

Szegedi Tudományegyetem
Szent-Györgyi Albert Orvostudományi Kar
Kísérletes és Megelőző Orvostudományi Doktori Iskola

**Fehérje-ligandum komplexek és peptidek
számítógépes vizsgálata molekulamechanikai
és gépi tanulós módszerekkel**

Ph.D. értekezés tézise



Resch Vivien Erzsébet

Témavezető:

Dr. Paragi Gábor

Szeged, 2026

1. Bevezetés

A komplex biológiai rendszerek megértése többértű perspektívát igényel. Mivel a kvantummechanika (QM) alkalmazása nem kivitelezhető nagy biomolekulák esetében, a klasszikus molekulamechanika (MM) és a mesterséges intelligencia (MI) modellek (például neurális hálózatok (FFNN), a támogató vektor gépek (SVM vagy SVR), a véletlenszerű erdő (RF) vagy a logisztikus regresszió (LR)) hatékony alternatívákat kínálnak. Ezeket a számítógépes módszereket olyan különböző biológiai rendszerek vizsgálatára alkalmaztam, amelyek eltérő szinteken kapcsolódnak a központi idegrendszerhez (KIR).

Kutatásom egyik fő vonala a tubulin fehérje vizsgálata volt, amely a sejtvázas (citoszkeleton) alkotóelemeként hozzájárul többek között az intracelluláris transzportozhoz és a sejtosztódáshoz. A mikrotubulusok működése a dinamikus instabilitáson alapul. Ezt a mechanizmust különböző mikrotubulus-asszociált fehérjék, mikrotubulus-célzó ágensek (pl. taxén) és szteroid molekulák képesek modulálni, ezáltal szabályozva a sejtciklust vagy apoptózist indukálva (mint például a gyorsabban osztódó rákos sejtekben).

A szteroidok biológiai jelentőségéhez szorosan kapcsolódva vizsgáltam az aldo-keto-reduktázok (AKR), kifejezetten az AKR1C alcsalád szerepét, amelyek a szteroid metabolizmus kulcsfontosságú szabályozói. Ezek az enzimek nemcsak a perifériás szervek patológiás folyamataiban (például a hormonérzékeny daganatok kialakulásában) játszanak szerepet, hanem a KIR neuroszteroid homeosztázisát is befolyásolják.

Az immunrendszerrel kapcsolatban ismert, hogy saját, nem saját epitópok T-sejteken történő bemutatása (pl: MHC-I révén) immunválaszt vált ki. Azonban megjelentek olyan tanulmányok, ahol az amyloid toxicitást az immunrendszer működésével is összefüggésbe hozták azáltal, hogy A β oligomerek képesek aktiválni a mikroglia sejteket, amelyek gyulladáskeltő (proinflammatorikus) citokineket bocsátanak ki. A β oligomerek pedig úgy termelődhetnek, ha a neuroprotektív amiloid prekursor protein feldolgozása eltolódik a patológiás „amiloidogén” útvonal irányába, ami toxikus A β 40/42 peptidek termelődéséhez vezet. A disszertáció bemutatja, hogy a genetikai mutációk hogyan változtatják meg az A β aggregációhoz kapcsolódó értéket valamint hogy ez a változás számítógépes módszerekkel analizálható-e.

Végezetül, neuroprotektív hatóanyagok, mint például a kinurénsav (KYNA) és analógjainak vér-agy gáton keresztül történő célba juttatásának nehézségeivel foglalkoztam. A kutatás a humán és szarvasmarha-szérumalbumint (HSA, illetve BSA; együttesen SAs) potenciális hordozófehérjeként való alkalmazását vizsgáltam mely a közeli jövő gyógyszerfejlesztést segítheti.

