

# Célpont-ligandum komplexek számítása fragmens alapú molekulatervezési eljárásokkal

*Ph.D. értekezés*

**Horváth István**

*okl. vegyész, programozó matematikus*

Témavezető

Dr. Habil. Hetényi Csaba

*egyetemi docens*

*PTE ÁOK Farmakológiai és Farmakoterápiai Intézet*



Kémia Doktori Iskola

SZTE TTIK

Szeged

2019.

# 1. A szerző tudományos közleményei

## 1.1. Az értekezés alapját képező referált szakcikkek\*

**Horváth, I.**; Jeszenői, N.; Bálint, M.; Paragi, G.; Hetényi, C., A fragmenting protocol with explicit hydration for calculation of binding enthalpies of target-ligand complexes at a quantum mechanical level. *Int. J. Mol. Sci.* **2019**, 20, 4384. IF<sub>2018</sub>: 4,183

Bálint, M.; **Horváth, I.**; Mészáros, N.; Hetényi, C., Towards unraveling the histone code by fragment blind docking. *Int. J. Mol. Sci.* **2019**, 20, 422. IF<sub>2018</sub>: 4,183

Jeszenői, N.; Schilli, G.; Bálint, M.; **Horváth, I.**; Hetényi, C., Analysis of the influence of simulation parameters on biomolecule-linked water networks. *J. Mol. Graph. Model.* **2018**, 82, 117-128. IF: 1,863

Az értekezés alapját képező referált szakcikkek összesített impakt faktora: 10,229

## 1.2. Az értekezés témájához kapcsolódó referált szakcikkek

Bálint, M.; Jeszenői, N.; **Horváth, I.**; van der Spoel, D.; Hetényi, C., Systematic exploration of multiple drug binding sites. *J. Cheminform.* **2017**, 9, Article number: 65.

Jeszenői, N.; Bálint, M.; **Horváth, I.**; van der Spoel, D.; Hetényi, C., Exploration of interfacial hydration networks of target–ligand complexes. *J. Chem. Inf. Model.* **2016**, 56 (1), 148-158.

Jeszenői, N.; **Horváth, I.**; Bálint, M.; van der Spoel, D.; Hetényi, C., Mobility-based prediction of hydration structures of protein surfaces. *Bioinformatics.* **2015**, 31 (12), 1959-1965.

## 1.3. Az értekezés témájához kapcsolódó konferenciaközlemények

Hetényi, C.; Bálint, M.; Schilli, G. ; **Horváth, I.**, Computational blind docking of ligands to drug targets: Methodology and applications (**2018**), 2nd Edition of Global Conference on Pharmaceutics and DrugDelivery Systems 2018, Róma, Olaszország.

Bálint, M.; Hetényi, C.; Schilli, G. ; **Horváth, I.**, Discovering multiple drug binding sites and modes with Wrap ‘n’ Shake (**2018**), 2nd Edition of Global Conference on Pharmaceutics and DrugDelivery Systems, Róma, Olaszország.

Mészáros, N.; **Horváth, I.**; Bálint, M.; Hetényi, C., Unraveling the histone code by fragment blind docking (**2017**), ECBS 2017 5th European Chemical Biology Symposium, Budapest, Magyarország.

Hetényi, C.; Bálint, M.; Jeszenői, N.; **Horváth, I.**, Improvements of computational optimization of drug-target interactions, 93 p. (**2017**), Global Conference on Pharmaceutics and Drug Delivery Systems (PDDS 2017), Valencia, Spanyolország.

Jeszenői, N.; Bálint, M.; **Horváth, I.**; Hetényi, C., Komplex szerkezetek vízhálózatai (**2016**) Előadás, KeMoMo–QSAR Szimpózium (**2016**), Az MKE Szerves- és Gyógyszerkémiai Szakosztályának QSAR és Modellezési Szakcsoportja és az MTA Szegedi Akadémiai Bizottságának Kemometria és Molekulamodellezés Munkabizottsága által szervezett kétnapos szimpózium, Miskolc, Magyarország.

Jeszenői, N.; Ábrahám, I.; Bálint, M.; **Horváth, I.**; Hetényi, C., Az ösztrogén receptor alfa nem-klasszikus ösztrogén hatásokért felelős kötőhelyének azonosítása: in silico szerkezeti biokémiai vizsgálatok (**2016**), III. Magyar Neuroendokrin Szimpózium, Magyar Endokrinológiai és Anyagcsere Társaság XXVI. Kongresszusának szatellit rendezvénye, Szeged, Magyarország.

Hetényi, C.; Jeszenői, N.; Bálint, M.; **Horváth, I.**; Lopata, A. (szerk.) Fehérjék hidrátszerkezete: modellezzem vagy mérjem? (**2015**), Előadás, KeMoMo–QSAR Szimpózium.

#### **1.4. Egyéb közlemények**

Bálint, M.; Jeszenői, N.; **Horváth, I.**; Ábrahám, M. I.; Hetényi, C., Dynamic changes in binding interaction networks of sex steroids establish their non-classical effects. *Sci. Rep.* **2017**, 7, Article number: 14847.

Hetényi, C.; Jeszenői, N.; Bálint, M.; **Horváth, I.**, Answers to current challenges in target-based drug design, In: Laszlo, Gelencser; Zoltan, Horvath (szerk.) (**2016**), BJMT Conference of Applied Mathematics 2016: Book of Abstracts.

Solymosi, F.; Bugyi, L.; Oszkó, A.; Horváth, I., Generation and Reactions of CH<sub>2</sub> and C<sub>2</sub>H<sub>5</sub> Species on Mo<sub>2</sub>C/Mo(111). *J. Catal.* **1999**, 1, 160-169.

*\*Az 1.1 alatti saját közleményekre az értekezésben név szerint történik hivatkozás.*

## 2. Az értekezés főbb eredményeinek összefoglalása (tézispontok)\*

- I. Kidolgoztunk egy új, fragmentáló eljárást, a Fragmentert, amelynek segítségével molekulamechanikai szinten optimalizált fehérje-ligandum komplex-szerkezetek interfész régiója standardizált és automatizált módon extrahálható kvantummechanikai számításokhoz (Horváth et al. 2019).
- II. Extrahált interfész szerkezetek felhasználásával, szisztematikus vizsgálatok során kimutattuk, hogy a szemiempirikus kvantumkémiail szinten számolt reakcióhők és a kísérleti fehérje-ligandum kötési entalpiaváltozások között a legjobb korrelációt egy hibrid vízmodell segítségével érhetjük el. Ezen új modellben a COSMO implicit vízmodellt és a korábban általunk publikált MobyWat módszerrel számolt explicit interfész vízmolekulákat ötvöztük (Horváth et al. 2019).
- III. Kiszámítottunk egy skálázó faktort, amely a szemiempirikus kvantumkémiail szinten számolt reakcióhők és a kísérleti fehérje-ligandum kötési entalpiaváltozások között egyszerű lineáris kapcsolatot teremt tengelymetszet alkalmazása nélkül. Ezáltal egy új, direkt kapcsolat jött létre a molekulaszerkezet és kötési termodinamika területei között (Horváth et al. 2019).
- IV. Kidolgoztuk és hisztonpeptid-komplexek felhasználásával teszteltük a fragment blind docking (FBD) algoritmust, amely egy új módszerben ötvözi a blind docking megközelítésünket és a fragmens dokkolás technikát (Bálint et al. 2019).
- V. Az FBD általános algoritmusán belül a célpont molekulát a csomagolási lépés után betérítő peptid fragmensek összekötésére és összeforrasztására alkalmazható eljárást dolgoztunk ki, kódoltunk és kalibráltunk be (Bálint et al. 2019).
- VI. Kiszámítottuk hidratációban részt vevő vízmolekulák hálózatainak szerkezetét és feltártuk azok hőmérsékletfüggését komplex interfészek és teljes fehérje felszín esetében. A hálózati megfontolások segítségével értelmezést adtunk a vízszerkezet molekuladinamikai számításának sikerére vonatkozólag (Jeszenői et al. 2018).

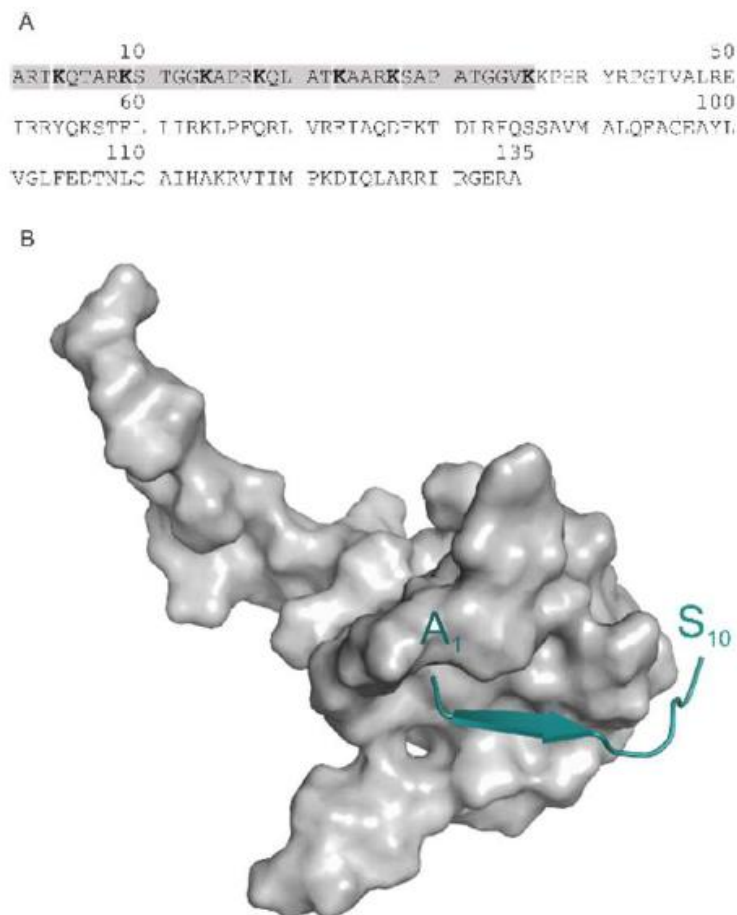
*\*A tézispontokra az értekezésben római számmal történik hivatkozás.*

### 3. Bevezetés

A fragmens alapú molekulatervezési eljárások elterjedéséért sok esetben a biokémiai komplexek nagy mérete és szerkezeti összetettsége okolható. A bevezetésben felvázolom a munkánk során vizsgált komplexek szerkezeti biokémiai hátterét, kitérve azon szerkezeti sajátosságaikra, amelyek kihívást jelentenek a szerkezetvizsgáló és -számító technikák számára. Ezt követően a dolgozat közvetlen előzményét jelentő, dokkolási és hidratációs számítási módszerek és a fragmentálás alap gondolatainak ismertetésére kerül sor.

#### 3.1. A hiszton-komplexek szerkezetmeghatározásának nehézségei

Ebben a dolgozatban több, terápiásan és epigenetikailag fontos komplexen végzett számításokat és azok eredményeit mutatom be. A bevezető részben ezek közül a hisztonok komplexeire fókuszálunk, mivel talán ezek jelentik a legnagyobb kihívást a szerkezetmeghatározási módszerek számára. Az elmúlt évtizedekben az epigenetika új utakat nyitott a gyógyszertervezésben [1]. Az epigenetikus folyamatok közül a hisztonfehérjék poszttranszlációs módosításai (PTM) különösen érdekesek [2-4]; ezeknek a PTM-eknek az epigenetikus szerepe tisztázható a kromatin szerkezetét befolyásoló módosító hatással, amely a kromatinszálak nukleoszómáiban a hiszton-DNS és a hiszton-hiszton kapcsolatokat befolyásolja. A hisztonok kis, konzervált eukarióta fehérjék, nagyon rugalmas N-terminális farokkal (**1a. ábra**) amelyek jellemző hossza a hiszton H3 esetében kb. 35 aminosav [5]. A farok kovalensen módosítható a K, R, T és S aminosavak oldalláncaiban. A kapott PTM-ek sokféle kémiai reakciókban vehetnek részt, mint például metilezés, acetilezés, foszforiláció. Ha például a metilezést vizsgáljuk, három (mono-, di- és trimetilezett) PTM-et nyerhetünk ki a K oldalláncának töltéssel rendelkező aminocsoport hidrogénatomjainak helyettesítésével. Az N-terminális hiszton H3 peptidnek hét K helyzete van (**1a. ábra**), így a metilezés 47 PTM-variációt eredményezhet (négy a három PTM-ből és a nem módosított K-ból származik). Ily módon hatalmas számú PTM-variáció származhat, ha az összes fent említett aminosavat és módosító csoportot figyelembe vesszük.



**1. ábra** (a) A H3 szekvenciája a 36 aminosav hosszúságú N-terminális peptiddel, szürke szín jelöli (b) A H3 fark (szürke) AIRE-PHD-hez kötődve (2ke1-es rendszer). Forrás: Bálint et al. (2019).

Az azonosításra a genetikai kódon túl a PTM-variációk nagy számán alapuló hisztonkódot javasoltak [6], amely egy nagyon komplex, térben és időben dinamikusan változó mintázat. A hisztonkód alapvető fontosságú a különböző betegségek kromatinhoz kapcsolódó patomechanizmusainak epigenetikájában [2, 3], kromatinhoz kapcsolódó olvasófehérjékkel [5-7] „dekódolható és fordítható”. A hiszton-olvasó (író) komplexek atomi felbontási szerkezetei kulcsfontosságúak a hisztonkód kibogozásában és a gyógyszertervezésben. A hisztonok PTM-variációinak nagy száma hasonlóan nagyszámú lehetséges komplexet eredményez, az ilyen nagyméretű komplexek szerkezeti meghatározása még nagy teljesítményű kristálytani [8, 9] technikák esetében is kihívást jelent.

A lehetséges hisztonkomplexek variációinak nagy száma igazi kihívást jelent a kísérleti szerkezeti meghatározási technikák számára. Ezen komplexek számolása a peptidek jelentős mérete és

rugalmassága, valamint az olvasók sekély kötőfelületei miatt nehéz feladat. Ezenkívül a kötési helyek gyakran nem ismertek, ami szükségessé teszi az egész célpontfehérje felületén a blind docking (BD) keresési módszer alkalmazását.

### **3.2. A dokkolási probléma és eddigi megoldásai**

Az elmúlt fél évszázadban a számítógépes kémiai eljárások lehetővé tették, hogy proteinek és ligandumjaik szerkezetét és azok kölcsönhatásait atomi szinten gyorsan modellezzük. Hogy mely módszereket, illetve mely módszerek kombinációit alkalmazzuk, függ a vizsgált rendszer méretétől, az elérhető számítási kapacitástól, illetve hogy milyen pontosságot várunk el.

A molekulamechanikai módszerek (MM) a vizsgált molekulákat az atomok szintjén kezelik, szemben a kvantummechanikai (QM) eljárásokkal, amelyek az elektronok és atommagok kölcsönhatásainak leírására is képesek [10]. Egy erőteret (force field, FF) a kölcsönhatásokat leíró egyenletek matematikai alakja és az ezekben szereplő paraméterek, valamint a topologikus leírás (atomtípusok, kötéstípusok, stb.) jellemeznek. Az erőtér parametrizálásának az a kritériuma, hogy minél pontosabban reprodukálni tudjuk kísérletileg meghatározott jellemzőit. A kísérleti módszerek fejlesztésével egyre több adat állt rendelkezésre a kisméretű molekulákról, így ismertté váltak az atomok közti kötéstávolságok, kötésszögek és egyéb molekulageometria és energetikai tulajdonságok. Felmerült, hogy egy nagyobb molekula a kisebbekkel modellezett szerkezeti egységek kombinációjaként leírható-e, ebből fejlődött ki az erőtér eljárás alapján alapuló MM számítási módszer, amely Andrews mechanikai modelljén [11] alapul, amely a molekulákat merev, gömbszerű atomokból álló rendszerként kezeli és az atomok közti kötések rugókkal szemlélteti. Napjainkban biomolekuláris rendszerek MM-szintű számítására használt, kedvelt FF az AMBER [12], a CHARMM [13], az OPLS [14] és a GROMOS [15].

Dokkolásnak azokat a MM-szintű módszereket nevezzük, amelyekkel megjósoljuk egy molekula preferált kötési helyét egy másik molekula felületén, miközben egy stabil komplex alakul ki. Az előnyben részesített kötőhely ismeretében - amely lehet kísérleti eredmény [16-18] vagy korábbi számolás eredménye [19] - pedig meghatározhatjuk a kötése erősséget a két molekula között pl. pontozó (scoring) függvények segítségével. A molekuladokkolást általában kisebb ligandum és nagyobb célpont molekulák közötti kölcsönhatások vizsgálatára alkalmazzák [20]. Fókuszáltan történik a dokkolás, ha a kötőzsebről van hozzávetőleges ismeretünk és vakon (blind docking, BD),

amennyiben a célpont molekula teljes felületen végezzük a keresést. A BD a kötőzsebek elhelyezkedéséről és a kötési módokról egyidejűleg ad információt a teljes felületen [21, 22] előzetes ismeretek nélkül.

A gyógyszertervezésben használt dokkoló eljárások, amilyen pl. az AutoDock [23, 24] gyorsasága megköveteli, hogy megközelítéseket használjanak, mint pl. a kölcsönható felszínek nem megfelelően meghatározott hidratációja, vagy a célpont molekula flexibilitásának figyelmen kívül hagyása; viszont ezek a megközelítések nem mindig adnak megfelelő eredményeket. A dokkolás vízmolekulákkal feltöltött kötőzsebekbe is irányíthatja a ligandumot, amennyiben nem vesszük figyelembe a határfelületi vízmolekulák pozícióit [22, 25].

A dokkoló módszerek globális keresési algoritmusokat is használnak, de heurisztikusak, és nem determinisztikusak, mint a molekuladinamikai szimulációk. A dokkolás során több közelítést is alkalmaznak a számítások felgyorsítása érdekében, míg a molekuladinamika sokkal pontosabb, de számítási szempontból drágább a MM-módszer, mint a dokkolás.

A molekuladokkolás az egyik legszélesebb körben elterjedt MM-módszer, amit protein-ligandum komplex szerkezetek jellemzésére és előrejelzésére [26] használnak a számítógépes gyógyszertervezésben [27]. De a nagy sebességű számítások bizonyos korlátozásokat követelnek meg, amint azt egy korábbi tanulmányban már összefoglaltuk [28]:

- a) a keresési hely korlátozása a kötőhely környezetére [16]
- b) a nem megfelelő hidratáció, vagy annak teljes hiánya a ligandum-célpont interfészen [29, 30]
- c) a célpont rugalmasságának teljes, vagy részleges figyelmen kívül hagyása [27, 31-33]
- d) véletlenszámgeneráláson alapuló nemdeterminisztikus kereső algoritmus [26, 34] alkalmazása.

Az első korlátot a BD-megközelítés [21, 35] képes kiküszöbölni, mivel a keresést a célpont teljes felületén végzi anélkül, hogy bármilyen információval rendelkezne a protein felületéről. Mivel BD-ban nincs szükség az elsődleges kötőhely előzetes ismeretére vagy erre vonatkozó korlátozásra, így az több kötőhelyet is megtalálhat. Ezért a BD nagy népszerűsége tett szert [36-38] és alloszterikus [39-41], vagy többszörös kötőhelyek [42-47] megkeresésére is sokan használják.



A BD-t eredetileg az AutoDock [24] programcsomagra adaptálták, az utóbbi években azonban újabb dokkolóprogramokat, szervereket is használnak [48]. Viszont ki kell emelni, hogy az AutoDock az egyetlen olyan programcsomag, amely valóban a teljes fehérjefelületen [48] végzi el a keresést. Gyakran használnak Pocket Search (PS) módszereket, hogy megtalálják a kötőhelyeket, viszont a PS után helyi fókuszált keresést is el kell végezni, hogy megtaláljuk a ligandum szerkezetét a kötő részekben [49, 50].

