindulási szerkeztre molekuladinamikai szimulációval számolt – konformációs eloszlás alkotja. Az eljárás egy előre kiválasztott farmakoforra illeszti rá a molekulák egyes konformációit. Ez a módszer annyiban hasonlít a CoMFA eljáráshoz, hogy a számítás során térbeli rácspontokat használ. A rácspontokhoz ugyanakkor a CoMFA módszertől eltérően, nem egy meghatározott próbacsoportra számolt kölcsönhatási értékeket, hanem egy ún. rács betöltöttégi értéket (grid cell occupancy) rendel. Az így kapott érékekből PLS módszer felhasználásával építi fel a modellt.

Vedani és munkatársai számos publikációban[47-50] közölték az indukált illeszkedési modellen alapuló 4D QSAR eljárásuk működését és eredményeit. A nemrég megjelent, ún. 5D QSAR, ahol az ötödik dimenzió az egyes ligand molekulák különböző protonáltsági állapotának a figyelembevételére utal[51]. Az eljárásuk – meglehetősen ügyes ötlet alapján – a ligand molekulák körüli teret nem kockarács alapján írja le, hanem az indukált kötődéskor lejátszódó folyamatokat modellezi. Ennek első lépése hogy, az egyes molekulák van der Waals felszínére bizonyos tulajdonsággal rendelkező pontokat szór (hidrofób+semleges, +pozitív, +negatív töltés, H kötés akceptor, donor stb). Vedani ezt a felszínt belső burkolónak (inner envelope) hívja. Az egyes molekulák belső burkolóiból genetikus algoritmus segítségével kialakít egy minden molekulára közös burkolót (mean envelope), amely lényegében maga a kötőhely modellje lesz. A genetikus algoritmus az indukált illeszkedési folyamathoz számított kötődési szabadenergiát összeveti a mért értékekkel, és ennek alapján építi föl a közös burkolót. A módszer a közös és a belső burkoló felszínek geometriai különbségével jellemzi az egyes molekulák indukált illeszkedésének folyamatát, illetve predikciót is ez alapján végez. A kapott prediktív r^2 érték nagyon jónak mondható.

Az fent említett példában 65 darab neurokinin-1 receptor antagonistát vizsgáltak. Konformációs analízis segítségével feltérképezték a ligandok alacsony energiájú

[†] A dolgozat készítésének időpontjáig (2007 április) az idézett publikációra 118 hivatkozás érkezett.

konformereit. Az egyes ligandok „gerincét” öt forgatható torzió mentén definálták. Azt a gerinctorziók által leírt, minden ligand esetén megtalálható konformert fogadták el közös bioaktív konformerként, amelyik a legalacsonyabb energiával rendelkezett. Az egyes ligandok bioaktív konformerein csak kisebb változtatásokat végztek, (aromás gyűrű forgatása, protonáltság változtatása) így egy ligandra 1-6 konformáció adódott. A modellt az összes, így kiválasztott konformáció alapján számították ki.

Martinek és munkatársai konformációs flexibilitást leíró „3+3D” deszkriptorokat vezettek be. Az előre kiválasztott farmakoforpont-párok[52] távolságának eloszlását vizsgálták, majd kiralitástól függő flexibilitást leíró deszkriptorokat[53] dolgoztak ki. Mindkét esetben az egyik példaként 37 endomorfin analógból álló könyvtárat vizsgáltak. Az egyszerű, távolság alapú deszkriptorok felhasználásával a modell LOO q^2 (ld. 5.4.2.1. fejezet) értéke 0,52 és 0,58 között változott, míg a kiralitás figyelembe vételével ez az érték 0,62-re nőtt. Ugyanakkor a prediktív r^2 értéke a kiralitás figyelembe vételével nem javult (bár mindkét esetben nagyon jónak mondható 0,74-0,81 értékeket kaptak).

Bernard és munkatársai ugyancsak flexibilis opioid ligandok konformációs sokaságában belső koordináták segítségével kerestek farmakoforokat[54, 55].

5.4. A QSAR modellek statisztikai ellenőrzése

Egy adott szerkezet-hatás összefüggés elkészítésekor ún. *setubal*i[‡] irányelveket érdemes figyelembe venni:

1. A szerkezet-hatás összefüggésnek jól meghatározott végeredményt (endpoint) kell adnia, azért hogy a különböző kísérleti adatok alapján kapott modellek összehasonlíthatóak legyenek.

2. A modell felépítését végző algoritmusnak egyértelműnek, könnyen alkalmazhatónak, illetve az összehasonlíthatóság és a reprodukálhatóság miatt hozzáférhetőnek kell lennie.

3. Ideális esetben a modell felépítését végző algoritmusnak automatizálhatónak kell lennie.

4. A kémiai alkalmazhatósági területet (chemical domain of applicability) megfelelően definiálni kell, ugyanis a modell korlátaiból kifolyólag csak egy jól definiált,

[‡] „Workshop on Regulatory Use of (Q)SARs for Human Health and Environmental Endpoints”, Setubal, Portugália, 2002

kémiai szerkezetekkel vagy fiziko-kémiai paraméterekkel leírt térben ad megbízható predikciót.

5. Belső validálással a modell illeszkedésének jóságát,

6. külső validálással pedig a predikciójának jóságát kell ellenőrizni.

A felsorolt alapelvekből jól kitűnik, hogy egy szerkezet-hatás összefüggés felállításának általában három fő lépcsője van[56]: az adatok előkészítése, modellépítés, modellvalidálás. A dolgozatban egy olyan speciális, általam kifejlesztett QSAR módszert mutatok be, amely intenzíven épül a modellvalidálási[57-60] (ellenőrzési) eljárásokra, ezért e fejezetben ezekről részletesen lesz szó.

5.4.1. Korrelációs együttható és a determinációs együttható

QSAR modellek jóságának leírására általánosan használják az r^2-et, az ún. determinációs együtthatót (coefficient of determination). Az alábbiakban az r^2 szemléletes jelentését az egyváltozós lineáris modell esetén vezetem le.

Ha ξ és η valószínűségi változók kovarianciája létezik és szórásuk pozitív (nem nulla) a korrelációjukat az (1) képlettel definiáljuk, ahol a kovariancia (*cov*) a ξ és az η együttváltozásának mértéke.

$$\rho = \frac{\mathrm{cov}(\xi,\eta)}{\sqrt{\mathrm{var}(\xi)\,\mathrm{var}(\eta)}} \qquad (1)$$

$$\mathrm{cov}(\xi,\eta) = E\big((\xi - E(\xi))(\eta - E(\eta))\big) = E(\xi\eta) - E(\xi)E(\eta) \qquad (2)$$

$$\mathrm{var}(\xi) = E\big((\xi - E(\xi))^2\big) = E(\xi^2) - E^2(\xi) \qquad (3)$$

Ahol a (2) és (3) kifejezésekben az „E” a változó várható értékét jelenti. Ha például az egyik változó a várható értéke fölött, a másik alatta mozog, a kifejezés negatív lesz. Ha a két változó véges és független, azaz $E(\xi\eta)=E(\xi)E(\eta)$ teljesül, akkor $cov(\xi,\eta)=0$.

A korrelációval, amely nem más, mint az egységnyi varianciákra vonatkoztatott kovariancia, a két valószínűségi változó közötti kapcsolat szorosságát fejezhetjük ki. A korreláció értéke 1 növekvő, -1 csökkenő lineáris kapcsolat esetén, és $-1>\rho>1$ bármilyen más esetben. Minél közelebb van az értéke ±1-hez, a lineáris kapcsolat a két változó között annál erősebb. A korreláció csak a két változó lineáris kapcsolatát jellemzi. Ha a két változó független, korrelációjuk 0. Ez azonban fordított esetben nem igaz, például ha a ξ változó [-1, 1] intervallumban egyenletesen vesz fel értékeket és $\xi=\eta^2$, akkor a két változó nem független, a korrelációjuk mégis 0.

Jellemezzük a két valószínűségi változót egy-egy *n* mérésből álló sorozattal (x_1, x_i,…x_n és az y_1, y_i,... y_n). Ekkor, ha a változók normális eloszlásúak, ρ értékének legjobb becslését a Pearson-féle, vagy totális korrelációs együttható adja (4). A dolgozatban egy sorozat értékeinek, például (x_1, x_i,…x_n) átlagát $\bar{x}$-el jelölöm, és a szummázás minden esetben a szokásos indexeléssel történik (*i*, *j*…). A korrelációs együttható lényegében az (x_i- $\bar{x}$) és az (y_i- $\bar{y}$) vektorok által bezárt szög koszinuszát jelenti.

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \tag{4}$$

A determinációs együttható, amely kiszámítása alapvető követelmény a QSAR modellek validálásakor, a Pearson-féle korrelációs együttható négyzetének általánosításaként is felfogható. A következőkben a két együttható közötti különbséget mutatom be. A levezetésben „*s*”-sel a „sum of square” (négyzetösszeg) angol kifejezést rövidítem, emellett az alább feltüntetett s_{yy} az s_{xx}-el analóg módon adódik. E jelölésekkel a korrelációs együttható (4) négyzete az alábbi alakban adható meg (7).

$$s_{xx} = \sum (x_i - \bar{x})^2 = \sum x_i^2 - n\bar{x}^2 \tag{5}$$

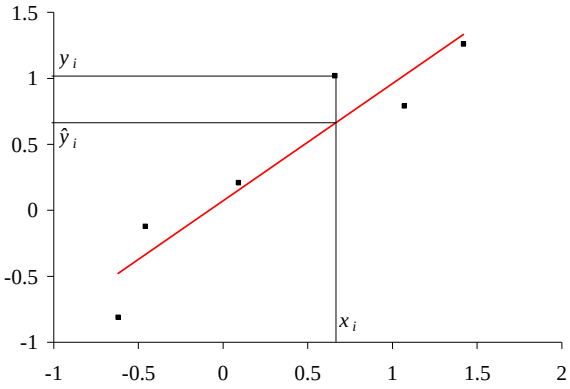
$$s_{xy} = \sum (y_i - \bar{y})(x_i - \bar{x}) = \sum x_i y_i - n\bar{y}\bar{x} \tag{6}$$

$$r^2 = \frac{s_{xy}^2}{s_{xx} s_{yy}} \tag{7}$$

Mivel a korrelációs együttható a változók lineáris összefüggésének a mértéke, illesszünk egyenest a pontokra a legkisebb négyzetek módszerével. A továbiakban ezt, az egyváltozós regressziós egyenest tekintem modellnek. Ha az egyenes egyenletét $y=a+bx$ alakban írjuk fel, a regressziós együttható a (9) alakban adható meg.

$$a = \bar{y} - b\bar{x} \tag{8}$$

$$b = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - \left(\sum x_i\right)^2} = \frac{s_{xy}}{s_{xx}} \tag{9}$$



1. ábra *Lineáris regresszió az adatpontokra. x_i a levezetésben használt független y_i a függő, $\hat{y}_i$ pedig az y_i becslését jelőli.*

Ekkor a (9)-es összefüggés felhasználásával felírhatjuk a „maradék" négyzetek összegét (residual sum of squares, s_{res}) (10), amelyből r^2 kifejezhető (11). Ekkor az x_i a független y_i a függő, $\hat{y}_i$ pedig az y_i becslését jelöli (**1. ábra**).

$$
\begin{aligned}
&s_{res} = \sum (y_i - \hat{y}_i)^2 = \sum (y_i - \bar{y} + b\bar{x} - bx_i)^2 = \\
&\sum (y_i - \bar{y} - b(x_i - \bar{x}))^2 = \\
&\sum (y_i - \bar{y})^2 + b^2 \sum (x_i - \bar{x})^2 - 2b \sum (x_i - \bar{x})(y_i - \bar{y}) = \\
&s_{yy} + b^2 s_{xx} - 2b s_{xy} = s_{yy} + \frac{s_{xy}^2}{s_{xx}^2} s_{xx} - 2\frac{s_{xy}}{s_{xx}} s_{xy} = \\
&s_{yy} - \frac{s_{xy}^2}{s_{xx}} = s_{yy}\left(1 - \frac{s_{xy}^2}{s_{xx} s_{yy}}\right) = s_{yy}(1 - r^2)
\end{aligned}
\qquad (10)
$$

$$
r^2 = 1 - \frac{s_r}{s_{yy}} = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \qquad (11)
$$

Ha tehát egyváltozós lineáris regressziót használunk modellként, és e modell építésekor az összes adatpontot felhasználjuk, a Pearson-féle korrelációs együttható négyzete (11) szerint felírható. Ha azonban bármilyen más modellt szeretnénk alkalmazni $\hat{y}_i$ számításához, akkor az r^2-et a determinációs együtthatónak (coefficient of determination) nevezzük. Gyakran használják az r^2-et a kísérletileg meghatározott és a QSAR modell által becsült akitivitások összehasonlítására.

Érdemes megjegyezni, hogy ha az összes adatpontot felhasználjuk a modell építésekor, és a modellünk az $y=a+bx$ egyenes, akkor az $r^2 \geq 0$. Ha azonban bizonyos számú pontot kihagyunk, az r^2 negatív értéket is fölvehet. QSAR modell jóságának becslésére gyakran használják ezt az r^2-et és az r^2-tel analóg kifejezéseket (5.4.2.1. fejezet).

A Pearson-féle korrelációs együttható négyzetének további néhány érdekes tulajdonságát a A. Függelék tartalmazza.

A dolgozatban, ha másként nem említem, a r^2-t alatt a determinációs r alatt a Pearson-féle korrelációs együtthatót értem.

5.4.2. Belső ellenőrzés

A belső ellenőrzés (internal validation) során többféle eljárás végezhető a tanító halmazra (training set) felállított modell robosztusságának és belső előrejelzési képességének (internal predictivity) becslésére.

A szakirodalomban a modell jóságának becslésére leggyakrabban a determinációs együttható (11) -es egyenlettel definiált összefüggését használják. Emellett megadható még az úgynevezett „*s*” értéke (12) is, amely a maradékok négyzetösszegének és a rájuk eső szabadsági fokok számának hányadosának négyzetgyöke v.ö. (16), ahol n az adatpontok és p a modellhez használt független változók száma (például $y=a+bx$ esetén $p=1$).

$$s = \sqrt{\frac{\sum_i \left(\hat{y}_i - y_i\right)^2}{n-p-1}} \tag{12}$$

Sok esetben alkalmazzák még a keresztellenőrzést (5.4.2.1.), a randomizációs tesztet (5.4.2.3.), ritkábban az úgynevezett bootstrap eljárást (5.4.2.2), és az F-próbát (5.4.2.4.).

5.4.2.1. Keresztellenőrzés

A keresztellenőrzést[56, 61] (cross validation magyar elnevezése[32]) gyakran használják a regressziós modell minőségének belső ellenőrzésére. Ez a technika általánosságban úgy működik, hogy a modell elkészítésekor bizonyos számú, ismert aktivitással rendelkező molekulát kihagynak a tanító halmazból, és az így kihagyott értékeket a modellből becsülik. Ezt az eljárást sokszor megismételve, statisztika készíthető a becsült és a mért értékek összehasonlításával. Ilyen módon, az ismételt felosztásokkal megbecsülhető a teljes tanító halmazra elkészített modell belső prediktivitása.

A k darab érték kihagyásával történő keresztellenőrzés[32] (k-fold cross validation) esetén a mintát k darab nagyjából egyenlő nagyságú alhalmazra osztják. Ezután k-szor felépítik a modellt, kihagyva mindig egy alhalmazt, majd a kihagyott elemekre kiszámítják a kívánt hibakritériumot. Így $\binom{n}{n/k}$ lehetőség állhat elő, ahol n a minta elemszáma.

Ha $k=n$, akkor egy érték kihagyásával történő keresztellenőrzésről beszélünk (leave one out crossvalidation, LOO). Ebben az eljárásban tehát minden értéket egyszer hagynak ki, és a modellt a maradék alapján építik fel. Ez a módszer széles körben elterjedt validálási módszer, de két esetben hibásan becsülheti meg a modellünk jóságát. Túlbecsülhet abban a gyakori esetben, amikor például egy ígéretes vezérmolekulát, és szerkezetileg kismértékben eltérő analógjait válogatnak össze a tanító halmazba. Ekkor ebbe a halmazba túl sok hasonló molekula kerülhet, így hiába hagynak ki egyet-egyet a keresztellenőrzés során, a modell mégis tartalmazni fog a kihagyotthoz közel azonos szerkezetet. Egy másik probléma az LOO keresztellenőrzés esetén akkor léphet fel, ha a tanító halmaz egy-egy kiugró pontot, a többitől jelentősen eltérő molekulát tartalmaz. Ekkor a maradékra illesztett modell a kiugró pontot extrapolációval, gyakran jelentős hibával becsli. Ez a becslés a végleges modell jóságát is jelentősen ronthatja.

Robosztusabb keresztellenőrzés végezhető úgy, hogy az adathalmazt csoportokra osztjuk (tipikusan négy, öt vagy tíz csoportra) és mindegyik csoportot egyszer kihagyva keresztellenőrzést végzünk (leave many out, LMO, például L10%O). Ezt az eljárást többször, általában százszor szokták elvégezni, véletlenszerűen összeválogatott csoportokkal. Az LMO során tehát több pontot is kihagynak egy ciklusban, de a kihagyott pontok száma egy keresztellenőrzés során állandó.

Keresztellenőrzés során leggyakrabban az alábbi hibafüggvényeket számolják ki. A modell belső becslőképességének (internal predictivity) egyik mértéke az úgynevezett prediktív „maradék” négyzetek összege (predictive residual sum of squares, a szakirodalomban „*PRESS”*-nek jelölik, így én is ezt a jelölést használom) (13), ahol $\hat{y}_{CV,i}$ az i-dik molekulára, az azt nem tartalmazó modellből becsült aktivitásérték. Kiszámíthatjuk az úgynevezett keresztellenőrzéssel előálló r^2-et (cross validated r^2, q^2) is. A q^2 értéke általában alacsonyabb, mint a vonatkozó r^2, és akár negatív értékeket is fölvehet (ld. 8.2. fejezet **9. ábra** és **10. ábra**). Míg az r^2 az illeszkedés jóságát, a q^2 értéke a belső predikció jóságát jellemzi. A q^2 definiáló képlete (14) analóg az r^2-re kapott kifejezéssel (11).

$$PRESS = \sum_i \left(\hat{y}_{CV,i} - y_i\right)^2 \tag{13}$$

$$q^2 = 1 - \frac{PRESS}{\sum_i \left(y_i - \bar{y}\right)^2} = 1 - \frac{\sum_i \left(\hat{y}_{CV,i} - y_i\right)^2}{\sum_i \left(y_i - \bar{y}\right)^2} \tag{14}$$

$$MSE = \frac{\sum_i \left(\hat{y}_{cv,i} - y_i\right)^2}{n} \qquad (15)$$

A (15)-ös kifejezésben definiált kifejezés az átlagos hibanégyzet (mean squared error), ahol *n* azon adatpontok száma (a szakirodalomban a kifejezés négyzetgyökét is használják, RMSE).

5.4.2.2. Bootstrap

E módszerben[62] az alapfeltevés az, hogy a teljes populációt a rendelkezésünkre álló *n* elemű adatsor megfelelően leírja. A feltétel nélküli (unconditional) bootstrap[§] eljárással azt vizsgáljuk, hogy ebből az egyetlen adatsorból vett bootstrap mintákkal (bootstrap samplings) hogyan tudjuk leírni a kiindulási populáció bizonyos tulajdonságait. A bootstrap minták a kiindulási *n* elem véletlenszerűen előállított n-ad osztályú ismétléses kombinációi. Vagyis az eredeti adatsorból egy elemet többször is beválogathatunk az új, ugyancsak *n* dimenziós bootstrap mintákba, és bizonyos értékek ki is maradhatnak.

A bootstrap mintákra egyenként, egymástól függetlenül kiszámoljuk a kívánt statisztikai paramétereket (pl.: q^2), majd ezeket átlagoljuk. Az átlag és az eredeti mintára számított érték különbségével az eredeti minta statisztikai torzítását, és/vagy különböző paraméterek (pl. regressziós együttható) konfidencia intervallumát is becsülhetjük.

Létezik egy, a bootstrapping módszerhez nagyon hasonló, ún. *Jackknife* eljárás, ahol az *n* darab (*n*-1) értékből álló mintát az LOO keresztellenőrzéshez hasonlóan, egy adat kihagyásával állítják elő, és az eredeti adatsorral összehasonlíva végezik el a statisztikai torzítás konfidencia intervallumát becslését[62].

5.4.2.3. Randomizációs teszt

A randomizációs teszt[57] arra használható, hogy az egyes megfigyelések (mérési pontok) függetlenségének mértékét ellenőrizzük. Ekkor a független változó értékeit randomizáljuk, azaz például a biológiai aktivitás értékeit véletlenszerűen újra hozzárendeljük a molekulákhoz, és így építünk modellt, amelyekre az r^2 és a q^2 értékeket kiszámoljuk. Ezt az eljárást sokszor (pl 200-szor) megismételjük. Ha a random modell predikciós képessége szignifikánsan rosszabb, mint az eredeti modellünké, biztosak lehetünk abban, hogy eredeti

[§] A bootstrapping angol szó, töve a bootstrap, jelentése a lovaglócsizma szárának tetején megtalálható hurok, amelyet fogantyúként használva a csizmát könnyebb felhúzni. A származtatott igét úgy fordíthatjuk, hogy „felépíteni külső segítség nélkül”. A továbbiakban az eljárás angol nevét fogom használni.

modellünk jól megalapozott, és nem csak a véletlen korreláció műve. Ha azonban a random modellek r^2 és q^2 értéke összevethető az eredetiével, akkor az eredetit nem szabad elfogadnunk.

5.4.2.4. A modell szignifikanciája – F-próba

QSAR modellek esetén az F-próba számolt értéke (16) egy olyan hányados, ahol a számláló a „becslésből adódó" négyzetösszegek (estimated sum of squares) és a rá jutó szabadsági fokok hányadosa, a nevező pedig a „maradék" négyzetösszegek (residual sum of squares) és a rá jutó szabadsági fokok hányadosa. Ahol „*p*"-vel a modell független változóinak, „*n*"-el a mért adatpontok számát jelöltem.

$$F = \frac{\sum_i (\hat{y}_i - \bar{y})^2 / p}{\sum_i (\hat{y}_i - y_i)^2 / n - p - 1} \qquad (16)$$

Gyakori az az eset, amikor az *x* tengelyen a mért, az *y*-on a számolt aktivitást ábrázolják. Ekkor a két változó közötti összefüggés egy 0-n átmenő 1 meredekségű egyenessel írható le, így (p=1) a becslésből adódó négyzetösszegekre 1, a „maradék" négyzetösszegekre n-2 szabadsági fok jut. Így (11) és (16) felhasználásával az alábbi összefüggés adódik F-re.

$$F = (n-2)\frac{r^2}{1-r^2} \qquad (17)$$

Az F-próba esetében általános esetben a nullhipotézis (H_0) az, hogy két normális eloszlású populáció varianciája egyenlő, így feltételezhető, hogy a populációk azonos eredetűek. Jelen körülmények között H_0 úgy fogalmazható meg, hogy a modellünkkel nem készíthető jobb becslés, mint a mért (aktivitás) értékek átlagával. A H_1 hipotézis a H_0 ellentéte. Az F-próbához tartozó táblázatból kikeressük (vagy az F eloszlás alapján ki is számíthatjuk) az aktuális két szabadsági foknak és a kívánt szignifikancia szintnek (α) megfelelő $F_{1,n-2,\alpha}$ értékét. Ha $F > F_{1,n-2,\alpha}$, H_1-et elfogadjuk. Vagyis F értéke minél nagyobb, annál jobb a modellünk. Mind (16) és (17) gyakran előfordul az irodalomban. A H_0 és H_1 közti határ szignifikancia szintet az aktuálisan adatainkra számolt F értékből (18)-as kifejezésből kifejezhetjük, ahol $F_{1,n-2,\alpha}$ a kívánt szabadsági fokok melletti F eloszlás.

$$F = F_{1,n-2,\alpha} \qquad (18)$$

5.4.3. Külső ellenőrzés

A modell igazi predikciós képessége csak egy olyan külső teszthalmazon ellenőrizhető le, amelyet a modell építéséhez nem használtunk fel. Ha nem állnak rendelkezésre kiegészítő mért értékek (általában ez az eset), a rendelkezésre álló adatsorunkat felosztjuk egy tanító- (training) és egy teszthalmazba (prediciton, validation vagy test set). A modellt a tanító halmazon készítjük el, és külső előrejelző képességét (external predictivity) a teszt halmazon határozzuk meg.

Ilyen ellenőrzésre használható a prediktív r^2 (más néven a prediktív q^2) (19). Ahol a számlálóban a teszthalmaz egyes molekuláira a modell alapján becsült és a valós érték különbségének négyzetösszege található. Mivel a modellben nem használtuk fel a teszthalmaz molekuláit, a (19)-es összefüggés alapján számolt külső prediktív r^2 sokkal jobban tükrözi a modell becslő teljesítményét (predictivity), mint a belső ellenőrzésből származó q^2. A (19)-es kifejezésben $y_{i,test}$ és $\hat{y}_{i,test}$ a teszthalmaz egyes molekuláira vonatkozó mért ill. számított akitivitásoknak, az $\bar{y}$ a tanítóhalmaz molekuláinak mért aktivitásának az átlagát jelenti. Meg kell jegyezni azonban, hogy $\bar{y}$-ot némely esetben a teszthalmaz aktivitásaira számolják. (Az első[39], és a második[63] esetre is van példa a szakirodalomban). A (13)-mas és a (15)-ös kifejezést külső ellenőrzésre is szokták használni.

$$r^2_{pred} = 1 - \frac{\sum \left(y_{i,test} - \hat{y}_{i,test}\right)^2}{\sum \left(y_{i,test} - \bar{y}\right)^2} \qquad (19)$$

Előfordulhat, hogy a módszerünk a rendelkezésre álló molekulák halmazából automatikusan felépíti a legkisebb, még jól prediktáló tanítóhalmazt, és az elkészült modell jóságát a maradék molekulákból álló teszthalmazon vizsgálja. Ekkor az elkészült modell ugyan közvetlenül nem tartalmaz információt a teszthalmazról, közvetve minden molekula figyelembe lett véve a tanító- és teszthalmaz kijelölésekor. Ezért ilyenkor egy olyan, úgynevezett *teszt-2* molekulahalmazt szoktak alkalmazni a modell valós predikciójának vizsgálatára, amely molekulái nem vettek részt a tanító- és teszthalmaz elemeinek kialakításakor.

5.4.4. A Kubinyi paradoxon

A keresztellenőrzés során kapott magas q^2 értékek a modell jó robosztusságát és belső prediktív képességét mutatják. Feltételezhetnénk, hogy ilyen esetben jó prediktív r^2-et kapunk. A szakirodalomban azonban gyakran tapasztalták, hogy nem lehet korrelációt

felfedezni a belső ellenőrzéssel kapott q^2 és a külső ellenőrzéssel kapott prediktív r^2 érték között[56, 64-66].

A korreláció hiánya a belső és a külső ellenőrzés eredményei között visszavezethető többek között a felhasznált „független" változók kollinearitására, túlillesztésre, vagy extrapolációra.

Kollinearitásnak nevezzük azt az esetet, amikor a függetlennek hitt változók között lineáris kapcsolat áll fenn, így bizonyos redundancia lép fel a rendszerben. Ekkor előfordulhat, hogy a modellnek csak látszólagos prediktív ereje lesz. Egy lehetséges mód a választott deszkriptorok (multi)kollinearitásának vizsgálatára az u.n. QUIK (Q Under Influence of K) szabályok használata[67].

Túlillesztési[68] (overfitting) hiba fordulhat elő, ha például túl sok deszkriptort alkalmazunk a modell építéséhez és/vagy a modellünk flexiblisebb annál, mint amilyennek lennie kellene. Ilyen eset például, ha valamilyen „görbe vonalú" (curvilinear) összefüggéssel modellezünk egy amúgy lineáris kapcsolatot. Másik ilyen eset, ha a modellt úgy építjük fel, hogy jó belső ellenőrzésből származó értékeket kapjunk (q^2-re optimáljuk), teszthalmazon ez a modell még mutathat nagyon rossz predikciós eredményt. Túlillesztést történhet akkor is, ha irreleváns deszkriptorokat is tartalmaz a modellünk.

A modellből extrapolált eredmények lehetnek teljesen megbízhatatlanok is, ezért törekednünk kell olyan becslések előállítására, amelyek interpolációval készülnek. Ezért fontos eldönteni, hogy az aktuális predikció éppen inter- vagy extrapolációnak számít-e. A döntést segíti a leverage (befolyás, továbbiakban az angol szót használom) függvény, amely többdimenziós modell esetében is használható. Például többváltozós lineáris regresszió (multiple linear regression) esetén az úgynevezett *Hat* mátrix (H) az alábbi formában írható fel (20), ahol X nxk méretű mátrix sorai tartalmazzák a deszkriptorokat *(k)*, ahol n a minta elemszáma, a deszkriptorok száma.

$$H = X\left(X^T X\right)^{-1} X^T \qquad (20)$$

A *Hat* mátrix tehát egy projektor, és dimenziója (n x n) lesz. Ha a deszkriptorok lineárisan függetlenek, akkor X^TX invertálható, egyébként az általánosított inverze használható. A H mátrix egyes diagonális elemeit (h_{ii},) nevezik az aktuális mintaelemre vonatkozó *leverage*-értéknek. Egydimenziós lineáris regresszió esetén, amikor k=1, (20)-as kifejezés a (21)-es formára egyszerűsödik[68], ahol x_i a minta elemei, $\bar{x}$ a mintaelemek értékének átlaga, és x_j a becsülendő elem.

$$h = \frac{1}{n} + \frac{(x_j - \bar{x})^2}{\sum_i (x_i - \bar{x})^2} \qquad (21)$$

Adott $p \leq k$ dimenziós modell esetén, a h minimális értéke $1/n$, maximális értéke 1.0 és az átlagos értéke p/n lesz. Ha $p=k$ akkor a deszkriptorok lineárisan függetlenek, más esetben $p<k$. Alacsony h érték esetén interpolációról, magasak esetén extrapolációról beszélünk, határnak az átlagos érték két- háromszorosát szokták használni. Ha a *leverage* értéket a tanító halmaz elemeire kiszámítjuk, megkaphatjuk, hogy az egyes elemek mennyivel járulnak hozzá a regressziós modellünkhöz (h minél nagyobb annál inkább). Ha nagy *leverage* értéket kapunk a teszthalmaz bizonyos elemeire, akkor ezeket az elemeket a modellünk extrapolációval, megbízhatatlanul becsli.

Alapvető cél, hogy mind a tanító- és a teszthalmaz külön-külön lefedje a teljes, deszkriptorok által leírt azon kémiai teret, amelyet a kiindulási adatsor is felölel. Az ideális felosztás esetén a tanítóhalmaz minden eleméhez találunk egy megfelelően „közeli" elemet a teszthalmazból[69, 70]. A legegyszerűbb ilyen eljárás a véletlenszerű felosztás módszere, amelytől kifinomultabb – többek között – a „D-optimális kísérleti tervezés" (D-optimal experimental design)[71] és a Kohonen mesterséges ideghálózatok (Artificial Neural Network) alkalmazásával működő módszer[33].

Ha tehát a modellt prediktív (előrejelző) céllal is akarjuk használni, akkor a belső ellenőrzés szükséges, de önmagában nem elégséges, külső ellenőrzést (external validation) is tanácsos végezni.

5.4.5. Összefoglalás

Általánosan egy QSAR modell építéséhez minél kevesebb és lehetőleg független deszkriptort érdemes felhasználni. A kiindulási adatsorból úgy válogassuk ki a tanító és a teszt halmaz molekuláit, hogy mindkét halmazban megfelelően diverzek legyenek, viszont az egyes szerkezetekhez találunk hasonlót a másik halmazból. Ekkor az r^2_{pred} értékek jó eséllyel interpolációval állnak elő. A modellünk megfelelő[72] ellenőrzéséhez általános esetben érdemes alkalmazni a randomizációs tesztet, a több érték kihagyásával történő keresztellenőrzést, és a külső validálást.

6. Célkitűzés

Doktori munkám során első feladatom az volt, hogy egy mucin epitóp pentapeptideket tartalmazó molekulatár esetén határozzam meg a kötődésért felelős konformációt. A molekulatár egyes tagjai kísérletileg meghatározott kötődési aktivitás értékkel rendelkeztek, de a célfehérje háromdimenziós térszerkezete ismeretlen volt. Miután a szakirodalomban nem volt olyan módszer, amely e feladatot megnyugtató módon képes lett volna megoldani, célul tűztük ki, egy általánosan használható, ilyen eljárás megvalósítását. E célt az alábbi lépések során kívántam elérni:

1. a módszer elméleti hátterének kidolgozása – a használhatóságához szükséges feltételek vizsgálata.
2. A számítást végző algoritmusok kidolgozása és megvalósítása.
3. A bioaktív-konformáció kereső eljárás elvégzése és a számítási eredmények értékelése a mucin epitóp peptidkönyvtáron.
4. A módszer tesztelése további példákon.

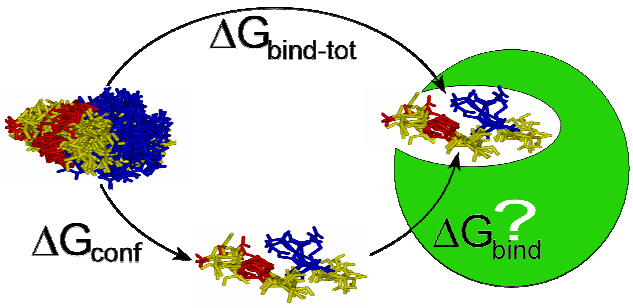
7. A 3D QSAR elméleti háttere flexibilis molekulákra

A 3D eljárások általában olyan, könnyen magyarázható modellt adnak, amelyek segítségével a kötőhely térszerkezetére következtethetünk. Ugyanakkor a molekulák egymásra illesztése ilyen módszerekben legtöbbször nagyon nehézkes és gyakran emberi beavatkozást is igényel. Ha a közös farmakofort (amely alapján összeforgatnánk a molekulákat) nem ismerjük, a módszerek automatizálhatósága nehezen oldható meg[37, 39, 73]. Megállapíthatjuk tehát, hogy a bioaktív konformáció (farmakofor) előzetes ismeretével a ligand alapú 3D VS és 3D QSAR módszerek hatékonysága jelentősen növelhető.

Ebben a fejezetetben a ligand molekulák kísérletileg meghatározott kötődési aktivitás értékei és a farmakofor közötti összefüggést vezetem le. E levezetés azon túl, hogy a ligand molekulák e két tulajdonsága közötti kapcsolatot feltárja, megmutatja a használhatóságához szükséges feltételeket.

7.1. A ligand konformációs flexibilitásának hatása a kötődésre

Tekintsük egy molekula kötődését egy ismeretlen térszerkezetű kötőhellyel rendelkező célfehérjéhez. A kötődési folyamatot felbonthatjuk két részfolyamatra; 1. a bioaktív konformáció (BC) kialakulása az oldatban, szabad állapotban, 2. ennek a konformációnak a rigid kötődése az aktív helyhez (**2. ábra**).



2. ábra *Balról jobbra haladva: A molekula lehetséges konformációi szabad állapotban, bioaktív konformáció ugyancsak szabad állapotban, és a bioaktív konformáció az ismeretlen térszerkezetű kötőhelyen. Az ábrán feltüntetett molekulák csak reprezentációs célt szolgálnak.*

Így a kötődéssel járó teljes szabadentalpia változást ($\Delta G_{bind-tot}$) felírhatjuk két tag összegeként[74] (22).

$$\Delta G_{bind-tot} = \Delta G_{conf} + \Delta G_{bind} \qquad (22)$$

ΔG_{conf} a BC oldatbeli kialakulásával, ΔG_{bind} pedig a BC elméleti, rigid kötődésével járó szabadentalpia változás. Jelöljük a fehérjéhez kötött molekula (ligand) egyensúlyi koncentrációját [l_{bound}]-dal, az oldatban a bioaktív konformer egyensúlyi koncentrációját [l_{BC}]-vel, a legstabilabb konformer egyensúlyi koncentrációját pedig [l_0]-lal. Fontos megjegyeznem, hogy a levezetésben a fiziko-kémiai aktivitásokat koncentrációértékekkel közelítettem. Ekkor:

$$[l_{bound}] = [l_0] e^{-\beta\left(\Delta G_{conf} + \Delta G_{bind}\right)} \tag{23}$$

$$[l_{BC}] = [l_0] e^{-\beta\left(\Delta G_{conf}\right)} \tag{24}$$

ahol β hőmérsékletfüggő állandó (statistical beta).

$$\frac{[l_{bound}]}{[l_{BC}]} = e^{-\beta \Delta G_{bind}} = \frac{\frac{[l_{bound}]}{l_{tot}}}{\rho_{BC}} \tag{25}$$

A (23)-as és (24)-es egyenlet szerint a kötődött és a szabad állapotú bioaktív konformer hányadosát a (25)-ös egyenlet szerint írhatjuk fel. A hányadost a ligand teljes bemérési koncentrációjával bővítve (l_{tot}) a nevezőben az oldatban megvalósuló bioaktív konformáció relatív koncentrációját kapjuk (26).

$$\rho_{BC} = \frac{[l_{BC}]}{l_{tot}} \approx \frac{f_{BC}}{n_{traj}} \tag{26}$$

Tegyük fel, hogy tudunk végezni egy olyan számítási eljárást, amely a molekula oldatbeli konformációs eloszlását képes leírni. Ebből az eloszlásból egyenletesen mintavételezünk n_{traj} darab konformációt, ekkor f_{BC} a bioaktív konformáció gyakorisága. Definiáljuk magát a bioaktív konformációt, egy előre meghatározott konformációs térben egy véges, konformációs térfogatként (V_{BC}). Ekkor ebbe a térfogatba a konformációk pontosan f_{BC}/n_{traj} hányada esik. A továbbiakban az ebbe a térfogatba eső relatív konformációs sűrűségértéket ρ_{BC}-vel jelölöm.