2. Célkitűzések

1. Tubulin fehérjének taxén kötőhelyébe kötődő módosított szteroidmolekulák számítógépes vizsgálata (MM/GBSA és umbrella sampling (US) módszerekkel).
2. AKR1C1 és AKR1C2 izoenzimekhez kötődő EM1062 és EM1082 módosított szteroidok számítógépes vizsgálata (dokkolási vizsgálatok).
3. KYNA és származékainak SA-hoz való kötődésének jellemzése kísérleti és számítógépes módszerekkel.
4. MI modellek (FFNN, SVM, RF, LR) hogyan képesek 9 aminosav hosszúságú (9-mer) epitópok T-sejtes aktivációs értéket előjelezni, két különböző tesztkészleten.
5. Különböző MI modellek (FFNN, SVR, RF) használata a mutált A β peptidek nukleációs értékének (NS)

előrejelzéséhez különböző reprezentációk (ESM-2, Z-scale, AAC, pw.BLOSUM62) felhasználásával.

3. Eredmények és diszkussziói

Mivel számos kemoterápiás szer nemkívánatos mellékhatásokkal jár és a rezisztencia kialakulása tovább korlátozhatja a kezelés hatékonyságát, a szteroid vázszerkezet vonzó tulajdonsága, hogy még a legkisebb szerkezeti módosítások is jelentős változásokat eredményezhetnek a farmakológiai profilban. Doktori tanulmányaim során az **első kutatási irányvonal** az MM/GBSA és US módszerek alkalmazása volt annak érdekében, hogy molekuláris szintű magyarázatot adjunk három kiválasztott módosított szteroidvegyület kötődési preferencia-sorrendjére a tubulin fehérjéhez, amelyek ígéretes biológiai aktivitást mutattak a HeLa és a SiHa sejtvonalakban (IC_{50} : 13AES3 > 13AES2 > 13AES1).

Az US-számítások szerint a 13AES3 ligandum mutatta a legerősebb kölcsönhatást a kötőhellyel:

PMF_{13AES3} : -14 kcal/mol, PMF_{13AES2} : -6.5 kcal/mol és PMF_{13AES1} : -5.2 kcal/mol szabadenergia profillal.

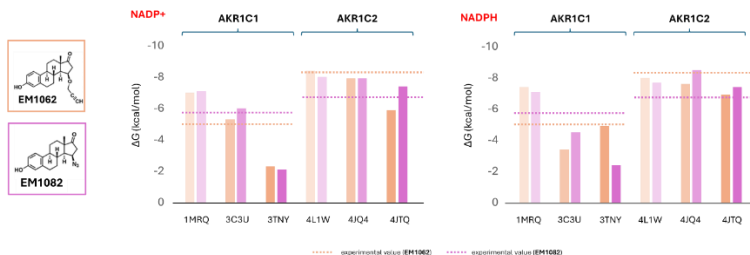
Az MM/GBSA-számítások is nagyon hasonló kötődési értékekkel támasztottak alá a kötése erősségi sorrendet:

ΔG_{13AES3} : -69.38 ± 2.14 kcal/mol, ΔG_{13AES2} : -50.85 ± 2.85 kcal/mol, ΔG_{13AES1} : -42.76 ± 2.35 kcal/mol értékkel.

A számítógépes szimuláció eredménye összhangba van a kísérleti megfigyelésekkel, mely szerint a 13AES3 C-2-es klorofenilgyűrűje hozzájárul a szteránváz kötődési helyzetének stabilizálásához, nevezetesen a taxén-kötőhelyen található PHE272 aminosavval való hidrofób kölcsönhatás révén.

Ismert tény az is, hogy az aldo-keto-reduktáz izoformák kulcsfontosságú enzimek a szteroid anyagcserében, és fontos célpontok hormonfüggő rákos megbetegedésekben. Az AKR1C izoformák szelektív inhibitor molekuláinak fejlesztése meglehetősen kihívást jelent. Különösen az AKR1C1 és AKR1C2 esetében, mivel mindössze egyetlen aminosavban különböznek az aktív helyen. Így tanulmányaim **második iránya** az EM1062 és EM1082 kötődésének vizsgálata volt AKR1C1 és AKR1C2 izoenzimekhez. Dokkolási számításokat végeztem annak érdekében, hogy atomi szintű betekintést nyerjek a két módosított szteroid lehetséges kötődési módjába. Megbízhatóságának értékeléséhez a dokkolási pontszámokat

(Glide score) összehasonlítottam a kísérleti gátlási adatokból származó ΔG értékekkel.