A fent említett korlátozások miatt a gyors módszereket használó dokkoló eljárások sok esetben nem adnak megbízható eredményeket [51], ami a ligandum helyzetén, a számolt kötési energián kívül legfőképpen a célpont dinamikus megközelítése miatt van. Az entrópiás hatások és a solvatáció nem megfelelő paraméterezése nem ad megfelelő sorrendet a kapott ligandumszerkezetekben. A gyors módszerek ezen hiányosságait molekuladinamikai (MD) szimulációkkal lehet kezelni. A MD egyre nagyobb szerepet kap a gyógyszertervezésben [52-54] és több tanulmány is szólt arról, hogy molekuladinamikát használtak a ligandumkötési folyamat nyomon követésére [18, 55-57], mindezt atomi szinten [18, 55, 58-61]. A dinamikus BD molekuladinamikával egyesítve lehetővé teszi, hogy a ligandum kötésének reakcióútját atomi szinten kövessük, miközben a teljes célpont és a ligandum is megőrzi rugalmasságát. A molekuladinamika lehetővé teszi explicit vízmolekulák használatát és azok jelentőségének vizsgálatát a ligandumkötés során [18] úgy, hogy közben a célpont rugalmasságát is figyelembe veszi [62], amely így a négy korlátozásból a másodikat és harmadikat kiküszöböli. A MD segítségével új lehetőségek nyílnak a BD javítására rugalmas rendszerek esetén is [32, 63, 64], így az allosterikus kötőhelyek tanulmányozása is lehetséges atomi szinten. Ha ismert a ligandum kötő szerkezete, akkor a fordított folyamatot is lehet tanulmányozni [65, 66], amit dynamic undocking néven [67] említ az irodalom és sok olyan köztes állapotú kölcsönhatást lehet vele vizsgálni, ami más módon nem lehetséges [68, 69].

### **3.3. A hidratációs probléma és eddigi megoldásai**

A gyógyszertervezésben részt vevő molekulák vizes közegben hatnak egymásra, így szükség van a hidratáció megfelelő kezelésére is a modellezések során. A vízmolekulák a reaktáns molekulák felszínén, a molekulák között, a célpont zsebében, vagy akár a két felszín közötti összekötő szerepben is előfordulhatnak, valamint távol a komplextől, azzal kölcsönhatást nem kialakítva. Az

oldószer modellezése történhet implicit, explicit és hibrid módon, mindegyik módszernek van előnye és hátránya is. Az implicit modellek hatékonyak és gyorsak, ha az oldószer viselkedését kell modellezni, viszont az oldott molekulák körül nem adnak jó megoldást. Az explicit modellek kevésbé hatékonyak számítási szempontból, viszont jobban leírják az oldott molekulák körüli viselkedést.

A hidratáció nagymértékben befolyásolja a különböző biomolekulák és komplexeik szerkezetét és működését [70]. A komplex interfész vízmolekulái hidrogénkötések hálózatának kialakításával [71, 72] járulnak hozzá a célpont-ligandum stabilitásához és a célpont-ligandum specifitásához [73-80], meghatározva az atomok közötti távolságokat és kitöltve az üregeket [70, 81]. Fontosságuk ellenére a határfelületi vízmolekulák pozícióinak meghatározása nem triviális [25]. A rendelkezésre álló vízpozíciókat többnyire [82] röntgenkristallográfiával határozták meg, azonban még ez a jól bevált technika is számos korlátozással bír. Az elektronsűrűség csúcsainak meghatározása a lehetséges interfészvíz-pozíciókhoz még mindig nem rutinmunka a víz természetes mozgása és a szabadságfokok nagy száma miatt [83] és a szerkezet minősége függ az oldott komponensek méretétől is. A kristályban kialakuló hidratáció nem ugyanolyan, mint oldatban [84], amit tovább bonyolít a fagyasztás módszertana [84]. Az elektron-sűrűségadatok túlillesztése és a vízpozíciók félrevezető azonosítása gyakran rossz gyakorlatnak bizonyul [77]. Egyéb kísérleti technikák, mint például a magmágneses rezonancia spektroszkópia vagy a krioelektronmikroszkópia viszonylag kisszámú szerkezetet eredményezett, amelyek vízpozíciókat is tartalmaznak. A fenti kísérleti módszerek korlátainak leküzdése érdekében elméleti megközelítéseket dolgoztak ki a vízpozíciók meghatározására. Ezek a megközelítések vagy csak az oldott komplex szerkezetén [85], vagy kizárólag a víz-víz kölcsönhatások dinamikájának kiszámításán [86-93] alapuló vízpozíciókat adnak meg. Mi a molekuladinamikai alapon működő MobyWat [25] programot alkalmaztuk a célpont-ligandum interfészek hidratációs szerkezetének meghatározására.

*Implicit vízmodellek.* Az implicit modellekben az oldószer molekuláit egy homogén polarizálható közeggel helyettesítjük, ami jó közelítéssel állandó tulajdonságokkal rendelkezik [94]. Tehát nincsenek explicit oldószermolekulák explicit koordinátákkal, azaz az oldószert kevés paraméterrel le lehet írni kellő pontossággal. A legfontosabb paraméter a dielektromos állandó

( $\epsilon$ ), ami egyéb paraméterekkel, mint pl. a felületi feszültséggel is kiegészülhet. A dielektromos állandó határozza meg az oldószer polarizálhatóságának mértékét. Ezek a folyamatos (continuum) modellek széles körben elterjedtek, beleértve az erőter- és kvantumkémiai módszereket. A kvantumkémiaiában a töltéeloszlásokat ab initio módszerekkel számolják (Hartree-Fock, DFT), a matematikai megközelítésről részletesebben az alábbi tanulmányok írnak: [95-99]

Ezek a modellek jól használhatók, mivel egy egyszerű függvénnyel írják le az oldószert, nem változik számottevően a tiszta oldószerben, ha az oldószer maga nem vesz részt a reakcióban. Valamint akkor is jól alkalmazható módszert adnak, ha a számítógépes erőforrások korlátozottak. Implicit oldószermodeellt használtak reakciók vizsgálatában és előre tudták jelezni a hidratációs Gibbs-energia ( $\Delta_{\text{hyd}}G$ ) nagyságát is [99]. Több ilyen modell létezik, mint pl. a Polarizable continuum model (PCM), amely több változatban is létezik [96]. A modell a Poisson-Boltzmann-egyenletre épül, amely az eredeti Poisson-egyenlet kibővítése. A COSMO szolvatációs modell, amelyet mi is használtunk a számolásokhoz, népszerű implicit modell [100], a „scaled conductor boundary” határfeltételt használja, amely gyors és robusztus közelítés a pontos dielektromos egyenletekhez és csökkenti a külső töltési hibákat a PCM-hez képest [101]. A közelítések a valós értékektől 0.07 kcal mol<sup>-1</sup> nagyságrendű RMSD-értékkel (root mean square deviation) térnek csak el [102].

*Explicit vízmodellek.* Ezek a modellek minden egyes oldószermolekula molekuláris tulajdonságait figyelembe veszik a számolás során. Jóval valóságosabb modellek, amelyben a vízmolekulák az implicit modellekkel ellentétben az oldott anyaggal közvetlen, specifikus kölcsönhatásokban vannak. Ezek a modellek általánosan megjelennek MM-, MD- vagy Monte Carlo (MC)-szimulációkban és egyes kvantumkémiai számolások is alkalmazzák őket. MD-szimulációkban diszkrét időpontokban vizsgálható a rendszer időbeli fejlődése, ahol gyakran molekulamechanikai erőtereket használnak, amelyek empirikus, paraméterezett függvények; segítségükkel hatékonyan számolhatók nagy rendszerek tulajdonságai is [97, 98]. A paraméterezés magasabb szintű elmélet, vagy kísérleti adatok alapján történik. A MC-szimulációkkal a rendszer potenciális energiafelületeit a rendszer perturbálásával és a perturbáció utáni energia számolásával

kaphatjuk meg. Előzetes feltételek szükségesek annak eldöntésére, hogy az algoritmus elfogadja az újonnan perturbált rendszert vagy nem.

Az erőtermmódszerek általánosságban a kötés menti rezgéseket, torziókat, a repulziót és a diszperziót, mint a Buckingham- vagy Lennard-Jones-potenciált tartalmazó energiafüggvényeket használják. A gyakran használt oldószerek, mint amilyen a víz is, általában idealizált modellekkel írhatók le, amelyek lehetővé teszik, hogy a számolás során csökkentsék a szabadsági fokok számát anélkül, hogy számottevően romlana a pontosság; igaz, így bizonyos modellek csak meghatározott körülmények között használhatók. A leggyakrabban használt modellek a TIPXP, ahol az X egy egész szám, amely az energiaszámolásban használt oldalak számára utal [103] és a simple point charge model (SPC).

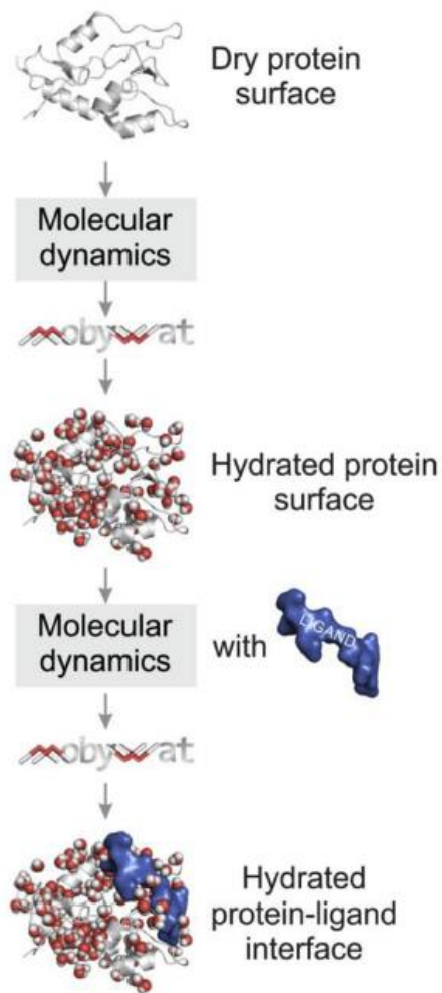
A polarizálható erőterek újabban már képesek a molekuláris töltéseloszlás változásait is számolni, amit a multipólusos momentummal rendelkező molekuláknál használnak, amely jobban tükrözi azok anizotrópiáját. Az egyik ilyen módszer az Atomic Multipole Optimised Energetics for Biomolecular Applications (AMOEBA) erőter [104], amelyet az ionok oldatbéli dinamikájának tanulmányozására használnak [94].

*Hibrid vízmodellek.* A hibrid modellek az implicit és explicit modellek „között” helyezkednek el, általában közelebb az egyikhez vagy a másikhoz. A vegyes kvantum- és molekulamechanikai (QM/MM) modellek közelebb állnak az explicit modellekhez. Pl. egy QM központi rész explicit oldószermolekulákat tartalmaz, amelyet körbevesz egy MM-réteg, azon kívül pedig egy harmadik réteg, amely a végtelen (bulk) oldószert jelképezi. A Reference Interaction Site Model (RISM) modell viszont közelebb áll az implicit modellekhez, amely lehetővé teszi, hogy az oldószert sűrűsége ingadozzon és így írja le az oldószert viselkedését [94, 96, 105].

QM/MM-módszerrel lehetőség van arra, hogy a rendszer egy részét, pl. egy biológiailag aktív helyet QM-, a rendszer többi részét MM-erőter segítségével modellezzük. Ha a harmadik réteget, a bulk oldószert implicit módszerrel modellezzük, „olcsóbban” tehetjük meg, mintha az összes oldószermolekulát explicit modelleznénk. Lehetséges megoldás még, hogy a QM-régióhoz explicit oldószert adunk, a többit pedig implicit módon kezeljük. Vegyes eredmények születtek azokban a tanulmányokban, ahol explicit oldószermolekulákat adtak implicit oldószertre. Vagy egyik, vagy

másik módszert használva jó közelítések születtek, de a vegyes modell vegyes eredményeket adott, ami főként az explicit hozzáadott oldószermolekulák számától függött [106].

*A hidratációs szerkezet szerepe és számítása.* A víz szerkezete és szerepe a technológia és tudomány egyik központi kérdése. Példákat találunk arra, hogy vízmolekulák között hidrogénkötések hálózata stabilizál aktív fehérjekonformációkat [107, 108] és elősegítik a feltekeredés folyamatát [109-111]. A vízmolekulák közvetítik az Alzheimer-kór b-amyloid protofilamentjeinek összekapcsolódását [112-115], részt vesznek a protein-RNS összekapcsolás folyamatában [116] és alapvető szerepet játszanak a protonátadási reakciókban [117-125], a fehérje belsejébe zárva pedig stabilizáló hatásuk van [126, 127]. A precíz szerkezetalapú gyógyszertervezésben figyelembe kell venni a vízmolekulák ligandumkötő hatását is [78, 89, 93, 128-137]. Statikus és dinamikus számítási módszerek is léteznek a fehérje felületén és az interfészben elhelyezkedő hidratációs hálózat előrejelzésére. Statikus megközelítésű módszer lehet például tudás alapú [138, 139], szerkezeti [140, 141] és dokkolás alapú [142]. A molekuladinamika az explicit vízmodellekkel együtt képes kezelni a vízmolekulák minden kölcsönhatását, beleértve nemcsak az oldott víz, hanem a víz-víz kapcsolatokat is, valamint kezeli az oldószer és a kötött vizek közti kicserélődést is. A MD egy hatékony motorrá vált a fehérjecélpontok vízszerkezeteinek és a ligandumokkal alkotott komplexeinek hidratációs szerkezetének meghatározásában [143-146], valamint széles körben alkalmazzák gyógyszertervezésben [18, 63, 147, 148] és a fehérjék közti kölcsönhatások vizsgálatában [55] is. A hardver- és szoftvertechnológiák folyamatos fejlődése következtében lehetőségessé vált a nagy molekuláris rendszerekben időben bekövező szerkezeti változások modellezése. A kovalensen nem kötő kölcsönhatások GPU-számolása [149-151] és a dedikált hardverek megjelenése [152, 153] ugyancsak kiszélesítették az MD-számolások határait.



**2. ábra** Fehérjék és a fehérje-ligandum interfészek hidratációs szerkezetének standardizált meghatározása MD-szimulációval és MobyWat-lépésekkel. Az interfésznél a MD- és MobyWat-lépéseket a ligandum molekula jelenlétében végezzük. Forrás: Jeszenői et al. 2018.

A hidratációs szerkezet jelenlegi MD-alapú prediktorait tovább lehet kategorizálni. Vannak megközelítések, amelyekkel a fehérje körüli hidratációs héj előre jelezhető, amely az átlagos oldószersűrűséget és a proximális sugárirányú eloszlásfüggvényt használja MD-ból származó különböző atomtípusok körül [144, 154-159], vagy amelyek elemzik a makromolekulák körüli hidratációs szerkezeteket [157, 159]. Más, foglaltságon alapuló algoritmusok az összes atomot tartalmazó MD-t alkalmaznak explicit oldószermolekulákkal, így vizsgálja [160] vagy jelzi előre a hidratációs helyeket [93, 143, 161-166]. Az újabb MD-algoritmusok a hidratáló vízmolekulák egyedi pozícióit számolják az átlagos sűrűség helyett, és különféle foglaltság-alapú értékelési sémákat alkalmaznak a hidratációs helyek meghatározására, mint például az időben átlagolt

pozíciók [165] és/vagy az azonosításon alapuló klaszterezés esetén a mobilitási pontszámok (score) [167]. Ez utóbbi megközelítés hasznosnak bizonyult a fehérjefelületek teljes hidratációs szerkezetének [167] és a célpont-ligandum komplexek interfész hidratációs hálózatainak előrejelzésében [71], ami egy önálló programban (MobyWat) valósult meg [25] (**2. ábra**). A program által megjósolt hidratációs szerkezetek átlagosan 81%-os egyezést mutattak a kristályos vízpozíciókkal és 90%-os egyezést a fehérjefelületek és az interfészek esetében.

Bármelyik megközelítést is vizsgáljuk, a megjósolt hidratációs szerkezetek pontosságát nagymértékben befolyásolhatják a MD-szimulációk különböző paraméterei. Az intenzív paraméterek, mint pl. a nyomás és a hőmérséklet, a helyes erőter kiválasztása vagy az explicit vízmodell mind meghatározzák az előrejelzés pontosságát. Például az explicit vízmodelleket széles körben alkalmazzák biomolekuláris szimulációkban és számos ilyen oldószermodellt fejlesztettek ki [168]. Gyakran azonban nem triviális, melyik vízmodell a legmegfelelőbb bizonyos szimulációkhoz és a víz-fehérje és a víz-víz energetikáinak kiszámításához. Kimutatták, hogy a vízmodell szignifikánsan befolyásolhatja az A-RNS duplexek [169-173], a peptid- és fehérjetekeeredés [169-173] szimulálását, a hidratáció termodinamikai számolását [174], valamint a ligandumok fehérjékhez történő kötési szabadenergiájának eredményeit [174, 175].

### **3.4. Fragmens alapú számítások a kvantummechanikában**

A hidratáció mellett a rendszer mérete is kihívás lehet, az ilyen vizsgálatok QM-szinten nagy számítógépes erőforrást igényelnek; ha nagyméretű célpont molekulát szeretnénk számolni, a célpont-ligandum komplex kezelhető alrendszerékké, kisebb alkotóelemekre bontása oldhatja meg ezt a problémát. Legalább két megközelítés létezik: az első megközelítés kvantummechanikát alkalmaz a kötési helyre és molekulamechanikai szimulációt a komplex többi részére [176-181]. A módszerek egy másik ága a célzott fragmentáláson alapul és kvantummechanikát alkalmaz a célpont fragmensek és a ligandum alrendszerére. Zhang és Zhang [182] például egy olyan eljárást dolgoztak ki a molekuláris fragmentálásra, ahol a fehérjét egyedi zárt fragmensekké bontották és ab initio Hartree-Fock- és DFT QM-számításokat hajtottak végre a célpont-ligandum komplexeken. Nikitina és munkatársai [183, 184] a célpont nehéz atomjait vágta ki, amelyek 5 Å, vagy annál kisebb távolságra vannak a ligandum nehéz atomjaitól. Röntgen-analízissel meghatározott szerkezeti vízmolekulákat is alkalmaztak, újakat helyeztek el az oldott molekulák

H-kötő vegyértékei alapján [184] és az in silico hidratáció iteratív módszerét is alkalmazták [183]. Korrelációt kaptak a kötési entalpiára ( $\Delta H_b$ ) tizenkét [183] és nyolc [184] komplex esetében. A komplexek fehérjecélpontokat és kis molekulatömegű (700-as tömegszámig) ligandumokat tartalmaztak és a számításokat szemiempirikus QM-szinten végezték PM3 paraméterezéssel. Dobes, Hobza és munkatársai [185] a kis molekulájú purin inhibitor Roscovitine-t tanulmányozták CDK2-vel alkotott komplexében B3LYP/6–31G\*\* és MP2 elméleti szinteken. A kináz célpont láncait néhány aminosav kis fragmenseivé vágta a  $C_{\alpha}$ -N kötés mentén, a peptidkötést megtartották és csak a ligandumtól 5 Å-ön belül található aminosavakat és kristályvíz-molekulákat vették figyelembe a számolások során.