Vizsgáljuk meg egy bizonyos célfehérjéhez kötődő molekulák halmazát. Ha ezek a molekulák a kötődésért felelős konformációt szabad állapotukban is felveszik, akkor a (27)-es kifejezés igaz a halmaz bármely *i* és *j* molekulájára.

$$\frac{\frac{[l_{bound}(i)]}{l_{tot}(i)}}{\frac{[l_{bound}(j)]}{l_{tot}(j)}} = \frac{\rho_{BC}(i)}{\rho_{BC}(j)} e^{-\beta\left(\Delta G_{bind}(j) - \Delta G_{bind}(i)\right)} \tag{27}$$

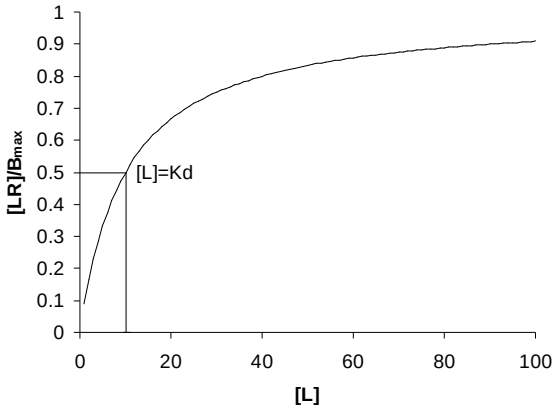
A (27)-es egyenletben a bal oldalon a mérhető kötődési aktivitás értékekből képzett hányados található. Azt, hogy a különböző standard kísérleti módszerek által szolgáltatott eredményekből hogyan kaphatjuk meg ezt a relatív koncentrációt 7.2-es fejezetben fejtem ki.

A dolgozat témája, a kötődési aktivitás függése a konformációs flexibilitástól, megkívánja, hogy a vizsgált molekulatár bizonyos feltételeknek megfeleljen, e feltételeket a 7.3-as fejezetben tárgyalom.

7.2. Relatív koncentráció meghatározása kísérleti módszerekkel

7.2.1. Telítési kötődési vizsgálat

A telítési kötődési vizsgálat esetén telítési görbét (**3. ábra**) vesznek fel úgy, hogy a célfehérjéhez fokozatosan adagolják a jelölt kötődő molekulát (ligand) és a kötődéssel arányos jelet mérnek[75]. Ekkor a (28) egyenletnek megfelelő reakció játszódik le. A telítési görbe egyenlete (29), ahol B_{max}=[R]+[LR] a kötőhelyek maximális koncentrációja.



3. ábra *Telítési görbe.*

$$\mathrm{L+R} \leftrightharpoons \mathrm{LR} \quad K_d = \frac{[L][R]}{[LR]} \tag{28}$$

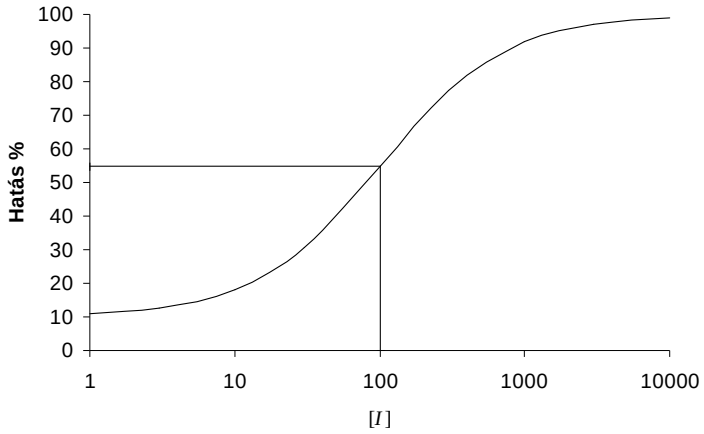
$$\frac{[LR]}{B_{\max}} = \frac{[L]}{[L]+K_d} \tag{29}$$

Telítési kötődési vizsgálat esetén a két standard végeredményből (B_{max}, K_d) a keresett koncentrációhányados megkapható (30). A (30) egyenlethez felhasználtam az [L]≈[L_0] közelítést és a relatív koncentrációt az [L]=K_d pontban számoltam ki. Ekkor ugyanis az [R]=B_{max}/2=[LR].

$$\frac{[\mathrm{l}_{\mathrm{bound}}]}{\mathrm{l}_{\mathrm{tot}}} = \frac{[LR]}{L_0} = \frac{[L][R]}{K_d[L]} = \frac{B_{\max}}{2K_d} \tag{30}$$

7.2.2. Dózis-hatás görbe felvételén alapuló módszerek

Az agonista, antagonista, illetve az inhibítor kötődési erősségét leggyakrabban az EC_{50} IC_{50} vagy a K_i értékek valamelyikével jellemzik, így célszerű a keresett relatív koncentráció értékét ezen, standard módon nyerhető mért adatokból levezetni. A mérést dózis-hatás görbék (**4. ábra**) felvételével végzik. A görbe általános egyenlete (31) négy paramétertől függ, az alapvonalként adott minimális (min), és maximális (max) választól, a görbe meredekségétől (m) és az 50% hatáshoz tartozó koncentrációértéktől.



4. ábra *Sematikus dózis-hatás görbe. Inhibíció vagy antagonista esetén a görbe az ábrázolt tükörképe is lehet. Mivel általában széles koncentrációtartományban veszik fel az adatokat, a vízszintes tengelyen logaritmikus skálát használnak.*

$$válasz = \min + \frac{(\max - \min)}{1 + \left(\frac{EC_{50}}{[L]}\right)^m} \tag{31}$$

Ha egy ligand kötődését vizsgáljuk, az 50% hatáshoz tartozó koncentráció az EC_{50}. Ha a „hatás” komplex folyamatok összessége, akkor az EC_{50} nem direkt mértéke a molekula kötődési erősségének, mint például a K_d. Jelen körülmények között ezért ki kell kötnünk, hogy a „hatás” közvetlenül a kötődött ligand molekulák számától függ.

Az IC_{50} és K_i meghatározásakor az *in vitro* tesztekben két kompetitív reakció játszódik le, (32) és (33), ahol a „ * ” szimbolizálja a molekula radioaktív, vagy egyéb módon történő jelölését.

$$L^{*}+R \leftrightharpoons L^{*}R \quad K_d = \frac{[L][R]}{[L^{*}R]} \tag{32}$$

$$I+R \leftrightharpoons IR \quad K_i = \frac{[I][R]}{[IR]} \tag{33}$$

Konstans bemérési ligand (L) koncentráció és változó inhibitor (I) koncentráció mellett rögzítik a dózis-hatás görbét, azaz [I] függvényében [L*R]-el arányos jelet vesznek fel. Az IC_{50} az az inhibitor koncentráció, amely a jelen lévő ligand kötődését a felére csökkenti.

Az előző fejezetben a keresett relatív koncentrációt a ligand egyensúlyi koncentrációján [L] keresztül fejeztem ki, ebben az esetben erre az inhibitor egyensúlyi koncentrációját [I] használom.

$$B_{\max} = [R]+[L^{*}R]+[IR] = (1+\frac{[L^{*}]}{K_d}+\frac{[I]}{K_i})[R] \tag{34}$$

$$\frac{[\mathrm{l}_{\mathrm{bound}}]}{\mathrm{l}_{\mathrm{tot}}} = \frac{[IR]}{I_0} = \frac{IC_{50}[R]}{IC_{50}K_i} = \frac{B_{\max}}{(1+\frac{[L^{*}]}{K_d})K_i + IC_{50}} = \frac{B_{\max}}{2IC_{50}} = \frac{B_{\max}}{2(1+\frac{L_0}{K_d})K_i} \tag{35}$$

A (35)-ös egyenletben, a második, lépésben a $[I] \approx I_0$ közelítést alkalmaztam, ahol I_0 a teljes bemérési inhibítor koncentráció. Ugyanitt az 50% hatáshoz rendelt $I_0 = IC_{50}$ koncentrációértéket használtam. A harmadik lépésben behelyettesítettem a Cheng-Prusoff[76] egyenletet (36).

$$IC_{50} = K_i\left(1+\frac{[L]}{K_d}\right) \tag{36}$$

A 7.2-es fejezetben alkalmazott közelítések ($[L] \approx [L_0]$ és $[L] \approx [L_0]$) végeredményre gyakorolt hatását a B. Függelék tartalmazza.

7.3. A molekulatár speciális tulajdonságai

Ahhoz, hogy a (27)-es egyenletben molekulák kötődési aktivitása csak a konformációs flexibilitásuktól függjön olyan molekulatárra van szükségünk, ahol a (37)-es kifejezés teljesül. E kritériumot és következményeit vizsgálom meg részletesebben e fejezetben.

$$\Delta G_{bind}(i) = \Delta G_{bind}(j) \tag{37}$$

Egy általános molekulatár egy tagjának, a célfehérjéhez való kötődésével járó szabadentalpia változása felírható[48-51, 77];

$$\Delta G_{\text{bind-tot}} \approx E_{\text{lig}-\text{rec}} + \Delta E_{\text{lig}} + \Delta E_{\text{rec}} - \Delta E_{\text{solv,lig}} - T\Delta S_{\text{bind}} \tag{38}$$

ahol $E_{\text{lig-rec}}$ a ligand-receptor kölcsönhatás energia hozzájárulása, ΔE_{lig} a ligand, ill. ΔE_{rec} a receptor energiájának változása a kötődés során. A $\Delta E_{\text{solv,lig}}$ tag a ligandot körülvevő oldószer molekulák kötődés során bekövetkező eltávolításával járó energiaváltozás. ΔS_{bind} pedig a ligand kötődéskor bekövetkező entrópia változás, amelyet tovább bonthatunk (39), ahol ΔS_{rot} a ligand kötődésével járó rotációs, ΔS_{trans} a transzlációs és ΔS_{conf} a konformációs entrópiaváltozás. A (38)-as egyenletbe a (39)-et behelyettesítve, és szétbontva (22) szerint a (40) és (41) egyenleteket kapjuk. Célunk tehát egy olyan molekulakönyvtár összeállítása, ahol a (41)-es kifejezés (40)-hez képest konstansnak tekinthető.

$$\Delta S_{\text{bind}} = \Delta S_{\text{rot}} + \Delta S_{\text{trans}} + \Delta S_{\text{conf}} \tag{39}$$

$$\Delta G_{\text{conf}} = \Delta E_{\text{lig}} - T\Delta S_{\text{conf}} \tag{40}$$

$$\Delta G_{\text{bind}} = E_{\text{lig}-\text{rec}} + \Delta E_{\text{rec}} - \Delta E_{\text{solv,lig}} - T\left(\Delta S_{\text{rot}} + \Delta S_{\text{trans}}\right) \tag{41}$$

Tegyük föl, hogy egy tár minden molekulája ugyanazon a szubstruktúrán keresztül, ugyanolyan módon kötődik a célfehérje ugyanahhoz a kötőhelyéhez. Ezt a közös szubstruktúrát PCORE-nak neveztem el az angol **p**harmacophore[6] és **core** szavak összevonásával. A PCORE-t tehát a különböző molekulák azon atomjai alkotják, amelyek a kötődés szempontjából ekvivalensek, és részt vesznek a fehérje – kismolekula kölcsönhatás kialakításában. Azaz a PCORE atomjait farmakofor pontoknak tekintjük, magát a PCORE-t pedig olyan, adott geometriával rendelkező farmakofor n-letnek, amely az összes molekulában megtalálható. A PCORE bioaktív konformációja a definícióból adódóan a farmakofor. A molekulák azon szerkezeti elemeit, amelyek nem tartoznak a PCORE-hoz off-PCORE-nak neveztem el.

Tisztán a flexibilitás vizsgálatához olyan speciális molekulakönyvtárra van szükség, ahol a PCORE elegendően sok szabadon forgatható kötéssel (rotatable bond) rendelkezik, továbbá az off-PCORE csak közvetve vesz részt a kötődésben, azáltal, hogy a PCORE konformációs eloszlását befolyásolja. Szükséges továbbá, hogy minden molekula esetén ugyanaz a jól meghatározott PCORE ugyanolyan bioaktív konformációja alakítsa ki a kötődést, ilyenkor a kötődési mód várhatóan ugyanaz lesz[78]. Ebből következik, hogy a $E_{\text{lig-rec}}$ kölcsönhatási energia minden molekula esetében jó közelítéssel ugyanannak az értéknek

tekinthető. A kötőhely energiájának megváltozása, és a deszolvatációs hatás is állandónak vehető ugyan ezen ok miatt. A rotációs és a transzlációs entrópiaváltozást is konstansnak tekinthetjük[79]. Ekkor a konformációs flexibilitást jellemző tag (27) változásához képest a (41) egyenletben szereplő kifejezés is reményeink szerint állandónak tekinthető.

A PCORE – off-PCORE koncepció jól rímel arra az esetre, amikor az enzimek szubsztrátja vagy a receptorok ligandja peptid vagy fehérje (pl. angiotenzinogén, angiotenzin, fibrinogén mint a fibrin prekurzora, a HIV proteáz fehérje és egyéb HIV fehérjék prekurzorai, serpinek, enkefalinok, neurokininek, szomatosztatin, vitronectin…)[3]. A jelátviteli folyamatok esetén történő fehérje-fehérje kölcsönhatásokkal ellentétben a peptid ligandok kötődését leggyakrabban csak néhány aminosav határozza meg, maradék polipeptid vagy fehérje legtöbbször csak a megfelelő konformációját stablizálja. Jó példa erre az RGD (arginin, glicin, aszpartát) motívum, amely különböző integrinekkel különböző konformációban hat kölcsön.

7.4. A konformációs flexibilitás – hatás összefüggés

A relatív koncentráció a standard *in vitro* módszerek eredményeiből tehát számítható, és minden esetben partícionálható egy olyan tagra, amely egy molekulakönyvtárra történő mérés esetén állandónak vehető (k_b), és egy molekulaspecifikus tagra (**2. táblázat**). A molekulaspecifikus tagot (b) továbbiakban kötődési aktivitásnak fogom nevezni. A kötődési aktivitás értékét definíciószerűen zérusnak tekintem inaktív molekulák esetén.

A szakirodalomban a QSAR-nak megfelelő egyenltetet általában a molekulák kísérletileg meghatározott $\ln(1/K_i)$ (vagy $\ln(1/IC_{50})$) értékei és a különféle számolt deszkriptorok között állítják fel. Ezzel az egyenlettel analóg összefüggést vezethetünk le a (25)-ös egyenlet és a **2. táblázat**ban foglaltak segítségével (42).

$$\ln(b) = \ln(\rho_{BC}(m)) - \beta \Delta G_{bind}(m) - \ln(k_b) \qquad (42)$$

Az egyenlet bal oldalán a tár m molekulájára vonatkozó, kötődési szabadentalipiával arányos érték áll. Az m molekula bioaktív konformációjának relatív sűrűsége (valószínűsége) az oldatban $\rho_{BC}(m)$, ennek a konformációnak a rigid úton történő kötődési folyamatához rendelhető szabadentalpia változás $\Delta G_{bind}(m)$. Az egyenlet jobb oldalát próbálják különféle deszkriptorokkal (2D, 2.5D, 3D) azonosítani, majd ún. „puha” modellépítő (soft modelling) eljárásokkal az összefüggést felállítani. A „puha” modellépítő eljárásokra az a jellemző, hogy

céljuk a függő és a független változók közötti kapcsolat felépítése, még akkor is, ha e kapcsolat mögött nincs valós természettudományos magyarázat.

	Kötődési aktiviás (b)		Konstans tag (k_b)
$\frac{[\mathrm{l}_{\mathrm{bound}}]}{]_{\mathrm{tot}}} =$	$\frac{1}{K_d}$	*	$\frac{B_{\max}}{2}$
	$\frac{1}{EC_{50}}$	*	$\frac{B_{\max}}{2}$
	$\frac{1}{IC_{50}}$	*	$\frac{B_{\max}}{2}$
	$\frac{1}{K_i}$	*	$\frac{B_{\max}}{2(1+\frac{L_0}{K_d})}$

***2. táblázat**: A relatív koncentráció számítása különböző standard módszerek eredményéből.*

A különösen flexibilis molekulák kezelésére más utat választottam, a (42)-es összefüggésben a ΔG_{bind} tagot a tár egyes molekuláira közel állandónak tekintettem, ekkor a molekula kötődési aktivitása egyenesen arányos a $\rho_{BC}(m)$ értékével (43), ahol α az egyenes meredeksége.

$$\rho_{BC}(m) = \alpha * b(m) \tag{43}$$

Tehát az egyes molekulák kötődési aktivitás erőssége egyenesen arányos a bioaktív konformáció oldószerben, fehérje jelenléte nélkül szimulált konformációs sűrűségével. Így a fenti kikötések mellett a bioaktív konformáció oldatbeli megvalósulása és a kötődési aktivitás között egy zérus *y* tengelymetszetű, pozitív meredekségű lineáris összefüggés áll fenn (43).

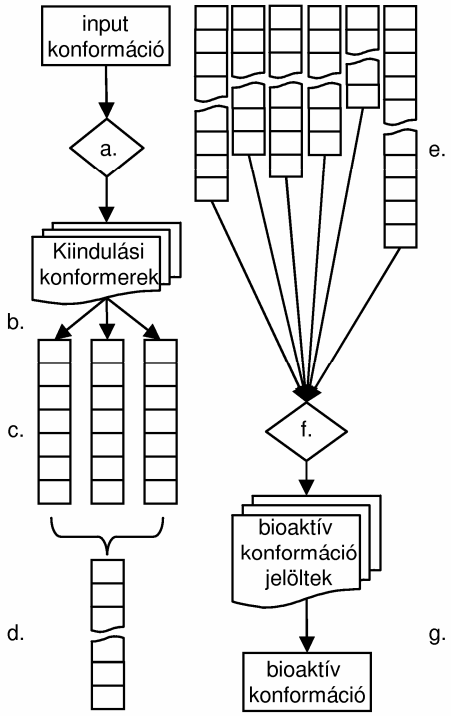
Ha az egyes molekulák konformációs eloszlását meghatározzuk, a bioaktív konformációnak megfelelő konformációs térfogat jellemzője lesz tehát, hogy az ebbe a térfogatba eső, a kötődő molekulákhoz rendelt konformációs-sűrűség értékek egyenesen arányosak a megfelelő mért IC_{50} vagy K_i értékek reciprokával.

A dolgozatban a különféle algoritmusok bioaktív konformáció jelölteket eredményeznek, azaz az összes megtalálható, megfelelő konformációs térfogatot (V) felsorolják. E térfogatokhoz tartozó relatív sűrűséget ρ_V-nek jelölöm. A jelöltek közül a bioaktív konformációnak megfelelőt mindig a megfelelő molekulatár tulajdonságait mérlegelve választottam ki.

8. Számítási eljárások

Doktori munkám során olyan számítási eljárást terveztem és valósítottam meg, amely egy molekulatár esetén előállítja az egyes molekulák számított konformációs eloszlását és megkeresi azokat a konformációs térfogatokat, amelyek a (43)-as összefüggés alapján kötődő konformációkat tartalmazhatnak.

Az eljárás[80-85] két független, a konformációs sokaság modellezését, és a bioaktív-konformáció keresését végző modulból áll. Az algoritmusok megvalósításakor szempont volt az automatizálhatóság. A módszer blokkvázlatát a **5. ábra** szemlélteti.



5. ábra *A konformációs eloszlás feltérképezését végző modul működésének blokkvázlata egy adott molekula esetén (a-d). a. kiindulási konformercsalád generálás. b. MD szimuláció indítása a kiindulási konformerekből. c. trajektóriák, d. összefűzött trajektória. Bioaktív konformációkereső modul blokkvázlata (e-g). e. a vizsgálandó molekulák összefűzött trajektóriái. f. Bioaktív konformáció jelöltek feltérképezése. g. klaszterezés, manuális szűrés.*

Az egyes molekulák egy tetszőleges bevitt (input) konformációjára alkalmaztam az általunk kifejlesztett konformációs analízis – kiindulási konformer kereső eljárást (**5. ábra**, a.). Az így kapott konformerek molekuladinamikai (MD) szimulációk (**5. ábra**, c.) kiindulási szerkezeteként szolgáltak. A kiindulási konformerektől elvártuk, hogy megfelően alacsony energiájuk legyen, ugyanakkor az egymáshoz viszonyított minimális konformációs

távolságuk egy küszöbérték felett maradjon. A konformerek keresésekor a számítási sebesség növelése céljából nem alkalmaztunk oldószermodellt.

A kiindulási konformerekből indított MD szimulációk hivatottak bejárni az egyes ligand molekulák szabad állapotban (amikor nincs a célfehérje a rendszerben) felvett konformációs terét. A szimulációs hőmérséklet minden példa esetén 300 K volt. Az oldószerhatást ilyenkor igyekeztünk a lehető legpontosabban, explicit (víz) molekulákkal figyelembe venni. Több MD szimuláció futtatása egy molekulára egyfajta párhuzamos „mérésnek" fogható fel (**5. ábra**, b., c.). Egy molekula kiindulási konformereiből indított trajektóriákat a szimuláció végén összefűztem (**5. ábra**, d.).

A bioaktív konformációkereső eljárás az egyes molekulák összefűzött trajektóriáiban keresi meg azt a térfogatot, amelyhez rendelt konformációs sűrűség értékek a molekulák mért kötődési aktivitásával arányosak. Két algoritmust is kidolgoztam a feladat megoldására (8.4.1, 8.4.2). Mindkét algoritmus több ilyen térfogatot eredményezhet, amelyeket a hozzájuk rendelhető konformációk alapján klaszterezek. Eddig a pontig az eljárás automatizált / automatizálható lépéseket tartalmaz. Amennyiben az eredmény több klaszter, a talált klaszterek közül a felhasználó választja ki a kötődő konformációnak megfelelőt.

Az algoritmusok bemutatása előtt, a módszerekhez szükséges bizonyos kiindulási paraméterek értékének meghatározását a 8.1-es és a 8.2-es fejezetben ismertetem.

8.1. A konformációs távolság definíciói

Két konformáció közti eltérést - a konformációs távolságot - *d*-vel, azt a távolság küszöbértéket, amely fölött a két konformációt már különbözőnek tekintem τ-val jelölöm a dolgozatban. Két tetszőleges molekula tetszőleges konformációjának összehasonlítása minden esetben úgy kezdődik, hogy meghatározzuk az ekvivalens atomok mátrixát (n x 2), majd az alábbi fejezetekben leírt számítás eredményeként megkapjuk a két konformáció eltérését jellemző valós számot.

8.1.1. Összeforgatás után, atomi pozíciók alapján számított távolság

Egyes molekulák közötti térbeli eltérést leggyakrabban összeforgatás után állapítják meg. Az összeforgatáskor leggyakrabban a kvaternionokon[86] és mátrix diagonalizáláson alapuló, nagyon gyors eljárást[87, 88] alkalmazzák. Az összeforgatás során a két struktúra ekvivalens atomjai távolságának négyzetösszegét minimalizáljuk. A két szerkezet távolságát

ebből a minimalizált távolságnégyzetekből képzett, átlagos eltérés-négyzetösszegek gyöke (root mean square deviation, a továbbiakban rmsd) adja.

A különféle algoritmusokban használt távolság-küszöbértékek beállítása a dokkolási eljárások esetén általánosan elfogadott 2 Å érték alapján történt[**]. A (43)-as összefüggés a bioaktív konformációnak megfelelő konformációs térfogat nagyságát nem szabja meg, ezért azt úgy definiáltam, hogy bármelyik két, ebbe a térfogatba eső konformáció távolsága legyen nagyobb, mint 2 Å.

8.1.2. A peptidgerinc ϕ és ψ torziószögei alapján számított távolság

A dolgozatban leírt módszereket peptideken fejlesztettük ki, ezért kézenfekvőnek tűnt e molekulák konformációs terét a peptidgerinc ϕ és ψ torziószögeivel definiálni. Két konformáció távolságát ekkor a torzószögekből álló vektor rmsd értéke adja. Meg kell jegyezni azonban, hogy a torziószögek által definiált tér nem lineáris (pl. -178° és +178° különbsége 4° és nem 356°), ezért az rmsd függvényt módosítani kellett (44).

$$d = \sqrt{\frac{\sum_i \Delta^2\left(t_{1,i} - t_{2,i}\right)}{n}} \tag{44}$$

A (44)-es kifejezésben a $t_{1,i}$ és $t_{2,i}$ az első ill. a második konformációt leíró i-dik torziószög, Δ operátor pedig a (45)-ben pszeudokóddal leírt műveletet végzi.

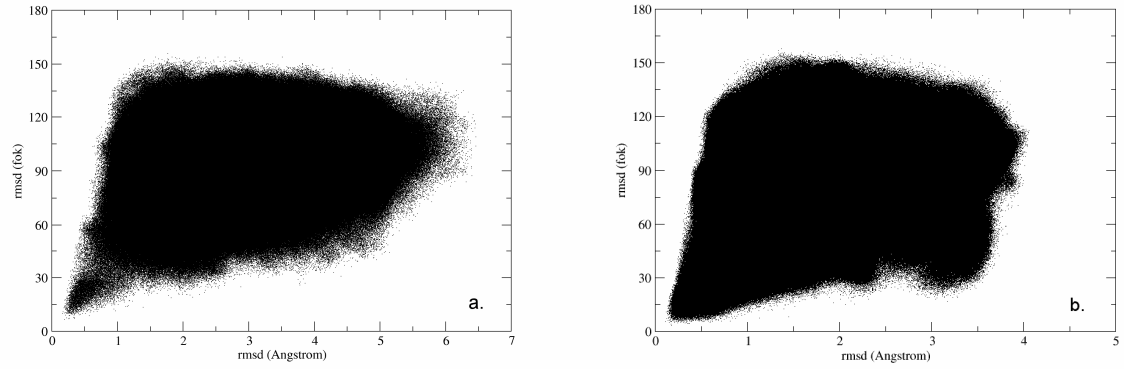
```
Δ(r) {
   Ciklus amíg ( r > 180 ) {
      r=r-360
   }
   Ciklus amíg ( r < -180 ) {
      r=r+360
   }
   Δ(r)=r
}
```
(45)

Ezen számítási mód egyik vitathatatlan előnye a gyorsaság. Egy konformációt kevés valós számmal (gerinc torziószögekkel) lehet jellemezni, így általában minden számítást megelőzött a Descartes trajektória fájlból egy, méretében annál sokkal kisebb, „gerinc torzió” trajektória fájl elkészítése. Sok esetben ez a trajektória már teljes egészében befért a memóriába, nem kellett konformerenként felolvasni a merevlemezről, így nagy sebességnövekedést lehetett elérni.

[**] Dokkolás folyamán egy feltételezett ligandot próbálnak beilleszteni az ismert 3D (röntgen) szerkezettel rendelkező fehérje kötőhelyére. A számításkor kapott, és a rendelkezésre álló komplexet a fehérje alapján összeforgatják. Ha a beillesztett ligand és a kísérletileg meghatározott ligand rmsd értéke 2 Å-nél jobban eltér, a beilleszett ligandot elvetik[89].

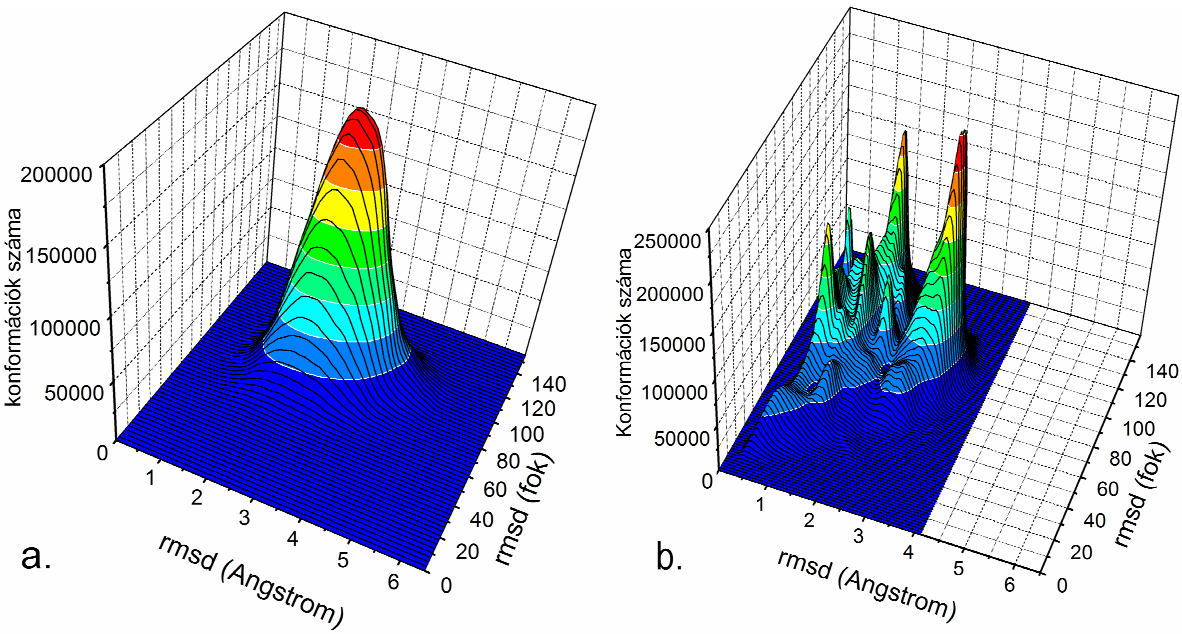
8.1.3. A torzióalapú és koordináta-alapú konformációs távolságmértékek összehasonlítása

A torzióalapú és koordináta-alapú konformációs távolságmértékek összehasonlítását két molekulatár petidjein (9.2. és a 9.3. fejezet) végeztem el. Csak a gerincatomokat használtam fel a koordináta-alapú konformációs távolság kiszámításakor. Az első összehasonlítás a 14 peptid szobahőmérsékleten felvett, explicit oldószeres közegben végzett szimulációjának eredmény-trajektóriáin történt. A 14 peptid esetében az atomi megfeleltetés mátrixát a közös FRHDSGY aminosavak gerinc atomjai alkották. A 14 trajektóriából, amelyek együttes hossza több mint 0,5 μs volt, véletlenszerűen 10 konformációt választottam ki. Az egyes kiválasztott konformációk és a trajektória összes konformációi között kiszámítottam mind a két távolságértéket, és ezeket egymás függvényében ábrázoltam (**6. ábra** a.). Hasonló vizsgálatot végeztem a másik, bioaktív-konformáció keresésnek alávetett, nyolc mucin epitóp pentapeptidet tartalmazó molekulatáron. A pentapeptidek közös TQTXT motívumot tartalmaztak. Az összehasonlítás ebben az esetben is a 300 K-en felvett, explicit oldószeres közegben végzett MD szimulációk (együttes hosszuk 1,9 μs) eredmény-trajektóriáin történt (**6. ábra** b.). A két molekulasorozatra a jelen fejezetben az „FRHDSGY tár” és a „TQTXT tár” kifejezést fogom használni.



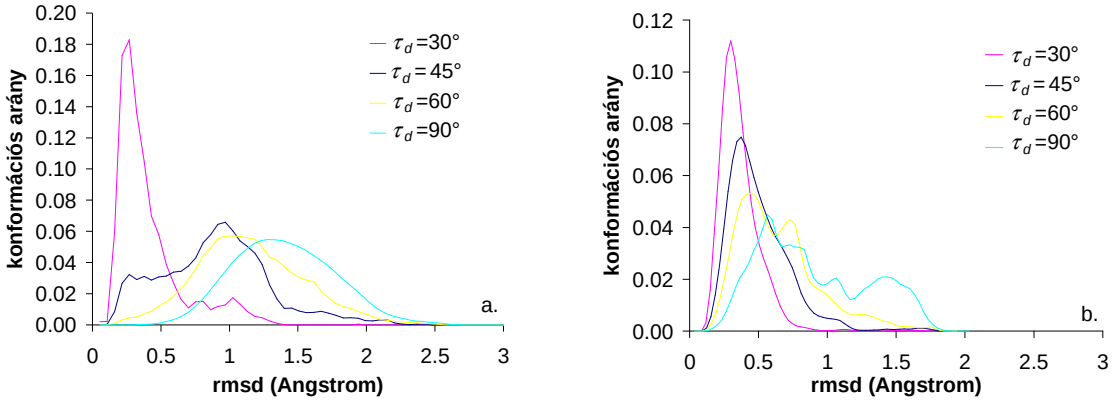
6. ábra *A torziószögeken alapuló konformációs távolságértékek az összeforatáson alapuló távolságértékek függvényében. a. FRGDSGY peptidtáron, b. TQTXT peptidtáron.*

Érdekes megfigyelni, hogy az FRHDSGY tár esetén a tetszőlegesen kiválasztott konformációktól mindkét számítási mód szerint vett konformációs távolságok egyetlen maximumot eredményeztek (**7. ábra** a.) (a távolságok átlaga 2,82 nm-nek 0,74 nm szórással és 96°-nak 15° szórással adódott) viszont a TQTXT tár esetén nem volt ilyen kitüntetett maximum (**7. ábra** b.).



7. ábra *A konformációs távolságok eloszlása. a. FRGDSGY peptidtáron, b. TQTXT peptidtáron*

Azért, hogy megállapítsam, mekkora torziószög küszöbértéket érdemes alkamazni később a módszerekben, metszeteket készítettem a fenti távolságeloszlásból és kigyűjtöttem a d<30°, d<45°, d<60°, és d<90° torziószög rmsd értékekhez tartozó Descartes-rmsd távolságértékek eloszlásait (**8. ábra**). Az egyes görbe alatti területeket normáltam, így közvetlenül összehasonlítható, hogy a választott torzió alapú távolséglimitek mellett az összeforgatáson alapuló, Descartes térben vett konformációs távolság eloszlása miként alakul.



8. ábra *Metszetek a konformációs távolság eloszlásokból. a. FRGDSGY peptidtáron, b. TQTXT peptidtáron*

Az FRHDSGY tár esetén, egy tetszőleges konformációnak 30°-os távolságküszöbön belül tehát a legvalószínűbb összeforgatás után kapott rmsd értéke 0,3 Å körülinek, 45° és 60° esetén 1 Å körülinek adódott. A TQTXT tár esetén a torziószög limitek növelésével a Descartes-rmsd értékek nem változnak ilyen arányban. E jelenségre magyarázatot adhat a két tár molekuláinak eltérő mérete, illetve, hogy az FRHDSGY szekvencia *N*- és a *C*-

„terminálisának” konformációs mozgékonyságát további, hozzákapcsolt más aminosavak csökkentik, míg a TQTXT tár esetén a *N*- és *C*-terminálisok szabadon mozoghatnak.

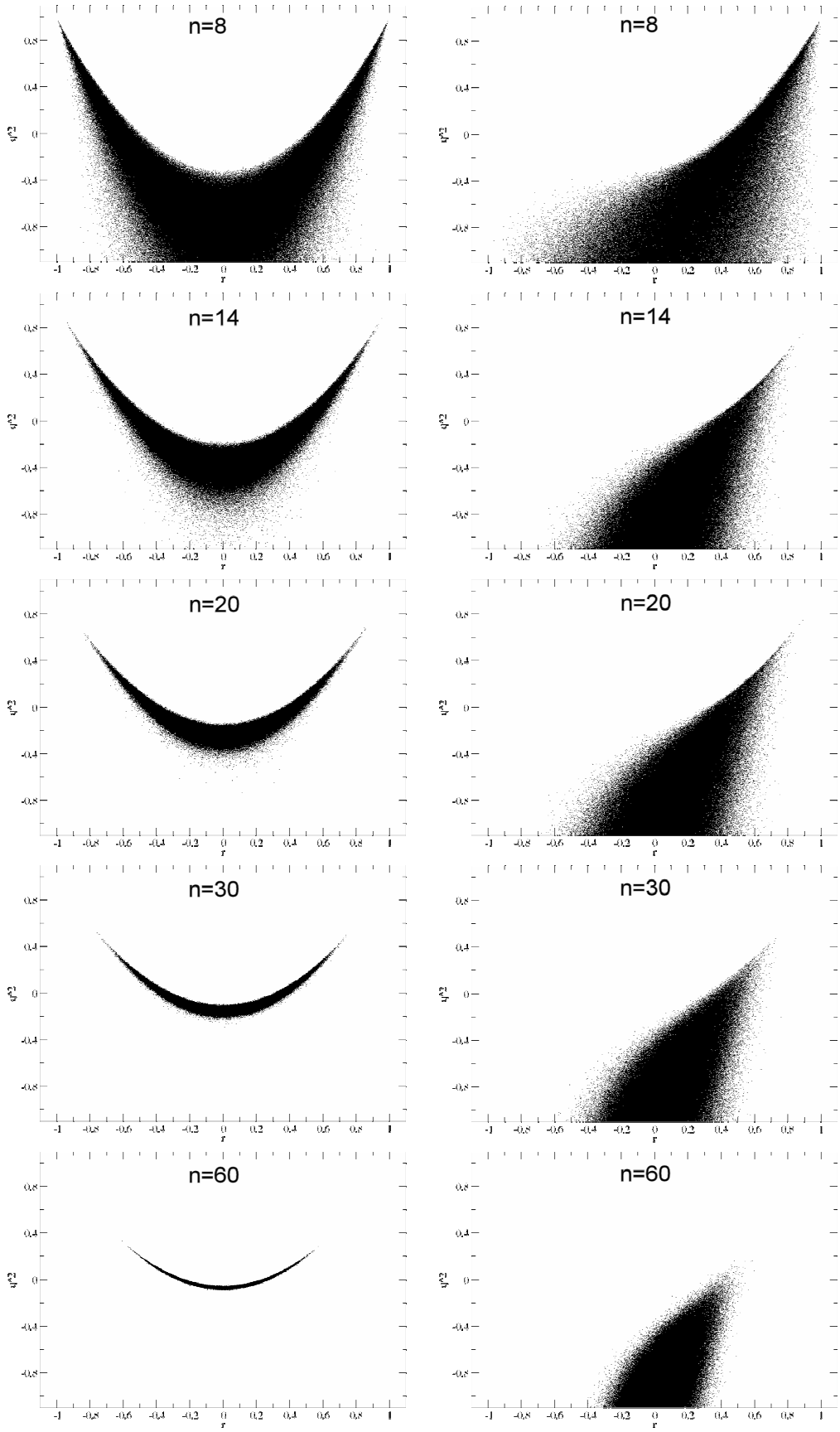
8.2. Az r és a q^2 összehasonlítása

Doktori munkám során olyan algoritmusokat készítettem, amelyek megfelelő modellt készítenek a lokális konformációs sűrűség és a kötődési aktivitás értékei között.