1.ábra Módosított szteroidok dokkolása az AKR1C1 és AKR1C2 kötőhelyekre különböző (AKR1C1 PDB-kódok: 1RMQ, 3C3U, 3TNY; AKR1C2 PDB-kódok: 4L1W, 4JQ4, 4JTQ) kristályszerkezetek felhasználásával. A pontozott vonal a kísérleti kötődési szabadenergiákat jelöli (a gátlás százalékos értékéből számítva).

A számítógépes eredmények azt mutatták, hogy mindkét ligandum erősebb kötődési affinitást mutat az AKR1C2 izoenzimhez mint az AKR1C1-hez (1. ábra). Ez a tendencia összhangban áll a kísérleti gátlási adatokkal, amelyek szerint mindkét vegyület nagyobb inhibíciót mutatott az AKR1C2 (kb. 97–99%-os gátlás), mint az AKR1C1 fehérjével szemben (kb. 32–63%-os gátlás).

Összefoglalva, a módosított szteroidok kötődési tulajdonságainak számítógépes vizsgálatai molekuláris szintű értelmezést nyújtottak a kísérleti eredményekhez és elméleti kiegészítést nyújtottak biológiai aktivitásuk megértéséhez.

Ezt követően a **harmadik** kutatási vonal keretében vizsgáltam a KYNA és egyes származékainak (SZR72, SZR104) kötődését a SA-hoz. Az US eredmények szerint a KYNA erősebben kötődik a BSA-hoz, mint a HSA-hoz, ami alátámasztja a BSA alkalmazását a vizsgálatokban. A C-2 módosított SZR72 analóg alacsonyabb kötődési hajlamot mutatott a HSA iránt, mint a KYNA, és csak gyenge preferenciát mutatott a BSA iránt, míg összességében viszonylag alacsony maradt a kötődési képességük. A morfolingyűrűt (C-3) tartalmazó SZR104 származékot figyelembe véve, mindkét SA-hoz gyengén kötődött, preferencia sorrend nélkül.

Az ITC mérések eredményeként a BSA_{KYNA} K_d értéke $158,9 \pm 81,4 \mu M$, a HSA_{KYNA} -hoz viszonyítva pedig $K_d = 5,7 \pm 6,0 \mu M$ disszociációs állandó, míg a progresszív szerkezeti módosítások eredményei (SZR72 és SZR104 esetében) az affinitás szisztematikus csökkenését mutatják $1 \text{ mM} < K_d$ értéknél (a kimutatható tartomány felett). Annak érdekében, hogy tisztázzuk ezen molekulák kötődését, STD-NMR kísérleteket végeztünk. STD-NMR kísérleteinkben a KYNA mutatta a legerősebb albuminkötést, míg az SZR72 gyengébb, de még mindig kimutatható kölcsönhatást mutatott, a kombinált szerkezetileg módosítást tartalmazó

származék (SZR104) csak minimális, vagy semmilyen kimutatható kötődést nem mutatott.

PhD tanulmányaim **negyedik kutatási iránya** a gépi tanulási technikák alkalmazása volt különféle peptidekkel kapcsolatos kérdésekre. Első esetben a T-sejt aktiváció változékonyságára összpontosítottunk, melyek a kölcsönható peptidek szerkezeti különbségeiből is adódhatnak, melyek epitóp-MHC-I komplexként a sejt felszínére kerülnek. A tézisfüzetben csak egy osztályozási értékelési metrikára szeretnék összpontosítani, nevezetesen a kiegyensúlyozott pontosságra (bACC), amely különösen értékes olyan valós helyzetekben, ahol gyakoriak a kiegyensúlyozatlan adathalmazok (ha a bACC értéke 1, akkor az osztályozási modell tökéletes). Gépi tanuláson alapuló regresszió esetén a meghatározási együtthatóra (R^2) koncentrálok, amely az MI-modell illeszkedésének minőségét jelzi (ha $R^2 = 1$, akkor a regressziós modell tökéletes).