A termodinamikai tulajdonságok, például a kötési entalpia ( $\Delta H_b$ ) szerkezeti alapú számítása a hatékony gyógyszerjelöltek tervezésének központi kérdése. Az új molekulák entalpiikus optimalizálása [186-188] a gyógyszertervezés jól ismert útja és szükségessé teszi a célpont-ligandum komplexek  $\Delta H_b$  meghatározását vagy előrejelzését. Annak ellenére, hogy ilyen adatokra, mint a  $\Delta H_b$ , nagy szükség van, csak néhány QM-vizsgálat létezik a célpont-ligandum komplexek termodinamikájának fragmens alapú számolására, ahogyan azt az előző bekezdésben említett cikkek is alátámasztják. Nem nagyméretű ligandum-molekulák vettek részt a rendelkezésre álló vizsgálatokban és a nagy (peptid) ligandumok meghatározott számú hidratációs helyet tartalmazó komplexeit nem vizsgálták széles körben. Ezen túlmenően az összetett interfészek szerkezeteinek kivágására és a megbízható hidratációs módszerekre szolgáló automatizált eszközök kifejlesztése szintén hasznos lenne az ilyen fragmens alapú QM-vizsgálatokhoz. Az új gyógyszerjelöltek entalpiikus tervezésének elősegítése és a nagy ligandumok komplexszerkezeteinek automatizált számolásával kapcsolatos fenti kihívások megoldása érdekében kifejlesztettünk egy új fragmentáló protokollt, amely magában foglalja az explicit vízmolekulák szisztematikus elhelyezését, majd egy teljesen automatizált fragmentálási módszert a komplex interfészek megfelelő kezeléséhez QM-számítások számára. A validálást kémiaiag különböző ligandumokkal és ismert kísérleti  $\Delta H_b$ -értékekkel rendelkező komplexek csoportjával végeztük el.



## 4. Módszerek

A jelen fejezetben kizárólag a tézispontokban összefoglalt eredményekhez közvetlenül csatlakozó módszereket ismertetem.

### 4.1. A módszerfejlesztés informatikai háttere

A fragmentáló algoritmus (Horváth et al. 2019, I. **tézispont**) programozási nyelvének és környezetének megválasztásakor fontos szempont volt, hogy úgy terveztük, egy weboldalon is közzéteesszük, bárki számára elérhetővé tesszük azt. Azaz valamilyen front end-programozásra alkalmas környezetet kellett keresni, ami lehetőség szerint olcsó (vagy ingyenes) és elterjedt, valamint megfelelő támogatás is van hozzá. Szerverként Windows vagy Linux jöhetett szóba, de mivel a Windows és a hozzá tartozó .NET-keretrendszerben használható fejlesztőkörnyezet nem ingyenes, vagy ingyenesen nem tud mindent, amire szükség lehet, így a Linuxra esett a választás. A fejlesztés első szakaszában tárhelyszolgáltatótól bérelt területen dolgoztunk, ahol nem volt teljes jogosultságunk, így a HTML-oldallal (**H**yper**T**ext **M**arkup **L**anguage) kényelmesen használható PHP és JavaScript tűnt jó megoldásnak, mert teljesen meg lehetett írni a háttérben fragmentáló algoritmust PHP-ban. A PHP ugyanis rengeteg függvénnyel támogatja a szövegfájlok, sztringek kezelését és mivel inputként PDB-fájlokat (Protein Databank [189]) olvas be az algoritmus, remek választás lett a PHP. Igaz, hogy időközben olyan funkciókat is beépítettünk, amelyek pl. programokat futtatnak a szerveren, így saját, magunk által konfigurált, teljes hozzáférést és jogosultságot nyújtó, felhőben futó Linux-szervert hoztunk létre és oda költöztettük a teljes rendszert.

A HTML és a JavaScript (amely tartalmazza a JavaScript könyvtárakat, mint például a jQuery vagy a Scriptaculous) kliensoldali nyelvek. Ez azt jelenti, hogy a böngésző felelős a HTML- és a JavaScript- kód végrehajtásáért. Ez az oka annak, hogy miért kell tesztelnünk weblapjainkat annyiféle böngészőben, amennyiben csak lehetséges, mivel a böngészőnek kell végrehajtania a HTML-t, a JavaScript-et és a CSS-t, így mindegyik böngésző más végeredményt adhat. A tesztelést Windows alatt Chrome-on, Internet Explorer 11-en és Firefox-on, Linux alatt Chromium-on és Firefox-on, Mac OS-en Safari és Chrome böngészőben végeztük.

A PHP-kód a gazdaszámítógépen vagy a kiszolgálón fut le, mielőtt az oldal a böngészőben megjelenik. A PHP felelős azért, hogy a dolgokat „áthúzza” a kiszolgálóról, legyen az egy adatbázis

vagy egy fájl és létrehozza és/vagy biztosítja a megfelelő HTML, JavaScript és/vagy CSS kódot. A HTML-, a JavaScript- és a CSS- kódok jórészt „hardkódolva” vannak, ami azt jelenti, hogy a kód nem változik. A fejlesztő létrehozza a HTML-t, a JavaScript-et és a CSS-t, majd feltölti azokat a szerverre. A szerver, ha ezt a weboldalt kéri le, csak elküldi az oldalt a böngészőnek, ahogyan azt a fejlesztő lekódolta.

Viszont ha dinamikus oldalra van szükségünk, ami azt jelenti, hogy a megjelenítendő tartalom függ a felhasználó által megadott inputtól, szerveroldalon olyan szkriptnyelvet kell alkalmazni, mint pl. a PHP. A PHP-t HTML-, JavaScript- és néhány esetben a CSS- kód generálására használják. A PHP összegyűjti és feldolgozza azokat az információkat, amelyekre szüksége van ahhoz, hogy a weboldalt HTML-ben és JavaScript-ben hozza létre, mielőtt elküldi a kész weblapot a böngészőbe. *HTML.* A weboldalak készítéséhez a HTML nyelvet használtuk, amely mára internetes szabvánnyá vált. A webböngészők webszerverről vagy helyi tárolóból HTML-dokumentumokat kapnak és a dokumentumokat multimédiás weboldalakon jelenítik meg. A HTML szemantikailag leírja a weblap felépítését és tartalmazza a dokumentumok megjelenítésére vonatkozó utasításokat.

A HTML-oldalak alapegységei az elemek. Képeket és más objektumokat, például interaktív űrlapokat lehet beágyazni a megjelenített oldalra. Olyan strukturált dokumentumok létrehozására szolgál amelyen fejlécek, bekezdések, listák, hivatkozások, idézetek és egyéb elemek lehetnek. A HTML-elemeket tag-ek határozzák meg, '`<`' és '`>`' jelek közé írva. Az olyan címkék, mint a `<img />` és a `<input />`, közvetlenül megjelenítik a tartalmat az oldalon. Más tag-ek, mint például a `<p>` információt tartalmazhatnak a dokumentum szövegéről és más tag-eket is tartalmazhatnak. A böngészők nem jelenítik meg a HTML-tag-eket, hanem felhasználják őket, hogy értelmezzék az oldal tartalmát.

A HTML beágyazhat olyan szkriptnyelvre írt programokat, mint a JavaScript, ami a weboldalak viselkedését és tartalmát módosíthatja és dinamikussá teheti, ellenőrizheti a felhasználó által megadott inputokat. A CSS használatával a tartalom megjelenését és elrendezését is megadhatjuk. A World Wide Web Consortium (W3C), amely mind a HTML, mind a CSS szabványok karbantartója, 1997 óta kifejezett ösztönözi a CSS használatát a HTML-ben.

*JavaScript.* A JavaScript (JS) egy magas szintű programozási nyelv, amely megfelel az ECMAScript-specifikációnak. Egy dinamikus, gyengén típusos, prototípus alapú programozási nyelv. A HTML

és a CSS mellett a JavaScript a világháló egyik legfontosabb technológiája. A JavaScript lehetővé teszi az interaktív weboldalak használatát és a webes alkalmazások lényeges részét képezi. A weboldalak túlnyomó többsége használja és a nagyobb webböngészőknek van egy külön erre a célra szolgáló JavaScript-motorja, amely végrehajtja azt.

A JavaScript támogatja az esemény-vezérelt, funkcionális és kényszerítő (beleértve az objektum-orientált és prototípus alapú) programozási stílusokat. A szövegek, tömbök és dátumok használatát API-kon keresztül támogatja, de maga a nyelv nem tartalmaz I/O-, például hálózati, fájlkezelési vagy grafikai eszközöket. A futtató környezetre támaszkodik, amelybe be van ágyazva, amin keresztül ezeket a funkciókat biztosítja.

Eredetileg a webböngészőkben kliensoldali szerepe volt, de a JavaScript-motorokat ma már sok más típusú host szoftverekbe ágyazzák be, beleértve a szerveroldalt a webszerverekben és adatbázisokban, valamint nem webes programokban, például szövegszerkesztőkben és PDF-szoftverekben és olyan futásidejű környezetekben, amelyek lehetővé teszik a JavaScriptet mobil- és asztali alkalmazások megírására, beleértve az asztali widgeteket is.

A magas szintű, sokféle struktúrát tartalmazó nyelvek közé tartozó Java programozási nyelvet választottuk az FBD-algoritmus (Bálint et al. 2019, **V. tézispont**) összekötő algoritmusának implementálásánál, ugyanis a teljes számolás túlnyomó részét az egyes kötések mentén végzett forgatás alkotja. Ehhez a forgatáshoz szükséges ismerni a forgástengelyen kívül minden egyes kapcsolódó atom koordinátáit és kapcsolódását a szomszédos atomokhoz, amit molekulagráf struktúra/típus segítségével könnyen kezelni tudtunk.

## **4.2. A Fragmenter eljárás és a kötési entalpiaváltozás számításának módszerei**

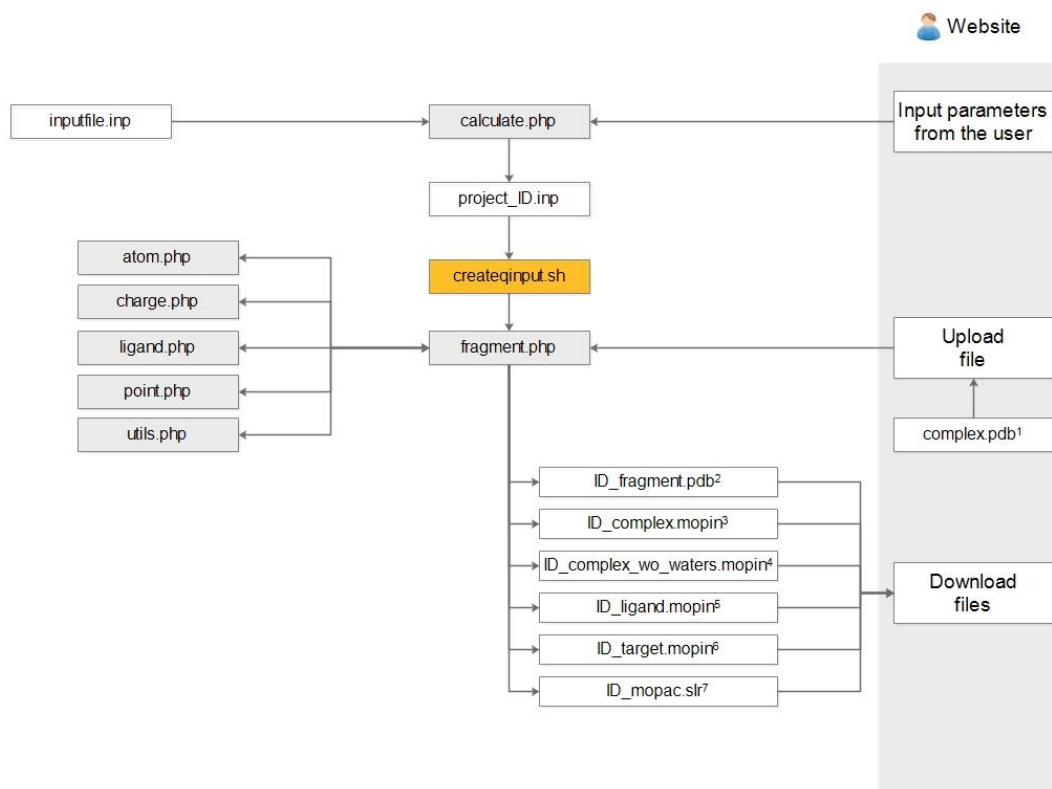
Ebben a szakaszban az első publikáció (Horváth et al. 2019) **I-III. tézispontjainak** egyes módszereiről írok.

### **4.2.1. A Fragmenter technikai megvalósítása**

A Fragmenter automatikusan vágja ki a nagyméretű komplexek célpont-ligandum interfészeit, ez az eljárás a [www.fragmenter.xyz](http://www.fragmenter.xyz) webhelyen bárki számára szabadon elérhető. Az algoritmus részleteit, az input, az algoritmus, a megvalósítás, az implementáció és az output szkriptek közötti kapcsolatokat a **3.** és az Eredmények fejezetben az **5. ábra** mutatja be: a komplex szerkezetet célpontra, ligandumra és a vízmolekulákra választja szét az algoritmus. A **3. ábrán** látható

adatfolyam mutatja, hogy hogyan történik a felhasználó által megadott inputok feldolgozása, illetve milyen folyamatok után kapja meg a számára szükséges outputokat.

A Fragmentert parancssorból a createqinput.sh shell szkripttel lehet indítani, de a webszerveres megoldásban a felhasználó által megadott adatokat az oldalról át kell irányítani, így ott a calculate.php állítja össze a felhívás paraméterlistáját és hívja meg a Fragmenter-t. Mivel a Fragmenter eljárás a dolgozat eredménye, ezért további részleteit az Eredmények részben mutatom be.



**3. ábra** Adatfolyam a Fragmenterben (<sup>1</sup>A felhasználó által feltöltött komplex szerkezet, <sup>2</sup>Interfész szerkezet PDB, <sup>3</sup>Interfész szerkezet MOPAC input, <sup>4</sup> Interfész szerkezet MOPAC input vízmolekulák nélkül, <sup>5</sup>Ligandum PDB, <sup>6</sup>Célpont PDB, <sup>7</sup>SLURM input fájl)  
Forrás: Horváth et al. 2019.

#### 4.2.2. A képződéshő számítása

A MOPAC 2012 [190] programcsomagot használtuk szerkezeti relaxációra és a kivágott komplex szerkezetek, valamint külön a ligandum, a vízmolekulák és a célpont fragmentek képződéshőjének kiszámítására Parametric Method number 7 (PM7 [191]) Hamiltonian módszerrel. A kilépési feltétel 1.0-s gradiens érték volt, amit a MOPAC Support Team ajánlása

alapján állítottunk be és amely kisebb, mint a Manual [190] által ajánlott 10-es érték volt. A számítási költség további csökkentése érdekében a Mozyme [192] LMO-megközelítését alkalmaztuk, a molekulák teljes nettó töltését az aminosavak egyéni nettó töltéseiből számítottuk ki (**5. táblázat**). A töltéseket a parancssorban is jeleztük, manuálisan és automatikusan a GEO-OK kulcsszóval ellenőriztük. A számolások nem kívánt megszakítását a PREC kulcsszóval értük el, az alapértelmezett geometriaoptimalizáláshoz Eigenvector Followig-ot [193], a peptidkötések molekulamechanikai korrekciójára az MMOK kulcsszót használtuk. A vákuumos, explicit vizes számolások kivételével a COSMO (CONductor-like ScreeningMODEL) implicit vízmodell [100] alkalmaztuk, ehhez 78.3-as értéket állítottunk be az EPS kulcsszóra, amely a víz dielektromos állandója 293.15 K-en és 101325 Pa nyomáson. A víz  $\Delta_f H$ -értékét is kiszámoltuk vákuumban és COSMO-moddal is a fenti kulcsszavak használatával, a 3ptb\_ben, 3ptb\_pad, 3ptb\_pam, 3ptb\_pme és 3ptb\_pmo rendszerekben a diszulfidhidak S atomjainak koordinátáit a COSMO számítások során rögzítettük, így tudtuk megőrizni a szerkezetek integritását.

#### 4.2.3. Célpont–ligandum intermolekuláris kölcsönhatási energia

Az energiaminimalizált célpont-ligandum komplexekre az intermolekuláris kölcsönhatási energiákat is kiszámoltuk, amelyeket a Lennard-Jones (LJ) és a Coulomb-potenciálok [194] összegeként írhatunk fel. Mind a LJ-, mind a Coulomb-potenciáloknál az Amber99sb-ildn erőter paramétereit használtuk [195]:

$$E_{inter} = \sum_{i,j}^{N_T N_L} \left[ \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} + \frac{q_i q_j}{4\pi\epsilon_0\epsilon_r r_{ij}} \right]$$

$$A_{ij} = \epsilon_{ij} R_{ij}^{12}$$

$$B_{ij} = 2\epsilon_{ij} R_{ij}^6$$

$$R_{ij} = R_i + R_j$$

$$\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j}$$

$$\epsilon_r = C + \frac{D}{1 + k e^{-\lambda B r_{ij}}}$$

$$C = \epsilon_0 - D; \epsilon_0 = 78.4; D = -8.5525; k = 7.7839; \lambda = 0.003627$$

ahol  $\epsilon_{ij}$  a potenciálgödör mélysége egyensúlyi állapotban az i-edik és a j-edik atomok között,  $\epsilon_0$  a víz dielektromos állandója 25° C-on;  $R_{ij}$  a az atommagok közötti távolság egyensúlyi állapotban az

i-edik (ligandum) és a j-edik (célpont) atomok között;  $q$  az atom parciális töltése az AMBER99SB-ILDN erőterben;  $r_{ij}$  az i-edik (ligandum) és j-edik (célpont) atomok aktuális távolsága;  $N_T$  a célpontok atomjainak száma;  $N_L$  a ligandum-atomok száma.

#### 4.2.4. Statisztika

Egyszerű lineáris regressziót végeztünk a számított  $\Delta_r H$  és a kísérleti  $\Delta H_b(\text{exp})$  értékek között az Eredményekben a **6-8. táblázat, Függelék A. és B. táblázat** alapján. A  $\Delta H_b(\text{exp})$  értékeket a **2. táblázatban** felsorolt különböző publikációkból nyertük. A regressziók statisztikai paraméterei, beleértve a regressziós együtthatókat ( $\alpha$  és  $\beta$  a **3. egyenletben**),  $R^2$ -,  $t$ -,  $F$ -értékeket, reziduálisokat és RMSD-értékeket a **6. táblázatban** mutatom meg (Horváth et al. 2019, **II. tézispont**). A korrelációs görbékhez  $\Delta H_b(\text{calc})$  értékeket számítottunk a  $\Delta_r H$  értékek és a regressziós együtthatók segítségével.

#### 4.3. A fragmens blind docking eljárás módszerei

Ebben a szakaszban a második publikáció (Bálint et al. 2019) **IV-V tézispontjainak** egyes módszereiről írok.

##### 4.3.1. A Célpont–ligandum intermolekuláris kölcsönhatási energia (Einter) számítása

Ahogy a 4.2.3. fejezetben, az FBD-eljárás során kapott szerkezetek intermolekuláris kölcsönhatási energiáját is számoltuk.

##### 4.3.2. Összekapcsolás és hegesztés

Az összekapcsolás és összefűzés algoritmusai Java 1.8-as verzióban készültek FragmentMerge néven. A parancsfájl az **1. táblázatban** leírtak szerint futtatható és a csomagolásból származó PDB-fájlokat használja inputként. Van egy bemeneti szövegfájl is, amely tartalmazza a rendszer nevét (ez majd a riportban lesz felhasználva) és a fragmensek sorrendjét is (**1. táblázat**).