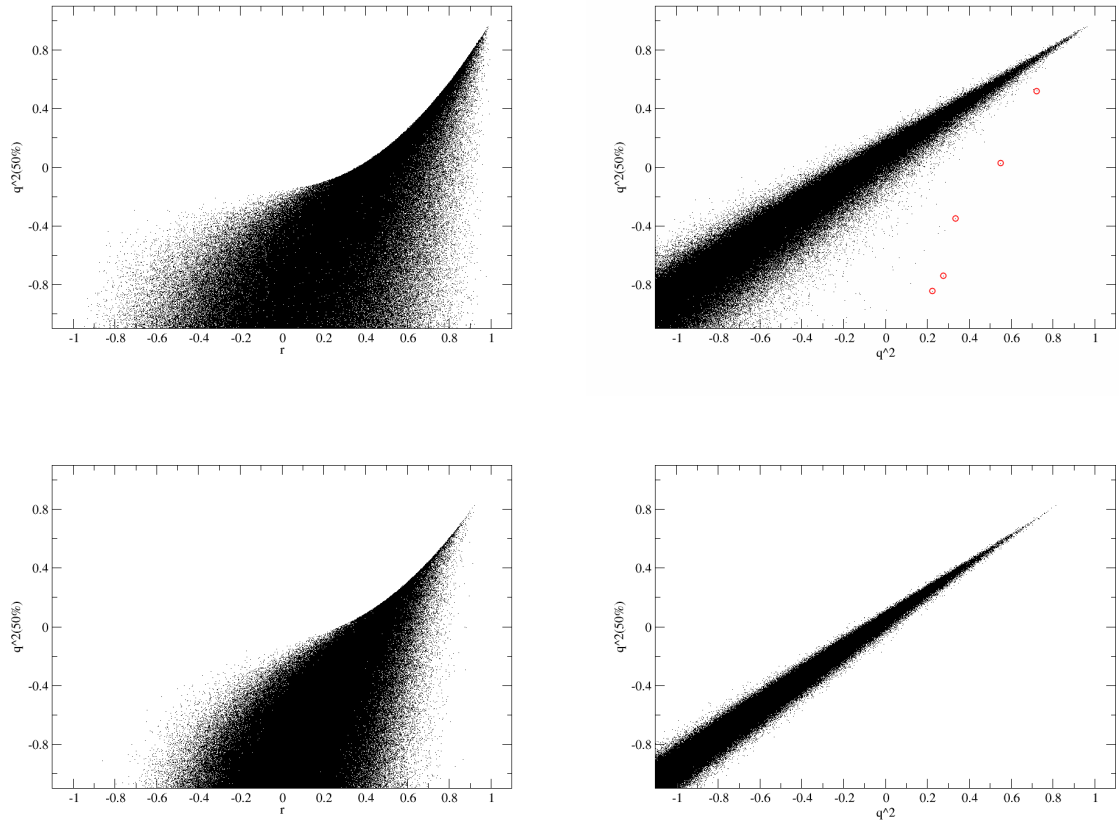
Az eljárások során a felhasználható, korrelációt jellemző függvények kiválasztása céljából elvégeztem azok összehasonlítását általános esetben. Ekkor előállítottam 1000000 darab *x* és *y* vektor párt, mindkét vektor *n* méretű, az egyes elemek 0 és 100 közötti valós, véletlen számok. Minden párra kiszámítottattam az *r*, az LOO q^2 és az L50%O q^2 értékeket. A q^2 számításakor lineáris regressziót és zérus *y* tengelymetszetű egyenesillesztést használtam a modellből kivett értékek becslésére. A TQTXT tár nyolc, az FRHDSGY tár 14 molekulát tartalmaz, ezért az r - q^2 összefüggést n=8, 14, 20, 30, 60 értékek mellett vizsgáltam (**9. ábra**).

A LOO q^2 definíciójából adódóan $r^2 \geq q^2$, emellett kis elemszám esetén (n=8) találhatunk olyan *x*, *y* párokat, amelyek elég magas $|r|$ érték (pl 0.8) mellett sem eredményeznek megfelelő q^2 értéket. Ekkor egy-egy kiugró (outlier) pont javítja föl az *r* értékét, de a q^2 értéke természetesen ettől még nem lesz magasabb. Az elemszám (n) növelése során két hatás befolyásolja a grafikon alakját. Az *x* és *y* random értékekből áll, ezért egyre kisebb a valószínűsége a magas $|r|$ -ű pároknak. A kiugró értékek q^2-re gyakorolt hatását a növekvő elemszám egyre tompítja, így az (r^2-q^2) érték csökken, és a max(q^2(r=0)) is növekszik, vagyis, ha az *x* és *y* véletlen számokból áll, akkor $r(n=\infty)=0$ és $q^2(n=\infty)=0.5$.

A fentiekhez hasonlóan megvizsgáltam az L50%O q^2 függését az *r*-től és az LOO q^2-től. Az L50%O q^2 számításakor a dolgozatban minden esetben figyelembe vettem az összes kombinációt a sorozat felének kihagyásakor. Ez a megközelítés jelentősen növelte a számításigényt, ezért az összefüggést csak az n=8 és n=14 mellett vizsgáltam. A modellt ebben az esetben csak a zérus *y* tengelymetszetű egyenes képezte.



9. ábra *A LOO q^2 r függvényében. Bal oszlopban egyszerű lineáris regressziót, a jobb oszlopban zérus y tengelymetszetű egyenest alkalmazva modellként. A sorokban rendre n=8, 14, 20, 30, 60*



10. ábra *A L50%O q^2 r és LOO q^2 függvényében, zérus y tengelymetszetű egyenest alkalmazva modellként. A sorokban rendre n=8, 14.*

Az *r* – L50%O q^2 és az *r* – LOO q^2 összefüggés hasonló lefutású (**10. ábra**). Az LOO q^2 és a L50%O q^2 között heteroszkedasztikus, nullán átmenő egy meredekségű lineáris összefüggés áll fenn, a jelen példában alkamazott random adatok esetén. Meg kell jegyezni azonban, hogy találhatóak olyan pontok ahol a megfelelően nagy LOO q^2-hez meglepően alacsony L50%O q^2 tartozik (**10. ábra** pirossal jelölve) például LOO q^2≈0.6 és L50%O q^2≈0.1, emellett található olyan *x,y* pár amire LOO q^2<L50%O q^2.

A q^2 függvény elvileg nem korlátos alulról, a jobb szemléltethetőség kedvéért értékeit mégis -1 és +1 között ábrázoltam.

Az eljárások során a modell felállítása Pearson-féle korrelációs együttható (*r*) és a keresztellenőrzés során elkészített r^2 (q^2) értékekek alapján történik. Ennek okai a következőképpen foglalhatók össze. Egyrészről kívánatos lett volna az aktivitás-konformációs sűrűség közti összefüggés modellezésére a determinációs együttható (11)-ben definiált képletét használni, mert ekkor zérus *y* tengelymetszetű modellt lehetne alkalmazni (43). Másrészről viszont ugyancsak a (43)-as összefüggés alapján, csak pozitív korrelációjú kapcsolatot fogadhatunk el. Ugyanakkor, ha csak a determinációs koefficienst (r^2-et) tekintenénk, nem tudnánk megadni az egyenes meredekségét. A korrelációs térfogat keresése

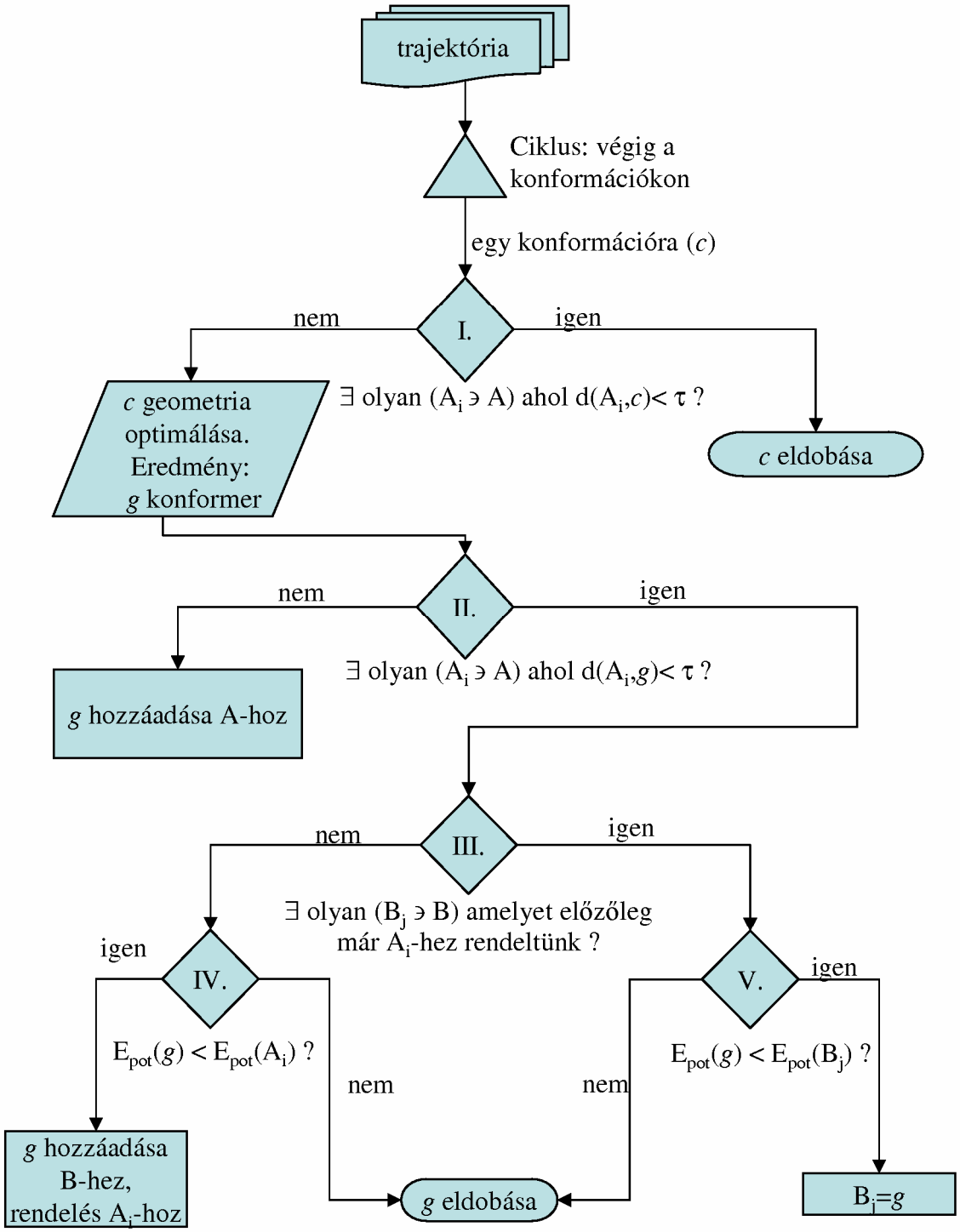
során a módszerekben az r mellett az r^2-hez hasonlóan viselkedő LOO q^2-et és L50%O q^2-et használtam, mert e függvények segítségével mind az egyenes pozitív meredeksége, mind a tengelymetszete figyelembe vehető, emellett a modell robosztusságát is képes megbecsülni.

8.3. A konformerkereső algoritmus bemutatása

Ebben a fejezetben az általunk kifejlesztett konformerkereső eljárás algoritmusát mutatom be. Ez a módszer a konformációs teret olyan MD szimulációkkal járja be, ahol a molekula kinetikus energiáját végig magas értéken tartjuk, és oldószerhatást nem veszünk figyelembe. Ekkor elvárható, hogy a molekula konformációja gyorsan változik, és a magas energiájú konformációs gátakon is át tud jutni. Az algoritmus működése az alábbi pontokba szedve összegezhető:

1. A molekula bevitele egy tetszőleges konformációban
2. Létrehozunk két, kezdetben üres listát („A” és „B”) amelybe a futás során konformereket gyűjtünk. Az „A” lista elemei azok a konformerek, amelyek egymástol vett mininmális távolsága τ. A „B” lista minden egyes eleme egyértelműen hozzárendelhető az „A” egy eleméhez (A_i), és az A_i τ sugarú körében talált legalacsonyabb energiájú konformert tartalmazza.
3. Geometria-optimálást végzünk a kiinduló (input) konformációra, majd futtatunk egy kezdeti molekuladinamikát, t_{init} szimulációs idővel. E lépés a kezdeti konformáció egyensúlyba hozására szolgál.
4. Újabb MD futtatása t szimulációs idővel.
5. E pont alábbi működését szemléltetem folyamatábrán is (**11. ábra**). A 4-es pontban nyert trajektória minden konformációján végigmegyünk. Ha az aktuális konformáció az „A” lista minden elemétől legalább τ távolságra van, végezzünk rajta geometria-optimálást (g, geometria-optimált konformer). Ha g az „A” lista minden elemétől továbbra is legalább τ távolságra van, hozzáadjuk A-hoz. Ha legalább egy elemhez (A_i) közelebb van, mint τ, de a potenciális energiája kisebb, mint az A_i konformeré, és a „B” lista még nem tartalmaz olyan elemet, amelyet ehhez az A_i-hez rendeltünk, a g-t adjuk hozzá a „B” listához. Ha tartalmaz ilyen elemet (B_j), és g energiája alacsonyabb, mint B_j energiája, akkor cseréljük le B_j-t g re. Az **11. ábra** I. elágazás azért szükséges, hogy csak egymástól a legalább megkívánt τ mértékben eltérő konformációkra végezzünk geometria-optimálást, élve azzal a jogos feltételezéssel, hogy egymáshoz hasonló konformációjú magas energiájú

szerkezetek geometria-optimálása ugyanahhoz a konformerhez vezet. A második elágazásban azért kell tovább vizsgálni az „igen” ágat, hogy az A_i τ sugarú környezetében megtaláljuk a lehető legalacsonyabb energiájú konformert.



11. ábra *„A” és a „B” lista konformereinek gyűjtése egy trajektóriából.*

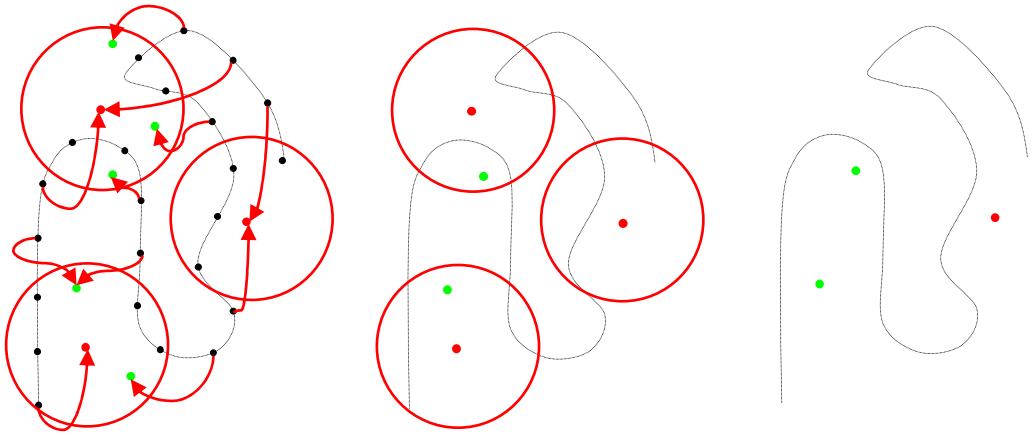
6. Ha az „A” lista elemszáma nőtt a teljes trajektória feldolgozása során, ugorjunk vissza a 4-es ponthoz. Ha nem, kétszerezzük meg a t szimulációs időt, és úgy térjünk vissza a 4-es

ponthoz. Ha egymást követő N darab trajektóriában nem találtunk elraktározható *g* konformert, vagy a dinamika lépésszáma elérte az előre beállított limitet lépjünk tovább.

7. „A” és „B” lista konformereit egy átmeneti közös listába (L) másoljuk, és e lista elemeit rendezzük potenciális energia szerint növekvő sorrendbe.

8. A kiindulási konformerek listáját (S), úgy hozzuk létre, hogy a L lista első elemét S-be másoljuk. Majd L-elemein sorrendben végigmenve, az aktuális L_i-t akkor adjuk az S-hez, ha az S minden elemétől τ legalább távolságra van.

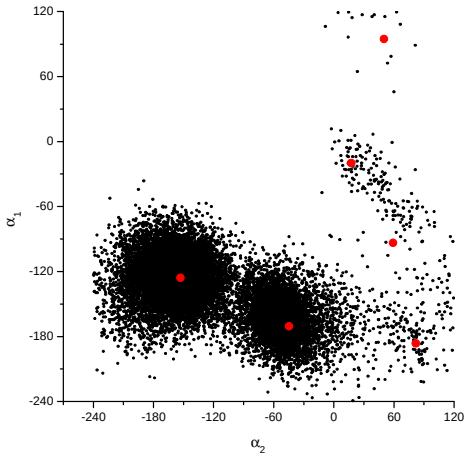
A fenti eljárás olyan S listát eredményez, ahol a lista elemei energia szerint növekvő sorrendbe rendezett, lehető legalacsonyabb energiájú konfomerek, amelyekkel a bejárt konformációs tér úgy lefedhető, hogy bármely két konformer minimális távolsága τ (pl. **13. ábra**).



12. ábra *A kiindulási konformereket kereső algoritmus működésének vázlatos bemutatása 2 dimenzióban: A fekete pontozott vonal a molekuladinamikai szimuláció lefutását mutatja. Feketével a trajektória konformációit, pirossal az A-, zölddel a B lista konformereit jelöltem. A geometria-optimálás során bekövetkező konformációváltozást piros nyilak, az A lista aktuális eleméhez rendelt referenciasugarat (τ) piros kör szimbolizálja. Ezen a sugáron belül maximum egy B lista elem megengedett, amely az aktuális környezetben talált legalacsonyabb energiájú konformer (középső ábra). Az A és B listából, konformereinek energia szerinti növekvő sorbarendezése után a legalacsonyabbtól kezdve a kiindulási konformereket úgy választom ki, hogy egymástól legalább τ távolságra legyenek (jobb oldal).*

Az algoritmus szemléltető leírását a **12. ábra** mutatja. A trajektória első konformerén geometria-optimálás történik, majd végighaladva a trajektórián csak azokon a konformációkon végzünk optimálást, amelyek a referenciasugáron kívül esnek (bal oldal). Ekkor gyűjtjük össze az „A” és a „B” lista konformereit. A két listára a robosztusság miatt van szükség. Ha a konformációs teret, amelyben a keresés történik, nem a molekula összes atomja segítségével definiálom, a „maradék” atomok mozgásával előállhat olyan konformáció, amely az előzőleg találttól a definiált térben csak kismértékben különböznek, az energiája mégis alacsonyabb. Például peptidek esetén, ha a teret a gerinc atomok alapján

definiálom, az oldalláncok konformációjának átrendeződésével alacsonyabb energiák is előállhatnak.



13. ábra *Egy valós molekula két tetszőlegesen kiválasztott belső koordinátája (α_1, α_2 tortziószög) által leírt 2D térben (fekete pontok) szemléltetési céllal végzett kiindulásikonformer (piros) keresés.*

Az algoritmus előnye, hogy lineáris és (makro) ciklusos molekulák esetén is használható, és alkalmas globális energiaminimum keresésre is. A molekulák konformációs flexibilitására (a bejárható konformációs tér nagyságára) jó becslést adhat az S lista konformereinek száma. A módszer az „A” és a „B” lista alkalmazása miatt kevéssé érzékeny a konformációs tér definíciójára. A szimulált megeresztés[32] (simulated annealing) jellegű konformációs kereséssel összehasonítva elmondhatjuk, hogy ebben az esetben a hosszabb lehűtési szakaszokat a gyorsabb energiaminimalizálással helyettesítettük, amely a térszerkezeti változástól függően és nem bizonyos idő elteltével történik.

8.4. Bioaktív-konformáció kereső algorimus

Amennyiben a 7.3-as fejezetben megfogalmazott állítások igazak a vizsgált molekulatárra, a (43)-as egyenlet alapján kidolgozható egy olyan számítási eljárás, amely az egyes, kötődésben nem lévő (szabad) molekulák konformácós eloszlásából az aktív hely ismerete nélkül kiválszthatja a kötődő (bioaktív) konformációt (BC).

A számítási eljárások kidolgozásakor nem alkalmazhattam CoMFA-típusú háromdimenziós rács alapú vizsgálatot, mert a molekulák jelentős flexibilitása miatt lehetetlen *a priori* kiválasztani egyetlen olyan közös bioaktív konformációt, amelyre az összes molekula összes konformerét jó átfedéssel rá lehetett volna forgatni. Így nem egyetlen közös konformáción, hanem egy *a priori* közös konformációs téren – a PCORE téren – alapuló

módszert kellett kidolgozni. PCORE térnek neveztem azt a konformációs teret, amelyet a molekulatár összes tagjának a PCORE-ja bejár. A molekulák konformációs viselkedését ebben az egy közös térben írom le, így két különböző molekula konformációja is közvetlenül összehasonlítható. Ebben az esetben elég „csak” ezt a konformációs teret definiálni, nem kell magát a bioaktív konformációt ismerni. Ekkor a számítás pontosan ezt, a kötődésért felelős konformációt eredményezi.

E tér megállapításakor – vagyis azoknak az atomoknak a kiválasztásakor, amelyek a teret jellemzik – nem alkalmaztam változókiválasztás jellegű módszereket. Ezek az eljárások nem kívánt flexibilitást vihetnek a modellbe azáltal, hogy megkeresik azokat az atomokat, amelyek által leírt térben maximális korrelációt kapunk. Ehelyett a PCORE-t alkotó atomokat kísérleti adatok alapján, a kiinduláskor explicit definiálni kell. Ilyen kísérlet lehet például a minimális epitóp meghatározása, amelyet egyfajta „kémiai úton történő változókiválasztásnak” is tekinthetünk.

Két algoritmust dolgoztam ki, készítettem el és teszteltem. Mindkét eljárás célja az, hogy az összes olyan konformációs régiót megtalálja, ahol az egyes molekulák konformációs sűrűsége megfelelően aránylik a mért kötődési aktivitásokhoz, vagyis a Pearson-féle korrelációs koefficiens legyen pozitív és a keresztvalidált r^2 (q^2) legyen minél nagyobb.

8.4.1. Algoritmus 1

Az eljárás első lépésben horgonykonformációkat (anchor conformation, továbbiakban AC) gyűjt úgy, hogy a horgonykonformációk listájához csak akkor adja hozzá a trajektória aktuális konformációját, ha az, az összes, már elrakott AC-től legalább τ_{AC} távolságra van.

Itt említem meg, hogy a 300 K-en végzett MD szimulációk teljes szimulációs idejét kétféleképpen lehet beállítani az eljárásban; egyszerűen megadjuk az egy kiindulási konformerre vonatkozó szimulációs időt (t_{sum}); vagy a rövidebb, t_s hosszúságú szimulációkat addig ismétli, amíg talál AC listához új elemet.

Miután az összes molekula összes kiindulási konformációjára elkészültek az AC listák, egyesítem azokat az egységes AC listává (uAC lista), a következő módon. A legtöbb AC-t tartalmazó molekula listája lesz a kiindulási uAC lista. Ehhez a többi molekulához tartozó AC-ket akkor adja hozzá, ha minden, már elrakott uAC-től legalább τ_{AC} távolságra vannak. A τ_{AC} küszöböt úgy választottam meg, hogy a bejárt konformációs teret az uAC-k megfelelően nagy sűrűségben lefedjék.

Minden uAC-hoz rendelek egy molekulaszám hosszúságú, valós számokból álló lokális konformációs sűrűség vektort. A *i*-dik horgonykonformációhoz rendelt vektor *m*-dik

molekulához tartozó elemét a (46)-os kifejezéssel definiáltam. A számlálóban az *m* molekula azon konformereinek száma, amelyek *i*-dik horgonykonformációhoz esnek a legközelebb, a nevezőben pedig az *m* molekulához tartozó összes konformáció található.

$$\rho_{uAC(i)}(m) = \frac{n_{m,uAc(i)}}{n_m} \tag{46}$$

Az algoritmus több bioaktív konformációhoz rendelhető térfogatot (*V*) is találhat, úgy, hogy a szomszédos uAC-k unióját képezi, a következő módon:

Válassza egyszer az összes uAC-t a leendő *V* térfogat első elemének. A *V* térfogathoz adja hozzá egyenként az összes szomszédos uAC-t. Egy szomszéd hozzáadásakor képzett unió esetén számítsa ki a kumulatív sűrűségvektort (ρ_V). A ρ_V és a kötődési aktivitás vektor (47) között határozza meg a Pearson *r* értékét. Ezek után törölje ezt, a legutóbb hozzáadott szomszédot *V*-ből.

$$b(m) = \frac{1}{IC_{50}(m)} \tag{47}$$

Ha a ciklus az összes szomszédon végigért, azt az uAC-t adja hozzá véglegesen *V*-hez, amelyik a teszt *V* térfogatok közül a maximális *r* értéket produkálta. Ha ez az *r* nagyobb, mint nulla két másik korrelációs mértéket, a LOO q^2-et és a L50%O q^2-et is határozzon meg. Ha mindkét fajta q^2 nagyobb, mint egy előre beállított küszöbérték (κ_{LOOq}, $\kappa_{L50\%Oq}$), ezt a *V* térfogatot mentse el későbbi klaszterezés céljából. Ilyen formán dolgozza fel az összes szomszédot és növelje *V* értékét addig, amíg az el nem éri az előre beállított konformációs térfogat limit értékét (τ_V). Fontos megjegyezni, hogy olyan V térfogatokat, amelyek csak egyetlen uAC-t tartalmaznak, véletlen korrelációnak tekintem, és nem fogadom el.

Mivel a szakirodalomban az általánosan elfogadott[90] határ a még megkérdőjelezhető és a már éppen elfogatható LOO q^2-re 0,6, a κ_{LOOq}, $\kappa_{L50\%Oq}$ paraméterek értékének 0,6-ot választottam.

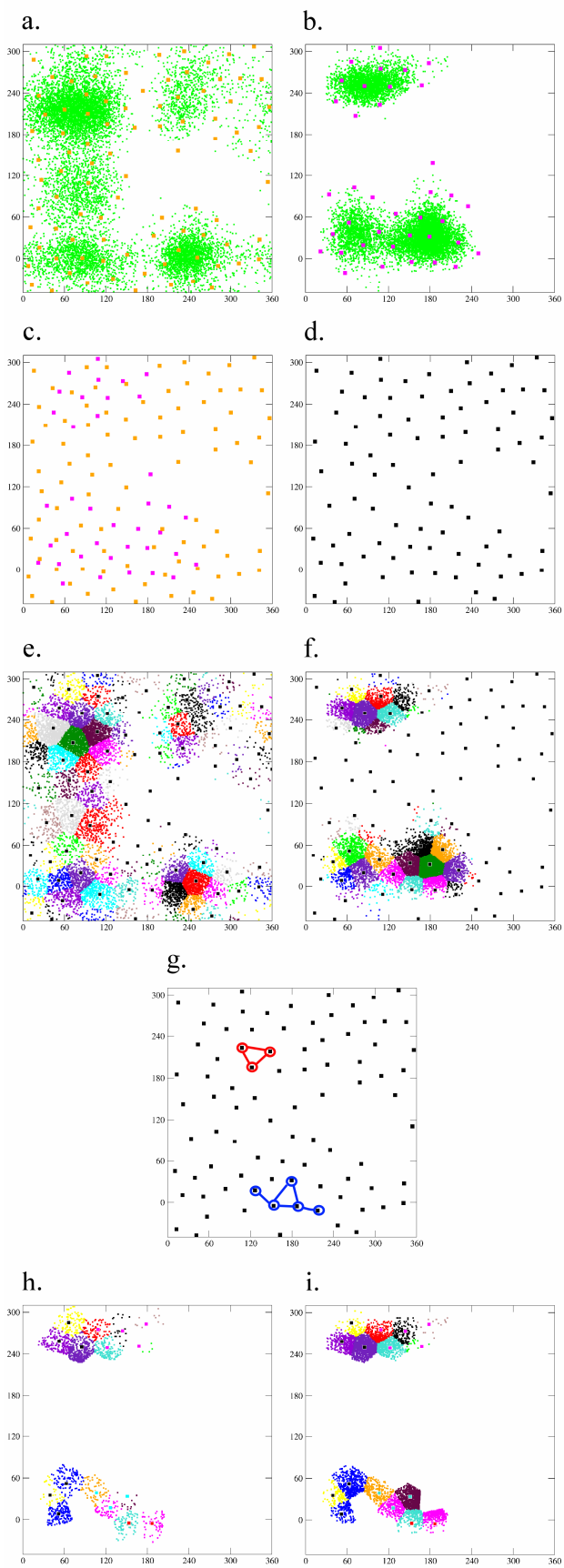
A *V* térfogat nagyságát a girációs sugárral analóg módon számítottam (48), ahol *n* a *V*-t alkotó uAC-k száma, *d(i,j)* pedig az *i*-dik és a *j*-dik, *V*-ben szereplő uAC közötti távolság. A különbség csak annyi, hogy centrumként nem a komponensek átlagos értékét, hanem használtam, hanem azt az uAC-t, amelyik a legkisebb rmsd éréket adta. Az így definiált térfogatnak választottam felső korlátként a 2 Å-nek megfelelő τ_v=30°-ot.

$$v = \min\left(\sqrt{\frac{\sum_{j=1}^{n} d^2(1,j)}{n}}, \sqrt{\frac{\sum_{j=1}^{n} d^2(2,j)}{n}}, \ldots, \sqrt{\frac{\sum_{j=1}^{n} d^2(n,j)}{n}}\right) \tag{48}$$

Az eljárást kettő darab 2D-ben, kettő belső koordinátával rendelkező szimbolikus PCORE-on mutatom be (**14. ábra**). A PCORE-okhoz rendelt kötődési aktivitás értékek arányát 1:3-ra állítottam. Az algoritmus eredménye a jelen esetben két olyan konformációs térfogat, amelyben az egyes molekulák konformerei is 1:3 arányban szerepelnek. A bemutatáshoz használt példában a térfogatot nem korlátoztam.

Az elrakott *V* térfogatok között lehetnek olyanok, amelyek átfednek, vagyis közös uAC-ket tartalmaznak, ezért ezeket *V* térfogatokat Jarvis-Patrick[91, 92] klaszterezésnek vetettem alá. Minden klaszterből azt a konformációs térfogatot jelölöm lehetséges bioaktív konformációnak, amelyiknek a legmagasabb a LOO q^2 értéke. A Jarvis-Patrick klaszterező algoritmust ugyancsak elkészítettem, és a szoftver egy modulját képezi.

Ha az eredmény több bioaktív-konformáció jelölt, akkor a kiválasztás a felhasználó feladata.



14. ábra *Az algoritmus bemutatása 2D konformációs térben. a., b.: a két „molekula” konformációs trajektóriái (zölddel), az egyes molekulákhoz tartozó AC-k (narancssárga és rózsaszín). c., d.: AC-kből (c.) az uAC-k (d.) előállítása (fekete). e., f.: lokális konformációs sűrűségek hozzárendelése a két PCORE esetén. g. térfogat építő algoritmus. h., i. Eredmény térfogatok.*

8.4.2. Algoritmus 2

A 8.4.1. fejezetben bemutatott algoritmus elvileg eredményezhet nem folytonos *V* konformációs térfogatot is, ha például egy nem beválasztott uAC-t körbevesznek a szomszédai, amelyek egy elfogadott *V* térfogat részét képezik. Ekkor a bioaktív konformációért felelős térfogatból a korreláció javítása során egy-egy kis régió kimaradhat, amely nem magyarázható kötődéskor fellépő szerkezeti (konformációs) viszonyokkal.

A probléma megoldására az algoritmust továbbfejlesztettem úgy, hogy csak homogén térfogatokat eredményezzen. E második bioaktív konformációkereső eljárás lépéseit sematikusan két dimenzióban a **15. ábra** mutatom be. Az eljárás első lépésben (**15. ábra** a.) horgonykonformációkat (AC) gyűjt a molekulákra, amelyekből előállítja az egyesített AC-ket (uAC), akárcsak az „Algoritmus 1" esetén. Ekkor azonban a τ_{AC} távolság limitnek nem szükséges olyan kicsinek lennie.

Ezután minden egyes uAC-hoz τ_d sugarú sűrűségi régiót rendel (**15. ábra** a.). A régió nagysága nem változik a keresés során. A régiót definiálhatjuk úgy, hogy csak azok a konformációk járulnak hozzá az uAC-hez rendelt sűrűségéhez, amelyek τ_d -n belül találhatók (cut off) (49), vagy úgy, hogy normális eloszlással súlyozva (50) $d_{uAC(i),m,c}<2\tau_d$ értékig figyelembe veszek minden konformációt, ahol $d_{uAC(i),m,c}$ az *m* molekula *c* konformerének i-dik uAC-tól vett távolsága. Minden uAC-hoz lokális konformációs sűrűség vektort rendelek (51), ahol $n_{traj,m}$ az *m* molekula összes konformációjának száma.

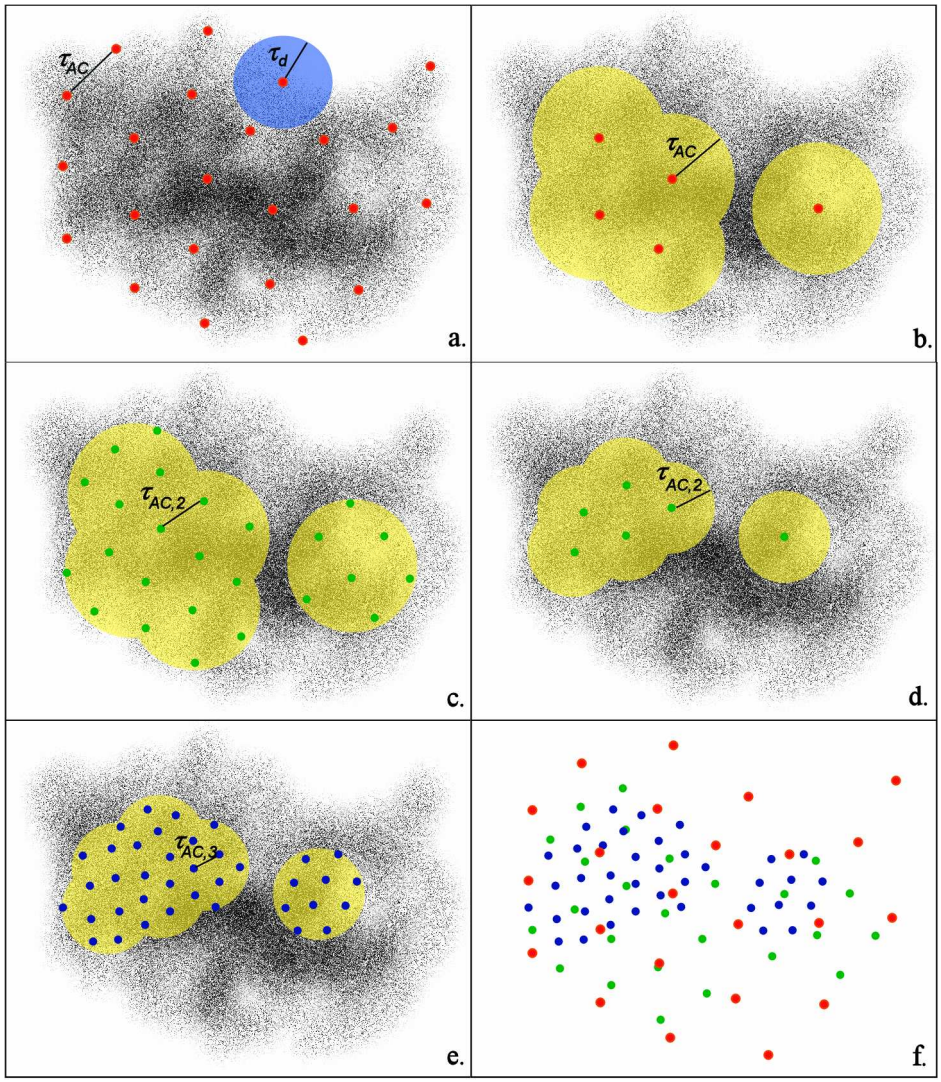
$$f_{uAC(i),m,c}(d_{uAC(i),m,c},\tau_d)=\begin{cases}\text{ha } \left(d_{uAC(i),m,c}\geq\tau_d\right)\Rightarrow f=0\\ \text{ha } \left(d_{uAC(i),m,c}<\tau_d\right)\Rightarrow f=1\end{cases} \tag{49}$$

$$f_{uAC(i),m,c}(d_{a,m,c},\tau_d)=\begin{cases}\text{ha } \left(d_{uAC(i),m,c}\geq 2\tau_d\right)\Rightarrow f=0\\ \text{ha } \left(d_{uAC(i),m,c}<2\tau_d\right)\Rightarrow f=\dfrac{1}{\tau_d\sqrt{2\pi}}e^{-\frac{d_{a,m,c}^2}{2\tau_d{}^2}}\end{cases} \tag{50}$$

$$\rho_{uAC(i)}(m)=\frac{\sum f_{uAC(i),m,c}}{n_{traj,m}} \tag{51}$$

Minden $\rho_{uAC(i)}$-ra és a kötődési aktivitás vektorra (47) kiszámítom a Pearson-féle korrelációs koefficienst, amely ha pozitív, a vonatkozó LOO q^2-et is. Azokat a régiókat fogadom el, ahol a q^2 értéke az összes talált érték a felső x_q százalékába esik. Az x_q előre beállított paraméter (tipikusan 10% vagy 20%). Az elfogadott uAC-k körüli teret (**15. ábra** b. sárgával jelölve) kisebb távolságküszöb alkalmazásával új uAC-kkel térképezem fel, ld. **15. ábra** c., ahol $\tau_{AC,2}=\tau_{AC}*x_\tau$ és x_τ ugyancsak egy előre beállított paraméter. Az így keletkezett

uAC-kre is kiszámítom a lokális konformációs sűrűség vektort, majd a r és a q^2 értékeket, és a távolságküszöböt újra csökkentem. N_{iter} számú iterációt végzek addig, amíg $\tau_{AC,Niter}$ értéke túl kicsivé nem válik (pl. τ kisebb, mint a trajektóriában két egymást követő konformáció átlagos távolsága).