A konszenzusalapú validációs (30%-os hold-out) eredmények erős összteljesítményt mutattak, ahol a $bACC_{FFNN}$, $bACC_{RF}$, $bACC_{SVM}$, $bACC_{LR}$ értékek körülbelül 0,9 vagy magasabbak, ami mind az aktív, mind az inaktív

epitópok robusztus osztályozását jelzi. A mesterséges intelligencia módszerek fő kihívása azonban továbbra is a külső adatkészletekre való alkalmazhatóságuk. Az osztályozást egy független külső adatkészleten is végeztem, amely 85, a modell által nem látott epitópot tartalmazott. Az első külső tesztkészlet alátámasztotta a konszenzusalapú megközelítés teljesítményét, bár voltak osztályozófüggő különbségek. Az eredmények a következők: $bACC_{FFNN}$: 0,805, $bACC_{RF}$: 0,804, $bACC_{SVM}$: 0,562, $bACC_{LR}$: 0,718.

A modellek robusztusságának és adathalmazok közötti általánosíthatóságának további felmérése érdekében egy második külső tesztalmazon értékeltem a teljesítményüket. A második külső tesztalmaz 172 további epitópot tartalmaz, amelyet más kutatócsoport publikált, és három T-sejt receptorban (TCR1, TCR2 és TCR3) mértek. Így a második külső tesztalmaz egy kihívást jelentőbb validációs eljárást képviselt, amit az osztályozók közötti alacsonyabb és változékonnyabb teljesítmény is tükröz. Nevezetesen: $bACC_{FFNN}$: $0,497 \pm 0,037$, $bACC_{RF}$: $0,618 \pm 0,074$, $bACC_{SVM}$: $0,615 \pm 0,008$, $bACC_{LR}$: $0,508 \pm 0,023$. Összességében ezek az eredmények azt mutatják, hogy a második külső tesztalmazt lényegesen nehezebb volt osztályozni, mint az előzőeket (belső validáció (30% hold-out) és az első külső

teszthalmaz), kiemelve a különböző laboratóriumi körülményeket és a kísérleti variabilitás hatását az adathalmazok közötti modellek átvitelére.

PhD-s kutatásom **ötödik vonala** az A β -mutációk (A β 1–42) „nukleációs pontszám”ának (NS) előrejelzéséhez kapcsolódik MI-algoritmusok segítségével. A nukleációs folyamat kis A β -aggregátumok kezdeti kialakulását írja le, amelyek később fibrillákká és plakkokká fejlődhetnek. Ez az esemény számszerűsíthető NS értékkel, amely kísérleti vagy számítógépes módszerekkel származtatható. Az utóbbi időben a „deep mutation scanning assay” vizsgálatok nagyszámú egyszeres és kétszeres A β -variánshoz rendelnek NS-értékeket. Az A β -aggregátumhoz tartozó NS értékben a mutációs hatások megjelennek és értéke jelentősen megváltozhat akár egyetlen aminosav (sAA) vagy kétszeres aminosav (dAA) szubsztitúció után is. Napjaink jelentős kísérleti erőfeszítések ellenére a jelenleg elérhető mutációs adatkészletek hiányosak maradnak. Tekintettel arra, hogy a nagyobb tanulóhalmazok robusztusabb MI-eredményeket tudnak szolgáltatni, a dAA-adatkészletre fogok összpontosítani. A modell teljesítményét R²-vel számszerűsítettem. A modellt a dAA adathalmazon (~15 000

adatpont) képeztük ki, 70/30%-os arányban 5-szörös keresztvalidációval és 30%-os belső validációval, majd sAA (~500 adatpont) külső teszttel teszteltük.