Az algoritmus osztályhierarchiát használ, a fragmenseket, a fragmensek atomjait és kötéseit és a PDB-fájlokat külön osztályok reprezentálják. A hegesztéshez szükséges forgatáshoz az atomokat egy molekulagráfban tárolja az algoritmus a memóriában, ami segít az új koordináták kiszámításában a forgatási folyamat során; az outputokat külön könyvtárba menti párok, triádok, tetrádok, pentádok, stb. formájában. Az összekötött fragmenseket külön PDB-fájlokban mentjük

el, az összekapcsolási és hegesztési folyamatról egy riport minden információt (bemenetek, kimenetek, paraméterek és elérési utak) tartalmaz.

**1. táblázat** A FragmentMerge parancssori paraméterei\*

Paraméter	Leírás
-i	Input könyvtár
-p	Output és riport könyvtár
-m	Minimum távolság a nehéz atomok között
-x	Maximum távolság a nehéz atomok között
-n	Minimum távolság az összekötendő N- és C- atomok között
-a	Forgatás szöge

```
java FragmentMerge -p <string> -m <real> -x <real> -n <real> -a <real> -i <string>
```

A példa input textfájl tartalma a rendszer és a célpontok nevével:

C2

AR

TK

QT

AR

KS

A fragmensszám és az inputfájlok neve a REMARK sorokban található, a fájlok neve pedig a tartalomra vonatkozik. A hegesztő algoritmusnak az atomok közötti kötési információra is szüksége van, ehhez az Open Babel [196] segítségével a PDB-fájlokat újra PDB-fájlba konvertálva megkapjuk a kötési listát is.

#### 4.4. A vízhálózatok számításának módszerei

Ebben a szakaszban a harmadik publikáció (Jeszenői et al. 2018) **VI. tézispontjának** egyes módszereiről írok. A MobyWat NetDraw módját használtuk a fehérje felületének vagy a célpont-ligandum interfész régiója hidratációs hálózatának elemzésére és jellemzésére, amely meghatározza az interfész-víz-molekulák kölcsönhatásait egymással, az oldószer vízzel és az oldott partnerekkel (ligandum + célpont) és olyan hálózati gráfot eredményez, amelynek mobilitása az összes vízcúcsához hozzá van rendelve. Így egy vízmolekulát vagy a célpont vagy a ligandum egy aminosavát csúcsnak tekintjük. Az célpont és a ligandum energiaminimalizált szerkezetét tartalmazó PDB-fájlokból, beleértve az előrejelzett felszíni vagy az interfész és az oldószer víz

molekuláit, bemenetként a NetDraw elkészíti a felület kétdimenziós kölcsönhatási hálózati gráfját élek és csúcsok listájaként. A gráfok éleinek listáját a távolságfájlokból finomítjuk azáltal, hogy kiküszöböljük a redundanciákat és a szénatomoktól való távolságot. Csúcsonként az élek számát négyre korlátoztuk, csak a legelső négy legrövidebb élt használtuk.

A csúcsok listája az élek listájából, az összes lista oldószer vízcúccsal és nélkülük is elkészül, majd a NetDraw előállítja a csúcsok és élek osztályozását. Egy csúcs akkor statikus, ha:

- ligandum- vagy célpontcsúcs
- vagy négy bármilyen típusú csúcshoz csatlakozik
- vagy legalább három ligandum- vagy célpontcsúcshoz csatlakozik
- vagy 50-nél kisebb mobilitási értékkel rendelkezik

Egyébként a csúcs dinamikus. Egy él statikusnak minősül, ha két statikus csúcsot köt össze, különben dinamikus. Az alhálózatok élekből épülnek fel, a csatlakoztatott statikus/dinamikus élek statikus/dinamikus alhálózatokat eredményeznek. A fenti osztályozással a NetDraw segít megkülönböztetni a teljes felszíni statikus és dinamikus alhálózatokat vagy az interfaciális hálózatot. A MobyWat a vízhálózati gráfot a csúcsok és élek mobilitási információival látja el olyan fájlformátumban, amelyet hálózatmegjelenítő és elemző programok által általánosan használnak. Mi a Gephi programot [197] használtuk a NetDraw által létrehozott gráfok megjelenítésére, az elrendezést a ForceAtlas elrendezés opcióval hoztuk létre.



## 5. Eredmények és értékelésük

E fejezetben három, a dolgozat alapját képező közlemény fontosabb eredményeit tárgyalom. Amint a dolgozat címe is utal rá, e közlemények a célpontmolekula-ligandum komplexek számítására fókuszálnak. Az első közleményben (Horváth et al. 2019) a fragmentálós technikát a célpont molekula oldalán, a másodikban (Bálint et al. 2019) pedig a ligandum oldalán ötvöztük új eljárásainkban. Az első és a harmadik (Horváth et al. 2019, Jeszenői et al. 2018) közlemény is kiterjedten foglalkozik a komplexek hidratációjának számításával, amely a fragmens-alapú technikákat kiegészítve járult hozzá a számítások pontosságának növeléséhez és az eredmények atomi szintű értelmezéséhez.

### 5.1. Fragmenter eljárás és a kötési entalpiaváltozás számítása

A célpont-ligandum komplexek szerkezetének és kötési termodinamikájának meghatározása kulcsfontosságú lépés a gyógyszertervezésben [198]. A termodinamikai mennyiségeket kísérleti módszerekkel, például izotermikus titrálási kalorimetriával (ITC [199-208]) lehet mérni. A kísérleti méréseket gyakran korlátozza a nagy tisztaságú és koncentrált célpontfehérje-minták hiánya és magas költsége. A molekulaserkezetek és a kötés termodinamikája ugyancsak gyors és olcsó molekulamechanikai módszerekkel megjósolható [209-213]. Ugyanakkor a molekulamechanikának komoly korlátai vannak az összetett szerkezetek elektronikus hatásainak kiszámításában. Az ilyen hatások szinte minden intermolekuláris kölcsönhatásban jelen vannak, beleértve az „egzotikus” eseteket, mint például az aromás és töltött oldalláncok kation- $\pi$  kölcsönhatásai [201, 214], vagy a szerkezeti vízmolekulák polarizációs hatásai [215]. A QM-módszerek megfelelően tudják kezelni az intermolekuláris kölcsönhatások elektronikus hatásait, viszont a célpont-ligandum komplexek hidratációja és nagy mérete további kihívásokat vet fel a QM-módszerek számára, amint azt a következő részekben leírom.

Az első közleményben (Horváth et al. 2019) egy fragmensképző protokollt (Fragmenter) valósítottunk meg (**I. tézispont**) egy olyan webes szolgáltatásban, amely szabadon elérhető a tudományos közösség számára. A protokollt a célpont-ligandum komplexfelületek szerkezeteinek szabványosított és automatizált kivágására alkalmaztuk. A kivágott interfészszerkezetek energiáját minimalizáltuk és a képződéshőt szemiempirikus QM-módszerrel számoltuk. A

komplexek határfelületeinek csökkentése ésszerű számítási időn belül lehetővé tette a különböző vízmodellek szisztematikus vizsgálatát. Jelentős korrelációt sikerült elérni egy hibrid oldószer modellel (**II. tézispont**), amely szisztematikus meghatározott explicit és implicit vizeket is tartalmazott. Ennek eredményeként egyetlen, statisztikailag szignifikáns skálázó faktort kaptunk, amivel a számított képződéshőt kísérleti kötési entalpiás értékekké konvertálhatjuk (**III. tézispont**).

A vizsgált rendszerek csoportja különböző kémiai ligandumokat tartalmazott, beleértve kis aromás vegyületeket és nagy peptideket is, amelyek molekulatömege 121 és 3318 közötti tartományba esett és a kísérleti kötési entalpiák két nagyságrendben, 2.3 és 24.5 kcal mol<sup>-1</sup> között mozogtak. A különböző oldószermodellekkel végzett szisztematikus vizsgálat kimutatta, hogy a komplex felületén pontosan előrejelzett vízmolekulák jelenléte javította a számított és kísérleti adatok közötti korrelációt. Végző modellünkben egyetlen konstans tényezőt vezettünk le a szerkezeten alapuló, QM-mel számolt képződési entalpia és a kísérleti kötő entalpiák közötti összefüggésre (**III. tézispont**). A Fragmenter a komplex szerkezet kivágásával jó korrelációt ad a kísérleti és az elméleti eredmények között és egyidejűleg nagyban csökkenti a számítási időket. A protokollt egy olyan webszerveren valósítottuk meg, amely bárki számára rendelkezésre áll.

#### 5.1.1. Végpont-megközelítés

Egy ligandummolekulának a célponthoz történő kötését formálisan az **1. egyenlet** írja le. Mivel a reakció biológiai környezetben történik, T és L vízmolekulák hidratálják a célpontot és a ligandumot. A vizek maradhatnak a helyükön ( $s = 0$ ), csatlakoznak a komplexhez a környező oldószerből ( $s > 0$ ), vagy el is hagyhatják a komplexet ( $s < 0$ ) a kötés kialakulása során. Az **1. egyenlet** reakcióhőjét ( $\Delta_r H$ ) a Hess-törvény (**2. egyenlet**) alapján lehet kiszámolni [183, 184, 211, 212, 216-218], ahol  $\Delta_f H$  a reagens vagy egy termék számított képződéshőjét jelenti.



$$\Delta_r H = \Delta_f H(\text{Target:Ligand}[\text{H}_2\text{O}]_s) - \Delta_f H(\text{Target}) - \Delta_f H(\text{Ligand}) - s \Delta_f H(\text{H}_2\text{O}) \quad (2)$$

Ez a végpont-megközelítés egyszerű és sikeresen alkalmazták korábbi publikációkban [183, 184, 211, 212, 216-218], ezenkívül lehetővé teszi az egyes vízmolekulák hozzájárulásának vizsgálatát. Ugyanakkor a komplexekben a vízmolekulák eredetének világos meghatározása és a ligandumhoz kötött vízmolekulák előrejelzése meglehetősen bizonytalan a ligandumok viszonylag kis kötési

felülete miatt, így  $T = L = 0$  értéket állítottunk be és az összes interfész vízmolekulát úgy tekintettük, mintha a környezetből származnának ( $s > 0$ , az **1. és 2. egyenletben**). A cépont-, ligandum- és vízmolekulákat tartalmazó komplex hidratált szerkezeteket a Módszerekben leírtak szerint kaptuk. A végpont-megközelítésen belül a reakció résztvevőinek  $\Delta_f H$  számítása (**2. egyenlet**) és  $\Delta_r H$  lineáris skálázása (**3. egyenlet**) az ismert kísérleti  $\Delta H_b$  (exp) értékekre szükséges ahhoz, hogy  $\Delta H_b$ -t ki tudjuk számolni.

$$\Delta H_b(\text{exp})_i = \alpha \Delta_r H_i + \beta + \varepsilon_i = \Delta H_b(\text{calc})_i + \varepsilon_i, \quad (3)$$

ahol  $i = 1, 2, \dots, N$

Tizenöt rendszert ( $N = 15$ ) (**2. táblázat**) vizsgáltunk a regressziós együtthatók ( $\alpha$ ,  $\beta$ ) meghatározására, amelyek eredményei a  $\Delta H_b(\text{calc})$  számolt értékek és reziduálisok ( $\varepsilon$ ) voltak. A tizenöt rendszerből kilencet, amelyek kis ligandumokkal rendelkeznek 550 MW-ig, egy korábbi tanulmányban [184] már vizsgálták. Ebben a tanulmányunkban további hat, nagy peptid ligandumot tartalmazó rendszert is felvettünk a vizsgált rendszerek közé, mivel méretük és kiterjedt hidratációjuk miatt kihívást jelentenek az optimalizálás során. Így a tizenöt rendszer jelenlegi csoportja különböző ligandumokat tartalmaz, amelyek tömege 3321-ig terjed, ami két nagyságrenddel nagyobb, mint az eső csoportban. A megfelelő kísérleti  $\Delta H_b$ -értékek széles tartományt ölelnek fel -2.935 és -15.5 kcal/mol között (**2. táblázat**).

**2. táblázat** A vizsgált célpont-ligandum rendszerek

Rendszer <sup>a</sup>	Felb. <sup>b</sup> (Å)	Célpont	Ligandum		Vizek száma			$\Delta H_b(\text{exp})^d$
			Név	MW <sup>c</sup>	Shell 1	Shell 2	Shell 3	kcal mol <sup>-1</sup>
3ptb_ben	1.7	beta-trypsin	benzamidine	121.2	1	6	7	-4.507[199]
3ptb_pme	1.7	beta-trypsin	p-methylbenzamidine	135.2	1	5	6	-4.412[199]
3ptb_pam	1.7	beta-trypsin	p-aminobenzamidine	136.2	3	4	7	-6.417[199]
3ptb_pmo	1.7	beta-trypsin	p-methoxybenzamidine	151.2	1	6	7	-3.742[199]
3ptb_pad	1.7	beta-trypsin	p-amidinobenzamidine	164.2	2	8	10	-2.935[199]
1k1l	2.5	bovine trypsin	NAPe-piperazine	467.6	5	10	15	-7.863[201]
1k1m	2.2	bovine trypsin	NAP <sup>e</sup> -4-acetyl-piperazine	508.6	4	12	16	-8.222[201]
1k1i	2.2	bovine trypsin	NAP <sup>e</sup> -D-pipecolinic acid	508.6	2	13	15	-10.899[201]
1k1j	2.2	bovine trypsin	NAP <sup>e</sup> -isopipecolinic acid methyl ester	523.6	3	13	16	-9.465[201]
1jyr	1.55	Grb2 SH2 domain	APS-PTR <sup>e</sup> -VNVQN	1069.0	1	14	15	-7.94[203]
1rlq	NA	C-src tyrosine kinase SH3 domain	RALPLPRY	1084.3	2	25	27	-10.2[204]

Rendszer <sup>a</sup>	Felb. <sup>b</sup> (Å)	Célpont	Ligandum		Vizek száma			$\Delta H_b(\text{exp})^d$
2ke1	NA	autoimmune regulator	ARTKQTARKS	1150.3	12	15	27	-9.2[205]
2bba	1.65	EphB4 receptor	NYLFSPNGPIARAW	1606.8	12	15	27	-15.5[206]
1jgn	NA	human poly(A)-binding protein	VVKSNLNPNAKEFVPGVKY GNI	2389.8	14	34	48	-14.8[207]
2roc	NA	induced myeloid leukemia cell differentiation protein homolog	EEEWAREIGAQLRRIADDLN AQYERRM	3317.6	14	38	52	-14.3[208]

<sup>a</sup>A rendszer nevét a PDB-név alapján adtuk meg és ahol lehetett, a ligandum nevét is használtuk.

<sup>b</sup>Felbontás (kristallográfiás szerkezetekből).

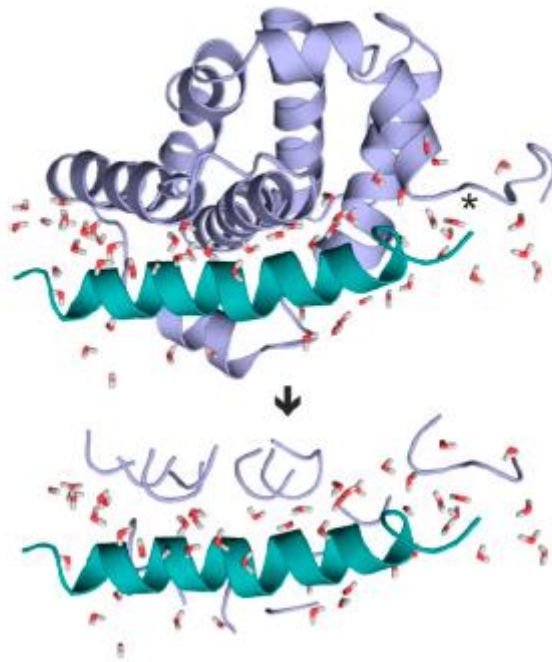
<sup>c</sup>Molekulatömeg.

<sup>d</sup>Kísérleti kötési entalpiaértékek az eredeti pontossággal megadva, ahol kJmol<sup>-1</sup>ből számolva, ahol 1 J = 4.184 cal. Az eredeti cikkek forrását is megadtuk.

<sup>e</sup>NAP: N-alpha-(2-naphthylsulfonyl)-N-(3-amidino-L-phenylalaninyl); PTR: o-phosphotyrosine

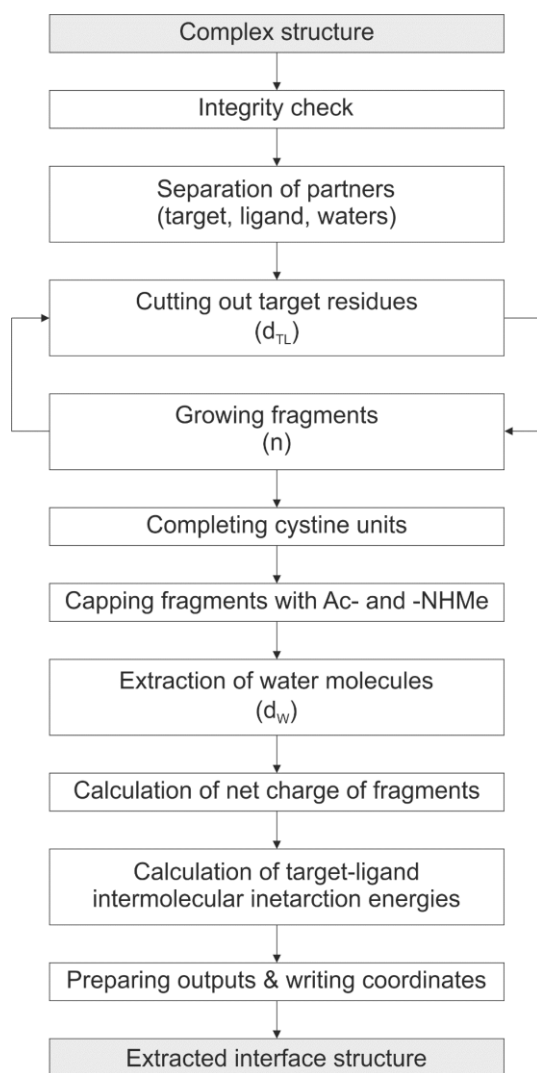
### 5.1.2. Fragmenter

A célpont teljes szerkezetét tartalmazó rendszer QM-számítását ésszerű számítási időn belül nem lehet megvalósítani, tehát a fenti  $\Delta_f H$ -értékek QM-számításához szükség lesz a célpont-ligandum komplex interfész régiójának kivágására. Azonban a komplex interfész kivágása és a célpontfehérje automatikus fragmentálása nem triviális, a probléma megoldására a Fragmenter nevű új fragmentáló módszert dolgoztuk ki (**I. tézispont**), hogy szabványosítsuk a célpont-ligandum interfészek kivágását (**4. ábra**).



**4. ábra** A Fragmenter kivágja a hidratált interfészt (lent) a célpont-ligandum komplexből. A célpont (fragmensek) és a ligandumok világoskék és zöldek. A 2roc rendszer tartalmazza a legnagyobb ligandumot, ebben a példában a Fragmenter  $d_{TL} = 5.0 \text{ \AA}$  távolságig vágja ki a célpont aminosavait, ami erősen lecsökkenti a rendszer méretét a QM-számításhoz. Az interfészvizeket ( $d_w = 5.0 \text{ \AA}$ ) ugyancsak tartalmazza a rendszer, ezeket pálcikák jelölik. A kivágás lépései a C-terminálison (csillagokkal jelölve) a **6. ábrán** láthatók. A Fragmenter webszerverként a [www.fragmenter.xyz](http://www.fragmenter.xyz) oldalon szabadon felhasználható. Forrás: Horváth et al. 2019.