15. ábra *A bioaktív konformerkereső eljárás vázlatos működésének bemutatása több molekula egy szimbolikus, 2D térbe vetített trajektórián. Az eljárás a keresés során a teljes trajektóriát feltérképezi úgy, hogy ha egy régióban, minél magasabb a q^2 értéke annál több horgonykonformációt helyezünk el.*

Így elérhető, hogy az uAC-k a megfelelő régiókban sűrűsödjenek (**15. ábra** f.). Mivel az uAC-k nagyban függenek attól, hogy a konformációk milyen sorrendben következnek a trajektóriában, a fenti, több iterációból álló ciklust, N_{rand}-szor megismétlem úgy, hogy minden ciklus előtt a konformációkat véletlenszerűen átsorrendezem[93]. Ebben a megközelítésben tehát egy megfelelő korrelációs tulajdonsággal bíró uAC megegyezik az 8.4.1-es fejezetben is használt, elfogadott V konformációs térfogattal.

Egy teljes futás eredményeként előálló uAC-ket több lépésben szűröm. Először az uAC-kből kiválasztom azokat, ahol a LOO q^2 és a L50%O q^2 is nagyobb volt, mint egy előre

megszabott limit (κ_{LOOq}, $\kappa_{L50\%Oq}$). A választáshoz opcionálisan hozzátartozhat egy olyan szűrés is, amely az aktuális uAC-t csak akkor fogadja el, ha a nem zérus kötődési aktivitással bíró molekulákhoz rendelt konformációs sűrűség sem zérus (zérus sűrűség kritérium). Majd a kiválasztott uAC-ket a konformációs távolságuk alapján fürtökbe (klaszterekbe) osztom. Kidolgoztam egy olyan egyszerű klaszterező eljárást, amelyik a jelen problémához jobban illeszkedik. Az eljárás eredményeként kapott egy-egy klaszter középpontja (centroid) a klaszter legmagasabb L50%O q^2 értékű eleme lesz, és olyan uAC-k tartoznak bele, amelyek a centroidtól maximum τ_{cl} távolságra vannak. Minden kiválasztott uAC csak egyetlen klaszterbe tartozhat.

A bioaktív konformáció kiválasztása a talált klaszterek centroidjai közül történik két tulajdonság vizsgálatával. Az egyik, a centroid uAC korrelációjának (L50%O q^2) jósága, a másik a kérdéses klaszter mérete. Ugyanis minél több horgonykonformációt tartalmaz egy klaszer, annál biztosabbak lehetünk benne, hogy a korreláció nem a véletlen eredménye.

8.5. IQsar

Az általános QSAR módszerek deszkriptorokat (független változókat) használnak fel arra, hogy meg tudják jósolni, egy új vegyület aktivitását (függő változó). Ezzel ellentétesen működenek az inverz QSAR eljárások, amelyek a molekuláris deszkriptoroknak (függő változó) keresnek olyan megfelelő értéket, amellyel a kívánt aktivitással vagy tulajdonsággal rendelkező (független változókat) molekula rendelhető.

A fent ismertetett módszer az egyes molekulák aktivitása (független változó) alapján a PCORE által leírt (deszkriptor, függő változó) konformációs terében a farmakofornak megfelelő konformációs térfogatokat kutat fel. A kidolgozott eljárás ezért egy farmakofor kereső inverz QSAR eljárásnak tekintető. A módszernek ezért az IQsar nevet adtam[††].

8.6. Az algoritmusok megvalósítása

A kiindulási konformációk előállítását végző algoritmust FORTRAN77 nyelven készítettem el, a molekuladinamikai szimulációkat a GROMACS programcsomaggal végeztem, a geometria-optimálásokat ugyancsak a GROMACS-ban implementált konjugált gradiens és l-BFGS eljárásokkal valósítottam meg. A kiindulási konformációkból a

[††] Az „iqsar” szó máltai nyelven rövidet jelent.

„realisztikus” MD szimulációk automatikus indítását unix shell script használatával és awk nyelven írt eljárások alkalmazásával végeztem. A trajektóriák egyes konformációit MySQL adatbázisban tároltam.

Mindkét bioaktív-konformáció kereső algoritmust JAVA nyelven írtam, kihasználva a nyelv objektum-orientált környezetét. A bioaktív konformációkereső eljárás párhuzamosítható részeit többszálú (multi thread) algoritmusokkal valósítottam meg. Ez a JAVA kód tartotta a közvetlen kapcsolatot a MySQL szerverrel is. A molekula leírására ChemAxon Kft. Marvin[94] csomagjában található molekula leíró osztályt használtam. Ugyancsak MySQL adatbázisba kerültek a horgonykonformációk, és a belőlük felépülő konformációs térfogatot reprezentáló objektumok. A Jarvis-Patrick klaszterező algoritmust is JAVA nyelven írtam.

9. Alkalmazások és eredmények

A 8-ös fejezetben ismertetett algoritmusokat három példára alkalmaztam. Először az oligotuftsin alapú hordozó molekulán található konjugálási helyek átlagos hozzáférhetőségét számoltam ki. Ebben az esetben az algoritmusok közül csak a kiindulási konformergeneráló programot használtam. Majd a teljes eljárást két peptidtáron, a mucin epitóp peptideket és a β-amiloid epitóp peptideket tartalmazó peptidtáron végeztem el.

9.1. Konjugálási hely hozzáférésnek becslése

9.1.1. Oligotuftsin: egy új gyógyszermolekula hordozó

A farmakológiailag aktív vegyületek sikeres terápiás alkalmazásának egyik legfontosabb feltétele a specifitás. A gyakorlatban azonban a legtöbb gyógyszermolekula nem csak a célzott fehérjékkel lép kölcsönhatásba, hanem rendszerint nemkívánatos mellékreakciók is lejátszódnak, amelyek közvetlenül befolyásolják a szer hatásosságát, valamint toxicitási tüneteket eredményeznek.

A specifitás növelésére, a toxicitás és a mellékreakciók kiküszöbölésére, a gyógyszerkutatásban alapvetően két irányzat alakult ki. Az egyik, még specifikusabb tulajdonságú analóg molekulák szintézisével a szerkezet–hatás összefüggések jobb megismerésével kívánja elérni az optimális gyógyszer szelektivitását. A másik elképzelés szerint a ható molekula célba juttatása egy megfelelően választott, célspecifikus hordozó segítségével oldható meg. A hordozóhoz kapcsolással (konjugálás) a hatóanyag kedvezőtlen biológiai (pl. toxicitás, szelektivitás hiánya) hatásain kívül a fiziko-kémiai (pl. oldékonyság), farmakokinetikai (pl. gyors kiürülés a szervezetből, előnytelen szervezetbeli eloszlás) tulajdonságait is előnyösen befolyásolhatjuk[95].

A polituftsin alapú hordozók a tuftsin, egy biológiailag aktív tetrapeptid polikondenzációjával készülnek $(Thr\text{-}Lys\text{-}Pro\text{-}Arg)_n$. Az "n" értéke szintézisúttól függően 40-90 között változhat. A polituftsin hordozók előnye, hogy a tetrapeptid kedvező hatásai a makromolekulában is megjelennek[96, 97]

A humán tuftsin szekvencia (Thr-Lys-Pro-Arg) az immunglobulin-G (IgG) fehérjék Fc régiójában található meg[98]. A tuftsin biologiai hatását szabad tetrapeptid formában fejti ki, miután két enzim, a tuftsin-endokarboxipeptidáz és a leukokináz kihasítja a fehérjeláncból. A

szabad tetrapeptid a vérben is megtalálható neutrofil granulocitákhoz és monocitákhoz specifikusan kötődik, és kötődése során immunmoduláló hatást vált ki. A tuftsin fontos hatása továbbá a ciklikus nukleotidok és a Ca^{2+} koncentrációjának szabályozása a sejtben. A tuftsinnak antibakteriális és antitumor hatása is van, mivel növeli a citotoxikus oxigén (H_2O_2, $O_2^{-\cdot}$, $OH^{\cdot}$) képződését, illetve a tumor nekrózis faktor (TNF/cachechtin) szintjét. A tuftsin fagocitózis stimuláló hatását megfigyelték emberen kívül kutyában, nyúlban, szarvasmarhában és egérben. A tuftsin alapú hordozók előnye lehet, hogy többé-kevésbé megtartják tuftsin-szerű biológiai aktivitásukat, és ezáltal hatékonyabbá tehetik a kapcsolt epitóp elleni ellenanyagválaszt[99, 100].

Különböző hosszúságú oligotuftsin analóg hordozómolekulákat (H-[Thr-Lys-Pro-Lys-Gly]$_n$-NH_2 (n=4,6,8)) állítottak elő[85]. A hordozómolekula alapegysége a kutyákban megtalálható tuftsin, Thr-Lys-Pro-Lys, amely a szerkezet-hatás összefüggések tanulmányozása során ugyanolyan hatékonynak bizonyult, mint a humán változat Thr-Lys-Pro-Arg[98]. A hatóanyag molekulák konjugálása lizin aminosavak ε-aminocsoportjaira peptidkötéssel történhet. A hordozóra különböző funkciójú vegyületek kapcsolása is megvalósítható a Thr-Lys-Pro-Lys peptid esetében, ha a két lizin oldalláncában az ε-aminocsoportokat szelektíven védjük, így a lizin oldalláncok a vegyület oligomerizációja után is szelektíven vihetők reakcióba. A tuftsinhoz képest erősebb immunstimuláló aktivitást tapasztaltak a *C*-terminálison glicinnel meghosszabbított pentapeptid esetében[98], ezért az oligomer ismétlő egységeinek *C*-terminálisára extra glicint építettek be.

Térszerkezeti kérdéseknek a hordozó esetében kiemelt jelentősége van, hiszen fontos, hogy a hordozón lévő konjugálási hely hozzáférhető legyen a konjugálandó farmakon számára, továbbá a szervezetbe jutva a konjugált hatóanyag ugyancsak jól hozzáférhető legyen. Ahhoz, hogy a kívánt konjugátum szerkezetét pontosabban leírhassuk fontos megvizsgálni, hogy a farmakonok milyen arányban, és hol kapcsolhatók hatékonyan a hordozóra.

A tuftsin és analógjainak konformációját kísérletileg (IR-, CD- és NMR-spektroszkópiával) és elméleti számításokkal is vizsgálták, vizes és nemvizes közegben egyaránt. Az IR spektrumban az amid I sáv abszorpciója alapján β-turn konformációra következtettek[101], de NMR mérések alapján ezt nem tudták igazolni[102]. A prolin jelenléte miatt a megfelelő peptidkötés cisz ill. transz izomerjének egyensúlyát figyelték meg[103].

Elméleti számításokat is végeztek az ikerionos végű Thr-Lys-Pro-Arg tetrapeptidre. A Lys és az Arg mindkét esetben protonált oldallácot tartalmazott, így azokhoz Cl^- ellenionokat rendeltek. Az 1 ns hosszú molekuladinamikai szimulációkat vizes közegben különböző NaCl

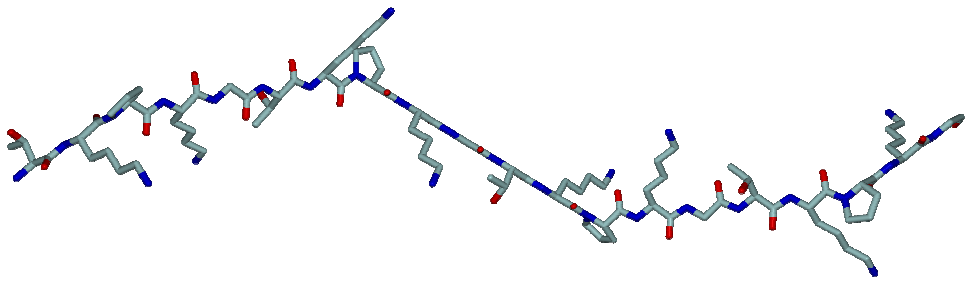
koncentráció mellett 300 K-en végezték. A NaCl koncentrációjának változtatása során nem tapasztaltak változást a tuftsin térszerkezetében. A kiindulási izomer az egyik esetben cisz, más esetben a transz konformer volt. A cisz konformert stabilizálta az ikerionos végek között fellépő elektrosztatikus vonzás, így viszonylag hamar egy IV-es típusú β-turn-szerű szerkezet stabilizálódott. A transz izomer esetén nem tudták megfigyelni az NMR adatok alapján feltételezett γ-turn szerkezetű konformer jelenlétét, amit az oldószer különbözőségével magyaráztak (DMSO és víz)[104].

A Gly-Gln-Pro-Arg tuftsin variáns molekulára konformerkeresést végeztek „hirtelen hűtéses" (quenched) molekuladinamikai szimuláció alkalmazásával, ekkor III típusú β-turn-szerű szerkezet találtak[105].

Az általam végzett számítás[85] célja a legrövidebb előállított peptiden (H-[Thr-Lys-Pro-Lys-Gly]$_4$-NH$_2$) található konjugációs helyek hozzáférhetőségének jellemzése volt. A konjugációra használandó lizin ε-aminocsoportjai kémiailag ekvivalensek, így reakcióképességüket jelentősen befolyásolhatja a hozzáférhetőségük, illetve eltemetettségük.

9.1.2. A kiindulási konformerek előállítása

A H-[Thr-Lys-Pro-Lys-Gly]$_4$-NH$_2$ molekula számos aminocsoportot tartalmaz, amelyek pH=7-en dominánsan protonált formában vannak jelen, ezért ahhoz, hogy a rendszer semleges maradjon ellenionokat kell a molekulához adni. A jelen esetben, a magas energiájú vákuumos szimulációk során az aminocsoportok semleges formáját használtam.



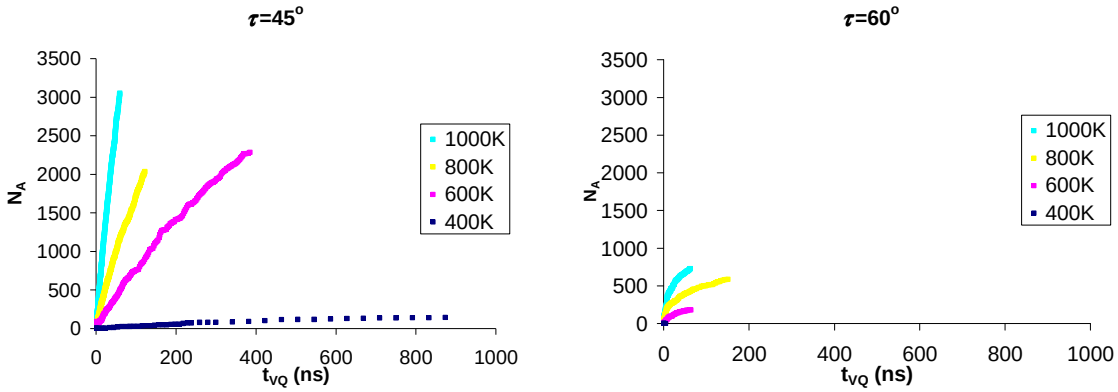
16. ábra *A H-[Thr-Lys-Pro-Lys-Gly]$_4$-NH$_2$ molekula kiindulási konformációja.*

A konformerkereső eljárás egy erőteljesen nyújtott konformációból indult (**16. ábra**). Az 8.3-as fejezetben leírt módszer esetén alkalmazott MD szimulációkat GROMACS[106-108] szoftvercsomag keretén belül végeztem. A felhasznált erőtér, a kifejezetten fehérjék, peptidek vákuumos szimulációjára kifejlesztett GROMOS96 43b1[107] erőtér volt. Az MD szimulációkhoz Berendsen[109] termosztátot (T=400 K, 600 K, 800 K, 1000 K) és barosztátot (p=10^5 Pa) használtam, a lépésközt 1 fs-ra állítottam. A többi, itt nem említett paraméter

értékét a programcsomag által javasolt alapbeállításon hagytam. Az alkalmazott nagy kinetikus energia gyakran átforgatta a peptidkötéseket cisz állásba. Azért, hogy e nem kívánt hatást kiküszöböljem, a megfelelő erőállandót tízszeresére növeltem, ekkor már a jelenség nem mutatkozott. Geometria-optimálásra a szoftvercsomagon belül implementált l-BFGS eljárást alkalmaztam.

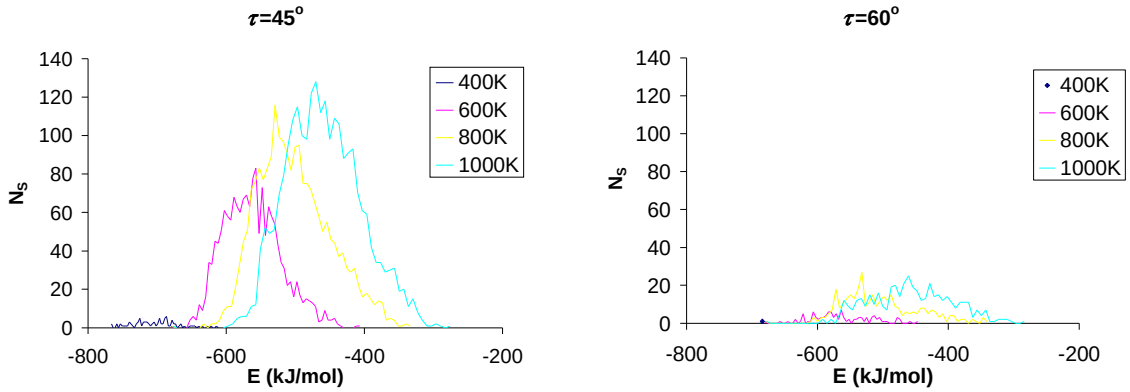
Két konformációs küszöbértéket vizsgáltam, τ=45° és τ=60°. A talált konformerek számát a szimulációs idő függvényében, különböző hőmérsékleten, a **17. ábra** mutatja. A négy hőmérsékleten végzett szimuláció párhuzamosan 4 processzoron futott, amikor az egyik hőmérsékleten a megállási kritériumnak (nem volt új konformer sem az aktuális trajektóriában, sem a megduplázott szimulációs idővel előálló trajektóriában sem) megfelelt mind a négy futást leállítottam.

A τ=45° esetben a 400 K-en végzett keresés érte el a megállási kritériumot. A τ=60° esetben a 400 K-en végzett keresés túl gyorsan befejeződött, a maradék három számításból a 800 K-en végzett érte el a megállási kritériumot, így a maradék kettő szimulációt ekkor leállítottam.



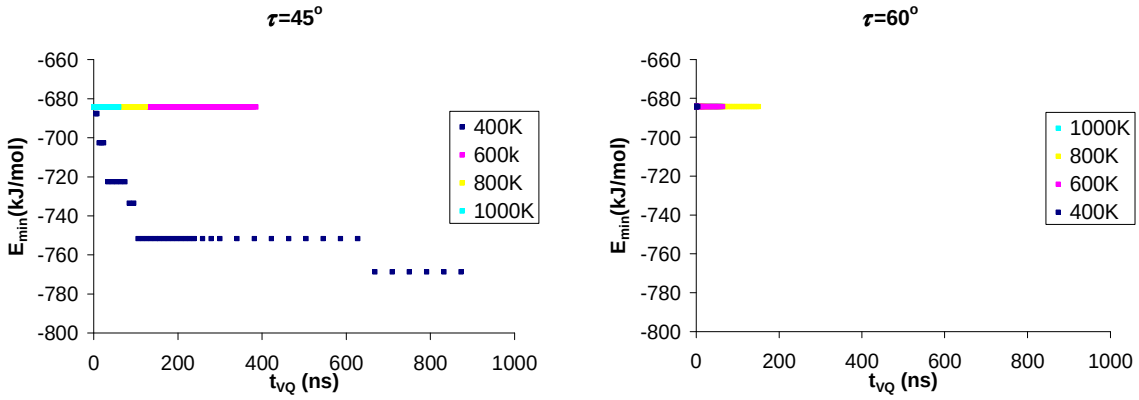
17. ábra *Konformációs keresés: A szimulációs idő függvényében a talált konformerek száma*

Lényeges számítási teljesítményt három feladat kötött le; 1. molekuladinamikai szimuláció, 2. geometria-optimálás 3. RMSD számítás. A τ=45° esetben jól látható, hogy 400 K-en kevesebb konformert talált, így kevesebb optimálás és RMSD számítás kellett, így egységnyi futásidő alatt hosszabb lehetett a szimulációs idő. Mindkét esetben megfigyelhető, hogy a talált konformerek száma jelentősen nő az alkalmazott hőmérséklet (kinetikus energia) függvényében. A keresés végén (leállításakor) talált referencia konformerek eloszlását az energiájuk függvényében ábrázolva megfigyelhető, hogy az egyes görbék maximum helye (azaz a legnagyobb valószínűséggel talált konformer energiája) nem függ a konformációs távolság limittől (τ) (**18. ábra**), ugyanakkor egyre nagyob energiájú konformereket találunk egyre magasabb kinetikus energia értéknél.



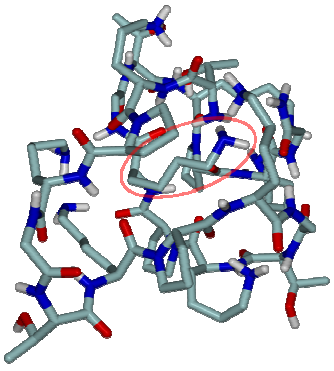
18. ábra *Talált kiindulási konformerek száma a potenciális energia függvényében.*

Ha módszert globális optimálásra akarjuk használni, célszerű alacsony konformációs küszöbértéket, és alacsony hőmérsékletet használni (**19. ábra**).

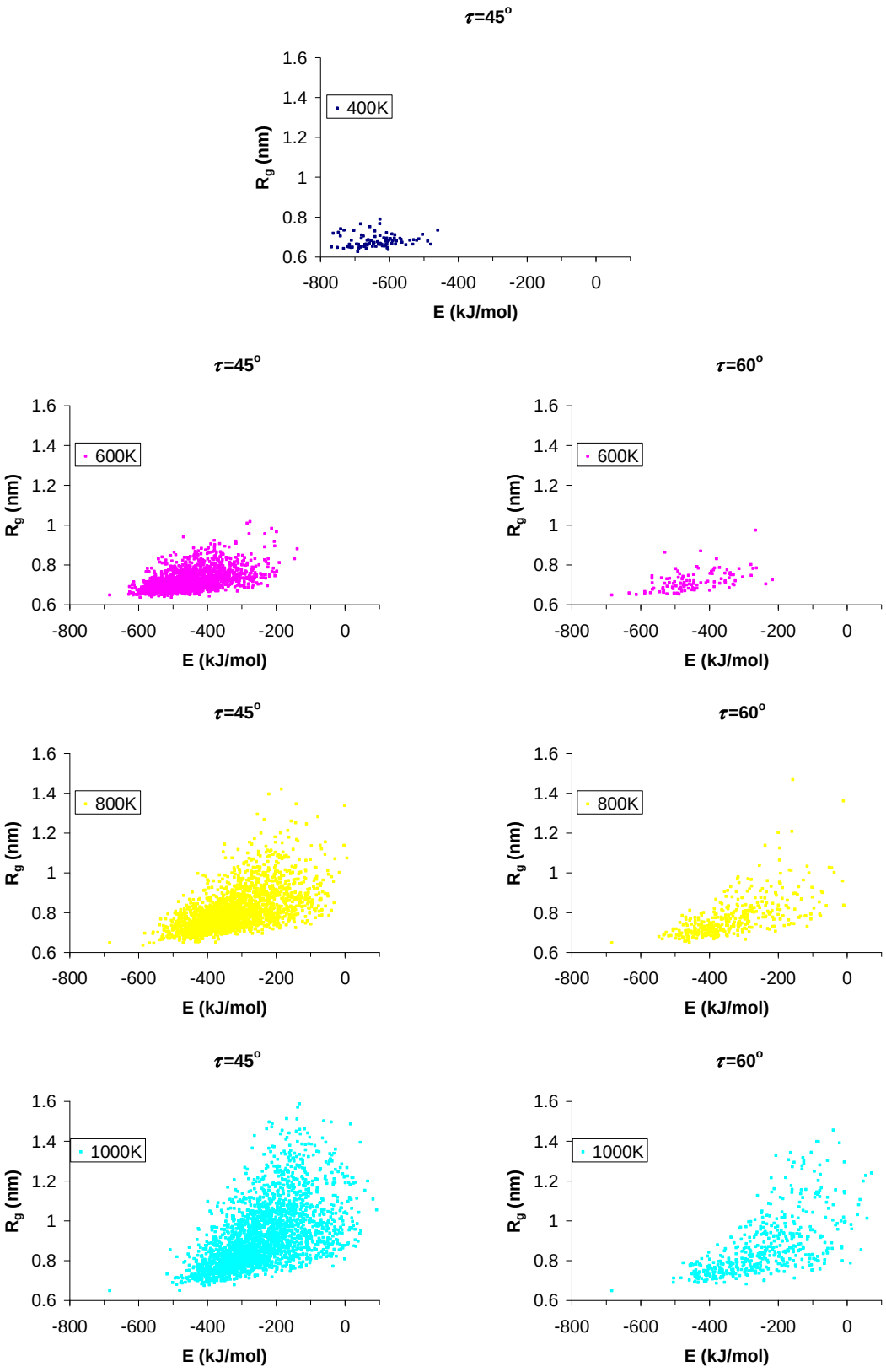


19. ábra *A legalacsonyabb energiájú konformer energiája a szimulációs idő függvényében.*

A talált legalacsonyabb energiájú konformáció esetében megfigyelhető (**20. ábra**), hogy az ábrán jelölt lizin oldallánc a molekula belseje felé néz, teljesen el van temetve.

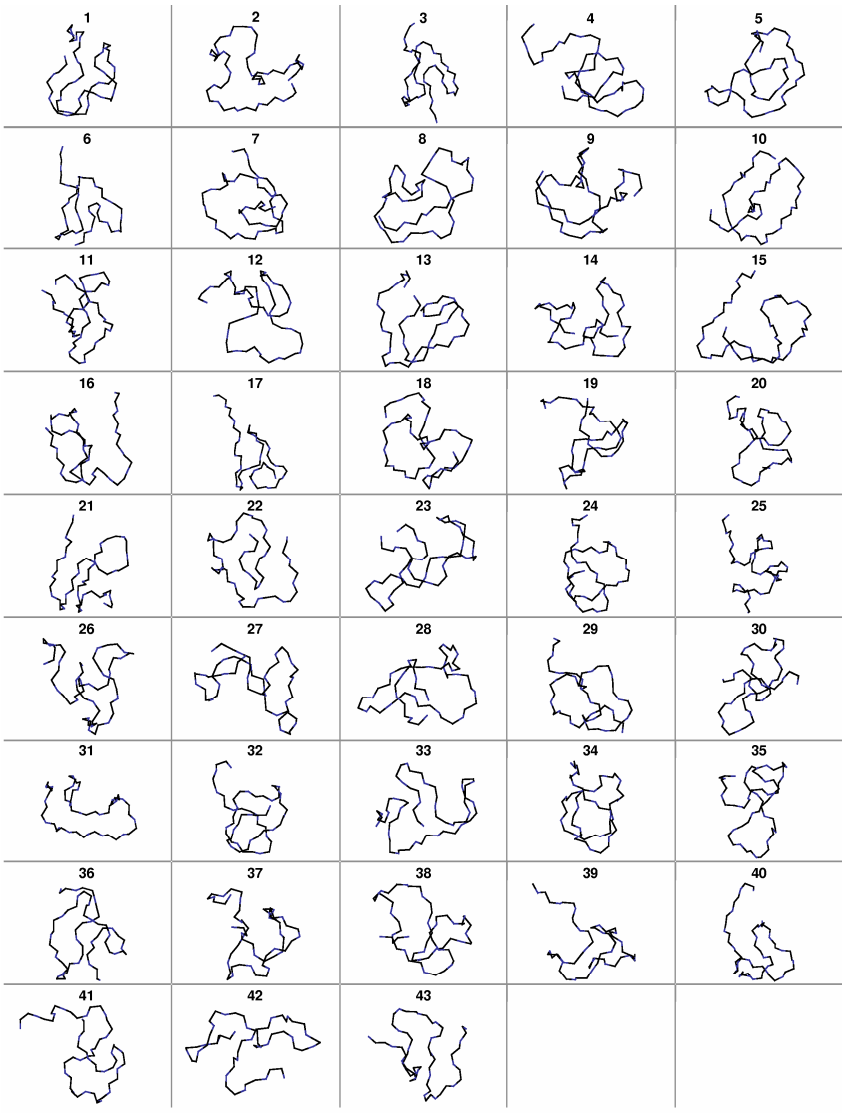


20. ábra *A talált legalacsonyabb energiájú konformáció*



21. ábra *A kiindulási konformerek girációs sugara az energia függvényében.*

Jól megvilágítja a szimuláció során végbemenő konformációs változásokat, ha a kiindulási (referencia) konformerek girációs sugarát ábrázoljuk az energia függvényében (**21. ábra**). Ekkor, mivel semmilyen közeg nem árnyékolja le az intermolekuláris kölcsönhatásokat, az input nyújtott konformáció hamar összegombolyodik. Ebből az összegombolyodott állapotból a különböző növekvő kinetikus energiájú szimulációk egyre inkább újra fel tudják nyitni a molekulát. Érdekes látni, hogy az algoritmus magasabb hőmérsékleten nem talál meg olyan lokális minimumokat, amelyeket alacsonyabban még megtalált. Vagyis 1000 K, 800 K, 600 K sorrendben tekintett szimulációk esetén a pontok egyre inkább „rálapulnak” az x tengelyre, azaz egyre nagyobb arányban vannak jelen az „összegombolyodottabb” konformációk, ezzel párhuzamosan a „foltok” egyre inkább az alacsonyabb energiaértékek felé tolódnak.



22. ábra *A kiválasztott kiindulási konformerek*

9.1.3. Az oldatbeli konformációs viszonyok feltérképezése

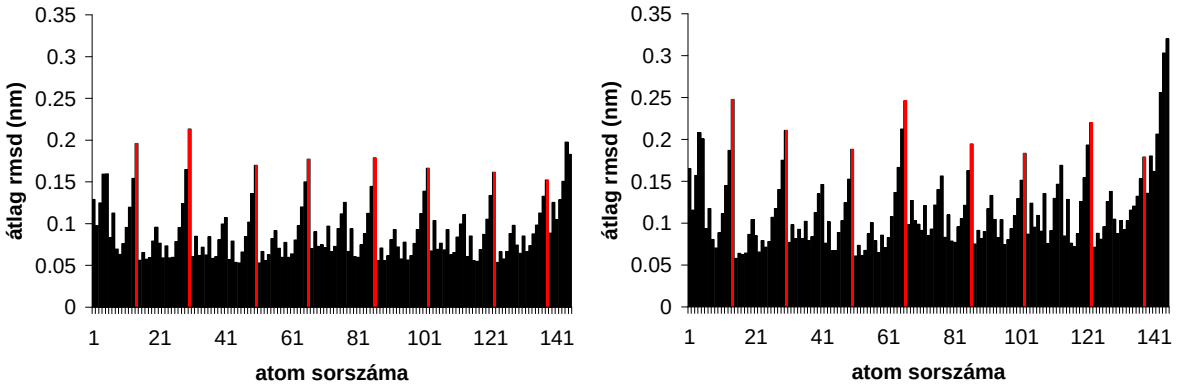
Mivel nem volt olyan számítási kapacitásunk, amely több ezer vagy több száz kiindulási konformerre megfelelő hosszúságú explicit oldószeres MD szimulációt belátható időn belül képes lett volna elvégezni, a τ=60° használatával és T=800 K-en végzett keresés eredményei közül az egész energiatartományt egyenletesen lefedve választottam 43 kiindulási konformert (**22. ábra**).

A kiválasztott 43 kiindulási konformerből molekuladinamikai szimulációkat indítottam; GROMACS erőtér, Berendsen[109] termosztát (T=300 K) és barosztát ($p=10^5$ Pa) használatával, 2 fs lépésközzel, a szimuláció többi paramétere ebben az esetben is a programcsomag által felkínált alapérték volt. Egy konformerből kiinduló dinamika szimulációs idejét 0,34 ns-ra állítottam, így a teljes futás szimulációs ideje 14,62 ns-nak adódott. A szimuláció során SPC vízmodellt alkalmaztam, periodikus határfeltétel mellett (periodic boundary box). Egy kiindulási konformert 1893 vízmolekulával vettem körbe, így minden egyes szimulációban 5863 atom szerepelt.

Érdekesnek tartottam megvizsgálni, hogy a kiindulási konformer vákuumban számolt energiája és az egyes szimulációk során a kiindulási konformertől vett maximális konformációs távolság hogyan viszonyul egymáshoz. Azt vártam, hogy a vákuumban alacsony energiájú (összegombolyodott) konformerek az oldószer hatására kinyílnak, de nem ez történt, nem kaptam értékelhető korrelációt (r=0,2). Ezek után a hét legalacsonyabb energiájú kiindulási konformerre a fenti beállításokkal egyenként 3 ns szimulációt futtattam, a vizsgálatot ekkor is elvégeztem, de ekkor sem találtam megfelelő korrelációt.

Összehasonlítottam még a kiindulási konformerhez viszonyított maximális konformációs távolságot a hét darab 3 ns hosszúságú szimuláció esetén és az ugyanazokból a kiindulási konformerekből indított 0,34 ns hosszúságú szimulációk esetén. A távolság meghatározásakor a konformációkat a gerinc atomok alapján összeforgattam és Descartes koordinátáik rmsd értékét kiszámoltam. A 3 ns hosszú szimulációk esetén a maximális változások átlaga 0,25 nm-nek adódott (0,07 nm szórással), a 0,34 ns esetben 0,21 nm-nek (0,04 nm szórással). Azaz elmondható, hogy a közel tízszeresére növelt szimulációs idő nem változtatja meg jelentősen a peptid oldatbeli konformációs viselkedését.

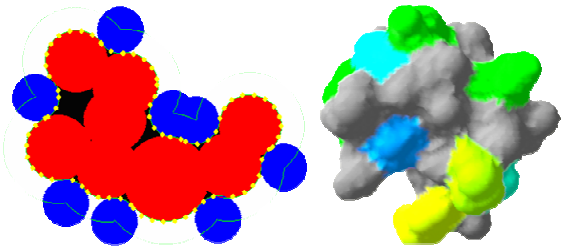
Minden egyes, kiindulási konformerből indított MD trajektóriára kiszámítottam az atomi pozíciók rmsd értékeit, majd ezeket atomonként átlagoltam. A 0,34 ns és a 3 ns hosszúságú szimulációra lizin ε-nitrogén atomok átlag rmsd értéke 0,18 nm-nek ill. 0,21 nm-nek, a molekula minden más atomjára 0,09 nm ill. 0,12 nm adódott. Így megállapítható, hogy a konjugációs hely atomjai nagyjából kétszer mozgékonyabbak, mint a molekula többi része.



23. ábra *Az atomi pozíciók szórásának eredményei az explicit vizes közegű molekuladinamikai szimulációkban; baloldalon a 43 konformerből indított, egyenként 0,34 ns hosszúságú szimulációk esetén, jobboldalon a hét darab 3 ns hosszúságú szimuláció esetén. A lizin ε-nitrogén atomjait pirossal, a molekula többi atomját feketével jelöltem, a hidrogénatomok nem szerepeltek a számításban.*

9.1.4. A molekulafelszín számítása

A 43 trajektória minden konformációjára kiszámítottam az oldószer hozzáférési felszínt Sanner és munkatársai által kidolgozott MSMS[110, 111] program segítségével. A módszer a vízmolekulákat – 1,5 Å sugarú – gömböknek tekinti, e gömböket addig közelíti a molekulához, amíg annak képzeletbeli (van der Waals) felszínéhez hozzá nem érnek. Ekkor kétféle felszínt, az oldószerhozzáférési (solvent accessible surface, SAS, a továbbiakban υ-el jelölöm) és az oldószer kizárási felszínt (solvent excluded surface, SES) lehet definiálni (**24. ábra**):

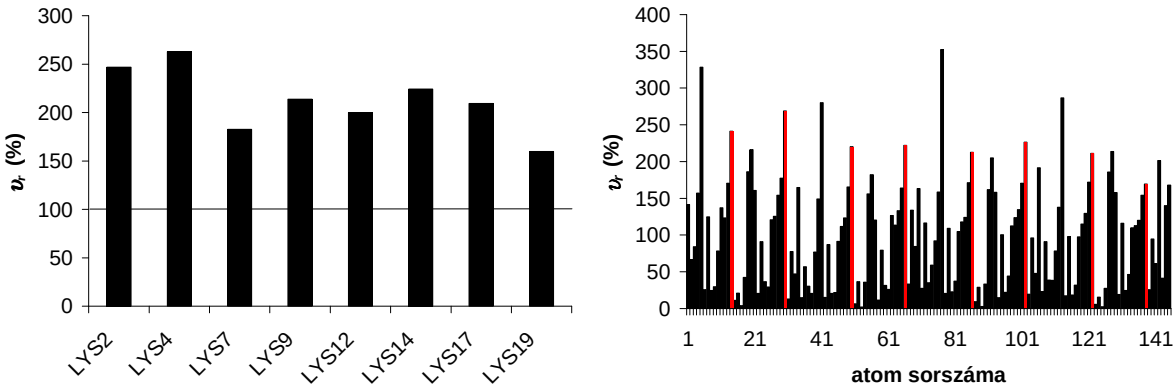


24. ábra *Baloldal: A piros körök az „oldott”, a kékek az „oldószer” molekulákat szimbolizálják.. A sárga pöttyökkel jelölt felszín a SES, a zöld vonallal jelölt a SAS. Jobboldal: az oligotuftsin egyik konformerének oldószer hozzáférési felszíne. A Lys ε-amino nitrogénatomokhoz rendelt felszín kék-sárga árnyalatúra színezve, ahol a kék színű kevésbé, a sárga jobban hozzáférhető.*

Bevezettük a relatív oldószerhozzáférési felszín (υ') fogalmát (52), ahol $\upsilon_{i,k}$ a k konformációban az i-dik atomhoz tartozó hozzáférési felszín, n az atomok, n_{traj} a konformációk száma. A υ_r szemléletes jelentése az, hogy az i-dik atom oldószer hozzáférési felszíne mennyivel nagyobb a molekula összes atomjára számított átlagos felszínnél, százalékban kifejezve (100% az átlagos atomi hozzáférés a molekulában).