Az FFNN következetesen a legmagasabb predikciós teljesítményt érte el szinte az összes vizsgált szekvencia reprezentáció esetében (30%-os hold-out teszt: R^2 : **0,737** és sAA külső teszt: R^2 : **0,580** az ESM-2-ben). Azonban továbbá, konszenzusanalízist végeztem annak vizsgálatára, hogy az integrált predikciók javíthatják-e a regressziós modellek robusztusságát és általánosíthatóságát.

A konszenzusmodell az ESM-2–FFNN, az ESM-2–SVR és a Z-skála–RF predikcióit kombinálta, amelyek a legmagasabb predikciós teljesítményt nyújtották. A konszenzusmodell érte el a legmagasabb validációs teljesítményt, **0,781** R^2 értéket eredményezve, míg eredetileg a validációs tesztben (ESM-2: FFNN) a legmagasabb teljesítmény 0,737 R^2 érték volt. Ezek az eredmények azt mutatják, hogy a három kiválasztott modell beépítése javította a predikciós pontosságot a tanulóhalmaz eloszlásából származó adatokon. Míg a külső teszt (sAA) esetében **0,582** R^2 értéket ért el.

4. Összefoglalás

PhD tanulmányaim alatt számos számítási és kísérleti megközelítést alkalmaztam a szerkezet-aktivitás és a szerkezet-affinitás kapcsolatok tisztázására a különböző biológiai rendszerekben. A molekula dinamikai szimulációk szerint a 13AES3 C-2 pozíciójában lévő klorofenil-szubsztituens kulcsszerepet játszik a szteroidváz stabilizálásában a taxén kötőhelyén belül.

Az AKR1C enzimek dokkolási és kölcsönhatási elemzése az azt mutatták, hogy mindkét ligand (EM1062 és EM1082) erősebben kötődik az AKR1C2-höz, mint az AKR1C1-hez, összhangban a kísérleti gátlási adatokkal. A szelektivitás a kötődési orientáció és az aktív hely szerkezetének finom különbségeiből, különösen a Leu54/Val54 szubsztitúcióból fakad.

A KYNA és származékainak HSA-hoz és BSA-hoz való kötődési tulajdonságainak számítási és kísérleti eredményei egyértelmű szerkezetfüggő trendre utalnak; nevezetesen az alapvegyület (KYNA) C-2 és C-3 módosulása gyengébb komplexet eredményez a SA-kal.

Az MHC-I epitóp aktivációs vizsgálatának MI-alapú vizsgálataiban az eredmények nem támasztottak alá egyetlen

univerzális osztályozót az összes validációban. Ehelyett az FFNN, RF, LR és SVM mindegyike kiegészítő predikciós erősségeket mutatott a külső tesztekől függően.

Végül az A β 1-42 peptid mutációs vizsgálatok arra utalnak, hogy maga az alkalmazott konszenzus modell is informatívnak bizonyult. Különböző regressziós modellek (FFNN, SVR, RF) és a peptidreprezentációk széles skálájának használata javította a konszenzus modell általánosítását az A β mutációk NS értékeinek előrejelzésére, míg az sAA külső tesztre történő alkalmazás hasonlóan moderált általánosítást eredményezett, mint nélküle.

Tézispontok

1. A módosított szteroidok tubulin fehérje taxén kötőhelyéhez való affinitásának számítógépes vizsgálata (MM/GBSA és US módszerek) kimutatta, hogy a 13AES3 klorofenil gyűrűje erősebb hidrofób kölcsönhatást alakít ki mint a 13AES2 és 13AES1, ezáltal növelve a kötési stabilitást. A kísérleti méréssel megállapított kötődési sorrendet a számítási eredmények is alátámasztották.
2. Az EM1062 és EM1082 módosított szteroidok erősebben kötődnek az AKR1C2 izoenzimhez, mint az

AKR1C1, ami azt jelzi, hogy az aktív hely finom szerkezeti különbségei jelentősen befolyásolják a szelektivitást, összhangban a kísérleti eredményekkel.