A kivágás egy olyan komplex szerkezeten történik, amely egy célpontot, egy ligandumot és több vízmolekulát tartalmaz. A fragmens aminosavakat az intermolekuláris távolsági határérték alapján választjuk ki ( $d_{TL}$ , **3. táblázat**), valamint a csatlakozó aminosavakat és a végcsoportokat (terminálisokat) is tartalmazza a fragmens. A Fragmenter és az adatfolyam rövid áttekintését a **3.** és az **5. ábra** mutatják.



**5. ábra** A Fragmenter folyamatábrája. Forrás: Horváth et al. 2019.

A Fragmenter a célpontfehérje azon szomszédos részeire összpontosít, amelyek kölcsönhatásba lépnek a ligandummal és az interfész-víz-molekulákkal. A komplexek interfészfókéinak teljes ligandum-molekuláját és a célpont fehérje-fragmenseit vágjuk ki, a célpont-fragmens aminosavait különálló aminosavak sorozata helyett peptidfragmensként kezeljük. A fő cél az, hogy a célpontfehérjéből a legrövidebb, de folyamatos peptidláncokat nyerjük ki standardizált módon, tehát jelentősen csökkentjük a célpont méretét, de a fragmens folytonosságát is megőrizzük, ahol ez csak lehetséges.

### 3. táblázat A Fragmenter input paraméterei, \*opcionális

Paraméter	Leírás
Projekt neve	A folyamat azonosítója és a mopac- és outputfájlok neve
Komplex	Az input PDB-fájl teljes elérési útja és neve
Célpont lánc(ok) jele	A célpont láncazonosító jele(i) a PDB-fájlban
Ligandum lánc(ok) jele	A ligandum láncazonosító jele(i) a PDB-fájlban
Ligandum töltése*	A ligandum töltése, alapértelmezésben 0.
Célpont-ligandum távolságlimit ( $d_{TL}$ )	Maximális távolság a célpont és a ligandum legközelebbi nehéz atomjai között (Å)
Víz-oldott partnerek távolságlimite ( $d_w$ )	Maximális távolság az oldott partnerek (célpont és ligandum) és a vízmolekulák legközelebbi nehéz atomjai között (Å)
Aminosavak közti toplogikus távolság (Inter-residual distance) (n)	Ez a paraméter határozza meg a peptidfragmens hosszát. Ha $n > 0$ , akkor n kapcsolódó aminosavat adunk a fragmens molekulához. Ha $n = 0$ csak a $d_{cls} \leq d_{TL}$ aminosavakat, ahol $d_{cls}$ a legkisebb távolság a partnerek nehéz atomjai között. Ha $n = 1$ , a sorrendben következő első, szomszédos aminosavat adjuk hozzá a végekhez, holott $d_{cls} > d_{TL}$ , és így tovább.
Megkötés	C-alfa atomok, gerinc atomjai, nehéz atomok lehetnek rögzítve
MOPAC kulcsszavak*	MOPAC paraméterek, amelyeket az input fájl első sorába ír be a Fragmenter. Ezek a paraméterek szerepelnek alapértelmezésben: PM7 PREC GRAD EF EPS=78.3 GNORM=1 MMOK GEO-OK DUMP=30M T=4W LET CYCLES=1000000000 MOZYME [190]

Az n paraméter azt adja meg, hogy hány szomszédos aminosavat adjunk a  $d_{TL}$  szerint extrahált aminosavak fragmensláncához, amelyek amúgy már kívül esnének a távolság miatt. Néhány kísérlet után (**4. táblázat**) megállapítottuk, hogy  $n = 0$  jó korrelációkat eredményez (lásd a következő bekezdésekben) és nem volt szükség az  $n = 1$  vagy nagyobb értékek alkalmazására.

### 4. táblázat Az aminosavak közötti távolság (n) hatása a kiválasztott aminosavakra (3ptb\_ben, $n = 0..5$ )

Célpont aminosavsorrendje	n=0	n=1	n=2	n=3	n=4	n=5
L166						L166
E167					E167	E167
G168				G168	G168	G168
G169			G169	G169	G169	G169
K170		K170	K170	K170	K170	K170
D171	D171	D171	D171	D171	D171	D171
S172	S172	S172	S172	S172	S172	S172
C173	C173	C173	C173	C173	C173	C173
Q174	Q174	Q174	Q174	Q174	Q174	Q174
G175		G175	G175	G175	G175	G175
D176		D176	D176	D176	D176	D176
S177	S177	S177	S177	S177	S177	S177
G178		G178	G178	G178	G178	G178
G179			G179	G179	G179	G179
P180				P180	P180	P180
V181					V181	V181
V182						V182
C183						
S184						
G185						
K186						K186
L187					L187	L187
Q188				Q188	Q188	Q188
G189			G189	G189	G189	G189
I190		I190	I190	I190	I190	I190

Célpont aminosavsorrendje	n=0	n=1	n=2	n=3	n=4	n=5
V191	V191	V191	V191	V191	V191	V191
S192	S192	S192	S192	S192	S192	S192
W193	W193	W193	W193	W193	W193	W193
G194	G194	G194	G194	G194	G194	G194
S195	S195	S195	S195	S195	S195	S195
G196	G196	G196	G196	G196	G196	G196
C197	C197	C197	C197	C197	C197	C197
A198	A198	A198	A198	A198	A198	A198
Q199		Q199	Q199	Q199	Q199	Q199
K200			K200	K200	K200	K200
N201		N201	N201	N201	N201	N201
K202	K202	K202	K202	K202	K202	K202
P203	P203	P203	P203	P203	P203	P203
G204	G204	G204	G204	G204	G204	G204
V205	V205	V205	V205	V205	V205	V205
Y206		Y206	Y206	Y206	Y206	Y206
T207			T207	T207	T207	T207
K208				K208	K208	K208
V209					V209	V209
C210						C210

A kivágás alapja a célpont és ligandum ( $d_{TL}$ ), valamint a célpont-ligandum és a vizek közötti távolságok ( $d_w$ ), illetve a csoportok közti távolságok ( $n$ ). A fő ciklusban a célponton egy aminosavcsoportot akkor választunk ki, ha legalább egy nehéz atomjának távolsága a ligandum valamely nehéz atomjától kisebb, mint a  $d_{TL}$  ( $d_{cls} \leq d_{TL}$ ), ahol  $d_{cls}$  a tényleges mért távolság a legközelebbi nehéz atomok a célpont és a ligandum között. A legkisebb engedélyezett távolságot a célpont és a ligandum nehéz atomjai között is állíthatja a felhasználó, aminek 3.5 Å az alapértelmezett értéke. Ugyanezt a távolságot a célpont, a ligandum és a vizek között is definiáltuk, amely a kivágott vizeket határozza meg. A peptidfragmens hosszát a maximális aminosavak közötti távolság ( $n$ ) is befolyásolja, amely megadja, hogy hány szomszédos aminosavat kell még a fragmensbe kiválasztani. Azaz ha  $n > 0$ , akkor  $n$  darab szomszédos aminosavat is hozzá kell fűzni mindkét oldalról a  $d_{TL}$  kritérium által kiválasztott aminosavhoz. Ha  $n = 0$ , akkor csak a  $d_{cls} \leq d_{TL}$  kritérium alapján kiválasztott aminosavak kerülnek be a fragmensláncba. Ha  $n = 1$ , akkor az első szomszédos aminosavakat fűzzük a láncvéghez, akkor is, ha  $d_{cls} > d_{TL}$ . A fő ciklusban (**5. ábra**) látszik, hogy a diszulfidhidakon keresztül kapcsolódó Cys-csoportokat is kiválasztjuk és hozzáadjuk a fragmensekhez. A szabad (nem-terminális) végeket acetil- (Ac-) és iminometil- (-NHMe) csoportokkal blokkoljuk, amelyeket a szükséges hidrogénatomokkal látjuk el. Miután az összes célpont fragmenst kialakítottuk (**5. ábra**), a  $d_w$



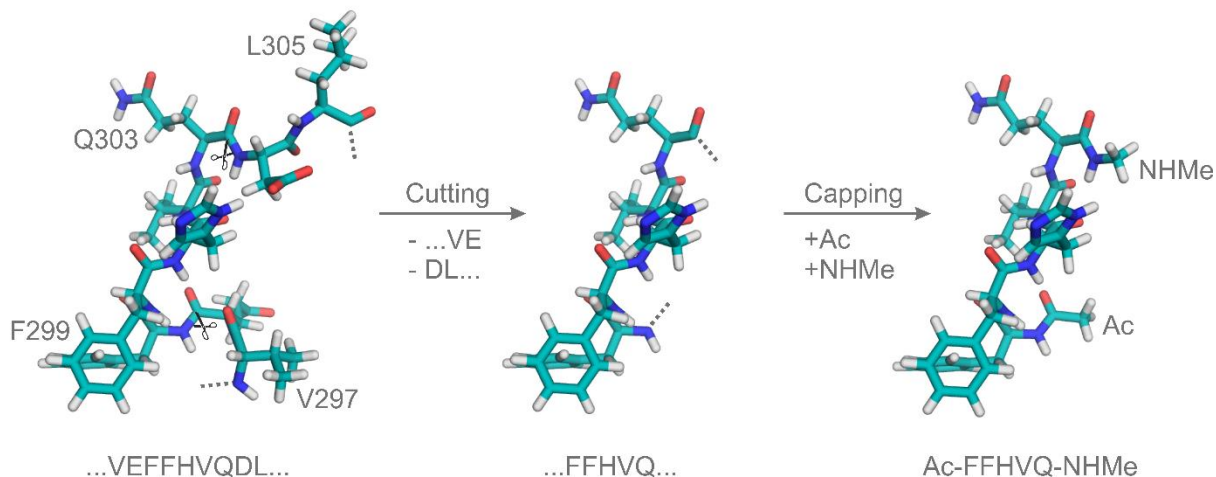
intermolekuláris távolság szerint a vízmolekulákat is (3. táblázat) kiválasztjuk. A célpont fragmensek teljes nettó töltését (5. táblázat) is számoljuk.

5. táblázat A rendszerek töltése és termokémiai adatai

Rendszer kódja	Ligandum töltése	Célpont fragmens töltése	Teljes töltés
3ptb_ben	1	0	1
3ptb_pme	1	-1	0
3ptb_pam	1	0	1
3ptb_pmo	1	0	1
3ptb_pad	1	-1	0
1k1l	2	0	2
1k1m	1	1	2
1k1i	0	-1	-1
1k1j	1	1	2
1jyr	-2	1	-1
1rlq	2	-3	-1
2ke1	4	-5	-1
2bba	1	2	3
1jgn	2	-1	1
2roc	-2	1	-1

Kísérleti termokémiai adatok ( $T$ ,  $\Delta H_b$ , és  $\Delta C_p$  K, kcal mol<sup>-1</sup>, és cal mol<sup>-1</sup>K<sup>-1</sup> egységekben. 1 cal = 4.184 J.  $T_1$  az eredeti mérés hőmérséklete. Minden  $\Delta H_b$  értéket – kivéve a 2ke1 rendszert - 298.15 K-en mértek vagy erre a hőmérsékletre konvertáltunk. A  $\Delta H_b$   $T_2$  hőmérsékletre konvertálása a Kirchhoff-törvény alapján történt:  $\Delta H_b(T_2) = \Delta H_b(T_1) + \Delta C_p [T_2 - T_1]$ . A  $\Delta H_b(T_2)$  oszlopot használtuk a számolások során. A  $T_1$  eredeti hőmérsékleten elért pontosságot vittük tovább kJ mol<sup>-1</sup>-ből, ahol három tizedesjeggyel számoltunk).

Rendszer kódja	$T_1$	$\Delta H_b(T_1)$	$\Delta C_p$	$T_2$	$\Delta H_b(T_2)$	Referencia
3ptb_ben	298.25	-4.517	-95.602	298.15	-4.507	[199]
3ptb_pme	298.25	-4.422	-100.382	298.15	-4.412	[199]
3ptb_pam	298.25	-6.429	-124.283	298.15	-6.417	[199]
3ptb_pmo	298.25	-3.752	-100.382	298.15	-3.742	[199]
3ptb_pad	298.25	-2.940	-54.971	298.15	-2.935	[199]
1k1l	298.15	-7.863	-525.813	298.15	-7.863	[4]
1k1m	298.15	-8.222	-611.855	298.15	-8.222	[4]
1k1i	298.15	-10.899	-619.025	298.15	-10.899	[4]
1k1j	298.15	-9.465	-621.415	298.15	-9.465	[4]
1jyr	298.15	-7.94	-146	298.15	-7.94	[6]
1rlq	298.15	-10.2	na	298.15	-10.2	[7]
2ke1	296.15	-9.2	na	296.15	-9.2	[8]
2bba	298.15	-15.5	na	298.15	-15.5	[9]
1jgn	298.15	-14.8	-253.2	298.15	-14.8	[10]
2roc	298.15	-14.3	na	298.15	-14.3	[11]



**6. ábra** A 2roc rendszer VEFFHVQDL szekvenciájának kivágása a C-terminális felől. Forrás: Horváth et al. 2019.

### 5.1.3. Száraz rendszerek és implicit vízmodell

A Fragmenter protokoll kidolgozásával és megvalósításával a (hidratált) célpont-ligandum komplex interfészek  $\Delta_f H$ -számítását egyszerűsített és szabványosított módon hajtottuk végre, az összetett interfészek kivágására a **2. táblázat** összes rendszerén fragmentálást alkalmaztunk.

A Fragmenter számára minden rendszert standard molekulamechanikai energiaminimalizálással és explicit hidratációs protokollal készítettük elő, ahogyan azt a Módszerekben leírtam. A  $\Delta_f H$  értékeket az egyes reaktánsok (ligandum és célpont fragmensek) és a termék (komplex interfész) szerkezeteire is kiszámoltuk. A számolásokat szemiempirikus szinten hajtottuk végre PM7 paraméterezéssel és Mozyme megközelítéssel, a Módszerekben leírt módon. A nyers értékeket a **Függelék A. táblázata** tartalmazza. Vizsgálataink első lépésében nem alkalmaztunk oldószermodelleket ( $s = 0$  a **2. egyenletben**). Ez azt jelenti, hogy a „száraz” input szerkezeteket explicit vízmolekulák nélkül és vákuumban számoltuk. A vízmodellek teljes hiánya nem mutatott korrelációt a számított és kísérleti  $\Delta H_b$  értékek között (Vákuum/száraz oszlop a **6. táblázatban**, **8. ábra**). Az implicit vízmodell alkalmazása (CONductor-like Screening MODEL, COSMO [100]) valamivel megnövelte a korrelációt (COSMO/Száraz oszlop a **6. táblázatban**), ez a korreláció azonban még mindig meglehetősen gyenge, ahogyan azt a keresztvalidáció is tükrözi.

**6. táblázat** Rendszerenkénti reziduálisok ( $\varepsilon$ ) és a különböző vízmodellekkel kapott lineáris regresszió statisztikai paraméterei

Rendszer	Vákuum					COSMO			
	Száraz	Shell 1	Shell 2	Shell 3	Száraz	Shell 1	Shell 2	Shell 3	Shell 3 <sup>b</sup>
	$ \varepsilon ^a$								
3ptb_ben	3.70	3.18	2.90	1.93	3.73	2.33	2.45	1.18	0.85
3ptb_pme	3.60	3.09	3.05	2.03	0.53	1.12	2.48	0.95	1.22
3ptb_pam	1.74	1.29	0.88	0.03	0.01	0.02	0.44	0.92	3.03
3ptb_pmo	4.45	3.91	3.71	2.64	3.59	3.25	3.32	2.24	0.33
3ptb_pad	5.65	5.11	5.04	4.25	3.03	3.12	4.32	3.22	1.37
1k1l	0.56	0.39	0.10	0.10	2.59	1.05	0.56	0.43	1.74
1k1m	0.16	0.56	0.36	0.79	2.71	1.17	0.75	1.50	3.12
1k1i	2.67	3.19	2.95	3.51	2.80	3.62	2.88	3.34	4.60
1k1j	1.43	1.81	1.60	2.10	0.57	2.60	1.47	2.32	3.75
1jyr	0.78	0.28	0.53	0.09	1.73	0.36	0.40	0.53	0.36
1rlq	0.67	0.64	0.27	0.33	0.37	2.13	0.61	0.24	0.16
2ke1	2.54	4.67	4.66	5.28	4.38	4.21	5.73	2.46	2.89
2bba	7.25	6.38	7.21	6.23	4.34	2.13	6.58	3.77	3.31
1jgn	5.88	5.24	4.75	2.56	1.61	0.27	3.25	0.25	1.36
2roc	4.96	4.11	3.74	1.26	2.76	1.24	3.05	1.71	3.92
R <sup>2</sup>	0.06	0.18	0.19	0.44	0.51	0.65	0.33	0.73	0.93
R <sup>2</sup> (cv) <sup>c</sup>	0.00	0.01	0.02	0.22	0.34	0.54	0.07	0.65	0.91
F	0.81	2.77	3.14	10.20	13.46	24.28	6.36	34.55	179.66
RMSE <sup>a</sup>	4.02	3.76	3.72	3.10	2.90	2.45	3.40	2.17	2.65
t <sub>α</sub>	0.90	1.66	1.77	3.19	3.67	4.93	2.52	5.88	13.40
t <sub>β</sub>	-5.56	-5.04	-4.68	-3.90	-2.18	-3.99	-4.24	-2.81	-

<sup>a</sup>Egység: kcal mol<sup>-1</sup>.

<sup>b</sup>Lineáris illesztés  $\beta=0$  (utolsó oszlop) és  $\beta \neq 0$  (többi oszlop).

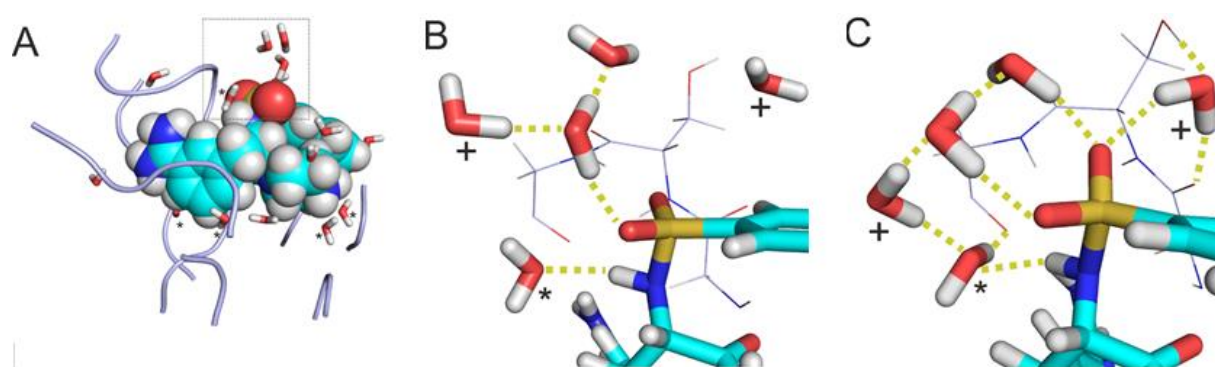
<sup>c</sup>Leave-one-out keresztvalidáció.

Általában az implicit COSMO modell alkalmazása előnyösnek bizonyult a vákuum/száraz eredményekkel összehasonlítva (**Függelék A. táblázat**). A 2ke1 rendszer volt az egyetlen eset, ahol a  $\Delta H_b(\text{exp})$  nem alakítható át 298.15 K-re és az eredeti, 296.15 K-nél ismert értéket használtuk a regresszióhoz. A 2ke1 rendszer nélkül is elvégeztük a lineáris regressziót és valamivel jobb korrelációt kaptunk a Vákuum/Száraz rendszerben (**7. táblázat**).