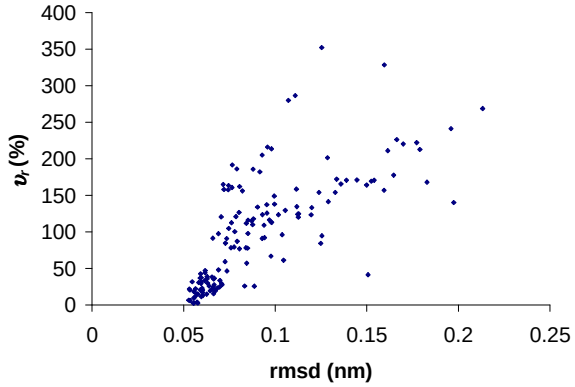
$$\upsilon_i^r = \frac{\sum_k^{n_{traj}} \upsilon_{i,k}}{\sum_k^{n_{traj}} \sum_j^{n} \upsilon_{j,k}} 100n \qquad (52)$$

A 43 darab, összesen 14,6 ns hosszúságú trajektóriából a konjugálási helyekre nyert υ_r értékeket kiszámítottam (**25. ábra**).



25. ábra *A konjugálási helyek υ_r értékei. Baloldalon a publikált[85] ábra, jobboldalon az összes atomra vonatkoztatott relatív felszín értékek is láthatók. A lizin ε-nitrogén atomjait pirossal, a molekula többi atomját feketével jelöltem, a hidrogénatomok nem szerepeltek a számításban.*

Korrelációt (r=0,73) lehetett felfedezni (**26. ábra**) az atomi mozgékonyság (**23. ábra** baloldal) és az atomi hozzáférési felszín között (**25. ábra** jobboldal).



26. ábra *Relatív atomi felszín az atomi mozgékonyság függvényében ábrázolva.*

9.1.5. Kísérleti eredmények

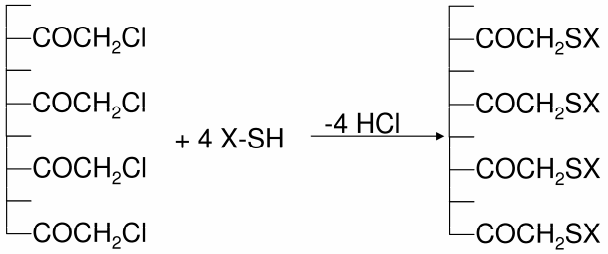
Felvették az általam is számolt oligotuftsin hordozómolekulák CD spektrumait különböző oldószerarányok mellett. Ebből kitűnik, hogy a molekula kitüntetett térszerkezetet nem vesz föl (random coil) sem hatóanyag nélkül, sem a konjugált β-amiloid epitóp peptiddel.

Ugyanakkor T sejt epitópokat tartalmazó konjugátum esetén a TFE koncentrációjának növelésével a CD spektrumban konformációs rendeződésre utaló jeleket találtak[85].

Mivel a hordozómolekulát csak önmagában vizsgáltam, elképzelhető, hogy egy-egy újabb hatóanyag konjugációja folyamán az új molekula térszerkezete, így a maradék konjugációs helyek hozzáférhetősége is megváltozik, ezért érdekes volt az alábbi kísérleti eredményekkel összevetni a számításaim eredményét.

Konjugálási reakciót végeztek[112] egy olyan módosított tetratuftsin hordozón, ahol minden második lizin aminosav ε-aminocsoportjára a ClAc-GFLG enzimlabilis (Chatepsin B bontja) peptid távtartó (spacer) szakaszt kötöttek fel (Ac-[TKPK(ClAc-GFLG)G]$_4$-NH_2).

Ebben az esetben a konjugálandó hatóanyag molekula, az Alzheimer-kór során kialakuló plakkokat alkotó β–amiloid 1-42 peptid epitóp szakasz, az Aβ(4-10) peptid volt. A konjugálást tioéter kötés kialakításával végezték úgy, hogy a hordozó klóracetil csoportjait reagáltatták az epitóp peptid *C*-terminálisához kapcsolt cisztein tiolcsoportjával (H-FRHDSGYC-NH_2), enyhén bázikus körülmények között (pH=8). A reakciót az (53)-as reakcióegyenlet szemlélteti, ahol X a konjugálandó epitópot jelenti.



(53)

A konjugációt HPLC kromatográfiával követték (D. Függelék, **43. ábra**), a kromatogramm elemzése során megállapítható, hogy a reakciók az egyes konjugálási helyeken párhuzamosan játszódnak le, megerősítve ezzel a számítások során kapott eredményt.

9.1.6. Az eredmények összefoglalása

A vizsgált molekulára a konformációs keresés során négy hőmérsékleten és két konformációs távolság-küszöbérték mellett összesen 1,7 μs szimulációt futtattam. A keresés során egyre magasabb kinetikus energia használatakor a talált konformerek egyre magasabb potenciális energia értéket mutattak, és egyre jellemzőbb volt a (nagyobb girációs sugárral rendelkező) nyújtott szerkezet. A választott 43 + 7 kiindulási konformerre összesen (14,62+21) 35,65 ns explicit oldószeres (közel 6000 atomon végzett) molekuladinamikai szimulációt futtattam. A 0,34 ns hosszúságú, oldószeres MD szimulációk során kapott

konformációk a kiindulási konformerhez viszonyítva 0,05-0.3 nm távolságtartományban mozogtak. Drasztikus változás a molekula globális alakjában az oldószer hatására általában nem következett be, bár a peptidgerinc bizonyos szakaszainak konformációja egyes esetekben jelentősen megváltozott.

Kiszámítottam, hogy a konjugálási helyek mozgékonysága a molekula többi atomjának körülbelül kétszerese. Számításaink alapján elmondható, hogy a konjugációs helyek egyformán, körülbelül ugyancsak kétszer jobban hozzáférhetőek, mint a molekula többi atomja. Az eredmények alapján elmondható, hogy a molekulák ugyan nem vesznek fel kitüntetett térszerkezetet, de a konjugálás szempontjából fontos aminocsoportok nagy valószínűséggel tartózkodnak a molekula oldószer hozzáférhetőségi felszínén.

Általános esetben a kidolgozott eljárás alkalmas flexibilis szerves molekulák, oligomerek oldatbeli konformer családjainak felderítésére, valamint a molekulák hozzáférhetőségének jellemzésére reakciók tervezése, vizsgálata céljából.

9.2. Mucin epitóp peptidek bioaktív konformációjának meghatározása

A mucinok olyan nagy molekulatömegű glikoproteinek, amelyek túltermelődhetnek vagy alulglikozilálódhatnak emésztőrendszeri rák esetén[113]. Ha a glikoziláció rendellenes, a molekula azon a részein, ahol a szénhidrát burok hiányzik a belső fehérje hozzáférhetővé válik. A mucinok e tulajdonsága kihasználható egy jövőbeli tumormarker elkészítése céljából. Számos, mucin glikoproteinre specifikus monoklonális antitestet készítettek[114], többek között az mAb 994-el jelöltet. A kifejeztetéshez a hibás glikoprotein hozzáférhető fehérjeláncában megtalálható VTPTPTPTGTQTPT peptidet használták. Szisztematikus kötődési vizsgálatokkal bebizonyították, hogy az mAb 994 olyan pentapeptideket ismer fel minimális epitópként, amelyek közös TX^1TX^2T motívumot tartalmaznak[115]. Ezután olyan kombinatórikus peptidtárat terveztek, ahol az X^1 és X^2 helyen minden természetes aminosavat beépítettek (kivéve cisztein)[116]. A leghatékonyabban kötődő peptideket a $TQTX^2T$ könyvtárban találták. Ennek a tárnak a molekuláit direkt szintézissel is előállították és kvantitatíve meghatározták a kötődési aktivitás értékeit. A natív fehérjében is megtalálható TQTPT szekvencia mellett még öt másik pentapeptid adódott elegendően aktívnak; a kötődési erősség sorrendjében: TQTPT > TQTAT > TQTYT > TQTWT > TQTFT > TQTST. Érdekes megjegyezni, hogy míg a TQTAT jelentős kötődési aktivitás erősséggel bírt, a TQTGT inaktívnak bizonyult. Emellett a TQTPT volt a legerősebben kötődő pentapeptid, ugyanakkor a *C*-terminálison amid tartalmú analógja (TQTPT-NH_2) ugyancsak inaktív volt. Feltételezték, hogy a peptidek ugyanahhoz a kötőhelyhez, és ugyanolyan módon kötődnek, illetve a kötődési erősségek közötti különbség főleg konformációs hatásoknak köszönhető. Ekkor merült fel az a kérdés, hogy milyen konformáció felelős a kötődésért.

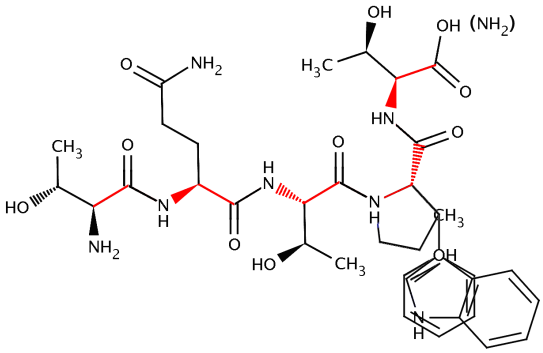
Célunk az volt, hogy megtaláljuk ezt, a kötődő konformációt[80, 81]. A molekulasorozat meglehetősen nagy kihívást jelent QSAR modellépítés szempontjából, a mért aktivitással rendelkező peptidek alacsony száma és nagy konformációs flexibilitása miatt. E feladat elvégzéséhez volt szükség a 7-es fejezetben leírt levezetésekre, amely rámutatott arra, hogy milyen összefüggés segítségével kereshetjük a kötődő, bioaktív konformereket, továbbá ehhez milyen feltételeknek kell teljesülniük.

A vizsgálandó molekulasorozatba a kötődési aktivitással rendelkező molekulák mellé még kettő inaktív pentapeptidet is beválogattunk úgy, hogy az aktív társaiktól szerkezetileg kevéssé különbözzenek, hogy inaktivitásukat egyértelműen konformációs viszonyokkal magyarázhassuk (**3. táblázat**).

	IC_{50} (μM)
TQTPT	3
TQTAT	11
TQTYT	40
TQTWT	70
TQTFT	88
TQTST	208
TQTPT-NH2	inaktív
TQTGT	inaktív

***3. táblázat**: A választott peptidek kísérletileg meghatározott IC_{50} értékei.*

A pentapeptidek konformációs terét a peptidgerinc ϕ-φ torziószögeivel jellemeztük (**27. ábra**).



***27. ábra** A TQTXT peptidtár molekulái két dimenzióban egymásra vetítve, a peptidgerinc ϕ-φ torziószögeinek megfelelő kötések pirossal ábrázolva.*

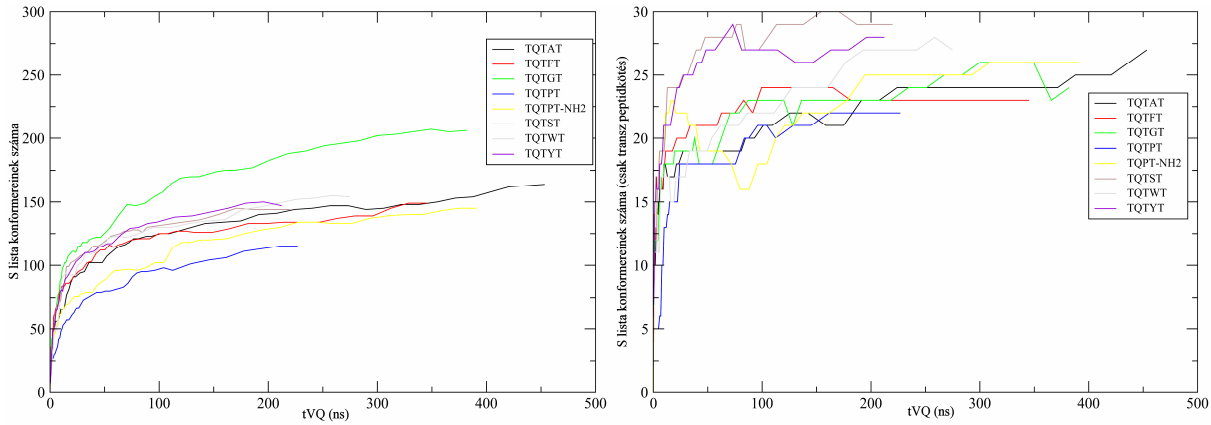
9.2.1. A kiindulási konformerek előállítása

A molekuladinamikai szimulációk ebben az esetben is a GROMACS 3.2.1 programcsomagban implementált gmx erőtérrel történtek. Ha nem a paraméterek alapbeállítását használtam, akkor azt a helyszínen megemlítem.

A kiindulási konformer előállításához alkalmazott magas kinetikus energiájú MD szimulációk esetén a molekula transzlációs és rotációs sebességkomponenseit dinamikai lépésenként kiküszöböltem. Nem alkalmaztam semmilyen oldószermodellt, vagy ellenionokat. Egy tetszőleges input konformációból 20 ps szimuláció után kezdtem el gyűjteni az „A” és a „B” konformerlista elemeit (8.3-as fejezet). A kezdő szimuláció hosszát, amelyből már konformációs mintavételezés történt, 4 ps-ra állítottam. Ez a szimulációs hossz duplázódott, ha már nem lehetett új konformereket felvenni az „A” listába. A konformációs távolságlimitet (τ) 60^{o}-ra állítottam.

A kinetikus energiát három különböző hőmérsékletnek megfelelően állítottam be; 600 K-re, 900 K-re és 1200 K-re. Az 1200 K-en futtatott szimulációk instabilnak bizonyultak, a 600 K-en futtatott szimulációk túl hamar elérték a konvergencia kritériumot, vagyis ebben az esetben viszonylag hamar megtörtént, hogy az algoritmus két egymás utáni szimulációban sem talált új elemet az „A” listába. A 900 K-en végzett szimulációk stabilnak bizonyultak és megfelelően intenzív konformációs mozgásokat eredményeztek. Az egyes molekulák legalacsonyabb energiájú, talált konformerei általában már a keresések első stádiumában fellelhetőek voltak.

Az S lista konformereit minden rész-szimuláció végén kiszámítottam, nem csak a teljes eljárás végén. Az S lista konformereinek száma a szimulációs idő függvényében (**28. ábra**) nem mutat monoton növekedést. Ez azzal magyarázható, hogy az aktuális S listát minden lépésben újra generálom a legutóbbi „A” és „B” lista konformereiből. A magas szimulációs hőmérséklet gyakran átbillentette a transz peptidkötéseket ciszbe és vissza. Csak azokat a konformereket választottam ki a további „realisztikus” szimulációk kiindulásaként, amelyek csak transz peptidkötést tartalmaztak (**28. ábra**, jobboldal).



***28. ábra**: Baloldalon: a talált konformerek száma a szimulációs idő függvényében. Jobboldalon csak azok a talált konformerek, amelyeknél csak transz peptidkötés található.*

9.2.2. Az oldatbeli konformációs viszonyok feltérképezése

A kiválasztott kiindulási konformereket vízmolekulákkal (SPC modell) vettem körbe. A vízmolekulák rendszerét 300 K-en végzett MD szimulációval hoztam egyensúlyba úgy, hogy a peptidet fixen tartottam. Ezután már a mozgását is szabadon engedtem, és ekkor is 300 K-nek megfelelő kinetikus energiát használtam, két hőmérsékleti csatolási csoporton keresztül (peptid, oldószer), Berendsen termosztát segítségével. A szimulációt 0,6 ns hosszúságúra állítottam, hogy a peptid konformációja alkalmazkodni tudjon az új környezethez. Ezt az eljárást minden molekula minden kiindulási konformerére elvégeztem.

Az így kapott oldószer-peptid rendszerből indulva térképeztem fel a „realisztikus” környezet konformációs viszonyait. A feltérképezés úgy történt, hogy viszonylag kicsi (τ_{AC}=30°) távolságküszöb alkalmazása mellett 1 ns hosszúságú dinamikai szimulációkat végeztem addig, míg új horgonykonformációt már nem találtam (ld.: 8.4.1 Algoritmus 1, második bekezdés). Ez a fajta számítási mód a nyolc pentapeptid összes kiindulási konformációjára összesen 1,9 μs teljes szimulációs időt eredményezett. E szimulációk során általánosságban elmondható, hogy a viszonylag kompakt szerkezetű kiindulási konformerek az oldószer hatására fellazultak, illetve bizonyos esetben felnyíltak. A molekuladinamikai szimulációk jellemzőit a **4. táblázat** mutatja.

	N_s	N_{wat}	t_{sol} (ns)
TQTPT	10	867	157
TQTAT	10	804	269
TQTYT	13	875	175
TQTWT	11	920	184
TQTFT	12	922	190
TQTST	15	843	269
TQTGT	11	787	315
TQTPT-NH_2	14	871	344

***4. táblázat**:Molekuladinamikai szimulációk jellemzői. ns a kiválasztott kiindulási konformerek száma, N_{wat} a rendszerhez adott oldószer-molekulák száma, t_{sol} szimulációs idő.*

9.2.3. Bioaktív-konformáció keresés

9.2.3.1. Algoritmus 1

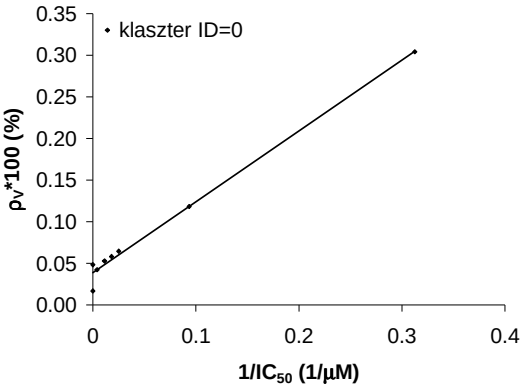
A szimulációs eredményekre először a 8.4.1-es fejezetben leírt módszert alkalmaztam, a már ott is említett τ_v=30°, κ_{LOOq}=0.6 és $\kappa_{L50\%Oq}$=0.6 értékekkel. A zérus *y* tengelymetszetű modellt alkalmazó korreláció jobban illik a 7. fejezetben kifejtett elméleti megközelítéshez ugyanis, ha egy molekula nem kötődik, a bioaktív konformáció relatív sűrűségének nulla értéket kellene felvennie. Ugyanakkor kipróbáltam, ha egyszerű lineáris regressziót alkalmazok modellként, a modell plusz flexibilitása (szabad *y* tengelymetszet) hogyan befolyásolja a kapott eredményt.

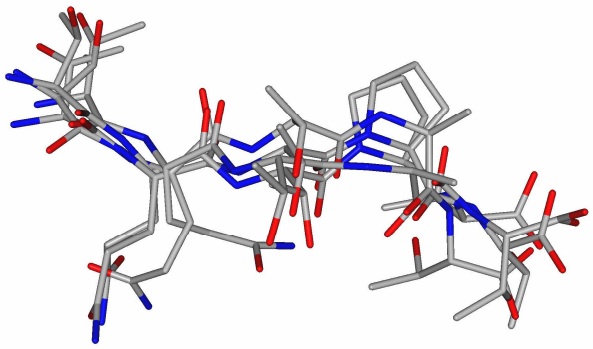
A horgonykonformációk egyesítése után 5937 uAC adódott (ld.: **14. ábra,** c., d.) ebből 90 darab konformációs térfogatot (*V*) találtam, ezek 8 klasztert alkottak. A modell ekkor az egyszerű lineáris regresszió volt. A klaszterezéshez a Jarvis-Patrick-féle eljárást alkalmaztam. Két *V* térfogat akkor szomszédja egymásnak, ha az őket alkotó uAC-k legalább 10%-ban megegyeznek, és akkor osztom őket egy klaszterbe, ha egymás szomszédjai, és legalább 10%

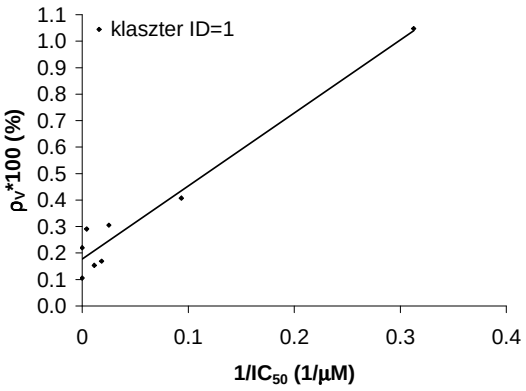
közös elemük van. Minden klaszterből azt a térfogatot fogadtam el, amelynek a legmagasabb a LOO q^2 értéke (**5. táblázat**). Egy elemű (singleton) klasztert ebben az esetben nem találtam.

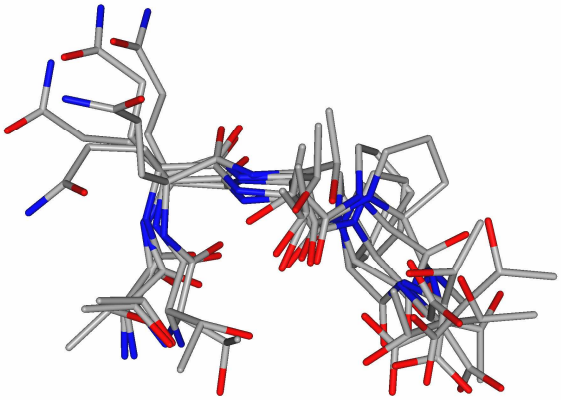
Klaszter		Klaszter legjobb tagja						
		korreláció			regresszió y= g +ax		V mérete	
ID	N_V	r	LOO q^2	L50%O q^2	g	a	rmsd (fok)	N_{uAC}
0	52	0,99	0,98	0,84	0,039	0,852	30	5
1	3	0,98	0,92	0,61	0,177	2,761	30	5
2	8	0,99	0,97	0,86	0,021	1,498	27	4
3	6	0,99	0,94	0,76	0,013	0,168	29	4
6	8	1,00	0,95	0,85	0,003	0,228	25	3
7	11	0,98	0,95	0,72	0,063	1,480	30	6

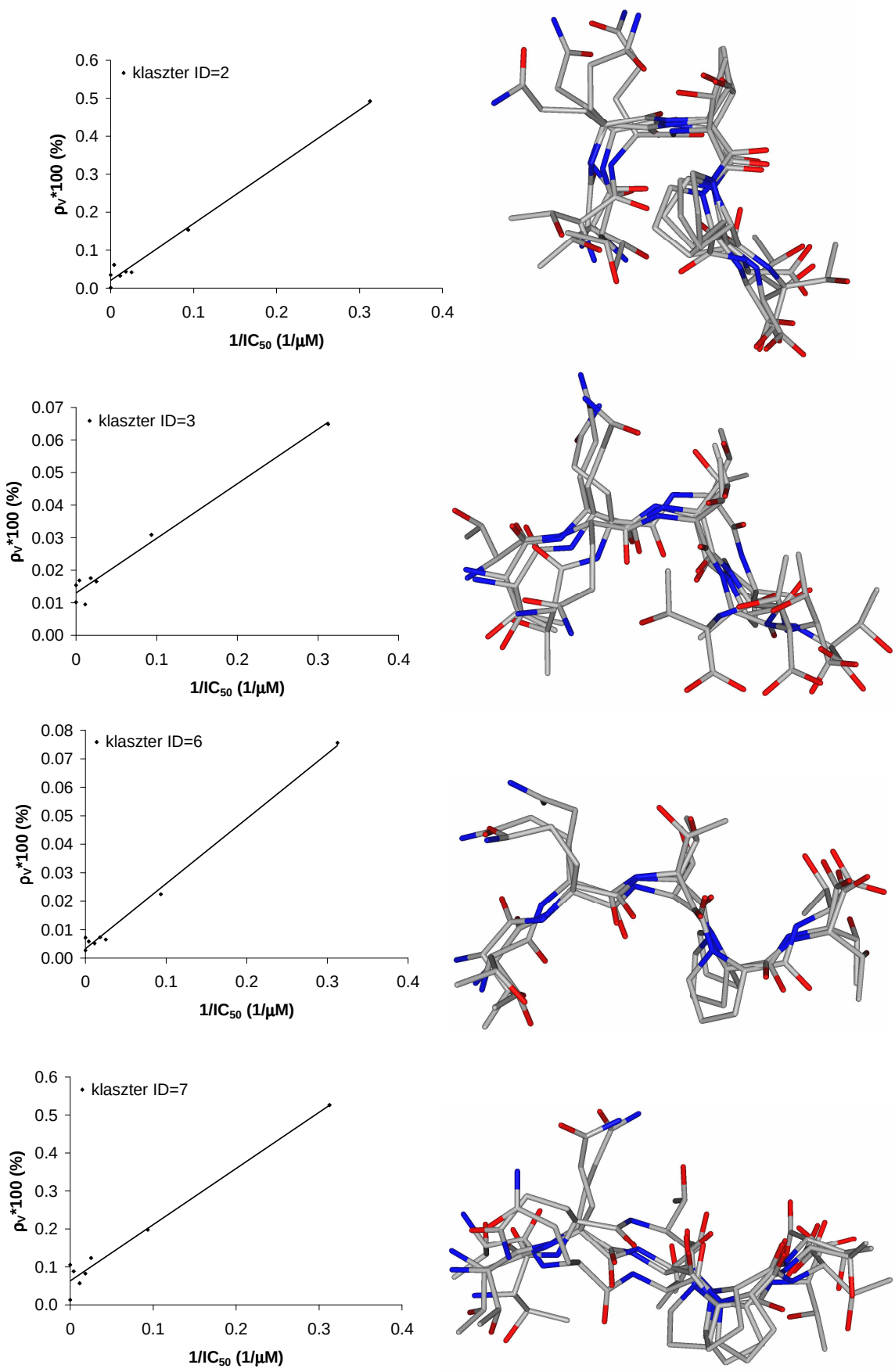
***5. táblázat**:A talált klaszterek a lineáris regresszió, mint modell alkalmazása esetén ($y=\alpha x+\gamma$). Ekkor az $x=1/IC_{50}$ és y az adott molekulához és adott térfogathoz rendelhető relatív konformációs sűrűség százalékban (ρ_V*100) kifejzve. N_v a klasztert felépítő V térfogatok számát, N_{uAC} a legjobb V térfogatot felépítő horgonykonformációk számát jelöli.*









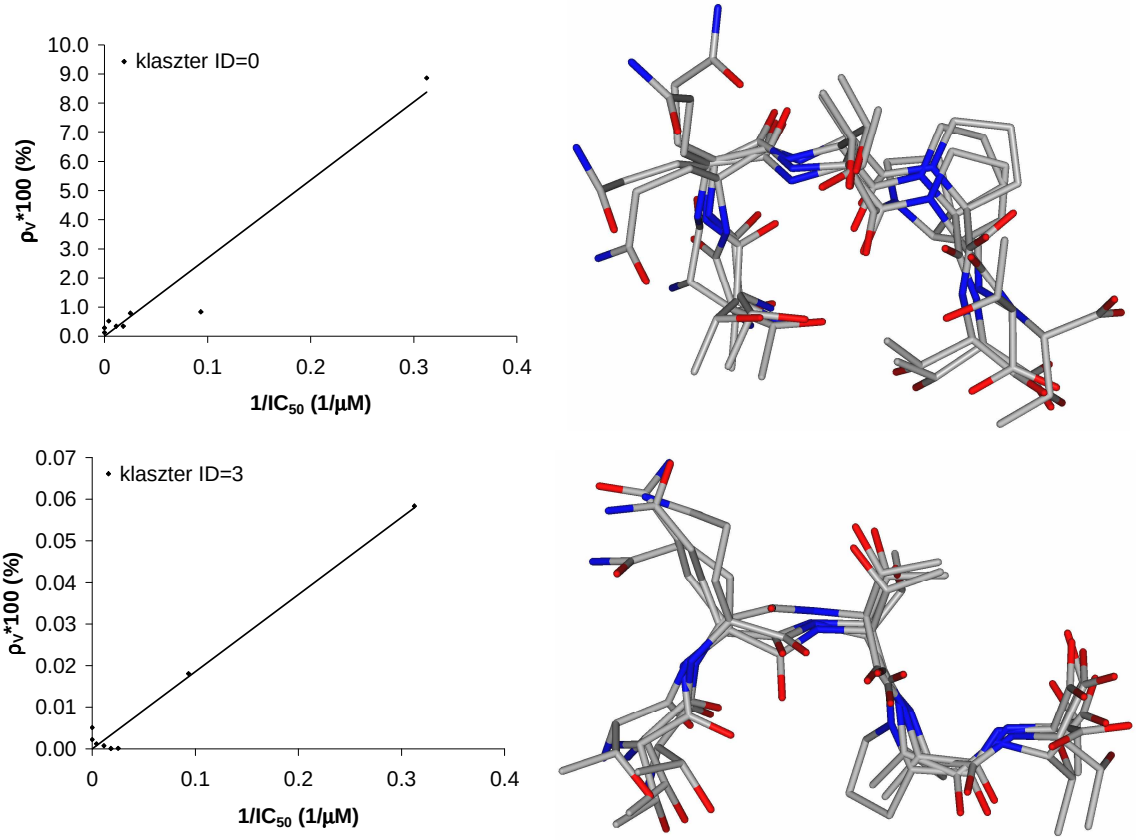


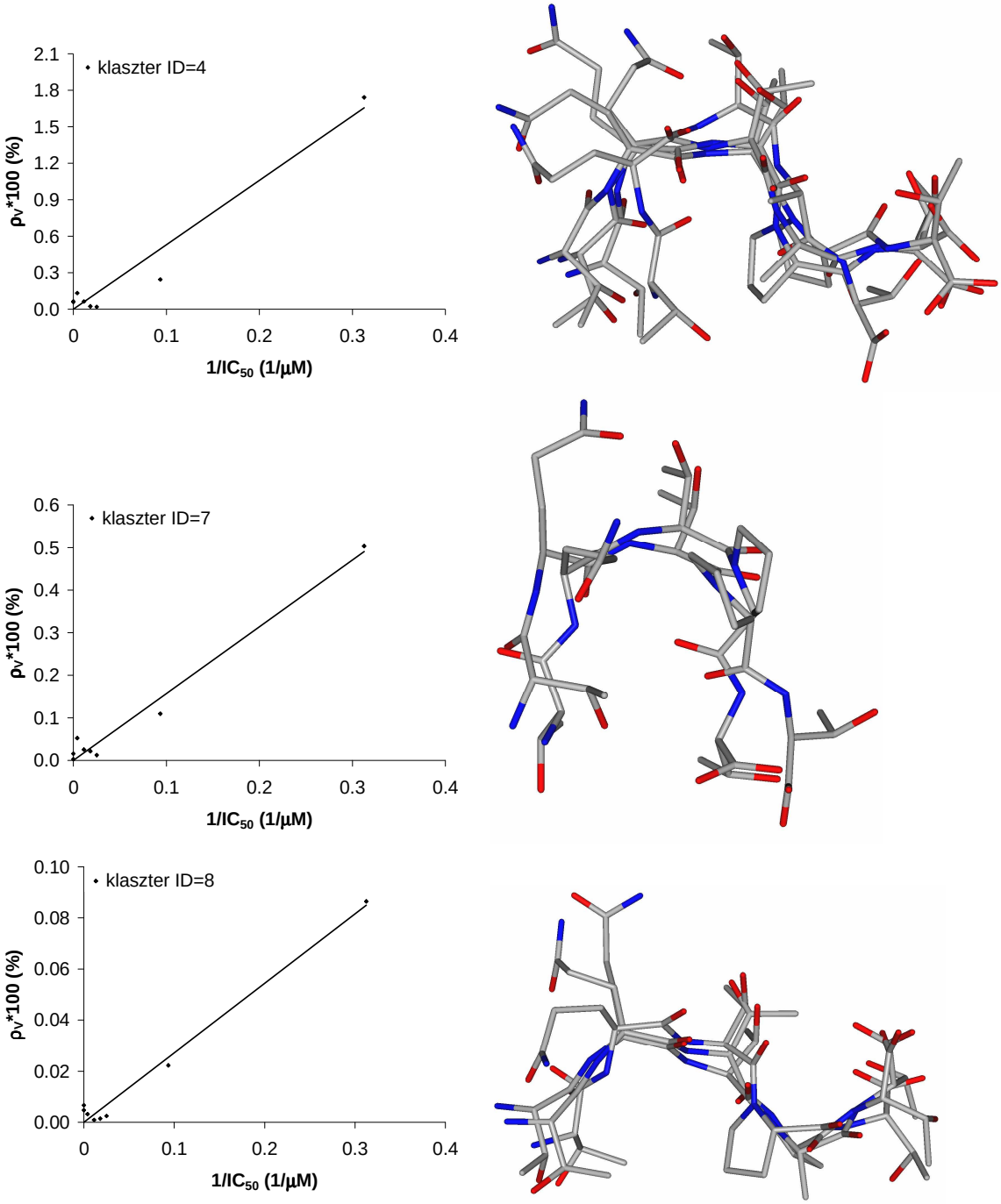
29. ábra *Baloldalon: A klaszter legjobb elemének kumulatív konformációs sűrűsége a kötődési aktivitás függvényében. Jobboldalon: A legjobb elemet alkotó horgonykonformációk.*

Amikor a zérus *y* tengelymetszetű egyenest alkalmaztam modellként, 50 Konformációs térfogatot kaptam, amelyek 18 klasztert alkottak. Az egy elemet tartalmazó klasztereket (N_V=1) véletlen korrelációnak minősítettem és kihagytam a további feldolgozásból, a maradék öt klaszter jellemzőit az **6. táblázat**, a klaszterek legjobb tagjait a **30. ábra** mutatja.

klaszter		klaszter legjobb tagja						
		korreláció			regresszió y= γ +αx		*V* mérete	
ID	N_V	r	LOO q^2	L50%O q^2	γ	α	rmsd (fok)	uAC száma
0	20	0,97	0,86	0,71	0	26,104	28	4
3	6	0,99	0,98	0,79	0	0,185	28	4
4	3	0,98	0,85	0,69	0	5,058	28	4
7	4	0,99	0,97	0,66	0	1,602	22	2
8	4	0,99	0,98	0,73	0	0,276	30	4

***6. táblázat**: Talált, nem egyelemű klaszterek a zérus y tengelymetszetű egyenes, mint modell alkalmazása esetén.*





30. ábra *Baloldalon: A klaszter legjobb elemének kumulatív konformációs sűrűsége a kötődési aktivitás függvényében. Jobboldalon: A legjobb elemet alkotó horgonykonformációk.*

A modellek magas korrelációs értékeienek (*r*, LOO q^2, L50% q^2) jelentőségét nem szabad túlértékelni, az egyenes meredekség jobban jellemzi a modell jóságát. Minél nagyobb az egyenes meredeksége, a modellt annál jobb minőségűnek tekintettem. A modelltípusokon ($\gamma=0$ és $\gamma\neq0$) belül az egyenes meredekségek összehasonlításához kiszámoltam a relatív meredekséget (α/α_{min}), ahol γ az aktuális modellhez tartozó egyenes meredeksége, és α_{min} a minimális meredekség.

Az eredmények áttekintéséhez elneveztem a jellemző peptidkonformációkat és a hozzájuk tartozó relatív meredekséget egész értékre kerekítettem. (**7. táblázat**).

A T_1 és az S konformációkat többször megtaláltam, függetlenül a modell típusától ($\gamma \neq 0$ vagy $\gamma=0$). A T_1 konformáció javasolható a kötődési konformációnak, de a teljesen biztos döntéshez szükséges lenne a megfelelő ciklopeptid származékok vizsgálatára.

$\gamma \neq 0$			$\gamma=0$		
klaszterID	peptidgerinc konformáció	α/α_{min}	klaszterID	peptidgerinc konformáció	α/α_{min}
0	L	5	0	T_1	156
1	T_1	16	3	S	1
2	T_1	9	4	S	30
3	T_2	1	7	S	10
6	L-S	1	8	T_3	2
7	S	9			

***7. táblázat**: A talált bioaktív konformáció-jelöltek kvalitatív típusai, és a kerekített relatív meredekségük. T-vel a kanyar (turn) jellegű, L-lel a lineárishoz közeli, S-sel az „s" szerű peptidgerinc konformációkat jelöltem.*

Meg kell jegyeznem, hogy noha a klaszterező eljárás algoritmusa tökéletesen működött mégis akár többfajta klaszter paraméter használata mellett is rakott azonosnak tekinthető konformációkat külön klaszterekbe. Ettől függetlenül a **7. táblázat**ból kiderül, hogy mindkét modelltípus ($\gamma \neq 0$ és $\gamma=0$) mellett a T_1 konformációnak adódott a legnagyobb meredekség értéke, sőt a fizikailag releváns, $\gamma=0$ esetben ez a hatás még inkább megmutatkozott.

Elvégeztem τ_v=45° beállításával is a fenti számításokat, és mivel a korreláció, az egyenes meredekségek és a kapott bioaktív konformációk a fenti értékekhez nagy hasonlóságot mutattak, ezeket az eredményeket külön nem mutatom be.