3. A KYNA kötődését és kötődési sorrendjét a HSA és BSA fehérjékhez kísérletileg és számítógépesen is megerősítettem. Azonban ugyanazok a kísérleti és elméleti technikák a módosított származékok esetében gyengébb affinitást mutatnak, mint az alapvegyület, ami arra utal, hogy a módosított vegyületek esetében eltérő gyógyszerbeviteli módszer szükséges.

4. A különböző ML osztályozókat figyelembe véve nincs egyetlen legjobb modell az epitóp aktiváció előrejelzésére az MHC-I komplexekben; Az FFNN, RF, LR és SVM algoritmusok mindegyike eltérő előnyöket kínál különböző validációk esetén. A különböző peptid reprezentációkat kombináló konszenzusmodell nyújtotta a legrobustusabb teljesítményt, mind a belső tesztalmazon, mind a független külső tesztalmazon.

5. Az A β 1-42 peptid dupla mutációval rendelkező variánsainak NS-értékeinek előrejelzésére használt MI algoritmusok (FFNN, SVM és RF) esetében azt tapasztaltam, hogy a legpontosabb előrejelzéseket a különböző

algoritmusokat és aminosav-reprezentációkat kombináló konszenzusmodellel kaptam. Azt is megfigyeltem, hogy a dAA-mutációkon betanított modell nem alkalmas az sAA-mutációk NS-értékeinek előrejelzésére.

A disszertáció alapjául szolgáló publikációk

1. Resch, Vivien, Marija Gjorgoska, Eva Hafner, Ildikó Bacsa, Benjamin Kovács, Tomaž Büdefeld, Attila Hunyadi, Ildikó Huliák, Mónika Kiricsi, Gábor Paragi, and et al. "*Targeting Steroid-Metabolizing Enzymes with 15 β -Substituted Estrone Analogues: Dual Discovery of AKR1C2/17 β -HSD1 Inhibitors and a Fluorescent 17 β -HSD1 Ligand*" *Cancers*, 2026; 18(12): 1889. doi:10.3390/cancers18121889. **SJR indicator Q1, IF: 4.8** (2025)
2. Ali, Hazhmat, Péter Traj, Gábor J. Szebeni, Nikolett Gémes, Vivien Resch, Gábor Paragi, Erzsébet Mernyák, Renáta Minorics, and István Zupkó. "*Investigation of the Antineoplastic Effects of 2-(4-Chlorophenyl)-13 α -Estrone Sulfamate against the HPV16-Positive Human Invasive Cervical Carcinoma Cell Line SiHa*" *International Journal of Molecular Sciences*, 2023; 24(7): 6625. doi:10.3390/ijms24076625. **SJR indicator D1, IF: 4.9**
3. Vivien Erzsébet Resch, László Tóth, Anita Rác, Dezső Virok, Gábor Paragi, "*Machine Learning-based prediction of Cross Immunity*" *Briefing in Bioinformatics*, 2026; **accepted SJR indicator D1, IF: 7.3** (2025)

4. Vivien Erzsébet Resch, Anita Rácz, Gábor Paragi
“*Mutation-Induced Changes in Nucleation Scores of A β -peptide: A Machine Learning Study*” Journal of Cheminformatics (in preparation) 2026
5. Vivien Erzsébet Resch, Edit Wéber, Zsófia Hegedűs, Ádám Juhász, Edit Csapó, Bálint Lőrinczy, István Szatmári, Gábor Paragi “*Binding properties of Kyuneric Acid Analogues to Serum Albumins*” Journal of Drug Delivery Science and Technology (submitted) 2026