#### 5.1.4. Explicit hidratáció és hibrid modell

Az előző fejezet korrelációinak további javítása érdekében szisztematikus vizsgálatot végeztünk explicit hidratáció segítségével. Ehhez a hidratációs szerkezetet [25]-ben leírt módon építettük fel

és a Fragmenterrel kivágtuk a célpont-ligandum interfészeket. A vízszerkezetet MobyWat [25] segítségével határoztuk meg, amely a rövid molekuladinamikai szimulációkból származó komplex interfészek teljes hidratációs szerkezetét elkészíti, explicit vízpozíciókat hoz létre a szimulációk sorozatképeiből és a víz-víz kölcsönhatásokat is figyelembe veszi az oldószer vízzel mellett. Kimutattuk, hogy a MobyWat számos esetben megtalálja a kísérleti referenciavíz-pozíciókat [25], így a vizezési eljárás során garantálja a felület teljes hidratációját [25] és a kísérleti méréssel nem érzékelhető vízpozíciókat is meghatározza [77, 82, 83, 219, 220]. Három réteget (Shell) definiáltunk  $d_w$  alapján (**2. és 3. táblázat**) az interfész-víz-molekulák felhasználásával (**7a. ábra**). A Shell 1 vízmolekulái vannak a legközelebb a célponthoz és a ligandumhoz ( $d_w = 3.5 \text{ \AA}$ ), a Shell 2 a közbenső pozíciójú vizeket tartalmazza ( $3.5 \text{ \AA} < d_w < 5.0 \text{ \AA}$ ), a Shell 3 héj pedig a Shell 1 és a Shell 2 vízmolekuláit tartalmazza, azaz  $d_w = 5.0 \text{ \AA}$ .



**7. ábra** Az 1k1l rendszer kivágott komplexe. A) A kiindulási vizezett és MM-energiaminimalizált szerkezet. A célpont fragmensek ribbon formában láthatók. A Shell 1-es vizek ( $d_w = 3.5 \text{ \AA}$ , csillagokkal jelölve) a célponthoz és a ligandumhoz közel helyezkednek el és híd szerepet töltenek be. A Shell 3 ( $d_w = 5.0 \text{ \AA}$ ) vizek többi része a Shell 2-vel van kapcsolatban, közel vannak az oldószer vízzel, az interfész szélén helyezkednek el. Shell 3 = Shell 1 + Shell 2. B) Az A szerkezet elforgatva a ligandum szulfonilcsoportját és a szomszédos célpont G216SG218 részét mutatja, ahol a számozás a követi a kristályszerkezet számozását (1k1l). A hidrogénkötéseket sárga szaggatott vonalak jelölik. C) A B szerkezet szemiempirikus optimalizálás után, amelyet PM7 parametrizálással végeztünk el Mozyne használatával. Azokat a vizeket, amelyek  $1.5 \text{ \AA}$ -nél messzebb vándoroltak, keresztek jelölik. Forrás: Horváth et al. 2019.

Az explicit vízmolekulák alkalmazása a vákuumban javította a „száraz” korrelációt 0.44-es  $R^2$ -re (Vákuum/Shell 3 oszlop a **6. táblázat**) és 0.65-re (2ke1 nélkül, **7. táblázat**) (**II. tézispont**).

**7. táblázat** Rendszerenkénti reziduálisok ( $\varepsilon$ ) és a különböző vízmodellekkel kapott lineáris regresszió statisztikai paraméterei a 2ke1 rendszer nélkül

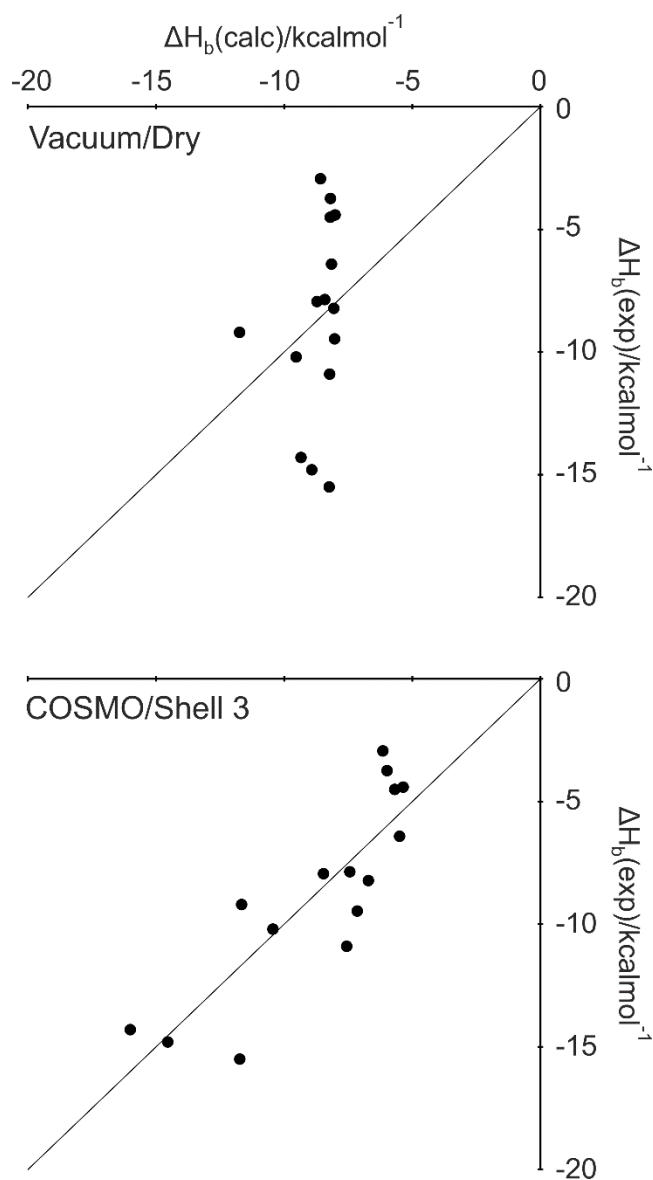
Rendszer	Vákuum					COSMO			
	Száraz	Shell 1	Shell 2	Shell 3	Száraz	Shell 1	Shell 2	Shell 3	Shell 3 <sup>b</sup>
	$ \varepsilon ^a$								
3ptb_ben	3.15	1.94	1.64	1.34	4.01	2.37	1.30	1.22	0.74
3ptb_pme	2.54	1.38	1.88	1.44	0.04	0.93	1.24	0.96	1.12
3ptb_pam	1.06	0.11	0.57	0.65	0.17	0.02	0.83	0.90	2.93
3ptb_pmo	3.88	2.58	2.53	2.02	3.64	3.31	2.29	2.29	0.46
3ptb_pad	6.12	4.77	4.77	4.05	2.72	3.02	3.53	3.27	1.51
1k1l	0.59	0.55	0.20	0.01	3.07	1.02	1.29	0.30	1.55
1k1m	1.08	1.85	0.84	0.85	3.13	1.10	1.28	1.42	2.95
1k1i	3.16	4.38	3.27	3.60	2.55	3.51	2.75	3.21	4.40
1k1j	2.44	3.14	2.08	2.20	0.11	2.56	1.37	2.21	3.57
1jyr	1.62	0.35	1.12	0.33	2.38	0.20	0.91	0.70	0.13
1rlq	2.36	2.76	2.87	1.89	1.26	1.88	1.43	0.52	0.49
2bba	7.69	4.09	6.94	5.33	3.29	0.96	5.34	3.42	2.93
1jgn	4.49	1.83	1.41	0.10	0.04	1.74	1.17	0.23	1.87
2roc	2.45	0.87	0.51	1.63	1.61	0.12	0.99	2.27	4.50
R <sup>2</sup>	0.19	0.57	0.50	0.65	0.63	0.75	0.69	0.76	0.93
R <sup>2</sup> (cv) <sup>c</sup>	0.03	0.45	0.39	0.56	0.52	0.68	0.62	0.67	0.91
F	2.79	15.76	12.03	22.29	20.15	36.47	27.23	37.41	173.54
RMSE <sup>a</sup>	3.88	2.83	3.05	2.55	2.63	2.14	2.38	2.12	2.61
t <sub>α</sub>	1.67	3.97	3.47	4.72	4.49	6.04	5.22	6.12	13.17
t <sub>β</sub>	-3.59	-3.04	-2.99	-3.37	-1.66	-4.06	-3.04	-2.77	-
3ptb_ben	3.15	1.94	1.64	1.34	4.01	2.37	1.30	1.22	0.74

<sup>a</sup>Egység: kcal mol<sup>-1</sup> (root mean square).

<sup>b</sup>Lineáris regresszió  $\beta=0$  (utolsó oszlop) és  $\beta\neq 0$  (többi oszlop).

<sup>c</sup>Leave-one-out keresztvalidáció.

A keresztvalidáció azt mutatja, hogy a korreláció javulása csak a 2ke1 rendszer nélkül stabil. Ezen a ponton ésszerűnek tűnt annak ellenőrzése, hogy a hibrid modell implicit és explicit hidratációt alkalmazva tovább javítja-e a korrelációt. Valóban, a hibrid (COSMO/Shell 3 oszlop a **6. táblázatban**) modell adta a legjobb egyezést a számított és kísérleti  $\Delta H_b$  között, melynek R<sup>2</sup>-értéke 0.73 (**8. ábra**) és amely az összes interfész-vízmolekulát tartalmazza. Figyelemre méltó, hogy a Shell 1 réteg vizei is nagy korrelációt eredményeztek (R<sup>2</sup> = 0.65). A korrelációk mindkét esetben teljesítették a keresztvalidáció kihívásait, a közbenső vízpozíciók önmagukban (Shell 2) stabil korrelációt csak 2ke1 rendszer nélkül eredményeztek (**7. táblázat**).



**8. ábra** Vízmentes (vákuum/száraz) és hibrid (COSMO/Shell 3) vízmodellek korrelációs egyenese. Forrás: Horváth et al. 2019.

A ligandumméret és a célpontok változatossága a fenti korreláció stabilitására gyakorolt hatásának vizsgálatára (COSMO/Shell 3) a **2. táblázatban** szereplő rendszereket a ligandum molekulatömege alapján két csoportra osztottuk. Az első csoport kilenc rendszert tartalmaz, amelyekben kis ligandumok ( $MW < 600$ ) vannak. Mindegyik ligandumnak közös célpontja van, nevezetesen a béta-tripszin. A második csoport hat rendszert tartalmazott nagy ligandummal, amelyeknek molekulatömege nagyobb mint 1000 és különböző célpontjaik voltak. A lineáris regressziókat külön-külön hajtottuk végre a két csoporton és a  $\Delta H_b(\text{calc})$ -értékeket a két regressziós egyenlet alapján számítottuk ki. A kapott statisztikai paraméterek összessége

(**Függelék C. táblázat**) összehasonlítható volt az összes rendszer regressziójával (COSMO/Shell 3 oszlop a **6. táblázatban**). A hibrid modell esetében tehát a korrelációk stabilitását nem befolyásolja a ligandumok mérete és a célpontok sokfélesége.

#### 5.1.5. A skálázó faktor

Ha megvizsgáljuk az általános regressziós paramétereit, a fenti COSMO/Shell 3 modell  $\beta \neq 0$  értékkel a **3. egyenletben** kellően robusztus. A  $t_\beta$  érték (**6. táblázat**) azonban azt mutatja, hogy a  $\beta$  regressziós együttható szignifikanciaszintje mérsékelt ( $p = 0,015$ ), így  $\beta = 0$  értékkel is újraszámoltuk a regressziót és a megfelelő statisztikai paramétereit a **2. táblázat** utolsó oszlopában (**Függelék B. táblázat**) soroltuk fel.

Ily módon az összes paraméter nagy szignifikanciájú ( $p < 0,01$ ) modelljét kaptuk és a **3. egyenletet** jelentősen egyszerűsítettük. A kapott **4. egyenlet (III. tézispont)** csak az  $\alpha$  regressziós együttható értékét tartalmazza, amely egyetlen, független skálázó faktorként szolgál a számított  $\Delta_r H$   $\Delta H_b$ -ba történő átalakításához.

$$\Delta H_b = 0.031(\pm 0.002)\Delta_r H \quad (4)$$

Hasonló nagyságú,  $0.032 (\pm 0,002)$  értéket kaptunk a skálázó faktorra, ha a 2ke1 rendszer nem vett részt a regresszióban. A QM-en keresztül ez a tényező közvetlen kapcsolatot teremt a nagy komplexek molekuláris szerkezete és termodinamikája között.

#### 5.1.6. Vízszerkezeti esettanulmányok

A vizsgált tizenöt rendszer kétharmadában a Shell 1 vagy Shell 3 explicit vízmolekulák alkalmazása a reziduálisok csökkenését eredményezte (COSMO modellek a **6. táblázatban**). A Shell 2 vizei az esetek egyharmadában hasonló hatást fejtettek ki, például az 1k1l rendszernél a reziduálisok 2.59-ről (száraz) 0.43-ra (Shell 3,  $\beta \neq 0$ ) és 1.74 kcal mol<sup>-1</sup>-ra csökkentek (Shell 3,  $\beta = 0$ , **6. táblázat**) és hasonló tendenciát figyelhetünk meg a vákuumos értékeknél is.

Az 1k1l rendszer interfészeiben a MM-energiaminimalizálás és a fragmentálás után (**7a. ábra**) a Shell 1 öt, a Shell 2 tíz vízmolekulát tartalmaz (**2. táblázat**). A Shell 1-ben lévő öt vízmolekula (**7a. ábra**) a célpont és a ligandum közötti interfészhíd alján, a Shell 2 vizei többnyire az interfész nyílásánál helyezkednek el a **7. ábra** jobb oldala felé. Ahogy vártuk, a töltött vagy poláris csoportok köré gyűltek össze a vizek nagyobb számban. Például a ligandum szulfonilcsoportját (**7b. ábra**)

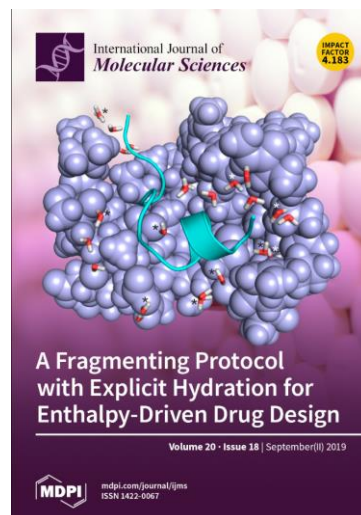
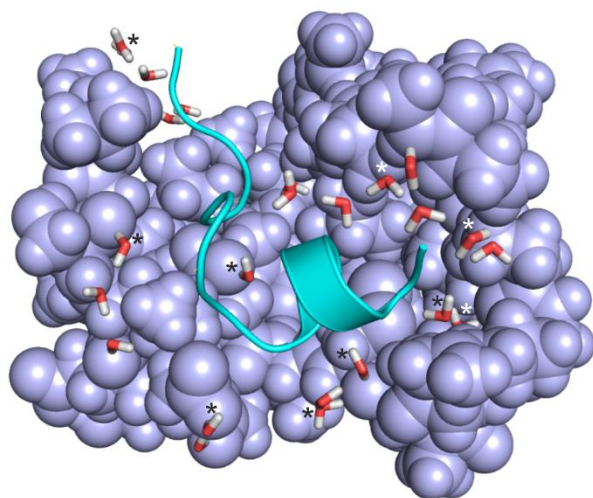
vízmolekulák csoportja veszi körül és csak egyikük tartozik a Shell 1-hez. A vizek és a legközelebbi célpont fragmens ( $G_{216}SG_{218}$ ) között nem figyeltünk meg kölcsönhatásokat.

A szemempirikus QM-relaxáció során (**7c. ábra**) az egyes vízmolekulák helyzete és orientációja jelentősen megváltozott, például két vízmolekula (**7c. ábra** keresztekkel jelölve) 3.2 és 1.8 Å-mel mozdult el. A Shell 1 vízmolekula irányultsága (**7c. ábra** csillaggal jelölve) úgy változott meg, hogy kölcsönhatásba tudjon lépni a célpont fragmenssel. Az ilyen változások eredményeként a vízmolekulák kiterjedt H-hídhálózata jött létre, amely stabilizálja a célpont-ligandum kölcsönhatást a szulfonilcsoport körül. Az új hidrogénhídkötések kialakulása azt jelenti, hogy a Shell 2 vízmolekulái közül néhány a Shell 1 rétegbe vándorolt (**7c. ábra**, nincs jelölve).

Míg a hidratációs szerkezet figyelemre méltó átalakuláson ment keresztül a szemempirikus QM-relaxáció során, a célpont fragmens szerkezete változatlan maradt. A fenti 1k1l (**7. ábra**) példája megmutatta, hogy a különböző rétegekben lévő vízmolekulák hozzájárulnak a célpont-ligandum interfész szerkezetének létrehozásához és ennek következtében a kiszámított  $\Delta H_b$  reziduálisainak csökkenéséhez.

A kis, merev ligandumok, mint például az 1k1l rendszer fenilalanin-származékai mellett nagy peptid ligandumokat is vizsgáltunk. Például a 2bba rendszer (**9. ábra**) viszonylag kiterjedt pentadekapeptid ligandumot (**2. táblázat**), a fragmentált interfészben huszonhét vízmolekulát tartalmaz. A 2bba rendszernél a legnagyobb a csökkenés, 4.34 kcal mol<sup>-1</sup>-ről 2.13 kcal mol<sup>-1</sup>-ra változott a reziduális (COSMO modellek a **6. táblázatban**) a Shell 1 vízmolekulákkal. A hidratációs szerkezet részletes vizsgálata azt mutatja, hogy a Shell 1 vízmolekulái (**9. ábra**, csillagok) többnyire a kötő zseb alján helyezkednek el és hídszerepet játszanak a célpont- és ligandumpartnerek között. Ebben az esetben a Shell 2 vizek alkalmazása a Shell 1 mellett nem volt hasznos, mivel megnövelte a reziduálist. Azonban a Shell 3 végső reziduálisa még mindig a száraz modell alatt van.





**9. ábra** A 2bba kivágott komplex interfésze szemempirikus PM7-es energiminimalizálás után, Mozyme módszert alkalmazva. A célpont fragmensek világoskék térkitöltésűek, a ligandumot zöld cartoon ábrázolja. A vízmolekulákat a Shell 1-ben csillagok jelölik, a nem jelölt vízmolekulák a Shell 2-höz tartoznak. Az ábrát és a tanulmányt, mint címlaptörténetet szerkesztőségi döntés alapján az újság a címlapjára emelte ki. Forrás: Horváth et al. 2019.

A hídkepzési és térkitöltési szerepeken túl (**7. és 9. ábra**) látható, hogy a MobyWat (lásd: Módszerek) által létrehozott határfelületi vízszerkezet valószínűleg javítja a vízmolekulák árnyékoló hatásának [221] kiszámítását. A hidratációs szerkezet fent említett szerkezeti jelentősége ellenére a krisztallográfiás mérésekből kapott szerkezetek nem tartalmazzák a lényeges vízmolekulákat, vagy hibásan határozták meg azok pozícióit.

A  $\Delta H_b$  szerkezeti alapú számítása csak az interfaciális hidratációs szerkezet helyes, elméleti meghatározása után volt lehetséges a MobyWat segítségével (lásd: Módszerek). A hidratációs szerkezeteken kívül négy rendszer hiányzó ligandum pozíciói (3ptb\_pad, 3ptb\_pam, 3ptb\_pme, 3ptb\_pmo) szintén modellezéssel készültek, így a kísérleti szerkezetet meghatározó atomi felbontási adatok és a modellezés vizsgálataink robusztus korrelációit eredményezte.