9.2.3.2. Algoritmus2

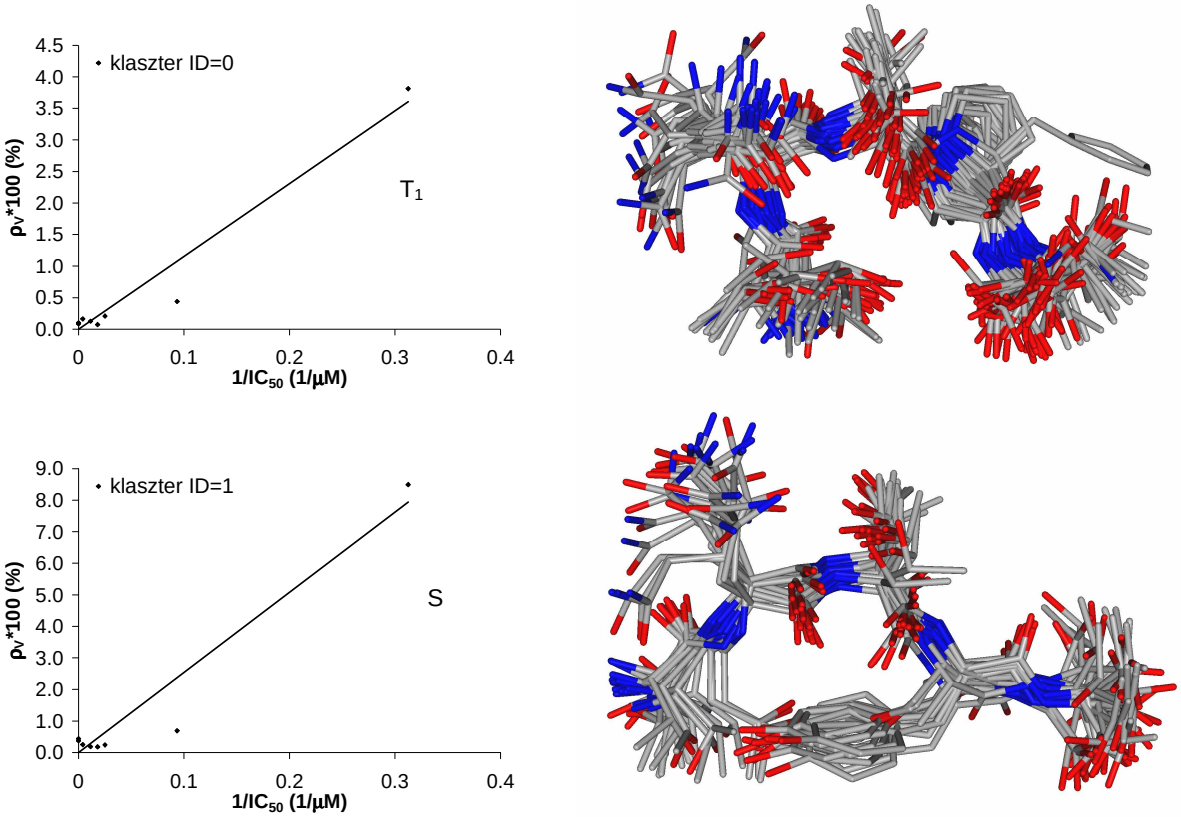
Az „Algoritmus1" eredményezhet nem folytonos *V* konformációs térfogatot is, például abban az esetben, ha egy el nem fogadott uAC-t körbevesznek a szomszédjai, amelyek egy elfogadott *V* térfogatot alkotnak. A konformációs „lyukak" a kötődés szempontjából nem értelmezhetőek, a korrelációs mértékek javításával működő modell építésével állnak elő. Folytonos konformációs térfogatok keresésére dolgoztam ki az „Algoritmus2"-t.

A horgonykonformációkat (uAC) a 9.2.2. fejezetben kapott trajektóriákon gyűjtöttem, úgy, hogy az első iterációban τ_{AC}=45° távolságküszöböt használtam. Egy randomizálási lépés után 5 iterációt végeztem (N_{iter}). Minden iterációban az új horgonykonformációkat 10%-kal

csökkentett (x_τ) távolságküszöb mellett gyűjtöttem, és a LOO q^2 értékük alapján szűrtem úgy, hogy csak a felső 20%-ot fogadtam el (x_q). A q^2 számításakor a zérus *y* tengelymetszetű egyenest alkalmaztam modellként. Egy horgonykonformációhoz rendelt sűrűségvektor (51) értékeinek meghatározásához az (49)-es összefüggést (cut off) használtam, a referenciasugár ekkor 45° volt (τ_d). Ez a referenciasugár a teljes futás során változatlan maradt. Összesen 20 randomizálási lépés történt (N_{rand}).

klaszter		klaszter legjobb tagja				
		korreláció			regresszió y= γ +αx	
ID	N_V	r	$LOOq^2$	$L50\%Oq^2$	γ	α
0	112	0,98	0,79	0,82	0	10,76
1	82	0,97	0,70	0,72	0	23,32
2	5	0,97	0,83	0,69	0	2,48
3	2	0,96	0,64	0,65	0	0,49

***8. táblázat**:Talált klaszterek. Az N_v ebben az esetben azt jelenti, hogy az aktuális konformációt hányszor találta meg az algoritmus.*



***31. ábra** Baloldalon: A klaszter legjobb elemének kumulatív konformációs sűrűsége a kötődési aktivitás függvényében. Jobboldalon: A legjobb elemet alkotó horgonykonformációk.*

Ez a gyűjtő algoritmus összesen 41012 darab horgonykonformációt eredményezett, amelyből kiválasztottam azokat, ahol a LOO q^2 és a L50%O q^2 is nagyobb volt, mint 0.6 (κ_{LOOq} , $\kappa_{L50\%Oq}$) és a térfogathoz nem tartozik olyan nem zérus aktivitás, amelynek zérus sűrűségérték felelne meg. Így 41012-ből 201 darab elfogadott uAC-t kaptam. Ezeket a

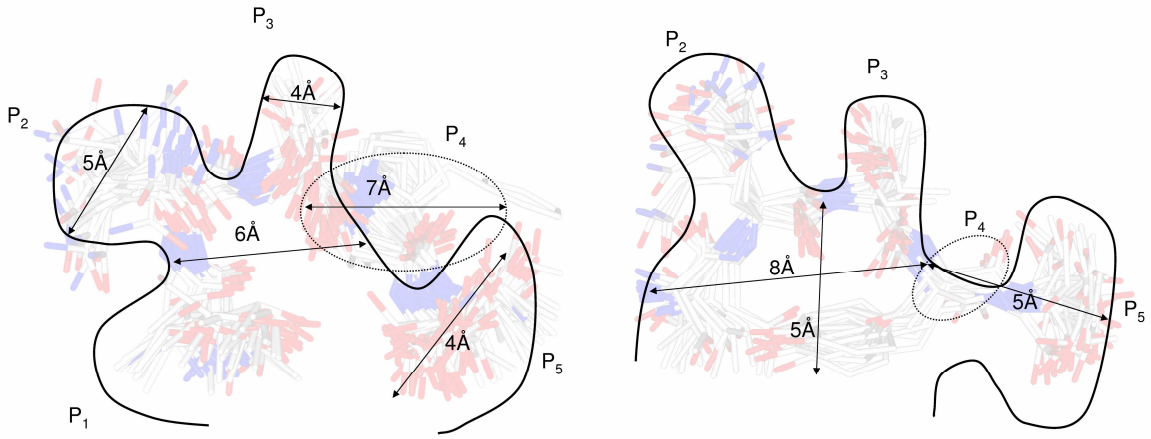
konformációkat a gerinc torziószögeik szerint klaszerekbe osztottam (τ_{cl}=60°). A klaszterezés konformációs szempontból egységes, megfelelően korrelált térfogatokat eredményezett (**8. táblázat**).

A klaszterekből a 0 és 1 jelű elegendően sok elemet tartalmazott ahhoz, hogy a véletlen korrelációtól eltekinthessünk. E klaszterek konformereit és a kapcsolódó aktivitás - konformációs sűrűség értékeket a **31. ábra** mutatja. E két konformáció megegyezik a 9.2.3.1. fejezetben T_1-el és S-el jelölt konformációkkal.

9.2.4. Az eredmények értelmezése

A bioaktív-konformáció kereső algoritmus lényegében két különböző bioaktív konformáció jelöltet eredményezett. A T_1-el jelölt konformáció Thr^1 hidroxil és a *C*-terminális oxocsoportja között feltételezhető, hogy hidrogénkötés alakul ki (**31. ábra**, felül). A peptidgerinc ilyenkor síkalkatú gyűrűt képez, e sík fölött található a Gln^2, Thr^3 és a Thr^5 a sík alatt helyezkedik el az X^4 aminosav oldallánca. Az S-sel jelölt konformációs térfogat meghatározó motívuma a Tyr^4 oldallánc aromás OH csoportja és a Thr^1 oldallánc alifás OH csoportja között fellépő hidrogénkötés.

A **32. ábra** a legvalószínűbb bioaktív konformációk jellemző méreteit mutatja. A peptid egyes aminosav oldalláncainak megfelelő térrészt P_1-től P_5-ig jelöltem. A pontozott vonal a peptidgerinc által képzett sík alatti csoportokat mutatja. Az S konformáció esetén kisebb gyűrű képződik, mit T_1 esetben. A P_2 P_3 és P_5 az S esetben egy egyenesbe esik. Ha a negyedik aminosav nem Tyr az oldalláncok a gyűrű síkja alatt helyezkednek (P_4).



32. ábra *A T_1 (baloldal) és S (jobboldal) bioaktív konformáció.*

Ugyan az eredményeket megfelelő óvatossággal kell kezelni, a peptidek kis száma és az aktivitásuk nem ideális eloszlása miatt, a módszer így is rámutatott arra, hogy a vizsgált

tárból hiányoznak a ciklopeptidek, így ilyen molekulák bevonásával jelentős többletinformációt nyerhetünk a kötődő konformációról és a kötőhelyről.

9.3. β-Amiloid epitóp peptid bioaktív konformációjának meghatározása

A β-amiloid (Aβ) egy 39-43 aminosavból álló peptid, ami káros folyamatokban felhalmozódhat, és oldhatatlan plakkokká aggregálódhat az agyban. Ezek a plakkok az Alzheimer-kórnak nevezett kognitív hanyatlást idézik elő[117-119]. Egy lehetséges terápiás megközelítés lehet az amiloid plakkok eltüntetésére az anti-Aβ-antitestekkel történő aktív vagy passzív immunizálás.

Számos kutatócsoport állított már elő az Alzheimer-kór transzgenikus egérmodelljeiben Aβ(1-42)-re és oligomerjeire kifejlesztett antitesteket, amelyekkel terápiás hatást lehetett kifejteni az Aβ szálképződésének csökkentésén keresztül[117-122]. Megállapították[123], hogy a terápiásan aktív antitestek specifikusan az Aβ *N*-terminális 4-10 régióját, az FRHDSGY aminosav szekvenciát ismerik fel. Az epitóp felismerésének specificitását kihasználva a jövőben Alzheimer-kór ellenes védőoltás vagy molekuláris diagnosztikai alkalmazás is elkészülhet.

Lineáris peptidből készült B vagy T sejt epitópok immunfelismerését és enzimatikus stabilitását többféle módon is lehet befolyásolni; például ciklizálással peptid kimérák alkamazásával[124-127], az epitóp *N*- és *C*-terminálishoz kapcsolódó lebegő (továbbiakban az szakirodalomban elterjedt angol „flank" szót használom) szakaszok módosításával[128, 129]; hordozó molekulához való konjugágással[130]. A fenti, Aβ epitópot is konjugálták polilizin alapú elágazó láncú SAK, EAK[131] és tetratuftsin[132] alapú hordozókhoz.

Kipróbálták különféle, *N*- és *C*-terminálishoz kapcsolt flank szakaszok hatását is. Ekkor azt találták, hogy a hordozó és az epitóp közötti távtartó (spacer), illetve a flank szakaszoknak jelentős hatásuk van az epitóp felismerése szempontjából. A β-alanin aminosavak bevitele az *N*- és *C*-terminálisra a kötődés erősségét, más esetben[128] D-aminosavak bevitele az epitóp *in vivo* enzimatikus stabilitását növelte meg. Az *N*- és *C*-terminálisra kapcsolt ciszteinek között kialakított diszulfid híd pedig a megfelelő bioaktív konformációt stabilizálhatja[127].

A fenti megfontolások alapján olyan peptidtárat terveztek, ahol a β-amiloid (4-10) epitóp *N*- és *C*-terminálisát α-L-, β- és α-D-alanint tartalmazó flank szakaszokkal szisztematikusan kiegészítették, emellett a peptidekből lineáris és ciklusos változatot is készítettek. (a D-alanint a továbbiakban „a"-val fogom jelölni). A peptidekhez minden esetben egy Biotin-GGGGG szakaszt kapcsoltak. A biotin az ELISA plate-hez kapcsolt avidin

fehérjéhez rögzítette a molekulát, az öt glicin aminosav pedig távtartó (spacer) szerepet töltött be. Indirekt ELISA kötődési vizsgálattal meghatározták a peptidek és az egér anti-Aβ(1-17) monoklonális antitest közötti kötődési aktivitásokat (**9. táblázat**).

Peptid No.	Peptid szekvencia	c ($x10^{-8}$ mol/l)
1	Biotin-GGGGGFRHDSGY-NH_2	1,10±0,12
3	Biotin-GGGGGAAFRHDSGYAA-NH_2	0,52±0,11
4	Biotin-GGGGGßAßAFRHDSGYßAßA-NH_2	0,71±0,25
5	Biotin-GGGGGaaFRHDSGYaa-NH_2	0,92±0,22
6	Biotin-GGGGGAAAFRHDSGYAAA-NH_2	0,70±0,16
7	Biotin-GGGGGßAßAßAFRHDSGYßAßAßA-NH_2	1,14±0,01
8	Biotin-GGGGGaaaFRHDSGYaaa-NH_2	0,82±0,20
16	Biotin-GGGGG[CFRHDSGYC]-NH_2	2,28±0,01
17	Biotin-GGGGG[CAAFRHDSGYAAC]-NH_2	0,61±0,06
18	Biotin-GGGGG[CßAßAFRHDSGYßAßAC]-NH_2	0,62±0,10
19	Biotin-GGGGG[CaaFRHDSGYaaC]-NH_2	1,07±0,39
20	Biotin-GGGGG[CAAAFRHDSGYAAAC]-NH_2	0,49±0,05
21	Biotin-GGGGG[CßAßAßAFRHDSGYßAßAßAC]-NH_2	0,52±0,03
22	Biotin-GGGGG[CaaaFRHDSGYaaaC]-NH_2	1,11±0,29

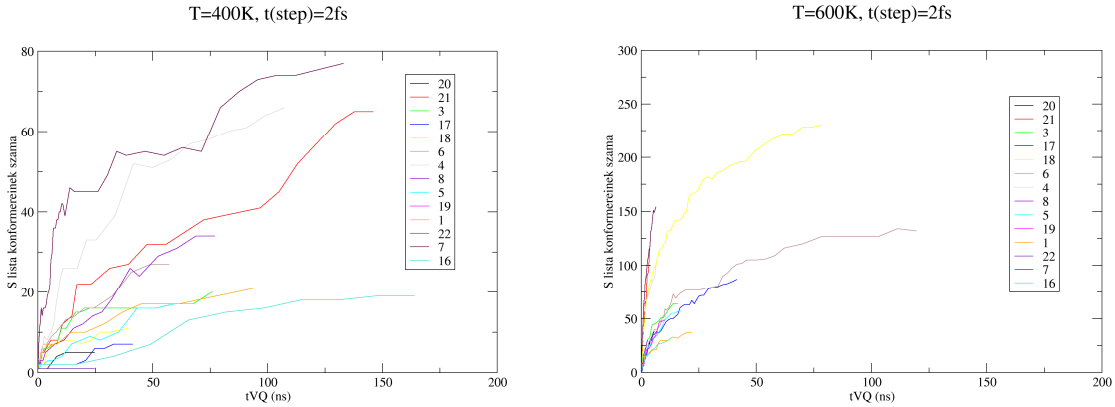
***9. táblázat**:A kötődő peptidek kötődési aktivitás értékei az ELISA kísérletben OD=1 értéknél. A peptidek sorszámozása (Peptid No.) az eredeti cikkből származik*

Célom az volt, hogy e peptidtár esetén is meghatározzam a kötődő konformációt[84]. A peptidek mérete és flexibilitása megnehezíti egy tradicionális QSAR vizsgálat elvégzését, de mivel a PCORE *a priori* ismert, a bioaktív konformációja meghatározható. Ebben az esetben ugyanis a PCORE az FRHDSGY aminosavakból felépülő szekvencia, amely közvetlenül kötődik az antitesthez, az off-PCORE rész pedig a flank és a ciklizáláshoz használt aminosavakból áll. A Biotin-GGGGG részt a számításhoz használt peptidekből kihagytam, mivel ezt a szakaszt az ELISA vizsgálatok elvégzése kedvéért vitték be a rendszerbe és feltételezhető, hogy a PCORE konformációját nem befolyásolja. Mindegyik peptid *N*-terminálisára acetil, *C*-terminálisára amid csoportot helyeztem el.

9.3.1. A kiindulási konformerek előállítása

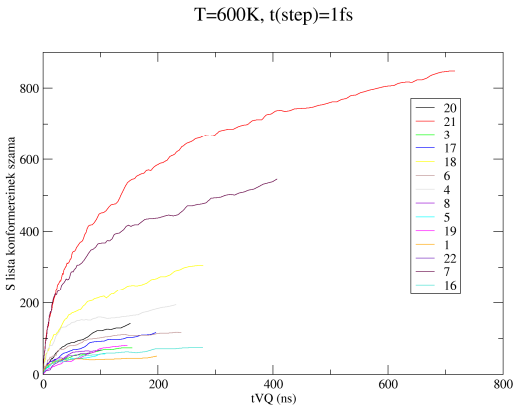
A konformerkereső eljáráshoz az eddigiekhez hasonlóan a GROMACS szoftvercsomag keretén belül implementált, kifejezetten fehérjék, peptidek vákuumos szimulációjára kifejlesztett GROMOS96 43b1 erőteret használtam (ld. 9.1.2). Az MD szimulációkhoz Berendsen[109] termosztátot (T=400 K, 600 K) és barosztátot ($p=10^5$ Pa) alkalmaztam. A peptidkötés torziószögéhez tartozó erőállandót ebben az esetben tízszeresére növeltem, azért, hogy az alkalmazott magas kinetikus energia ne legyen képes átbillenteni a transz peptidkötéseket cisz állásba. A többi, itt nem említett paraméter értékét a

programcsomag által javasolt alapbeállításon hagytam. A molekula transzlációs és rotációs sebességkomponensét minden szimulációs lépésben kompenzáltam, a kémiai kötések hosszát állandó értéken tartottam. Az arginin protonált, az aszparaginsav deprotonált formáját vettem figyelembe. A molekuladinamika időlépését először az alapbeállításon, 2 fs-on vizsgáltam T=400 K-en és T=600 K-en (**33. ábra**). A konformációs teret az egyes peptidek φ és ϕ gerinc torziószögeivel definiáltam. A távolság-küszöbértéket (τ) 60°-ra állítottam.



33. ábra *Talált konformerek száma a szimulációs idő függvényében, baloldalon 400 K-en, jobboldalon 600 K-en végzett MD szimulációkat használtam. A lépésköz mindkét esetben 2 fs volt. A jelmagyarázatban a molekulák kötődési aktivitás szerint csökkenő sorrendbe vannak rendezve.*

A 400 K-en végzett szimulációk nem eredményeztek megfelelő számú konformációt, valószínűsíthető, hogy ez a hőmérséklet nem tűnik elegendőnek a konformációs kinetikai energiagátak leküzdéséhez, viszont 600 K-en némelyik szimuláció instabillá vált. Ezért döntöttem úgy, hogy a MD időlépését 1 fs-ra csökkentem (**34. ábra**).



34. ábra *Konformerkeresés 600 K-en, 1 fs lépésköz mellett - a talált konformerek száma.*

Ebben az esetben a konformerkeresés már elfogadható eredménnyel zárult. Az egyes molekulák talált „S” konformereinek számát és a kereséshez felhasznált szimulációk hosszát a

10. táblázat „N_s” és t_{vq} oszlopa mutatja. A legtöbb molekula esetén a legkisebb energiájú konformert 200 ns szimulációs időn belül találta meg az algoritmus. A talált konformerek száma a bejárható konformációs tér nagyságáról, azaz a molekula flexibilitásáról nyújt kvantitatív információt. Megvizsgáltam azt, hogy a molekulák „S” konformereinek száma (flexibilitása) és a kötődési aktivitása között található-e összefüggés, de megfelelő korrelációt nem kaptam.

9.3.2. Az oldatbeli konformációs viszonyok feltérképezése

A 9.1-es és a 9.2-es fejezet példáiban az „S” listából választott konformerekre addig végeztem az MD szimulációkat, amíg e szimulációk új horgonykonformációkat eredményeztek. Ebben az esetben minden molekula összes kiindulási konformációjára a következő protokollt alkalmaztam. A kiindulási konformert vízmolekulákkal vettem körbe (SPC modell, periódikus határfeltétel alkalmazása mellett). Egy peptid különböző konformereihez különböző mennyiségű vízmolekulát adtam (ezek átlagát a **10. táblázat** N_{wat} oszlopa mutatja), úgy, hogy a peptid és az oldószerdoboz fala között legalább 8 Å távolság legyen.

Peptid No		t_{vq} (ns)	N_S	t_{sol} (ns)	N_{wat}
20	Ac-[CAAA**FRHDSGY**AAAC]-NH_2	151,63	142	28,40	1461
21	Ac-[CβAβAβA**FRHDSGY**βAβAβAC]-NH_2	716,05	848	169,60	1412
3	Ac-AA**FRHDSGY**AA-NH_2	154,53	74	14,80	1315
17	Ac-[CAA**FRHDSGY**AAC]-NH_2	196,11	117	23,40	1415
18	Ac-[CβAβA**FRHDSGY**βAβAC]-NH_2	278,29	306	61,20	1388
6	Ac-AAA**FRHDSGY**AAA-NH_2	239,79	117	23,40	1424
4	Ac-βAβA**FRHDSGY**βAβA-NH_2	230,93	194	38,80	1345
8	Ac-aaa**FRHDSGY**aaa-NH_2	101,13	68	13,60	1417
5	Ac-aa**FRHDSGY**aa-NH_2	107,53	57	11,40	1300
19	Ac-[Caa**FRHDSGY**aaC]-NH_2	146,47	81	16,20	1444
1	Ac-**FRHDSGY**-NH_2	197,67	51	10,20	1221
22	Ac-[Caaa**FRHDSGY**aaaC]-NH_2	81,73	61	12,20	1517
7	Ac-βAβAβA**FRHDSGY**βAβAβA-NH_2	406,96	545	109,00	1408
16	Ac-[C**FRHDSGY**C]-NH_2	277,52	75	15,00	1234

***10. táblázat**:A számításokhoz használt molekulaszerkezetek kötődési aktivitás szerint sorba rendezve. A t_{vq} a konformerkereséshez használt szimulációs időt, ns a kiindulási konformációk számát, N_{wat} a szimulációban felhasznált vízmolekulák darabszámát jelenti. A t_{sol} az explicit oldószeres közegben egy molekulára végzett szimulációknak az együttes hosszát jelenti, amelyből a bioaktív-konformáció keresés történt. A futtatott szimulációk teljes hossza ennek kétszerese.*

A peptid töltéssel rendelkező oldallácaihoz, a hisztidinhez és az argininhez Na^+ és Cl^- ellenionokat adtam. Erre a rendszerre úgy futtattam MD szimulációt, hogy a peptid és az ellenioinok mozgását „lefagysztottam” (freeze). Ebben az esetben a biomolekulák oldatbeli

szimulációjára a GROMOS96 53a6 erőteret[133] alkalmaztam. A lépésközt 2 fs-ra, a szimuláció hosszát 10 ps-ra állítottam. A szimulációs hőmérsékletet 300 K-re állítottam be három, a peptidhez, az ellenionokhoz és az oldószerhez rendelt Berendsen-féle termosztát segítségével. Ezután ugyanezekkel a beállításokkal már olyan szimulációt futtattam, ahol a peptid mozgását is szabadon engedtem, és a szimulációs időt 200 ps-ra növeltem. Az első szimuláció a vízmolekulák, a második a peptid egyensúlyba hozására szolgált.

Az ezt követő 200 ps hosszúságú MD szimulációt használtam fel a bioaktív-konformáció keresés céljára. Így az összes molekula minden konformációjára együttesen több mint 1 μs szimulációt futattam, és több mint 0.5 μs hosszúságú szimulációt vettem figyelembe a bioaktív-konformáció keresésekor (**10. táblázat** t_{sol} oszlopa).

9.3.3. Bioaktív-konformáció keresés

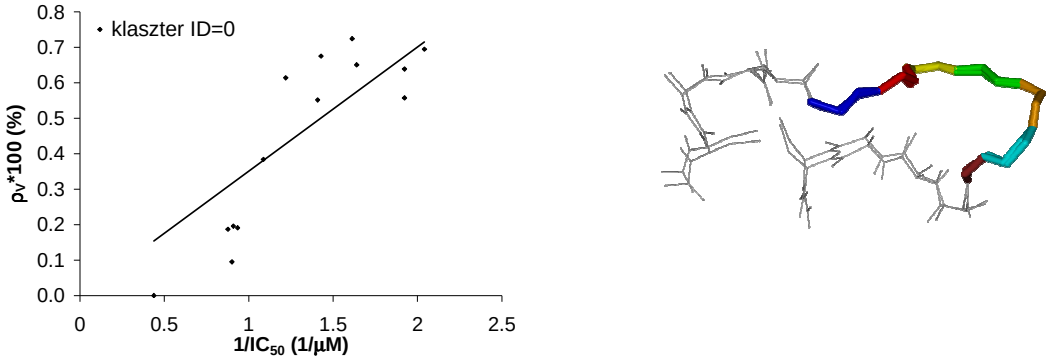
A bioaktív konformációt az *a priori* definiálható PCORE, az FRHDSGY peptidszakasz φ és ϕ gerinc torziószögeivel leírható konformációs térben kerestem. Erre a peptidtárra már csak az Algoritmus2-t alkalmaztam. A horgonykonformációkat (uAC) a 8.4.2. fejezetben leírt algoritmus alapján gyűjtöttem úgy, hogy az első iterációban τ_{AC}=45° távolságküszöböt használtam, amelyet iterációnként 10%-kal csökkentettem (x_τ), így 5 iterációt végeztem (N_{iter}). Az egyes iterációban gyűjtött új horgonykonformációkat LOO q^2 értékük alapján szűrtem, úgy, hogy a felső 20%-ot fogadtam el (x_q). A q^2 számításakor a zérus *y* tengelymetszetű egyenest alkalmaztam modellként. Egy horgonykonformációhoz rendelt sűrűségvektor értékeit az (49) és (51) szerint számoltam (cut off) a referenciasugár ekkor 45° volt (τ_d). Összesen 20 randomizálási lépés történt (N_{rand}).

klaszter		klaszter legjobb tagja				
		korreláció			regresszió y= γ +αx	
ID	N_V	r	LOOq^2	L50%Oq^2	γ	α
0	2	0,88	0,67	0,69	0	0,34

***11. táblázat**:A talált klaszter.(κ_{LOOq}=0.6 és $\kappa_{L50\%Oq}$=0.6, zérus sűrűségkritérium nélkül)*

A kapott 233356 darab uAC-t korreláció és konformációs hasonlóság szerint klaszteresztem. A κ_{LOOq}, $\kappa_{L50\%Oq}$ paraméterek értékét 0,6-ra állítottam. Abban az esetben amikor kikötöttem, hogy a nem zérus aktivitású molekulához ne tartozhasson zérus konformációs sűrűségérték, ahogy ezt a 9.2.3.2. fejezetben is tettem, elfogadható uAC-ket nem találtam. Ha ezt a kritériumot nem alkalmaztam két uAC-t találtam, amelyek LOO q^2

értéke 0,64-nek és 0,67-nek adódott. E két uAC a gerinc torziószögek szerinti konformációs távolságuk alapján egy klaszterbe (τ_{cl}=60°) tartozott (**11. táblázat**, **35. ábra**).

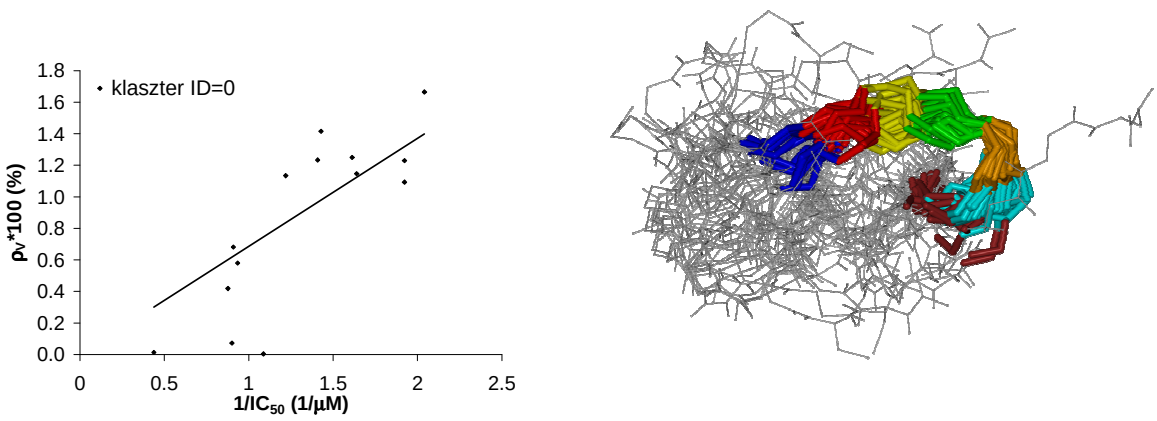


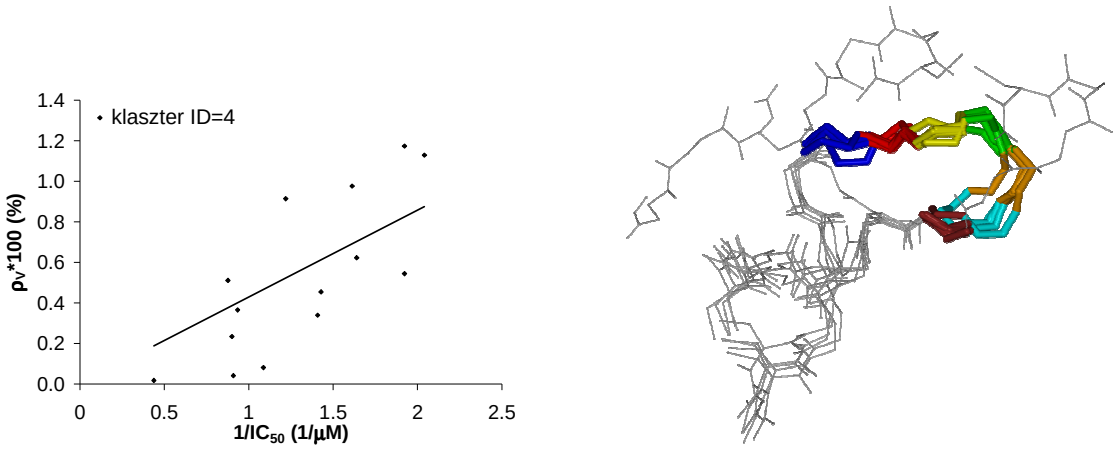
35. ábra *Baloldalon: A klaszter legjobb elemének konformációs sűrűsége a kötődési aktivitás függvényében. Jobboldalon: A 0-val jelölt klaszter tagjai (horgonykonformációk). A különböző alanin flank régiókat szürkével, a PCORE aminosavak peptidgerincét színessel (F- kék, R-piros, H-sárga, D-zöld, S-narancs, G-cián, Y-barna) ábrázoltam.*

A κ_{LOOq} és a $\kappa_{L50\%Oq}$ paraméterek értékét ezután 0,4-ra állítottam, zérus sűrűségérték kritérium alkalmazásával 88 uAC-t és 6 klasztert találtam (**12. táblázat**), amelyből kettőt, a 0-val és a 4-gyel jelöltet fogadtam el az elemszám alapján (**36. ábra**).

klaszter		klaszter legjobb tagja				
		korreláció			regresszió y= γ +αx	
ID	N_V	r	$LOOq^2$	$L50\%Oq^2$	γ	α
0	64	0,83	0,57	0,59	0	0,65
4	15	0,77	0,43	0,47	0	0,40
1	4	0,88	0,54	0,57	0	0,13
3	2	0,76	0,46	0,49	0	0,27
5	2	0,75	0,43	0,46	0	0,17
2	1	0,87	0,48	0,53	0	0,13

12. táblázat*: A talált klaszterek.(* κ_{LOOq}*=0,4 és a* $\kappa_{L50\%Oq}$*=0,4, zérus sűrűségkritérium használata mellett).*



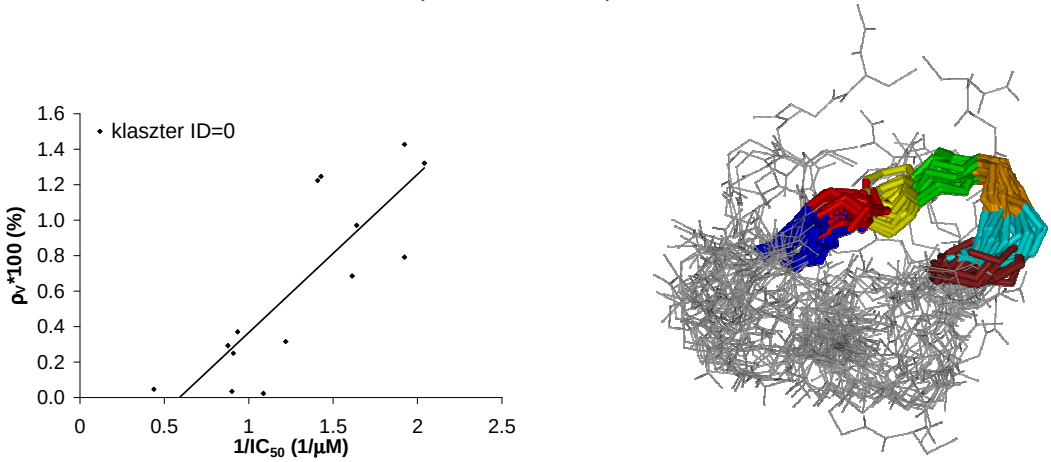


36. ábra *Baloldalon: A klaszterek legjobb elemének konformációs sűrűsége a kötődési aktivitás függvényében. Jobboldalon: A 0-val és a 4-gyel jelölt klaszterek horgonykonformációi. A különböző alanin flank régiókat szürkével, a PCORE aminosavak peptidgerincét színessel (F- kék, R-piros, H-sárga, D-zöld, S-narancs, G-cián, Y-barna) ábrázoltam.*

Összehasonlításképpen kipróbáltam a lineáris regressziót is, ahol az *y* tengelymetszet nincs megkötve a modell építésekor. A fenti paraméterek értékét változatlanul hagytam, ekkor összesen 148391 uAC-t találtam. Amikor a κ_{LOOq} és a $\kappa_{L50\%Oq}$ paraméterek értékét 0,6-ra állítottam, zérus sűrűségérték kritérium alkalmazása mellett nem találtam elfogadható horgonykonformációkat. Amikor a κ_{LOOq} és a $\kappa_{L50\%Oq}$ paraméterek értékét 0,4-re csökkentettem, összesen 84 elfogadható konformációs régiót találtam, ezek három klasztert alkottak (**13. táblázat**). Az elemszám alapján a 0-val jelölt klasztert fogadtam el (**37. ábra**).

klaszter		klaszter legjobb tagja				
		korreláció			regresszió $y=\gamma+\alpha x$	
ID	N_V	r	$LOOq^2$	$L50\%Oq^2$	γ	α
0	76	0,83	0,59	0,63	-0,53	0,89
1	5	0,77	0,45	0,51	-0,10	0,16
2	3	0,75	0,40	0,46	-0,36	0,50

13. táblázat*:A talált klaszterek.(* κ_{LOOq}*=0.4 és a* $\kappa_{L50\%Oq}$*=0.4, zérus sűrűségkritérium használata mellett).*



37. ábra *Baloldalon: A klaszter legjobb elemének konformációs sűrűsége a kötődési aktivitás függvényében. Jobboldalon: A 0-val jelölt klaszter horgonykonformációi.*

Az összeforgatott konformációk atomi koordináta-rmsd értékei ugyancsak használhatók lennének e célra, de az eljárás meglehetősen számításigényesnek bizonyult (C. Függelék)

9.3.4. Az eredmények értelmezése

Bioaktív konformációként olyan konformációk fogadhatók el, ahol a kanyar csúcsán a szerin aminosav helyezkedik el. A kanyar csúcsa általában a szárakkal egy síkban található, egyedül a 4-es klaszter esetén helyezkedik el kevéssel a szárak síkja fölött (**36. ábra**). A modell megbízhatóságát növelte az a tény is, hogy a PCORE-nak tekintett FRHDSGY peptidszakasz ebben a konformációban jól hozzáférhető az ellenanyag számára, a flank aminosavak nem temetik el a molekula belsejébe, ahogy ez a jelenség sok, rosszul korreláló uAC esetén megfigyelhető volt. A szűkebb ciklusba zárt peptidek flank szakaszai a PCORE *N*- és *C*-terminálisát gyűrűszerűen kötik össze, míg a nagyobb tagszámú ciklopeptideknél a flank régióban egy extra kanyar szerkezet figyelhető meg. Ilyen módon ki tud alakulni PCOREnak ugyanaz az aktív konformációja a kisebb és a nagyobb tagszámú ciklopeptideknél is.