#### 5.1.7. A Fragmenter az interneten

A fragmentálót ingyenes webszolgáltatásként (**10. ábra**) is rendelkezésre bocsájtjuk, amely a kivágott komplex interfész szerkezetét (célpont fragmensek, ligandum és vízmolekulák) interaktív képként jeleníti meg és amelyek PDB- és MOPAC-inputfájlként is letölthetők a Results lapon (**11. ábra**). A Fragmenter a célpont fragmensek és a ligandum közötti intermolekuláris kölcsönhatási energiát ( $E_{inter}$ ) is becsli, amely tartalmazza a Lennard-Jones (LJ) potenciálokat, a van der Waals-

és a Coulomb-potenciálokat az elektrosztatikus kölcsönhatásokban (Módszerek). Az  $E_{\text{inter}}$ -elemzés aminosavankénti listáját ugyancsak a Results lapon találjuk meg. A lista felhasználható a ligandummal ütköző célpont fragmensek azonosítására, mivel a nagy  $E_{\text{inter}}$  értékek nemkívánatos közeli kapcsolatot jeleznek. Ilyen esetekben további MM-energiaminimalizálásra lehet szükség a QM-vizsgálatokhoz megfelelő komplex szerkezet eléréséhez.

A Fragmenter a webszerveren a következő szakaszokban ismertetésre kerülő módon működik.

### 5.1.8. Input

A weboldal (**10. ábra**) Submit tab input formjának aktuális tartalmát egy project\_ID.inp nevű fájlba menti, amit egy inputfile.inp nevű template-ből generál (**Függelék A. ábra**).

The screenshot shows the FRAGMENTER web interface. At the top, the word "FRAGMENTER" is centered in a large, grey, sans-serif font, flanked by two blue and orange molecular icons. Below this, the text "INPUT DATA" is centered in a smaller, orange font. The form consists of several input fields and a dropdown menu, each with a label to its left. The labels are: "Project name", "Complex structure", "Target chain(s)", "Ligand chain(s)", "Ligand charge", "Target-ligand distance limit (Å)", "Water-solute distance limit (Å)", "Inter-residual distance", "Restraint", and "Mopac key words". The "Complex structure" field has a "Click to select file" button next to it. The "Restraint" field has a dropdown menu with "None" selected. At the bottom of the form, there are four buttons: "Submit", "Fill in test data", "Help", and "About".

**10. ábra** A Fragmenter webszerver főoldala. Forrás: Horváth et al. 2019.

Ez a template tartalmazza a rendszerváltozókat, a php elérési útvonalát, a createqinput.sh elérési útját és a weboldalon a felhasználó által beírt input paramétereket. A Submit tabon be lehet állítani a távolságokat ( $d_{\text{TL}}$ ,  $d_{\text{W}}$  és  $n$ ) és a **3. táblázat** többi paraméterét.

A Fragmenter lehetővé teszi az atomi pozíciók, mint pl. a gerincben elhelyezkedő  $C_{\alpha}$ -atomok, vagy az összes nehéz atom „befagyasztását” a MOPAC input fájlban. Ezeket a befagyasztásokat, az egyéb MOPAC paramétereket, távolságokat és a töltést a project\_ID.inp fájlba menti el az algoritmus. Az utóbbi paraméterek közé tartozik a komplex szerkezet elérési útja és fájlneve, a folyamat neve (a SLURM ütemező számára), a futtatható php-interpreter és a Fragmenter parancsfájlok elérési útja, a MOPAC licencfájl (a SLURM számára) és a MOPAC futtatható állomány

elérési útja. A fenti útvonal- és fájlinformációk használatával a rendszerváltozók beállítását a genqinput.sh szkript végzi el. A felhasználónak nem kell törődnie a szerver konfigurációjával (pl. a szerverspecifikus php.exe elérési útja), ez a kiszolgálón tárolódik. A Submit gombra kattintva a calculate.php szkript ellenőrzi a komplex szerkezet integritását (**3. és 5. ábra**), egy PDB-ből PDB-be történő konverzióval az OpenBabel [196] program használatával.

Konverziós hiba esetén a Fragmenter leáll és a hibákat kiírja egy külön oldalon. Ha a konverzió rendben lefut, összegyűjti és átalakítja a bemeneti paramétereket az inputfile.inp fájlból és a webhelyről a createqinput.sh parancsfájl számára, majd meghívja azt. A genqinput.sh parancsfájl csak egy input fájlt (project\_ID.inp) igényel, amely tartalmazza a futáshoz szükséges összes paramétert (**Függelék A. ábra**).

#### 5.1.9. Fő algoritmus

Miután a project\_ID.inp fájlban minden szükséges bemeneti adat szerepel, a createqinput.sh parancsfájl a Fragmenter fő motorját, a fragment.php-t hívja meg és más php-osztályokat felhasználva (output.php, atom.php, charge.php, ligand.php) hozza létre az output fájlokat. Az i) atom.php osztály az atomobjektumokat reprezentálja koordinátaikkal együtt, a ii) charge.php osztály számolja a ligandum és a generált fragmensláncok töltését, a iii) ligand.php kezeli a ligandum objektumot, tartalmazza az atomokat, kötések és olvassa, írja a PDB-fájlokat, a iv) point.php pedig egy kis osztály, amely az atomok koordinátáit kezeli. A utils.php a technikai paramétereket gyűjti össze, mint például az operációs rendszertől függő információkat és a konfigurációs fájlok kezelését végzi. A fragment.php szkript (**3. ábra**) végzi el az input feldolgozását, a fragmentálást és az output fájlok kezelését. A célpont- és ligandumobjektumokat és a vízmolekulákat a láncazonosító, illetve az aminosav típusa alapján (WAT, SOL, H2O) választjuk ki. Ennek megfelelően az inputszerkezetet ligandumra, célpontra és vizekre választjuk szét, az aminosavakat azonosítjuk alapján sorba rendezzük és csak a nehéz atomokat figyelembe véve kivágjuk őket. A fragment.php fő ciklusában a célpont aminosavakat  $d_{TL}$  és  $n$  szerint választjuk ki a ligand.php és point.php segítségével. Amennyiben csak 1 aminosav marad ki a láncból két kiválasztott aminosav között, azt ugyancsak beszúrjuk a fragmensbe.

Miután az előző lépésekben minden célpont fragmenst előállítottunk, mindegyiküket egy egységes eljárással zárjuk le, amelyet a fragment.php végez el (**3. és 5. ábra**). Szabad N-terminális

esetén a protonált aminocsoportot automatikusan úgy állítjuk elő, hogy megfelelő geometriával hidrogénatomokat adunk hozzá, C-terminális esetén a karboxilát aniont változatlanul hagyjuk. Miután a ligandumot és a vizeket egy új PDB-fájlba fűztük össze a célpont fragmensekkel, a komplex teljes töltését is kiszámoljuk és ezt egy megjegyzés sorban mentjük a fájlban. Az összes kivágott célpontláncon Ac- (N-terminális) és -NHMe (C-terminális) blokkoló csoportokat építünk a nem lezárt végekre és hozzáadunk három hidrogénatomot a metilcsoportokhoz (**6. ábra**).

Az összes fragmens létrehozása után az interfész vízmolekuláit a  $d_w$  szerint vágjuk ki, ezután a teljes nettó töltésüket a charge.php segítségével számítjuk ki az aminosavak egyéni töltéseit felhasználva (**5. táblázat**) pH 7 értéken. Különös figyelmet fordítottunk a cisztein aminosavak oldalláncai közötti diszulfid hidakra, mert diszulfidhidak vagy protonált szulfhidrilcsoportok esetében a Cys töltése automatikusan nulla lesz, egyébként -1. A His töltését az imidazolgyűrű protonációs állapotának (-1, 0, +1) megfelelően számítjuk ki.

#### **5.1.10. Output**

A Fragmenter tárolja az outputfájlokat egy könyvtárban és linket biztosít az összes létrehozott fájlra a Results oldalon (**11. ábra**).

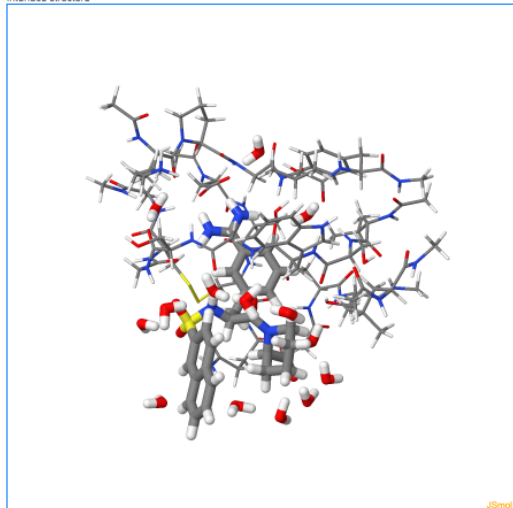
A ligandum, a kiválasztott vízmolekulák és a célpont fragmensek komplexként letölthetők PDB-formátumban, a komplex végső töltését a PDB-fájl a megjegyzések részben tárolja. A Fragmenter továbbá különálló PDB-fájlokat is biztosít és letölthető MOPAC inputfájlokká alakítja őket. Ezek a fájlok tartalmazzák a komplex szerkezetét vízmolekulákkal vagy nélkülük, külön a ligandumot vagy a célpont fragmenst.



## RESULTS

Input data	
PDB ID	1k1i_input.pdb
Project name	Test project
Target chain(s)	A
Ligand chain(s)	B
Ligand charge	0
Target-ligand distance (Å)	5
Water-solute distance (Å)	5
Inter-residual distance	0
Restraint	NONE
Additional mopac parameters	PM7 PREC GRAD EF EPS=78.3 GNORM=1 MMOK GEO-OK DUMP=30M T=4W LET CYCLES=1000000000 MOZYME
All charge	-1
Fragment charge	-1

Interface structure



Download

Complex structure (PDB)  
Interface structure (PDB)  
Interface structure (Mopac input)  
Interface structure without waters (Mopac input)  
Ligand structure (Mopac input)  
Target (Mopac input)  
Slurm input

Intramolecular interaction energy estimate per target residue

Residue #	Residue name	$E_{\text{total}}(\text{kJ})$	$E_{\text{total}}(\text{Coulomb})$	$E_{\text{total}}(\text{total})$
80	ACE	-0.04	0.59	0.54
81	LEU	-0.87	-0.16	-1.04
82	NME	-0.01	-1.18	-1.19
170	ACE	-0.04	0.93	0.89
171	ASP	1.23	-6.10	-4.87
172	SER	0.07	-12.64	-12.57
173	CYS	-4.42	0.03	-4.39
174	GLN	-7.43	-2.69	-10.12
175	GLY	-0.42	-0.03	-0.45
176	ASP	-0.60	1.12	0.52
177	SER	-1.00	0.22	-0.78
178	NME	-0.05	-1.39	-1.44
190	ACE	-0.03	0.49	0.46
191	VAL	-0.57	1.74	1.17
192	SER	-1.83	-0.29	-2.12
193	TRP	-5.82	-3.34	-9.16
194	GLY	-3.68	2.35	-1.33
195	SER	-2.07	-1.50	-3.56
196	GLY	3.14	-9.39	-6.25
197	CYS	-2.50	4.57	2.07
198	ALA	-0.37	2.31	1.94
199	NME	-0.07	-1.84	-1.90
201	ACE	-0.02	1.19	1.19
202	LYS	-0.43	-2.88	-3.31
203	PRO	-0.55	1.47	0.92
204	GLY	-0.68	5.29	4.61
205	VAL	-0.52	-2.01	-2.53
206	TYR	-0.42	-1.19	-1.61
207	NME	-0.01	-1.50	-1.51
Total		-30.29	-26.16	-56.45

Submit new project

Help

11. ábra Az eredmények oldal. Forrás: Horváth et al. 2019.

A Results lapon a fragmentált komplex egy kis ablakban is megjelenik, ahol a felhasználó a komplexet módosíthatja. A megjelenítést a JSmol [222] végzi a weboldalon, amely lehetővé teszi a molekulák forgatását, nagyítását, címkézését, színezését, az atomok összehasonlítását és a távolságok mérését. A jobb egérgombbal a felhasználónak lehetősége van több száz funkció elérésére: a stílus, a szín, a nagyítás kiválasztása, mérések; a szerkezet PDB- vagy képformátumban menthető és kis szkriptekkel programozható is.

## 5.2. Fragmens blind docking

A jelen dolgozathoz tartozó második tanulmányban (Bálint et al. 2019) egy új eljárást (FBD) vezettünk be és teszteltünk a hiszton H3- olvasó- és írófehérjék komplexeinek (**8. táblázat, IV. tézispont**) példáin, melynek során a H3 peptid ligandumokat fragmensekre vágtuk szét. Az ezen a területen végzett munka felgyorsítása érdekében új megközelítést adtunk a hiszton H3 peptidfarok szerkezetének előrejelzésére a célponton. Egy fragmentáló protokoll és egy szisztematikus BD-módszer alkalmazásával a H3 peptid jól pozícionált fragmenseinek összességét kapjuk, majd a fragmensek összekapcsolása után a célponthoz kötött H3 peptidszerkezetek kötőhelyeinek meghatározása válik lehetővé. Elsőként használtuk a blind- és fragmensdokkolás megközelítésének kombinációját, így az új módszerünket fragmens blind docking-nak (FBD) neveztük el (**V. tézispont**).

**8. táblázat** Tesztrendszer

PDB kód	Célpont	Ligandum (H3 hiszton peptid)*
2ke1	autoimmune regulator protein plant homeodomain (AIRE-PHD)	ARTKQTARKS
2pvc	DNA (cytosine-5)-methyltransferase 3-like (DNMT3L)	ARTKQTA
3qla	transcriptional regulator ATRX-ADD domain (ATRX-DNMT3-DNMT3L)	ARTKQTARK(Me <sub>3</sub> )S
4lk9	histone acetyltransferase KAT6A	ARTKQTARKSTGG
5tdw	Set domain containing protein 3	ARTK(Me <sub>3</sub> )QTARKST

A jelenlegi gyors dokkolási módszerek heurisztikus keresőmotorjai nem képesek kezelni a számos belső rotációval rendelkező peptid ligandumot [28, 223], így a peptid ligandumok nagy mérete és rugalmassága, valamint a célpontfehérjék sekély kötőfelületei nagy kihívást jelentenek a kísérleti szerkezeti meghatározási módszerek számára is. Továbbá a hisztonpeptidek C-terminális részének kölcsönhatása a célpontjaikkal gyakran gyenge, sőt a kísérletileg meghatározott szerkezetekben nem is látható. Megmutattuk, hogy a ligandumok apró peptidekké történő szétvágása ésszerű megoldást nyújt, még akkor is, ha a teljes fehérjefelületet is kell BD-módszerrel lefedni, ugyanis korábbi gyors dokkoló vizsgálatokban csak ismert kötő zsebre összpontosítva használtak fragmentáló módszert. Mi a fragmentálást és a BD-módszert egyszerre alkalmaztuk a keresési területre vonatkozó korlátozás nélkül, tehát még közelítő ismeret sem volt szükséges a kötő zseb helyéről. A fenti kihívások ellenére az N-terminális horgonyzó fragmenseit helyesen helyeztük el

és kapcsoltuk össze a szisztematikus BD-keresésünk eredményeivel (Wrapping, azaz csomagoló lépés). Mindent egybevetve, az FBD a szülő módszerek, a fragmentálás és a BD filozófiájának minden előnyét kihasználja. Az FBD jelenlegi korlátai és hibásan dokkolt esetei az egyszerűsített dokkoló erőtérből és az explicit vízmodell hiányából származnak, ezek a korlátok azonban számos esetben javíthatók a korábban már leírt módszerek alapján [28]. Az FBD szisztematikus megközelítése javítja a problémás komplexek nagy ligandumokkal, például hisztonpeptidekkel alkotott szerkezetének meghatározását **(V. tézispont)**.

A kísérleti technikák alternatívájaként elméleti megközelítések alkalmazása lehet célravezető. A hisztonpeptid számítógépes dokkolása [223-225] az olvasó (író) célponthoz ebben az esetben nyilvánvaló elméleti megközelítés lehet, ennek a megközelítésnek azonban gyakorlati problémái vannak. Először is, az egész N-terminális peptidfarok túl nagy a gyors dokkolási számolások számára [226], mert nagy a torziós rugalmassága és emiatt bonyolult keresési probléma alakul ki [227]. Másodszor, a hisztonpeptid kötőhelyeinek közelítő pozíciója a célpont felületén sok esetben nem ismert, a kötőhelyek közelítő, pre-docking pozícióját az olvasófehérjék mély zsebek nélküli, sekély kötőfelületei tovább akadályozzák.

A teljes hisztonpeptid dokkolása helyett a fragmens dokkolása az első problémát tudja kezelni. Korábbi vizsgálatokban a fragmens alapú megközelítést [228] sikerrel alkalmazták, a peptid-dokkolás esetén a fragmentálás a torziós szabadság csökkenéséhez vezet és ez megkönnyíti a keresést. A második problémát megoldja a BD [21, 22, 35], amely a kötőhely előzetes ismerete nélkül vizsgálja a célpont molekula teljes felületét. Egy nemrégiben megjelent tanulmányban [28] egy szisztematikus BD-eljárást adtunk közre a ligandum összes lehetséges kötőhelyének meghatározására a célponton.

Bevezettük a fragmens- és a szisztematikus BD- módszerek kombinációját és bemutattuk a tesztek eredményeit a hisztonpeptidek és célpontjaik komplexeire. Az eredményül kapott módszertant a szülő technikák alapján neveztük el fragmens blind dokkolásnak (FBD) **(IV. tézispont)**.

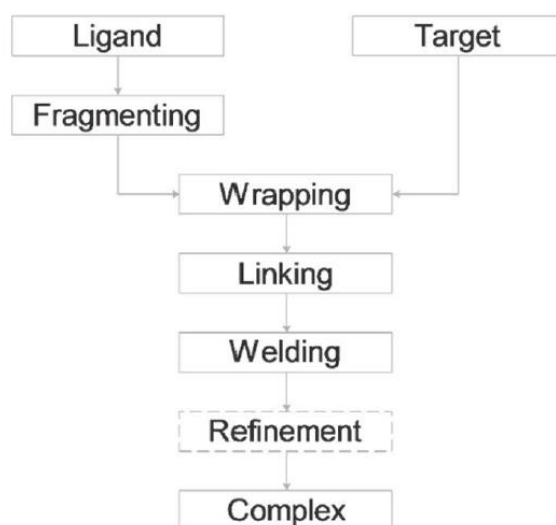
Az FBD kidolgozásához és teszteléséhez öt hiszton H3 peptid célpontkomplexet **(8. táblázat)** használtunk. A komplexek metilezett (3qla, 5tww) és nem-metilezett (2ke1, 2pvc, 4lk9) hiszton H3 farkakat tartalmaznak. A Protein Databank-ban (PDB) tárolt kísérleti komplex szerkezetek többnyire a farkok deka-peptid méretű részeit foglalják magukban [229], így nyilvánvaló kihívást

jelent a farok összes 36 aminosavának a meghatározása a jelenlegi szerkezeti meghatározási technikákkal (röntgen vagy NMR).

Ez a nehézség a mobilitással, valamint a farok C-terminális vége és a célpont fehérje közti kölcsönhatás hiányával, vagy gyenge kölcsönhatásával magyarázható, amelyet a következő bekezdésekben részletesebben elemzek. A fenti kísérleti nehézségek és a hisztoncélpont-komplexek meghatározásának szükségessége motiválta az FBD módszer kidolgozását (**IV. tézispont**). A következő szakaszokban az FBD főbb lépéseit (**12. ábra**) a H3 peptid autoimmun szabályozó fehérje növényi homeodomain-nel alkotott (2ke1 rendszer) komplex (**1. ábra**) szerkezetének segítségével mutatom be.