10. Következtetések, eredmények összefoglalása

Doktori dolgozatom témája, a konformációs flexibilitás és a kötődési aktivitás kapcsolatának vizsgálata egy gyakorlati példa, a mucin epitóp peptideket tartalmazó molekulatár kapcsán merült föl. Ezek a pentapeptidek jelentős konformációs flexibilitással rendelkeznek, így a standard, általában kevesebb szabadon forgatható torziót (rotatable bond) figyelembevevő 3D QSAR módszerek nem szolgáltattak volna megfelelő eredményt. Ezért tűztük ki célul egy olyan általános számítási eljárás kidolgozását és megvalósítását, amely használható nagy konformációs változékonysággal rendelkező molekulákból álló könyvtáron, és a kötőhely geometriájának ismerete nélkül a molekulák mért kötődési erősségének felhasználásával meghatározza a kötőhelyen felvett legvalószínűbb konformációt.

Levezettem, hogy az általánosan szerkezet-hatás összefüggésnek tekintett egyenlet hogyan partícionálható a konformációs flexibilitásnak és molekula rigid kötődésének megfelelő tagra. Ha elő tudunk állítani olyan molekulakönyvtárat, ahol az egyes molekulák a rigid kötődésének megfelelő tagját állandónak vehetjük, a mért (például IC_{50} vagy K_i) értékek reciproka egyenesen arányos lesz a bioaktiv konformáció oldatbeli relatív sűrűségével.

A rigid kötődést leíró tag állandóságának feltétele, hogy a molekuláknak azonos módon és azonos helyre kell kötődniük, és az egyes molekulákat felépítő atomoknak csoportosíthatónak kell lenniük egy úgynevezett PCORE, és esetlegesen egy off-PCORE halmazba. A PCORE halmaz olyan atomokat tartalmaz, amelyek lényeges szerepet játszanak a kötődésben, az off-PCORE atomjai pedig a kötődést közvetve, csak a PCORE konformációs viselkedésének módosításával befolyásolják. A PCORE halmaznak minden molekula esetén ugyanolyan típusú az atomokból kell állnia, és szükséges, hogy PCORE a kötődés során kialakuló konformációját (farmakofor) az oldatban is felvegye. Ideális példa a dolgozatban vizsgált egyik peptidtár, ahol minden molekulában megtalálható a PCORE-ként definiált közös FRHDSGY peptidszakasz. Ezt a közös régiót az *N*- és a *C*-terminálison változó hosszúságú lineáris és gyűrűvé záródó lebegő (flank) szakaszokkal vették körbe (off-PCORE).

Elkészítettem egy olyan, 3D farmakofor kereső inverz QSAR eljárást (IQsar), amely az egyes ligand molekulák oldatbeli konformációs eloszlását előállítja, majd egy közös térbe, a PCORE atomok által leírt térbe vetíti. Ebben a térben a fenti összefüggés alapján megkeresi azokat a konformációs térfogatokat, amelyekbe az egyes molekulák kötődési aktivitásával egyenesen arányos konformációs sűrűségértékek esnek. A levezetett összefüggés nem ad

tájékoztatást a keresendő konformációs térfogat méretével kapcsolatban, ezért azt irodalomban általánosan elfogadott érték alapján definiáltam. Megállapítottam, hogy a keresendő konformációs térfogatnak összefüggőnek és folytonosnak kell lennie.

A számítási eljárás két fő modulból áll. Az első modul az egyes molekulák konformációs terét térképezi fel molekuladinamikai szimulációk segítségével, és az általunk kidolgozott konformerkereső eljárást is tartalmazza. E konformerkereső eljárás ciklikus és lineáris molekulákra is alkalmazható, és globális energiaminimum keresésre is használható. Feltételezzük, hogy az alkalmazott konformációs analízis a ligand molekulák oldatbeli konformációs eloszlását megfelelően írja le. A második modul maga a bioaktív-konformáció kereső eljárás. Az első modult FORTRAN a másodikat JAVA nyelven készítettem el. Az elkészült módszernek az inverz QSAR nevet (IQsar) adtam, mivel e módszer ellentétben az általános QSAR eljárásokkal, az aktivitás értékek alapján választ szerkezetet.

Az IQsar-ral szemben a legtöbb 3D QSAR módszer megkívánja a bioaktív konformáció, vagy a farmakofor előzetes ismeretét azért, hogy a molekulákat ezek alapján összeforgatva a szerkezet-hatás összefüggés számítása elvégezhető legyen. A kidolgozott eljáráshoz nem kell ilyen előzetes ismerettel rendelkezni a farmakoforról, csak az őt (PCORE) alkotó atomokat kell előre definiálni és a kötődési aktivitás értékeket kísérleti úton meghatározni. Az eljárás eredménye pontosan ez a 3D farmakofor lesz.

Az első modult külön is felhasználtam egy úgynevezett oligotuftsin alapú hordozómolekula konjugációs helyeinek oldatbeli hozzáférhetőségének jellemzésére. A kapott eredmények rámutattak arra, hogy a konjugálási helyek nagyjából egyformán, a molekula többi részéhez képest jobban hozzáférhetőek.

Bioaktív-konformáció keresést két molekulatárra végeztem. Először egy mucin epitóp pentapeptidekből álló tár esetén arra a kérdésre kerestünk választ, hogy milyen konformációval magyarázható a kötődési erősség különbsége. A hat aktív peptidhez kettő inaktívat választottunk és feltételeztük, hogy a kötődésbeli különbségük a konformációs viselkedésüktől jelentősen függ. A konformációs analízis segítségével kapott alacsony energiájú konformerekből összesen 1,9 μs explicit oldószeres molekuladinamikai szimulációt futtattam. A kapott trajektóriákon kétféle algoritmussal bioaktív-konformáció keresést végeztem. Az első algoritmus eredményezhet nem folytonos konformációs térfogatot, ezért a hiba kiküszöbölésére elkészítettem az összefüggő konformációs térfogat keresésére alkalmas eljárást is. Mindkét módszer alapján ugyanazt a kettő, T_1-el és S-el jelölt kötődő konformációt

találtam meg. Ugyanakkor, a második módszer esetében egyértelműen szét lehetett választani a véletlen korrelációnak és a tényleges hatásnak megfelelő konformációs térfogatokat. Inkább a T_1-el jelölt fogadható el tényleges bioaktív konformációként, ahol a jellemző motívum kialakításában döntő szerepe van a Thr[1] oldallánca és a *C*-terminális között kialakuló hidrogénkötésnek. A kötődő konformációt ugyanakkor nagyobb biztonsággal tudnánk megjósolni, ha a ciklikus származékokat is fel tudnánk venni a peptidtárba.

Megállapítottam, hogy a módszer alapján kevés kiindulási információ esetén is képesek voltunk olyan új, vizsgálandó szerkezetcsaládra javaslatot tenni, amelynek segítségével a kötődéskor fellépő konformációs viszonyok jobban tisztázhatók.

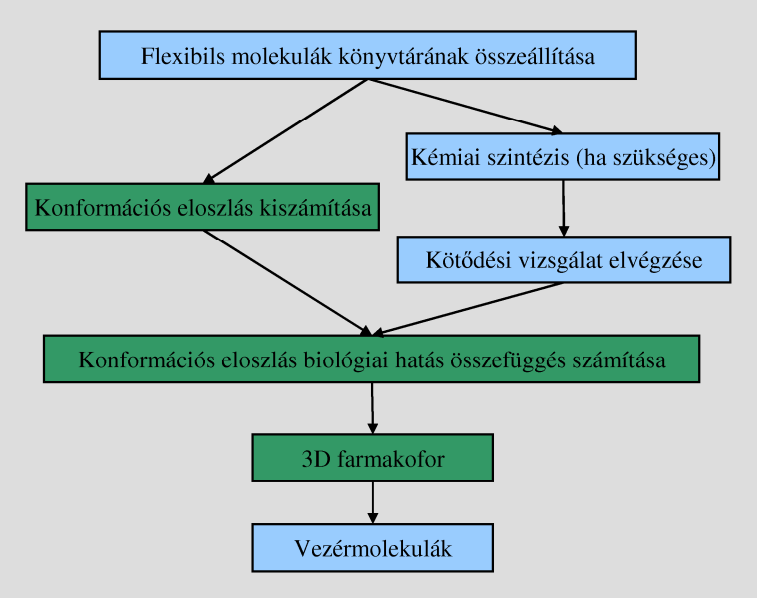
A számítások elvégzése során szerzett tapasztalatokat a következőképpen összegezhetem. A molekulatárnak sokkal több aktív molekulából kell állnia, úgy, hogy az aktivitásaik egyenletesen fedjék le a teljes intervallumot. Érdemes már eleve tár kialakításakor jól körülhatárolhatóan megalkotni a PCORE és off-PCORE részeket. A PCORE konformációs eloszlását ciklussá záródó off-PCORE régiókkal jelentősen befolyásolhatjuk, így a számítás pontosságát növelhetjük. E megfontolásoknak a figyelembe vételével esett a választás egy olyan peptidtárra, ahol a közös β-amiloid epitóp szekvenciát lebegő (flank) szakaszokkal egészítettek ki. Ekkor a PCORE régióként az epitóp szekvenciát, az off-PCORE régióként pedig a lebegő szakaszokat definiáltam. Ebben a peptidtárban már megfelelő számú molekulát és ciklikus peptideket is találunk, emellett az aktivitás értékek már egyenletesebben fedik le a teljes tartományt. A konformációs analízis segítségével kapott alacsony energiájú konformerekből kiindulva, az összes molekulára összesen 1 μs szimulációt futattam, és több mint 0,5 μs hosszúságú szimulációt vettem figyelembe a bioaktív-konformáció keresésekor. Egyetlen típusú kapott bioaktív konformációt kaptam. E konformáció egy olyan síkalkatú kanyar szerkezet, ahol a kanyar csúcsán a szerin aminosav áll, és a kanyar szárai egyenesen és párhuzamosan futnak.

A standard QSAR modellvalidálási eljárások során használható külső validálásra a fenti két esetben azért nem nyílt lehetőség, mert ha a kiindulási peptidtárakból teszt-molekulatárakat különítettem volna el, jelentősen csökkent volna a modellt felépítő molekulák amúgy még talán elfogadható száma. A módszert a közeljövőben széles körű tesztelésnek vetem alá, ahol megfelelő számú molekulákból álló halmazon már külső validálásra is lesz lehetőség. Ugyanakkor a modell validálásának legmeggyőzőbb módja, ha a kapott bioaktív konformációt kémiailag rögzítjük, és ekkor jelentős kötődési aktivitás növekedést érünk el[134].

Az elkészített számítási eljárás fizikai-kémiai összefüggésen alapul és nem alkalmaz úgynevezett „soft modeling” módszereket, ugyanakkor a kapott modellek statisztikai belső validálása alapján is elfogadhatónak bizonyultak olyan flexibilis molekulatárak esetén, melyek vizsgálatára más módszerekkel nem nyílna lehetőség.

11. Kitekintés

Az új gyógyszerek kifejlesztésének első szakasza - amennyiben egy megfelelő jelölt vagy a megcélzott fehérje, enzim szerkezete nem áll rendelkezésre - leginkább próbálgatáson alapul. A siker érdekében molekulák ezreit tartalmazó „könyvtárak” *in vitro* biológiai tesztelését végzik el, vizsgálva a molekuláknak a célfehérjéhez történő kötődését. Mivel a cél egy megfelelő gyógyszerjelölt megtalálása, ezért ezek elsősorban merev szerkezetű molekulák. A hatóanyag molekulák flexibitásának felső határát nagyjából 10 forgatható torziószöggel jellemzik[135]. A jelentősen flexibilis molekulákat nem kedvelik, mivel nem specifikus hatásokat válthatnak ki.



38. ábra *Vezérmolekulák keresése módszerünkkel. A zölddel jelölt részfolyamatokat a javasolt IQsar eljárás végzi.*

Ötletünk lényege, hogy a gyógyszerkutatás első fázisába egy átmeneti lépést iktatunk, azaz elsőként nem a gyógyszermolekulát, „csak” a megfelelő 3D farmakofort próbáljuk kísérleti és számítási eredmények alapján meghatározni. Erre a célra kihasználjuk a flexibilis molekulák szerkezeti változatosságát és, hogy könnyebben mutathatnak mérhető aktivitást biológiai tesztekben. Válasszuk meg ehhez egy flexibilis molekulákat tartalmazó tár tagjait úgy, hogy a kötődési erősségük különbsége jó közelítéssel csak a konfomációs eloszlásbeli különbségüknek legyen köszönhető. Vagyis a feltételezett, kötődéshez jelentősen hozzájáruló funkciós csoportokat nem változtatjuk, csak a közöttük meglévő flexibilis kapcsolatot. Ekkor tulajdonképpen a flexibilis molekulákkal letapogatjuk az ismeretlen térszerkezetű aktív helyet (scanning library). Ugyan a könyvtár molekuláinak kötési erőssége nem fog több

nagyságrendet fölölelni, és valószínüleg nem lesz közöttük nagyon erősen kötődő molekula sem, de ilyen speciális könyvtárra a dolgozatban ismertetett számítási eljárással megkaphatjuk a 3D farmakofort (**38. ábra**). A farmakofor alapján már végezhetünk vezérmolekula-optimálási lépéseket, a standard, 3D QSAR vagy 3D VS mószerekkel.

Fontos megjegyeznünk, hogy a farmakofort ilyen úton akkor is megkaphatjuk, ha nem áll rendelkezésre semmilyen *a priori* térszerkezeti információ a célfehérjéről. A hiányzó térszerkezeti információt ilyenkor más kísérleti adattal helyettesítjük. A megközelítés előnye, hogy egy, a kötőhelyre tervezett speciális, flexibilis molekulákból álló tár szintézise és a kötődési erősség meghatározása alternatív útvonalat kínál a röntgenkrisztallográfiai szerkezetmeghatározásra.

Előnyök	**Hátrányok**
A peptid alapú flexibilis ligandok esetén, a szintézis jól megalapozott, a molekulák gyorsan előállíthatók.	*Nem bitztos, hogy peptidek megfelelő ligandnak bizonyulnak. Egyáltalán az sem biztos, hogy megfelelő flexibilis ligand-jelölt található.*
Flexibilis ligand molekulák oldatbeli konformációs eloszlásának számítása elvégezhető feladat.	*Flexibilis ligand molekulák oldatbeli konformációs eloszlásának meghatározása komoly számításigényt támaszt.*
A vizsgált molekulák könnyen adhatnak mérhető kötődési aktivitás értéket.	*Mivel a molekulatárat úgy választjuk, hogy a kötődési erősségek különbsége lehetőleg csak a konformációs eloszlástól függjön, kötődés erőssége várhatóan sok nagyságrendet nem fog fölölelni, ezért szükséges azok pontos meghatározása.*
A peptidek általában nagyobb méretűek, ezért egy analógokból álló sorozat esetén valószínűsíthető, hogy ugyanaz lesz a kötődési mód. A ligand flexibilitása miatt feltételezhető, hogy a megfelelő kötődéskor nem ékelődnek be vízmolekulák.	*Viszonylag nagy hatástbeli különbséget kiváltó tulajdonságok (a szükséges csoportok megléte) konstans értéken tartása mellett próbáljuk a hatásért kisebb mértékben felelős tulajdonságot (kötődő csoportok térbeli helyzete) meghatározni.*

***14. táblázat**: A javasolt IQsar módszer előnyei és hátrányai.*

Peptidalapú molekulák megfelelően használhatók egy ilyen, speciális molekulakönyvtár kialakításakor. A peptideket és különböző analógjait viszonylag könnyen elő lehet állítani, akár párhuzamos szintézisekkel is, és köztük gyakran nagy affinitású szubsztrátokat vagy ligandokat is fel lehet fedezni. A következő lépés azonban, a peptidalapú vezérmolekulák kémiai átalakítása nem peptid (peptidmimetikum) szerkezetté messze nem triviális. Számos olyan molekulát találhatunk az irodalomban, amelyről azt feltételezték, hogy a peptidnek azt térszerkezetet utánozza, amelyik kölcsönhat a fehérjével, ugyanakkor aktív analógot mégsem lehetett belőle készíteni (sikeres példák közé tartoznak a benzodiazepinek, a neurokinin-1, -2 és szomatosztatin receptor ligandok és egyes HIV proteáz inhibitorok)[3]. A

jelenség egyik oka talán, hogy ha a kötőhely 3D szerkezete nem ismert, az igen flexibilis peptidek bioaktív konformációját és/vagy a farmakofort nagyon nehéz pontosan meghatározni. Ha azonban a bioaktív konformációt ismerjük, annak kémiai úton történő rögzítése jelentős aktivitásnövekedéssel járhat[134].

A kidolgozott módszer tehát olyan esetben szolgáltatja a vezérmolekula optimálásához szükséges farmakofort, amikor a kötőhely 3D szerkezetét nem ismerjük, viszont rendelkezünk egy speciális molekulatárra kísérletileg meghtározott kötődési aktivtások erősségével.

A. Függelék : r^2 személetes jelentésének levezetése

Levezethető, hogy az r^2 a „becslésből adódó” négyzetösszegek (estimated sum of squares) és a teljes négyzetösszegek (total sum of squares) hányadosa (59). A (54) összefüggést felhasználva kifejezhettük az illesztett egyenes x_i pontban felvett értékét (55).

$$A \equiv \left[\sum x^2 - n\bar{x}^2\right]^{-1} \tag{54}$$

$$\begin{aligned}&\hat{y}_i = a + bx_i = \bar{y} - b\bar{x} + bx_i = \bar{y} - b\left(\bar{x} + x_i\right) = \\ &A\left(\bar{y}\sum x^2 - \bar{x}\sum xy + x_i\sum xy - n\bar{x}\bar{y}x_i\right) = A\left(\bar{y}\sum x^2 + \left(x_i - \bar{x}\right)\sum xy - n\bar{x}\bar{y}x_i\right)\end{aligned} \tag{55}$$

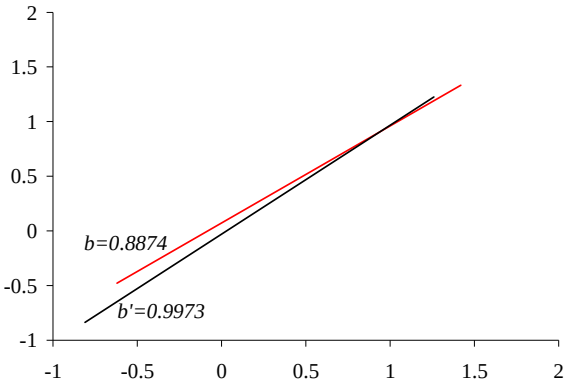
$$\sum \hat{y}_i = A\left(n\bar{y}\sum x^2 - n^2\bar{x}^2\bar{y}\right) \tag{56}$$

$$\begin{aligned}&\sum \hat{y}_i^2 = A^2\left[n\bar{y}^2\left(\sum x^2\right)^2 - n^2\bar{x}^2\bar{y}^2\left(\sum x^2\right) - 2n\bar{x}\bar{y}\left(\sum xy\right)\left(\sum x^2\right) + \right. \\ &\left. 2n^2\bar{x}^3\bar{y}\left(\sum xy\right) + \left(\sum x^2\right)\left(\sum xy\right)^2 - n\bar{x}^2\left(\sum xy\right)\right]\end{aligned} \tag{57}$$

Ekkor felírhatjuk a „becslésből adódó” négyzetösszeget (estimated sum of squares s_e) (58):

$$\begin{aligned}&s_e = \sum\left(\hat{y}_i - \bar{y}\right)^2 = \sum\left(\hat{y}_i^2 - 2\bar{y}\hat{y}_i + \bar{y}^2\right) = A^2\left(\sum xy - n\bar{x}\bar{y}\right)^2\left(\sum x^2 - n\bar{x}^2\right) = \\ &\frac{\left(\sum xy - n\bar{x}\bar{y}\right)^2}{\sum x^2 - n\bar{x}^2} = b \cdot s_{xy} = \frac{s_{xy}^2}{s_{xx}} = s_{yy}r^2\end{aligned} \tag{58}$$

$$r^2 = \frac{s_e}{s_{yy}} = \frac{\sum\left(\hat{y}_i - \bar{y}\right)^2}{\sum\left(y_i - \bar{y}\right)^2} \tag{59}$$



39. ábra *A b változó az egyenes meredekségét, b’ a felcserélt koordináták szerinti meredekséget jelöli.*

Az r^2 másik személetes jelentéséhez tekintsük az $x=a'+b'y$ egyenest. Ekkor a regressziós együttható (60) alakban írható fel.

$$b' = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum y_i^2 - \left(\sum y_i\right)^2} = \frac{s_{xy}}{s_{yy}} \tag{60}$$

A korrelációs együttható négyzete kifejezhető az eredeti és a „felcserélt koordinátájú” pontokra illesztett egyenes meredekségének szorzatával (61) (**39. ábra**).

$$r^2 = \frac{s_{xy}^2}{s_{xx} s_{yy}} = bb' \tag{61}$$

B. Függelék : Kiegészítés a relatív koncentráció számításához

A 7.2-es fejezetben a relatív koncentráció meghatározásakor az $[L]\approx[L_0]$ illetve a kompetitív kötődési vizsgálat esetén az $[I]\approx[I_0]$ közelítést használtam. Érdekesnek tartottam levezetni ezeket az összefüggéseket a közelítések alkalmazása nélkül.

A telítési kötődési vizsgálat esetén a (30)-as egyenlet (62)-es egyenletté alakul.

$$\frac{[\mathrm{l}_{\mathrm{bound}}]}{\mathrm{l}_{\mathrm{tot}}} = \frac{[LR]}{L_0} = \frac{[L][R]}{K_d([L]+[LR])} = \frac{B_{\max}}{2K_d + B_{\max}} \tag{62}$$

Kompetitív kötődési vizsgálat esetén az általában meghatározott IC_{50} érték az az inhibitor koncentráció, ahol a kötődő eredeti szubsztrát koncentrációja a felére csökken. Először tekintsük azt a szituációt, amikor az inhibitor még nincs a rendszerben, és határozzuk meg kísérletben használt bemérési koncentrációkhoz tartozó $[L^*R]_{\max}$ értéket. Az egyensúlyi állandó egyenlete (32) ekkor a jól ismert másodfokú kifejezéssé alakul. Ezt a kifejezést általában az $[L]\approx[L_0]$ feltételezéssel elsőfokúvá egyszerűsítik, azonban az $[L^*R]_{\max}<[L_0]$ és $[L^*R]_{\max}<[R_0]$ feltételek alkalmazásával csak az egyik gyököt fogadhatjuk el (63)(64).

$$[L^*R]_{\max} = \frac{R_0 + L_0 + K_d - \vartheta}{2} \tag{63}$$

$$\vartheta = \sqrt{(R_0 + L_0 + K_d)^2 - 4R_0L_0} \tag{64}$$

Amikor már inhibitort is adagolunk a rendszerhez, és eljutunk az $I_0=IC_{50}$ pontig a (32) és (33) kifejezés a következőképpen alakul.

$$K_d = \frac{\left(L_0 - \frac{[L^*R]_{\max}}{2}\right)\left(R_0 - \frac{[L^*R]_{\max}}{2} - [IR]\right)}{\frac{[L^*R]_{\max}}{2}} \tag{65}$$

$$K_i = \frac{(IC_{50} - [IR])\left(R_0 - \frac{[L^*R]_{\max}}{2} - [IR]\right)}{[IR]} \tag{66}$$

Ha a rendelkezésre álló adatsorunk IC_{50} értékekből áll, akkor (65)-be (63)-at helyettesítve $[IR]$-re meglehetősen hosszú összefüggést kapunk (67). Ez az összefüggés csak olyan tagokat tartalmaz, amelyeket egy molekulakönyvtár vizsgálatakor állandó értéken lehet tartani, így továbbra is jogos a 7.4-es fejezetben tett megállapítás (68).

$$[IR] = \frac{K_d^2 + 2K_dL_0 + L_0^2 + 2K_dR_0 + R_0^2 - (K_d + L_0 + R_0)\vartheta}{-2(K_d + 3L_0 + R_0) + 2\vartheta} \tag{67}$$

$$\frac{[\mathrm{l}_{\mathrm{bound}}]}{\mathrm{l}_{\mathrm{tot}}} = \frac{[IR]}{IC_{50}} \Rightarrow \frac{\{R_0, K_d, L_0\}}{IC_{50}} \tag{68}$$

Ha a rendelkezésre álló mért aktivitás adatsorunk K_i értékekből áll, akkor IC_{50}-et kell (66)-ból kifejezni és (67)-et felhasználva a (68)-ban felírt hányadosra az alábbi összefüggést kapjuk (69).

$$\frac{[IR]}{IC_{50}} = \frac{K_d(K_d + L_0 + R_0 - \vartheta)}{K_i(3L_0 - R_0 - K_d + \vartheta) + K_d(L_0 + R_0 - \vartheta) + K_d^2} \tag{69}$$

Nem meglepő módon a (69)-es kifejezésben ugyan szerepel a K_i^{-1}, de a konstans tag ebben az esetben, amikor közelítések nélkül számolunk, nem esik ki, mivel nevezőben összeg szerepel. Így, ha a relatív koncentrációt a K_i-ből számítjuk, tudatában kell lennünk, hogy a 7.4. fejezetben felállított összefüggés a közelítésekből adódóan hibával terhelt.

C. Függelék : Atomi pozíciók alapján számított konformációs távolság

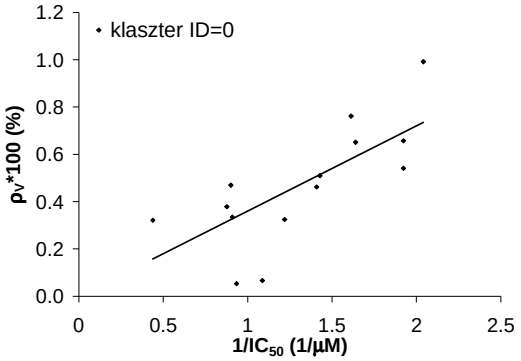
A horgonykonformációkat gyűjtésekor az Algoritmus 2-t használtam, τ_{AC}=3 Å, x_τ=10% N_{iter}=5, x_q=20%, τ_d=3 Å és N_{rand}=20 paraméterek használata mellett. A q^2 számításakor a zérus *y* tengelymetszetű egyenest alkalmaztam. Egy horgonykonformációhoz rendelt sűrűségeket az (49) és (51) szerint számoltam (cut off).

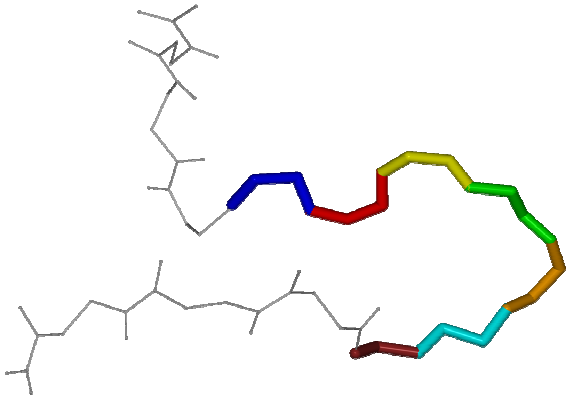
A beállítások szerint az algoritmusnak húsztor kellett volna végigmenni a randomizált trajektórián és minden esetben öt iterációt végezni. Ilyen beállítások mellett egy torziószög alapú számítás általában másfél nap alatt lefutott. Descartes-koordináta alapú konformációs távolság mértékkel összesen egyetlen véletlenszerűen átsorrendezett trajektórián négy iterációs ciklus elvégzése közel két hétig tartott. Az alábbi kiértékelés a részeredményeken történt, csak a teljesség kedvéért került a dolgozatba.

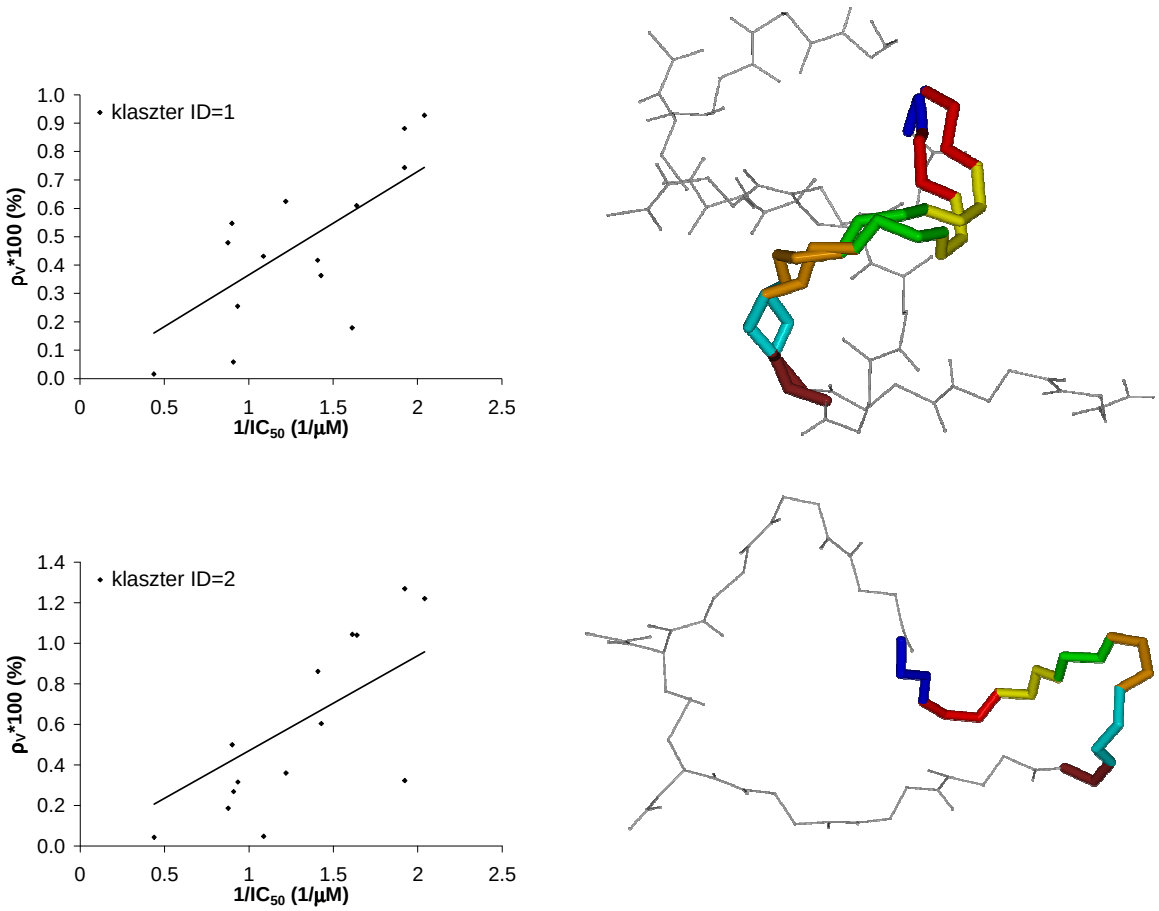
A négy iteráció során összesen 8952 darab uAC-t találtam, amelyet κ_{LOOq}=0,4 és $\kappa_{L50\%Oq}$=0,4 paraméterek, és zérus sűrűségkritérium használatával korreláció és konformációs hasonlóság szerint klaszteresztem. Ekkor három klasztert kaptam (**15. táblázat**, **40. ábra**).

Klaszter		Klaszter legjobb tagja				
		korreláció			regresszió	
ID	N_V	r	LOOq^2	L50%Oq^2	A	B
0	2	0,74	0,47	0,51	0	0,36
1	2	0,74	0,45	0,49	0	0,36
2	1	0,78	0,43	0,48	0	0,44

***15. táblázat**:A talált klaszterek.(κ_{LOOq}=0,4 és a $\kappa_{L50\%Oq}$=0,4, zérus sűrűségkritérium használata mellett).*



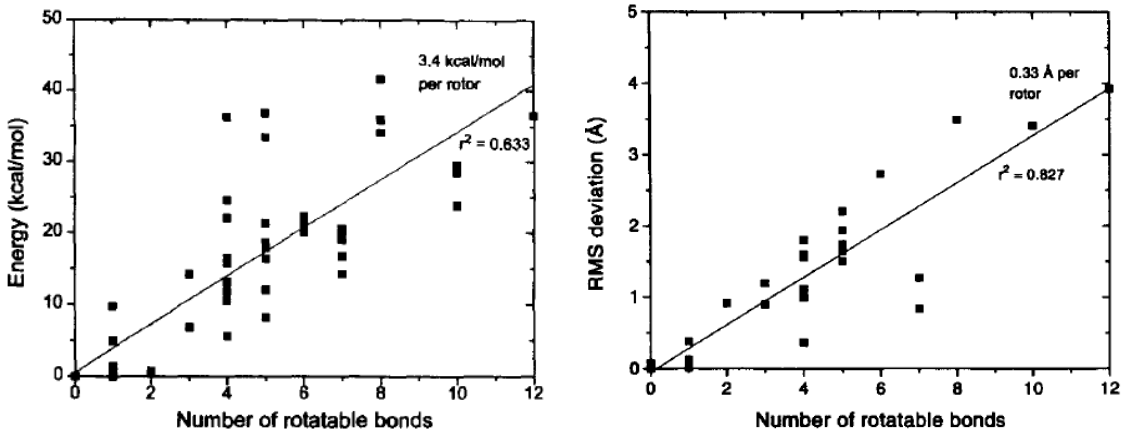




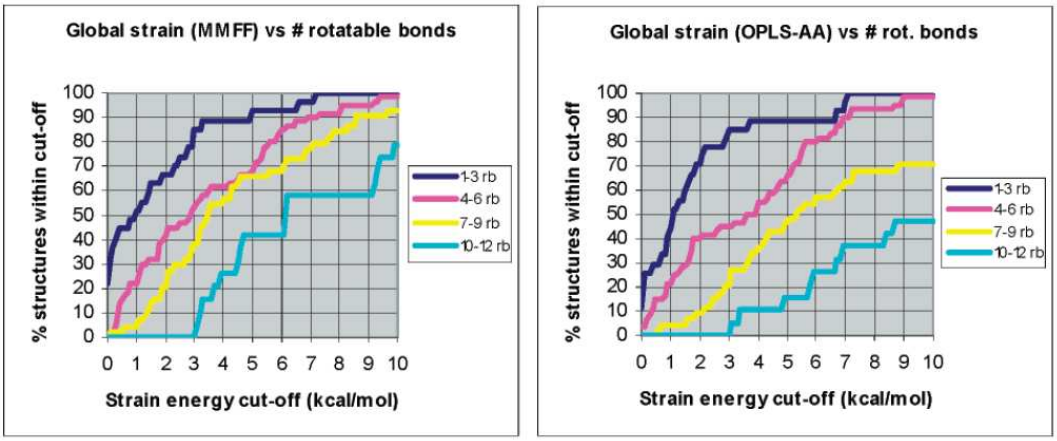
40. ábra *Baloldalon: A klaszterek legjobb elemének konformációs sűrűsége a kötődési aktivitás függvényében. Jobboldalon: A klaszterek horgonykonformációi.*

D. Függelék : Átemelt ábrák

Kiegészítés a 5.2. fejezethez



41. ábra *Nicklaus, 1995: A bioaktív konformáció és a talált globális energiaminimumnak megfelelő konformáció közötti energia (jobboldal) és konformációs különbség (baloldal) a 33 ligand szabadon forgatható kötéseinek függvényében*[20].

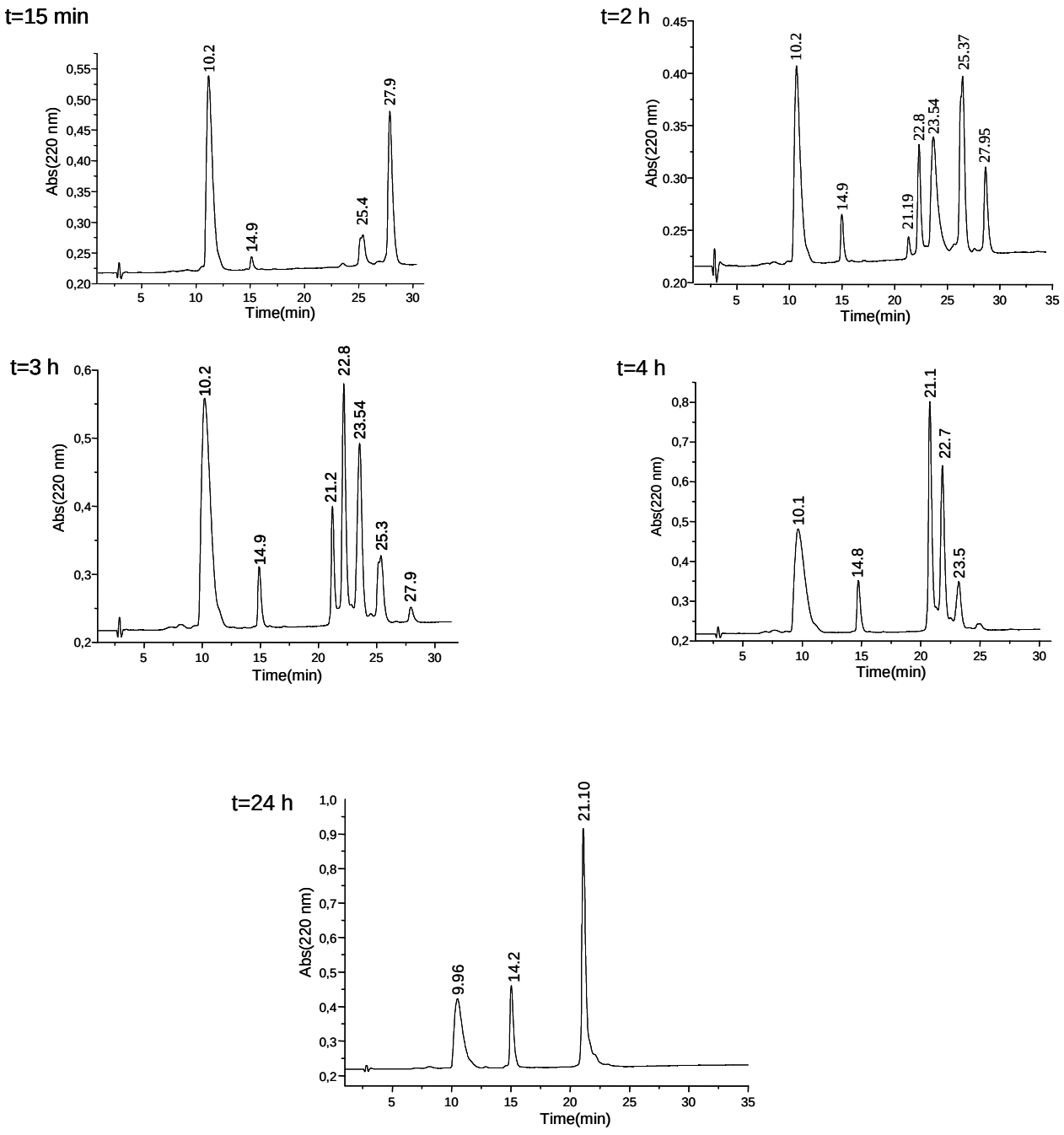


42. ábra *Perola, 2003: A kötődés során bekövetkező konformációs energiaváltozást a bioaktív konformáció és a talált legalacsonyabb energiájú konformer energiakülönbségével jellemezték (strain energy). A kapott eredményeket úgy ábrázolták, hogy a 150 ligand molekulát felosztották 5 csoportra a szabadon forgatható kötések száma szerint (1-3, 4-6, 7-9, 10-12), és feltüntették, hogy az egyes csoportok hány százaléka esik egy bizonyos konformációs feszültség-energiatag küszöbértéke alá*[21].

Kiegészítés a 9.1.5. fejezethez

A kromatogrammon 10,2 percnél az epitóp, 14,9 percnél az epitóp dimerjének, 27.9 percnél a szabad hordozónak a csúcsát találjuk. 25,4 percnél figyelhetjük meg azt a csúcsot, amelyik egy epitópot, 23,54 percnél, amelyik kettőt, 22,8 percnél, amelyik hármat, és 21,2 percnél amelyik négyet tartalmaz. A négy epitópot tartalmazó konjugátum csúcsa a legkeskenyebb, mivel ekkor már szerkezetileg egységes a konjugátum, a másik három csúcs

ennél szélesebb, ami bizonyítja azt, hogy a konjugálási reakciók párhuzamosan játszódnak le és a különböző szerkezetű konjugátumok keveréke képződik.



43. ábra *A konjugálási reakciót nyomon kísérő HPLC kromatogrammok (az ábrát Marilena Manea angol nyelvű doktori dolgozatából vettem át) .*

12. Irodalomjegyzék

1. Schachter, A. D.; Ramoni, M. F. From the analyst's couch - Clinical forecasting in drug development. *Nature Reviews Drug Discovery* **2007,** *6*, 107-108.

2. Rostami-Hodjegan, A.; Tucker, G. T. Simulation and prediction of in vivo drug metabolism in human populations from in vitro data. *Nature Reviews Drug Discovery* **2007,** *6*, 140-148.

3. Kubinyi, H. In search for new leads. *EFMC Yearbook* **2003,** 14-28.

4. Mestres, J.; Knegtel, R. M. A. Similarity versus docking in 3D virtual screening. *Perspectives in Drug Discovery and Design* **2000,** *20*, 191-207.

5. Sheridan, R. P.; Kearsley, S. K. Why do we need so many chemical similarity search methods? *Drug Discovery Today* **2002,** *7*, 903-911.

6. Mason, J. S.; Good, A. C.; Martin, E. J. 3-D Pharmacophores in drug discovery. *Current Pharmaceutical Design* **2001,** *7*, 567-597.

7. Ehrlich, P. On the present state of chemotherapy. *Berichte Der Deutschen Chemischen Gesellschaft* **1909,** *42*, 17.

8. Humblet, C. *Annu. Rep. Med. Chem.* **1980,** *15*, 267.

9. Abrahamian, E.; Fox, P. C.; Naerum, L.; Christensen, I. T.; Thogersen, H.; Clark, R. D. Efficient generation, storage, and manipulation of fully flexible pharmacophore multiplets and their use in 3-D similarity searching. *Journal of Chemical Information and Computer Sciences* **2003,** *43*, 458-468.

10. Debnath, A. K. Generation of predictive pharmacophore models for CCR5 antagonists: Study with piperidine- and piperazine-based compounds as a new class of HIV-1 entry inhibitors. *J. Med. Chem.* **2003,** *46*, 4501-4515.

11. Matter, H.; Potter, T. Comparing 3D pharmacophore triplets and 2D fingerprints for selecting diverse compound subsets. *Journal of Chemical Information and Computer Sciences* **1999,** *39*, 1211-1225.

12. Melani, F.; Gratteri, P.; Adamo, M.; Bonaccini, C. Field interaction and geometrical overlap: A new simplex and experimental design based computational procedure for superposing small ligand molecules. *Journal of Medicinal Chemistry* **2003,** *46*, 1359-1371.

13. Lemmen, C.; Lengauer, T. Computational methods for the structural alignment of molecules. *Journal of Computer-Aided Molecular Design* **2000,** *14*, 215-232.

14. Mayer, D. *J Comput Aid Mol Des* **1987,** *1*, 3.

15. Martin, Y. C.; Bures, M. G.; Danaher, E. A.; Delazzer, J.; Lico, I.; Pavlik, P. A. A Fast New Approach to Pharmacophore Mapping and Its Application to

Dopaminergic and Benzodiazepine Agonists. *Journal of Computer-Aided Molecular Design* **1993,** *7*, 83-102.

16. Accelrys. Catalyst scientific citations. http://www.accelrys.com/products/catalyst/refs.html . 2007. Ref Type: Electronic Citation

17. Kurogi, Y.; Guner, O. F. Pharmacophore modeling and three-dimensional database searching for drug design using catalyst. *Current Medicinal Chemistry* **2001,** *8*, 1035-1055.

18. Guner, O.; Clement, O.; Kurogi, Y. Pharmacophore modeling and three dimensional database searching for drug design using catalyst: Recent advances. *Current Medicinal Chemistry* **2004,** *11*, 2991-3005.

19. Chen, Q.; Higgs, R. E.; Vieth, M. Geometric accuracy of three-dimensional molecular overlays. *Journal of Chemical Information and Modeling* **2006,** *46*, 1996-2002.

20. Nicklaus, M. C.; Wang, S. M.; Driscoll, J. S.; Milne, G. W. A. Conformational-Changes of Small Molecules Binding to Proteins. *Bioorganic & Medicinal Chemistry* **1995,** *3*, 411-428.

21. Perola, E.; Charifson, P. S. Conformational analysis of drug-like molecules bound to proteins: An extensive study of ligand reorganization upon binding. *Journal of Medicinal Chemistry* **2004,** *47*, 2499-2510.

22. Markus, J. L.; Kolossvary, I.; Aszodi, A. Analyzing the performance of conformational search programs on compound databases. *Journal of Molecular Graphics and Modelling* **2007,** *25*, 700-710.

23. Hansch, C. Correlation of biological activity of plant growth regulators and chloromycetin derivatives with hammett constants and partition coefficients. *Journal of the American Chemical Society* **1963,** *85*, 2817.

24. Hall, L. H.; Kier, L. B. In *Reviews in Computational Chemistry II*; Lipkowitz, K. B.; Boyd, D. B. Eds.; VCH Publishers: 1991; pp 367-422.

25. Tripos Inc. Hologram QSAR. http://www.tripos.com . 2007. Ref Type: Electronic Citation

26. Wold, S.; Sjostrom, M.; Eriksson, L. PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems* **2001,** *58*, 109-130.

27. Wold, S.; Trygg, J.; Berglund, A.; Antti, H. Some recent developments in PLS modeling. *Chemometrics and Intelligent Laboratory Systems* **2001,** *58*, 131-150.

28. Cramer, R. D.; Patterson, D. E.; Bunce, J. D. Comparative Molecular-Field Analysis (Comfa) .1. Effect of Shape on Binding of Steroids to Carrier Proteins. *Journal of the American Chemical Society* **1988,** *110*, 5959-5967.

29. Klebe, G.; Abraham, U.; Mietzner, T. Molecular Similarity Indexes in A Comparative-Analysis (Comsia) of Drug Molecules to Correlate and Predict Their Biological-Activity. *Journal of Medicinal Chemistry* **1994,** *37*, 4130-4146.

30. Kubinyi, H.; Hamprecht, F. A.; Mietzner, T. Three-dimensional quantitative similarity-activity relationships (3D QSiAR) from SEAL similarity matrices. *J. Med. Chem.* **1998,** *41*, 2553-2564.

31. Kearsley, S. K. An Algorithm for the Simultaneous Superposition of A Structural Series. *Journal of Computational Chemistry* **1990,** *11*, 1187-1192.

32. Borosy, AP.; Héberger, K. H. Gy.; Kolossváry, I.; Lengyel, A.; PaksyL.; Rajkó, R.; Szepesvári, P. *Sokváltozós adatelemzés (kemometria)*; Nemzeti Tankönyvkiadó, Budapest: 2001.

33. Gasteiger, J.; Zupan, J. Neural Networks in Chemistry. *Angewandte Chemie-International Edition in English* **1993,** *32*, 503-527.

34. Baroni, M.; Costantino, G.; Cruciani, G.; Riganelli, D.; Valigi, R.; Clementi, S. Generating Optimal Linear Pls Estimations (Golpe) - An Advanced Chemometric Tool for Handling 3D-Qsar Problems. *Quantitative Structure-Activity Relationships* **1993,** *12*, 9-20.

35. Cho, S. J.; Tropsha, A. Cross-Validated R(2)-Guided Region Selection for Comparative Molecular-Field Analysis - A Simple Method to Achieve Consistent Results. *J. Med. Chem.* **1995,** *38*, 1060-1066.

36. Ajmani, S.; Jadhav, K.; Kulkarni, S. A. Three-dimensional QSAR using the k-nearest neighbor method and its interpretation. *J. Chem. Inf. Model.* **2006,** *46*, 24-31.

37. Kim, K. H.; Greco, G.; Novellino, E. A critical review of recent CoMFA applications. *Perspectives in Drug Discovery and Design* **1998,** *12*, 257-315.

38. Martin, Y. C. 3D QSAR: Current state, scope, and limitations. *Perspectives in Drug Discovery and Design* **1998,** *12*, 3-23.

39. Sutherland, J. J.; O'Brien, L. A.; Weaver, D. F. A comparison of methods for modeling quantitative structure-activity relationships. *Journal of Medicinal Chemistry* **2004,** *47*, 5541-5554.

40. Bravi, G.; Gancia, E.; Mascagni, P.; Pegna, M.; Todeschini, R.; Zaliani, A. MS-WHIM, new 3D theoretical descriptors derived from molecular surface properties: A comparative 3D QSAR study in a series of steroids. *Journal of Computer-Aided Molecular Design* **1997,** *11*, 79-92.

41. Pastor, M.; Cruciani, G.; Mclay, I.; Pickett, S.; Clementi, S. GRid-INdependent descriptors (GRIND): A novel class of alignment-independent three-dimensional molecular descriptors. *Journal of Medicinal Chemistry* **2000,** *43*, 3233-3243.

42. Silverman, B. D.; Platt, D. E. Comparative molecular moment analysis (CoMMA): 3D-QSAR without molecular superposition. *Journal of Medicinal Chemistry* **1996,** *39*, 2129-2140.

43. Silverman, D. B. The thirty-one benchmark steroids revisited: Comparative molecular moment analysis (CoMMA) with principal component regression. *Quantitative Structure-Activity Relationships* **2000,** *19*, 237-246.

44. Tuppurainen, K.; Viisas, M.; Laatikainen, R.; Perakyla, M. Evaluation of a novel electronic eigenvalue (EEVA) molecular descriptor for QSAR/QSPR studies: Validation using a benchmark steroid data set. *Journal of Chemical Information and Computer Sciences* **2002,** *42*, 607-613.

45. Wagener, M.; Sadowski, J.; Gasteiger, J. Autocorrelation of Molecular-Surface Properties for Modeling Corticosteroid-Binding Globulin and Cytosolic Ah Receptor Activity by Neural Networks. *Journal of the American Chemical Society* **1995,** *117*, 7769-7775.

46. Hopfinger, A. J.; Wang, S.; Tokarski, J. S.; Jin, B. Q.; Albuquerque, M.; Madhav, P. J.; Duraiswami, C. Construction of 3D-QSAR models using the 4D-QSAR analysis formalism. *Journal of the American Chemical Society* **1997,** *119*, 10509-10524.

47. Streich, D.; Neuburger-Zehnder, M.; Vedani, A. Induced fit - The key for understanding LSD activity? A 4D-QSAR study on the 5-HT2A receptor system. *Quantitative Structure-Activity Relationships* **2001,** *19*, 565-573.

48. Vedani, A.; Dobler, M.; Zbinden, P. Quasi-atomistic receptor surface models: A bridge between 3-D QSAR and receptor modeling. *Journal of the American Chemical Society* **1998,** *120*, 4471-4477.

49. Vedani, A.; Briem, K.; Dobler, M.; Dollinger, H.; McMasters, D. R. Multiple-conformation and protonation-state representation in 4D-QSAR: The neurokinin-1 receptor system. *J. Med. Chem.* **2000,** *43*, 4416-4427.

50. Vedani, A.; McMasters, D. R.; Dobler, M. Multi-conformational ligand representation in 4D-QSAR: Reducing the bias associated with ligand alignment. *Quantitative Structure-Activity Relationships* **2000,** *19*, 149-161.

51. Vedani, A.; Dobler, M. 5D-QSAR: The key for simulating induced fit? *J. Med. Chem.* **2002,** *45*, 2139-2149.

52. Martinek, T. A.; Otvos, F.; Dervarics, M.; Toth, G.; Fulop, F. Ligand-based prediction of active conformation by 3D-QSAR flexibility descriptors and their application in 3+3D-QSAR models. *Journal of Medicinal Chemistry* **2005,** *48*, 3239-3250.

53. Dervarics, M.; Otvos, F.; Martinek, T. A. Development of a chirality-sensitive flexibility descriptor for 3+3D-QSAR. *Journal of Chemical Information and Modeling* **2006,** *46*, 1431-1438.

54. Bernard, D.; Coop, A.; Mackerell, A. D. Conformationally sampled pharmacophore for peptidic delta opioid ligands. *Journal of Medicinal Chemistry* **2005,** *48*, 7773-7780.

55. Bernard, D.; Coop, A.; Mackerell, A. D. 2D conformationally sampled pharmacophore: A ligand-based pharmacophore to differentiate delta opioid agonists from antagonists. *Journal of the American Chemical Society* **2003,** *125*, 3101-3107.

56. Golbraikh, A.; Tropsha, A. Beware of q(2)! *Journal of Molecular Graphics & Modelling* **2002,** *20*, 269-276.

57. Wold, S.; Erikkson, L. Statistical validation of QSAR results. In *Chemometrics Methods in Molecular Design*; van de Waterbeemd, H. Ed.; VCH: Weinheim, 1995; pp 309-318.

58. Wold, S. Validation of Qsars. *Quantitative Structure-Activity Relationships* **1991,** *10*, 191-193.

59. Hulzebos, E.; Sijm, D.; Traas, T.; Posthumus, R.; Maslankiewicz, L. Validity and validation of expert (Q)SAR systems. *Sar and Qsar in Environmental Research* **2005,** *16*, 385-401.

60. Eriksson, L.; Jaworska, J.; Worth, A. P.; Cronin, M. T. D.; McDowell, R. M.; Gramatica, P. Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs. *Environmental Health Perspectives* **2003,** *111*, 1361-1375.

61. Leach, A. *Molecular modelling, Principles and applications*; Pearson Education Limited: Harlow, 2001.

62. Wehrens, R.; Putter, H.; Buydens, L. M. C. The bootstrap: a tutorial. *Chemometrics and Intelligent Laboratory Systems* **2000,** *54*, 35-52.

63. Gedeck, P.; Rohde, B.; Bartels, C. QSAR - How good is it in practice? Comparison of descriptor sets on an unbiased cross section of corporate data sets. *Journal of Chemical Information and Modeling* **2006,** *46*, 1924-1936.

64. Kubinyi, H.; Hamprecht, F. A.; Mietzner, T. Three-dimensional quantitative similarity-activity relationships (3D QSiAR) from SEAL similarity matrices. *Journal of Medicinal Chemistry* **1998,** *41*, 2553-2564.

65. Norinder, U. Single and domain mode variable selection in 3D QSAR applications. *Journal of Chemometrics* **1996,** *10*, 95-105.

66. van Drie, J. H. Pharmacophore discovery - Lessons learned. *Current Pharmaceutical Design* **2003,** *9*, 1649-1664.

67. Todeschini, R.; Consonni, V.; Maiocchi, A. The K correlation index: theory development and its application in chemometrics. *Chemometrics and Intelligent Laboratory Systems* **1999,** *46*, 13-29.

68. Hawkins, D. M. The problem of overfitting. *Journal of Chemical Information and Computer Sciences* **2004,** *44*, 1-12.

69. Golbraikh, A.; Tropsha, A. Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection. *Journal of Computer-Aided Molecular Design* **2002,** *16*, 357-369.

70. Golbraikh, A.; Shen, M.; Xiao, Z. Y.; Xiao, Y. D.; Lee, K. H.; Tropsha, A. Rational selection of training and test sets for the development of validated QSAR models. *Journal of Computer-Aided Molecular Design* **2003,** *17*, 241-253.

71. Marengo, E.; Todeschini, R. A New Algorithm for Optimal, Distance-Based Experimental-Design. *Chemometrics and Intelligent Laboratory Systems* **1992,** *16*, 37-44.

72. Tropsha, A.; Gramatica, P.; Gombar, V. K. The importance of being earnest: Validation is the absolute essential for successful application and interpretation of QSPR models. *Qsar & Combinatorial Science* **2003,** *22*, 69-77.

73. Doweyko, A. M. 3D-QSAR illusions. *Journal of Computer-Aided Molecular Design* **2004,** *18*, 587-596.

74. Janssen, L. H. M. Conformational flexibility and receptor interaction. *Bioorganic & Medicinal Chemistry* **1998,** *6*, 785-788.

75. Motulsky, H.; Christopoulos, A. PRISM: Fitting Models to Biological Data using Linear and Nonlinear Regression. 2003.
Ref Type: Generic

76. Yung-Chi Cheng; Prusoff, W. H. Relationship between the inhibition constant (Ki) and the concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic reaction. *Biochemical Pharmacology* **1973,** *22*, 3099-3108.

77. Vedani, A.; Dobler, M. Multidimensional QSAR: Moving from three- to five-dimensional concepts. *Quantitative Structure-Activity Relationships* **2002,** *21*, 382-390.

78. Bostrom, J.; Hogner, A.; Schmitt, S. Do Structurally Similar Ligands Bind in a Similar Fashion? *J. Med. Chem.* **2006,** *49*, 6716-6725.

79. Murray, C. W.; Verdonk, M. L. The consequences of translational and rotational entropy lost by small molecules on binding to proteins. *Journal of Computer-Aided Molecular Design* **2002,** *16*, 741-753.

80. Kalászi, A.; Farkas, O. Lead conformer prediction based on a library of flexible molecules. *Journal of Molecular Structure-Theochem* **2003,** *666*, 645-649.

81. Kalaszi, A.; Farkas, O. Predicting lead compounds using libraries of flexible molecules. *Biopolymers* **2003,** *71*, 331.

82. Kalaszi, A.; Imre, G.; Jakli, I.; Farkas, O. Conformational flexibility and 3D QSAR. *in preparation* **2007**.

83. Kalaszi, A.; Imre, G.; Jakli, I.; Farkas, O. Theoretical basis and a computational tool to identify binding conformations of flexible ligands. *European Journal of Medicinal Chemistry - submitted* **2007**.

84. Manea, M.; Kalaszi, A.; Mezo, G.; Horvati, K.; Horvath, A.; Farkas, O.; Przybylski, M.; Hudecz, F. Antibody recognition and conformational flexibility of a plaque-specific β-amyloid epitope modulated by non-native peptide flanking regions. *Journal of Medical Chemistry - submitted* **2007**.

85. Mezo, G.; Kalaszi, A.; Remenyi, J.; Majer, Z.; Hilbert, A.; Lang, O.; Kohidai, L.; Barna, K.; Gaal, D.; Hudecz, F. Synthesis, conformation, and immunoreactivity of new carrier molecules based on repeated tuftsin-like sequence. *Biopolymers* **2004,** *73*, 645-656.

86. Hamilton, W. R. On a new species of imaginary quantities connected with a theory of quaternions. *Proceedings of the Royal Irish Academy* **1844,** *2*, 424-434.

87. Flower, D. R. Rotational superposition: A review of methods. *Journal of Molecular Graphics & Modelling* **1999,** *17*, 238-244.

88. Karney, C. F. F. Quaternions in molecular modeling. *Technical report, Sarnoff Corp.* , **2005**.

89. Chen, Q.; Higgs, R. E.; Vieth, M. Geometric Accuracy of Three-Dimensional Molecular Overlays. J.Chem.Inf.Model. 7-11-0006.
Ref Type: In Press

90. Clark, M.; Cramer, R. D. The Probability of Chance Correlation Using Partial Least-Squares (Pls). *Quantitative Structure-Activity Relationships* **1993,** *12*, 137-145.

91. Barnard, J. M.; Downs, G. M. Chemical fragment generation and clustering software. *Journal of Chemical Information and Computer Sciences* **1997,** *37*, 141-142.

92. Brown, R. D.; Martin, Y. C. Use of structure Activity data to compare structure-based clustering methods and descriptors for use in compound selection. *Journal of Chemical Information and Computer Sciences* **1996,** *36*, 572-584.

93. Knuth, D. Knuth shuffle. In *The Art of Computer Programming* 1968.

94. ChemAxon MarvinBeans 4.1.0. *http://www. chemaxon. com* **2007**.

95. Duncan, R. Drug Polymer Conjugates - Potential for Improved Chemotherapy. *Anti-Cancer Drugs* **1992,** *3*, 175-210.

96. Ivanov, B. B.; Meshcheryakova, E. A.; Andronova, T. M.; Ivanov, V. T. Synthetic Carriers and Adjuvants Increase Immunogenicity of the Synthetic Peptide, B-Epitope from Cs-Protein of Plasmodium-Falciparum. *Bioorganicheskaya Khimiya* **1991,** *17*, 732-746.

97. Trudelle, Y.; Brack, A.; Delmas, A.; Pedoussaut, S.; Rivaille, P. Synthesis of A New Carrier for Immunization - Polytuftsin - 2 Examples of Its Use with Peptides

Selected in the Hepatitis-B Surface-Antigen. *International Journal of Peptide and Protein Research* **1987,** *30*, 54-60.

98. Fridkin, M.; Najjar, V. A. Tuftsin - Its Chemistry, Biology, and Clinical Potential. *Critical Reviews in Biochemistry and Molecular Biology* **1989,** *24*, 1-40.

99. Delmas, A.; Brack, A.; Trudelle, Y. Studies of the Influence of Different Cross-Linking Reagents on the Immune-Response Against A B-Epitope. *Bioconjugate Chemistry* **1992,** *3*, 80-84.

100. Gokulan, K.; Khare, S.; Rao, D. N. Increase in the immunogenicity of HIV peptide antigens by chemical linkage to polytuftsin (TKPR40). *Dna and Cell Biology* **1999,** *18*, 623-630.

101. Suchardasobczyk, A.; Siemion, I. Z.; Konopinska, D. Tuftsin and Its Analogs .8. Infrared Spectroscopic Investigations of Tuftsin and Its Analogs. *European Journal of Biochemistry* **1979,** *96*, 131-139.

102. Blumenstein, M.; Layne, P. P.; Najjar, V. A. Nuclear Magnetic-Resonance Studies on the Structure of the Tetrapeptide Tuftsin, L-Threonyl-L-Lysyl-L-Prolyl-L-Arginine, and Its Pentapeptide Analog L-Threonyl-L-Lysyl-L-Prolyl-L-Prolyl-L-Arginine. *Biochemistry* **1979,** *18*, 5247-5253.

103. D'Ursi, A.; Albrizio, S.; Fattorusso, C.; Lavecchia, A.; Zanotti, G.; Temussi, P. A. Solution conformation of a potent cyclic analogue of tuftsin: Low-temperature nuclear magnetic resonance study in a cryoprotective mixture. *Journal of Medicinal Chemistry* **1999,** *42*, 1705-1713.

104. Valdeavella, C. V.; Blatt, H. D.; Pettitt, B. M. Simulations of Conformers of Tuftsin and A Cyclic Tuftsin Analog. *International Journal of Peptide and Protein Research* **1995,** *46*, 372-380.

105. Ashish; Kishore, R. Folded conformation of an immunostimulating tetrapeptide rigin: High temperature molecular dynamics simulation study. *Bioorganic & Medicinal Chemistry* **2002,** *10*, 4083-4090.

106. Berendsen, H. J. C.; Vanderspoel, D.; Vandrunen, R. Gromacs - A Message-Passing Parallel Molecular-Dynamics Implementation. *Computer Physics Communications* **1995,** *91*, 43-56.

107. Lindahl, E.; Hess, B.; van der Spoel, D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *Journal of Molecular Modeling* **2001,** *7*, 306-317.

108. van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry* **2005,** *26*, 1701-1718.

109. Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. Molecular-Dynamics with Coupling to An External Bath. *Journal of Chemical Physics* **1984,** *81*, 3684-3690.

110. Sanner, M. F.; Olson, A. J.; Spehner, J. C. Reduced surface: An efficient way to compute molecular surfaces. *Biopolymers* **1996,** *38*, 305-320.

111. Upson, C.; Faulhaber, T.; Kamins, D.; Laidlaw, D.; Schlegel, D.; Vroom, J.; Gurwitz, R.; Vandam, A. The Application Visualization System - A Computational Environment for Scientific Visualization. *Ieee Computer Graphics and Applications* **1989,** *9*, 30-42.

112. Manea, M. Design, Structural and Immuno-analytical Properties of Antigenic Polypeptides comprising a β-Amyloid-Plaque Specific Epitope. 2006. Ref Type: Thesis/Dissertation

113. Kim, Y. S.; Gum, J. R.; Byrd, J. C.; Toribara, N. W. The Structure of Human Intestinal Apomucins. *American Review of Respiratory Disease* **1991,** *144*, S10-S14.

114. Price, M. R.; Sekowski, M.; Yang, G. Y.; Durrant, L. G.; Robins, R. A.; Baldwin, R. W. Reactivity of An Anti-(Human Gastric-Carcinoma) Monoclonal-Antibody with Core-Related Peptides of Gastrointestinal Mucin. *Cancer Immunology Immunotherapy* **1991,** *33*, 80-84.

115. Uray, K.; Price, M. R.; Majer, Z.; Vass, E.; Hollosi, M.; Hudecz, F. Identification and solution conformation of multiple epitopes recognized by a MUC2 mucin-specific monoclonal antibody. *Archives of Biochemistry and Biophysics* **2003,** *410*, 254-260.

116. Windberg, E.; Uray, K.; Illyes, E.; Skribanek, Z.; Price, M. R.; Sebestyen, F.; Hudecz, F. Heteroclitic recognition of combinatorial (TXTXT)-T-1-T-2 peptide mixtures by mucin-2 protein specific monoclonal antibody. *Journal of Peptide Science* **2004,** *10*, 56-65.

117. Weiner, H. L.; Lemere, C. A.; Maron, R.; Spooner, E. T.; Grenfell, T. J.; Mori, C.; Issazdeh, S.; Hancock, W. W.; Selkoe, D. J. Nasal administration of amyloid-beta peptide decreases cerebral amyloid burden in a mouse model of Alzheimer's disease. *Annals of Neurology* **2000,** *48*, 567-579.

118. Schenk, D.; Barbour, R.; Dunn, W.; Gordon, G.; Grajeda, H.; Guido, T.; Hu, K.; Huang, J. P.; Johnson-Wood, K.; Khan, K.; Kholodenko, D.; Lee, M.; Liao, Z. M.; Lieberburg, I.; Motter, R.; Mutter, L.; Soriano, F.; Shopp, G.; Vasquez, N.; Vandevert, C.; Walker, S.; Wogulis, M.; Yednock, T.; Games, D.; Seubert, P. Immunization with amyloid-beta attenuates Alzheimer disease-like pathology in the PDAPP mouse. *Nature* **1999,** *400*, 173-177.

119. Janus, C.; Pearson, J.; McLaurin, J.; Mathews, P. M.; Jiang, Y.; Schmidt, S. D.; Chishti, M. A.; Horne, P.; Heslin, D.; French, J.; Mount, H. T. J.; Nixon, R. A.; Mercken, M.; Bergeron, C.; Fraser, P. E.; St George-Hyslop, P.; Westaway, D. A beta peptide immunization reduces behavioural impairment and plaques in a model of Alzheimer's disease. *Nature* **2000,** *408*, 979-982.

120. Bard, F.; Cannon, C.; Barbour, R.; Burke, R. L.; Games, D.; Grajeda, H.; Guido, T.; Hu, K.; Huang, J. P.; Johnson-Wood, K.; Khan, K.; Kholodenko, D.; Lee, M.; Lieberburg, I.; Motter, R.; Nguyen, M.; Soriano, F.; Vasquez, N.; Weiss, K.;

Welch, B.; Seubert, P.; Schenk, D.; Yednock, T. Peripherally administered antibodies against amyloid beta-peptide enter the central nervous system and reduce pathology in a mouse model of Alzheimer disease. *Nature Medicine* **2000,** *6*, 916-919.

121. DeMattos, R. B.; Bales, K. R.; Cummins, D. J.; Dodart, J. C.; Paul, S. M.; Holtzman, D. M. Peripheral anti-A beta antibody alters CNS and plasma A beta clearance and decreases brain A beta burden in a mouse model of Alzheimer's disease. *Proceedings of the National Academy of Sciences of the United States of America* **2001,** *98*, 8850-8855.

122. Morgan, D.; Diamond, D. M.; Gottschall, P. E.; Ugen, K. E.; Dickey, C.; Hardy, J.; Duff, K.; Jantzen, P.; DiCarlo, G.; Wilcock, D.; Connor, K.; Hatcher, J.; Hope, C.; Gordon, M.; Arendash, G. W. A beta peptide vaccination prevents memory loss in an animal model of Alzheimer's disease. *Nature* **2000,** *408*, 982-985.

123. McLaurin, J.; Cecal, R.; Kierstead, M. E.; Tian, X.; Phinney, A. L.; Manea, M.; French, J. E.; Lambermon, M. H. L.; Darabie, A. A.; Brown, M. E.; Janus, C.; Chishti, M. A.; Horne, P.; Westaway, D.; Fraser, P. E.; Mount, H. T. J.; Przybylski, M.; St George-Hyslop, P. Therapeutically effective antibodies against amyloid-beta peptide target amyloid-beta residues 4-10 and inhibit cytotoxicity and fibrillogenesis. *Nature Medicine* **2002,** *8*, 1263-1269.

124. Drakopoulou, E.; Uray, K.; Mezo, G.; Price, M. R.; Vita, C.; Hudecz, F. Synthesis and antibody recognition of mucin 1 (MUC1)-alpha-conotoxin chimera. *Journal of Peptide Science* **2000,** *6*, 175-185.

125. Hudecz, F. Manipulation of epitope function by modification of peptide structure: A minireview. *Biologicals* **2001,** *29*, 197-207.

126. Mezo, G.; Drakopoulou, E.; Paal, V.; Rajnavolgyi, E.; Vita, C.; Hudecz, F. Synthesis and immunological studies of alpha-conotoxin chimera containing an immunodominant epitope from the 268-284 region of HSV gD protein. *Journal of Peptide Research* **2000,** *55*, 7-17.

127. Tugyi, R.; Mezo, G.; Fellinger, E.; Andreu, D.; Hudecz, F. The effect of cyclization on the enzymatic degradation of herpes simplex virus glycoprotein D derived epitope peptide. *Journal of Peptide Science* **2005,** *11*, 642-649.

128. Tugyi, R.; Uray, K.; Ivan, D.; Fellinger, E.; Perkins, A.; Hudecz, F. Partial D-amino acid substitution: Improved enzymatic stability and preserved Ab recognition of a MUC2 epitope peptide. *Proceedings of the National Academy of Sciences of the United States of America* **2005,** *102*, 413-418.

129. Wilkinson, K. A.; Vordermeier, M. H.; Kajtar, J.; Jurcevic, S.; Wilkinson, R.; Ivanyi, J.; Hudecz, F. Modulation of peptide specific T cell responses by non-native flanking regions. *Molecular Immunology* **1997,** *34*, 1237-1246.

130. Mezo, G.; de Oliveira, E.; Krikorian, D.; Feijlbrief, M.; Jakab, A.; Tsikaris, V.; Sakarellos, C.; Welling-Wester, S.; Andreu, D.; Hudecz, F. Synthesis and comparison of antibody recognition of conjugates containing herpes simplex virus type 1 glycoprotein D epitope VII. *Bioconjugate Chemistry* **2003,** *14*, 1260-1269.

131. Manea, M.; Mezo, G.; Hudecz, F.; Przybylski, M. Polypeptide conjugates comprising a beta-amyloid plaque-specific epitope as new vaccine structures against Alzheimer's disease. *Biopolymers* **2004,** *76*, 503-511.

132. Manea, M.; Hudecz, F.; Przybylski, M.; Mezo, G. Synthesis, solution conformation, and antibody recognition of oligotuftsin-based conjugates containing a beta-amyloid(4-10) plaque-specific epitope. *Bioconjugate Chemistry* **2005,** *16*, 921-928.

133. Oostenbrink, C.; Villa, A.; Mark, A. E.; Van Gunsteren, W. F. A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. *Journal of Computational Chemistry* **2004,** *25*, 1656-1676.

134. Khan, A. R.; Parrish, J. C.; Fraser, M. E.; Smith, W. W.; Bartlett, P. A.; James, M. N. G. Lowering the entropic barrier for binding conformationally flexible inhibitors to enzymes. *Biochemistry* **1998,** *37*, 16839-16845.

135. Veber, D. F.; Johnson, S. R.; Cheng, H. Y.; Smith, B. R.; Ward, K. W.; Kopple, K. D. Molecular properties that influence the oral bioavailability of drug candidates. *Journal of Medicinal Chemistry* **2002,** *45*, 2615-2623.

Kivonat

A gyógyszerkutatás egyik fontos lépése, a vezérmolekula keresése. E folyamat során egyre intenzívebben használják a virtuális szűrővizsgálatot (VS). A VS olyan *in silico*, modellező és szimulációs módszerek gyűjtőneve, amelyek a vizsgálandó molekulatárak szűrésére – szerkezeti diverzitásának növelésére, vagy éppenséggel fókuszálására – használhatók. A VS hatékonyságát jelentősen növelheti a farmakofor előzetes ismerete. Farmakofor meghatározása nem egyszerű feladat, még akkor sem, ha a kötőhely térszerkezete ismert, de még nagyobb kihívást jelent, ha nem rendelkezünk ilyen 3D információval.

Egy olyan kísérleti és számítási alapokon nyugvó eljárás kifejlesztésén dolgozunk, amely képes előállítani a farmakofort akkor, ha kiindulási térszerkezeti információval nem rendelkezünk a kötőhelyről. Ekkor, pusztán a ligand molekulák kísérletileg meghatározott kötődési aktivitás értékeit vesszük figyelembe. A doktori dolgozatom e kutatási folyamat első lépéseit foglalja össze.

A módszer kifejlesztésénél alapvető ötletünk az volt, hogy egy speciális, flexibilis molekulákból álló könyvtár segítségével térképezzük fel az aktív helyet. Ettől a tártól nem várjuk el, hogy hatékony gyógyszermolekulát tartalmazzon, ugyanakkor a tárat úgy kell összeállítani, hogy az *in vitro* meghatározott kötődési aktivitások különbsége lehetőleg csak a konformációs eloszlástól függjön.

A dolgozatban levezettem a ligand molekulák konformációs flexibilitása és kötődési erőssége közötti összefüggést. E levezetés azon túl, hogy a ligand molekulák e két tulajdonsága közötti kapcsolatot feltárja, megmutatja a használhatóságához szükséges feltételeket is.

Kidolgoztam és megvalósítottam egy farmakofor kereső inverz QSAR eljárást, amelynek vizsgálatát két peptidtárra végeztem el. Mindkét esetben a kapott farmakofort a QSAR modellekre általánosan használt, statisztikai belső validálási eljárásokkal ellenőriztem. A statisztikai külső validálás, és a meghatározott farmakoforok alapján rigid analóg keresése és kötődési vizsgálata nem képezte a dolgozat tárgyát, de rövid távú céljaink között szerepel.

Abstract

Virtual screening (VS) is intensively used today in drug research. The VS is the common name of *in silico* modeling and simulation approaches that are used to filter large molecular databases. The filtering is usually performed either to increase the structural diversity or to focus the library. The *a priori* knowledge of the pharmacophore may substantially increase the efficacy of the VS. However its identification is not a straightforward task even if the 3D structure of the binding site is known, and it is a significant challenge if such structural information is not available.

Our long term research objective is to elaborate a method, based on experiments and computations to find the pharmacophore in the absence of the 3D structure of the binding site. In the framework of this procedure the pharmacophore is identified via automatically operating computations, by using only the experimental binding affinities determined on a specially designed library. The present dissertation is the summary of the first steps in this field.

Our basic idea was to map the unknown binding site by a scanning library, designed specially from flexible molecules. We do not expect to find an appropriate drug molecule at once, but the *in vivo* binding affinity differences of the molecules should depend only on their conformational distributions.

In the present dissertation I deduced the theoretical connection between the binding affinity and the conformational flexibility of the ligands, which revealed the applicability and the limitations of the subsequent computational method. I designed and implemented computational algorithms, which generate the conformational distribution of the ligands in question and locates the bioactive conformation. Hence this procedure may be regarded as an inverse QSAR method for pharmacophore determination.

I performed the preliminary testing of the computational method on two peptide libraries, and the obtained bioactive conformations, were confirmed by the standard statistical internal validation techniques in both cases. The statistical external validation and the evaluation of rigid active analogs were beyond the scope the current study.