### 5.2.1. Fragmentálás

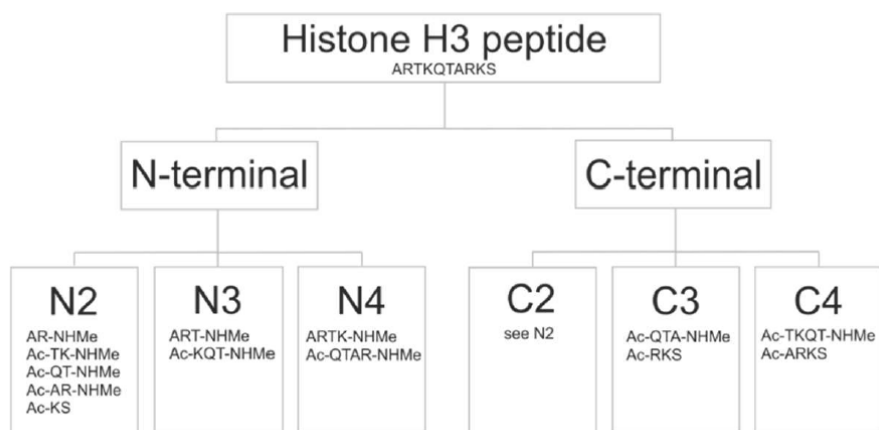
Ahogy az FBD neve is jelzi, a protokoll (**12. ábra**) az eredeti hiszton H3 peptid ligandum (H3) fragmensekre hasításán alapul. A fragmensek kisebb méretűek és kevesebb aktív torzióval rendelkeznek, mint az eredeti peptid. A szabadsági fokok csökkenése miatt várhatóan kevesebb kihívást jelentenek a dokkoló keresési algoritmus számára, amely a fragmenses megközelítés mögött rejlik. Azonban, mivel a fragmensdokkolás fókuszált módban van (lásd a Bevezetésben a referenciákat), nem világos, hogy milyen kis méretűre kell szétvágni a ligandumot, hogy a lehető legjobb eredményt adja a BD-keresés során a célpont teljes felületén.



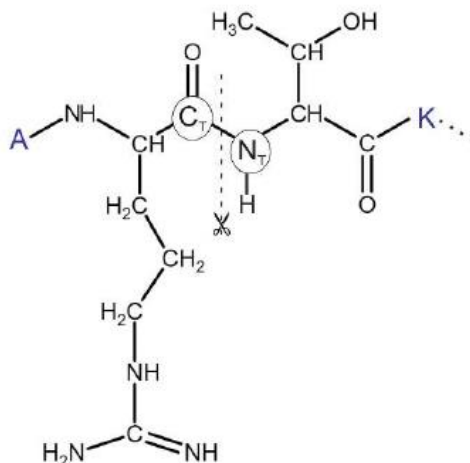
**12. ábra** Fragment blind docking (FBD). Forrás: Bálint et al. 2019.



A kérdés megválaszolásához a H3 fragmenseit egy szisztematikus rendszer szerint terveztük meg (**13. ábra**). A fragmensek mérete di- és tetrapeptidek között volt. Tetrapeptideknél nagyobb fragmenseket nem alkalmaztunk, mivel a fenti darabolás előnyei e méreten túlmenően csökkennének [223]. A fragmentálás kiindulási helyzetének megfelelően két sorozatot állítottunk elő, N- és C-terminális fragmenseknek nevezve el őket (**14. ábra**). A szimpla aminosav-fragmensek elkerülése érdekében dipeptides fragmentálást végeztünk a tri- vagy tetrapeptides fragmentálás során, ha a maradék C- vagy N-terminális szekvencia tetra- vagy pentapeptidek lettek volna.



**13. ábra** A fragmentálás menete. Forrás: Bálint et al. 2019.



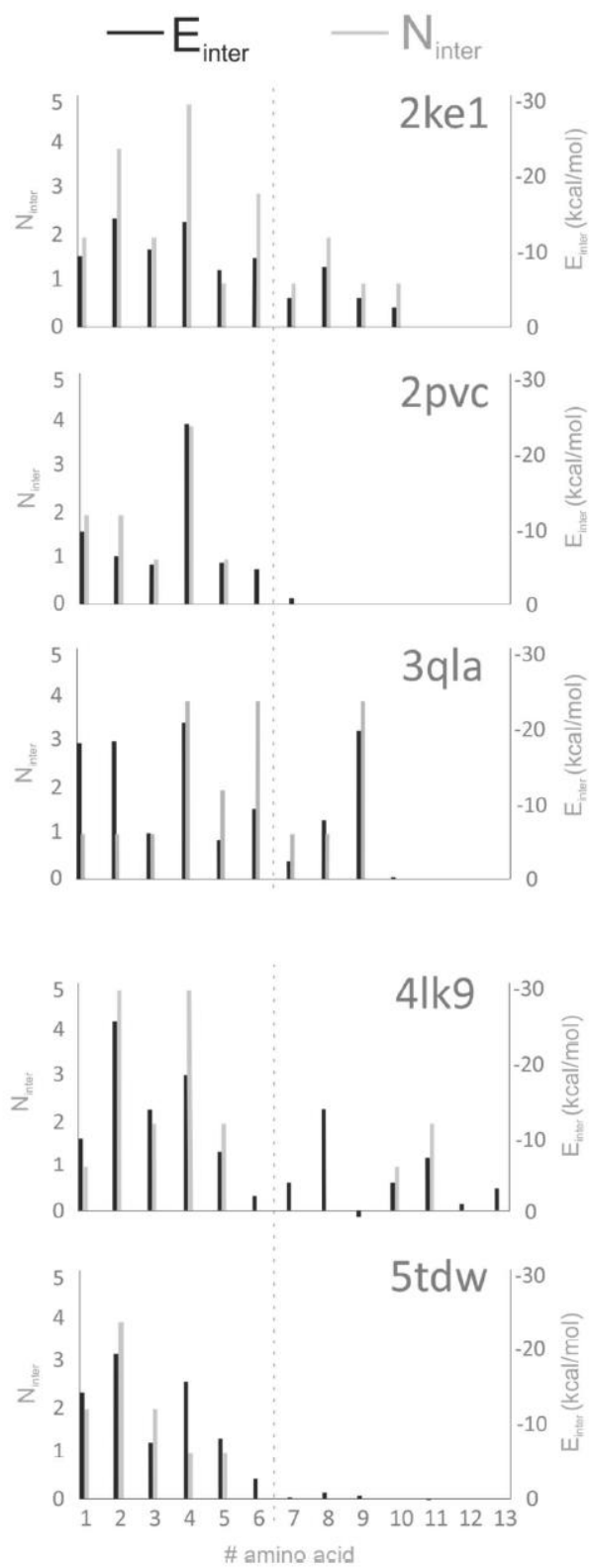
**14. ábra** A H3 hiszton N-terminálisa (ARTK). A peptidet dipeptiddé vágtuk a C<sub>T</sub> és N<sub>T</sub> atomok közötti peptidkötés mentén. Forrás: Bálint et al. 2019.

Ez a fragmentáló módszer szisztematikus, ugyanakkor diverzifikálja a dokkoláshoz használt fragmensek halmazait. Az eredeti H3 peptidet a szén- és a nitrogénatom közötti amidkötés mentén vágtuk el (**14. ábra**). Általánosan acetyl- (Ac-) és imino-methyl- (-NHMe) csoportokat használtunk a peptidek szabad N- és C-terminális levágott végeinek lezárására/blokkolására, amivel a gerinc láncát lehet utánozni. A 2ke1 példarendszer esetében a fragmentálás tizenhárom különböző peptidet eredményezett (**13. ábra**). Az összes fragmens szerkezetét elkészítettük (Módszerek) és összegyűjtöttük egy könyvtárba a Csomagolás lépés BD-ciklusaihoz.

#### Az intermolekuláris kölcsönhatások számítása

A célpont-ligandum kölcsönhatások mélyebb elemzéséhez az intermolekuláris kapcsolatok számát ( $N_{inter}$ ) és az egyes aminosavak közti intermolekuláris kölcsönhatás energiaértékeit ( $E_{inter}$ ) is kiszámítottuk. Ehhez a **8. táblázat** rendszereit használtuk fel, miután az eredeti krisztallográfiás szerkezeteket hidrogénatomokkal láttuk el és energiaminimalizálást hajtottunk végre rajtuk. A H3 peptid ligandumtól mért 3.5 Å-nél kisebb távolságra eső nehéz atomokat tartalmazó aminosavakat vizsgáltuk. A kölcsönhatásban részt vevő aminosavak listája megtalálható a **9. táblázatban**, az  $N_{inter}$  értékeket a **15. ábra** mutatja.

Az elemzés eredményei (**15. ábra** és **11. táblázat**) megmutatták, hogy az  $R_2$  valóban fontos rögzítő aminosav, amely a legnagyobb  $N_{inter}$  és a legjobb  $E_{inter}$  értékekkel rendelkezik. A 3qla rendszer esetében egy rosszul dokkolt AR-NHMe jött létre a negatív töltésű aminosavak hiánya és a célponttal kölcsönható ATRX-DNMT3-DNMT3L (ATRX-ADD) doménoldal miatt, ami egy egyszerű érintkezést eredményezett az  $R_2$ -nél (**15. ábra**). A 2pvc esetében az AR-NHMe fragmens 3.2 Å RMSD-értéknél találta meg a kötő szerkezetet a második horgonyzó (anchoring) ciklusban. Ennek oka az  $R_2$  viszonylag kicsi  $E_{inter}$  értéke  $K_4$ -hez viszonyítva, ami a fő horgonyzó aminosav (**15. ábra**). A  $K_4$  jelentőségét a sikeres dokkolás is tükrözi, amikor  $K_4$  a fragmentált szekvencia része volt ( $N_2$ ,  $N_3$  és  $C_2$  a **11. táblázatban**).



**15. ábra** A H3 hisztonligandum első 13 aminosavjának aminosavankénti célpont-ligandum intermolekuláris kölcsönhatási energiája ( $E_{inter}$ ) és az intermolekuláris kölcsönhatások száma ( $N_{inter}$ ). Forrás: Bálint et al. 2019

**9. táblázat**  $E_{inter}$  és  $N_{inter}$  értékei

H3 aminosav	2ke1		2pvc		3qla		4lk9		5tdw	
	$E_{inter}$	$N_{inter}$	$E_{inter}$	$N_{inter}$	$E_{inter}$	$N_{inter}$	$E_{inter}$	$N_{inter}$	$E_{inter}$	$N_{inter}$
A1	-10.4	2	-10.8	2	-20.2	1	-11.0	1	-15.7	2
R2	-16.0	4	-7.2	2	-20.5	1	-28.4	5	-21.5	4
T3	-11.4	2	-5.9	1	-6.8	1	-15.3	2	-8.2	2
K4	-15.5	5	-26.8	4	-23.2	4	-20.5	5	-17.4	2
Q5	-8.3	1	-6.2	1	-5.8	2	-9.0	2	-8.8	1
T6	-10.1	3	-5.3		-10.4	4	-2.2	0	-3.0	1
A7	-4.4	1	-0.9		-2.6	1	-4.2	0	-0.2	0
R8	-8.8	2			-8.7	1	-15.4	0	-0.9	0
K9	-4.4	1			-22.0	4	0.9	0	-0.4	0
S10	-2.9	1			-0.2	0	-4.2	1	-0.0	0
T11							-8.1	2	0.2	0
G12							-1.0	0		
G13							-3.4	0		

Korábbi dokkoló tanulmányok [223] megmutatták, hogy a kis, aktív torziókkal rendelkező peptidekhez megfelelően dokkolt szerkezetek nyerhetők, míg a nagy peptidek valódi kihívást jelentenek a legtöbb dokkolási eljárásban [223, 230, 231]. Ezekkel a korábbi eredményekkel összhangban alacsony RMSD-értékeket kaptunk, főleg az N2- vagy C2-sorozatú dipeptid fragmenseknél (**11. táblázat**). A fragmentálás kiindulópontjától (C- vagy N-terminális) függően különböző peptidkészleteket kaphatunk. Valószínűleg azonos fragmenskészletek állíthatók elő, ha például a dipeptid fragmenseket a H3 peptidből egyenlő számú aminosavakból állítjuk elő (**13. ábra**), azonban az N-terminális fragmentálása sikeres volt minden fragmenshossznál, mivel a horgonyzó aminosavak közel állnak a H3 N-terminálisához. C-terminálisú fragmentálás esetén az első néhány fragmens gyakran gyengén kötődik a célpontfehérjéhez (**15. ábra**, 4lk9 rendszer), vagy bizonyos esetekben teljesen elvándorolnak az oldószerbe (**15. ábra**, 5tdw rendszer). Ugyanakkor az utolsó fragmensek általában nagyobb (tri- vagy tetra-) peptidek, amelyek megnehezítik a referenciaszerkezetből ismert kötési szerkezetek megtalálását, amint azt a fentiekben is megmutattam.

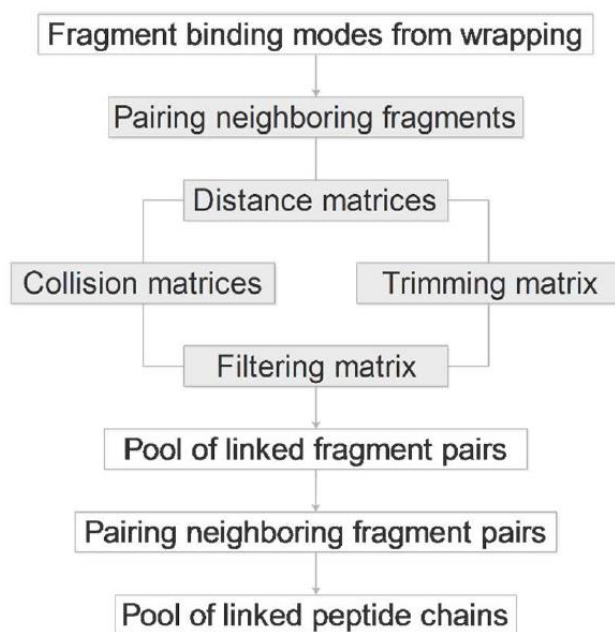
Bizonyos esetekben a tripeptid (2ke1) és a tetrapeptid (2pvc) fragmensek is jó RMSD-értékeket (2–3) adtak az első rank-ek esetén (**11. táblázat**). Ezért a tri- és tetrapeptidek hasznosak lehetnek

más rendszerek FBD-jében is, azonban az eredeti H3 ligandum (részleges) rekonstrukciójához a dipeptidek általában alkalmasabbak voltak, mint a hosszabb láncok.

Elemzéseink tehát a ligandum oldali aminosavak szintjén megmutatták, hogy dokkolási szempontból a H3 peptidek mely fragmensei különösen problematikusak és hogy melyek alkalmazhatók hatékonyan a következő fejezetben leírt összekapcsolási lépésben.

### 5.2.2. Összekapcsolás

A csomagolási lépésben dokkolt összes fragmens összes kötési módját figyelembe véve az azonos hosszúságú fragmensek párhuzamos összekapcsolása automatikusan történik az FBD-ben kidolgozott és a következő bekezdésekben leírt algoritmus segítségével (**IV. tézispont**). Az algoritmus megvizsgálja az összes lehetséges páros kombinációját és a leghosszabb lehetséges peptidet hozza létre a kapcsolódó aminosav-párokból. Az algoritmus folyamatát (**16. ábra**) a H3 hiszton ARTKQT N-terminális hexapeptidjére írtam le (**V. tézispont**).



**16. ábra** Az összekapcsoló algoritmus (a szürke dobozok utalnak arra, hogy minden fragmensre végre kell hajtani). Forrás: Bálint et al. 2019.

Tegyük fel, hogy ezt a hexapeptidet három dipeptidfragmensre (AR, TK és QT) hasítottuk, ahogyan azt a Fragmentálás fejezetben leírtam. Az FBD esetében a dipeptideket a hasítási helyeken Ac- és/vagy -NHMe csoportokkal blokkoltuk. Az összekapcsoló algoritmus először az  $n_1$  db kötési mód

AR-NHMe első dipeptidpárját és  $n_2$  db kötési mód Ac-TK-NHMe párját veszi be a feldolgozási sorba és eltávolítja a végekről a blokkoló csoportokat. Ily módon a szabadgyökök két sorát kapjuk az amidkötések létrehozásához rendelkezésre álló terminális szén- ( $C_T$ ) és nitrogénatomokkal ( $N_T$ , **14. ábra**). Azonban nem minden fragmenskötési mód van megfelelő helyzetben, hogy létrejöhessen az amidkötés. Bizonyos esetekben a  $C_T$  és  $N_T$  ( $d_{CN}$ ) közötti távolság túl nagy ahhoz, hogy a kovalens kötés kialakulhasson. A megfelelő távolságokkal rendelkező másolatok kiválasztásához minden ( $n_1 \times n_2$ )  $d_{CN}$  értéket kiszámítunk és egy mátrixba mentjük azokat (**10. táblázat**). Az összes dipeptid fragmenspár  $d$ -értékét kiszámoljuk, ezért a hexapeptides példánk esetében két  $d_{CN}$ -mátrixot állítunk elő. A felhasználó által meghatározott minimális távolságtolerancia  $d_{CN, \min} = 0.75 \text{ \AA}$ , ami egy amidcsoport  $C_T$ - $N_T$  kötésének félhosszának felel meg. Ha a tényleges  $d_{CN} \leq d_{CN, \min}$ , akkor az ütközési mátrix ( $c_{CN}$ ) megfelelő elemét nullára állítjuk, azaz a  $C_T$  és az  $N_T$  között ütközés történik. A  $6 \text{ \AA}$  maximális távolságtoleranciát ( $d_{CN, \max}$ ) is úgy definiáljuk, hogy kizárja azt, hogy a fragmenspárok egymástól túl messzire kerüljenek; ezen érték alapján egy  $C_N$  vágási mátrixot állítunk elő. A  $d_{CN, \max}$  nem lehet túl nagy, értelmes fizikai értékkel kell rendelkeznie.

**10. táblázat** Az összekapcsolási folyamat során használt mátrixok

Szimbólum	Leírás
$d_{all}$	A legkisebb távolságok a nehéz atomok között
$c_{all}$	Ütközési mátrix az összes nehéz atom között
$d_{CN}$	$C_T$ és $N_T$ atomok közötti számolt távolság
$c_{CN}$	Ütközési mátrix $d_{CN}$ -ből, $c_{CN} = 1$ , ha nincs ütközés; egyébként 0
$t_{CN}$	Kivágási mátrix $d_{CN}$ -ből, $t_{CN} = 1$ , ha nincs ütközés; egyébként 0
$f_{CN}$	Szűrőmátrix az ütközési és kivágási mátrixokból, $f_{CN} = 1$ , ha $c_{CN} = t_{CN} = 1$ ; egyébként 0

A fragmenspárok bármely atomja közötti általános ütközések elkerülése érdekében az összes nehéz atom közötti távolságokból létrehozunk egy távolságmátrixot (**10. táblázat**). Egy  $c$  ütközési mátrixot is kiszámítunk a  $d_{all}$ -ból, hogy azonosítsuk a fragmensek nehéz atomjai közötti szterikus ütközéseket. Egy ütközés akkor jön létre, ha a tényleges  $d_{all} \leq d_{all, \min}$ , ahol a  $d_{all, \min}$  a felhasználó által meghatározott minimális távolságtűrés;  $d_{all, \min} = 0.75 \text{ \AA}$ , ugyanaz, mint a fenti  $d_{CN, \min}$ , ami jól működik a peptid ligandumok esetében.  $c_{all}$  elemeit alapértelmezés szerint nullára állítjuk és

egyre, ha nincs ütközés az atomok között. Végül egy  $f_{CN}$  szűrési mátrixot állítunk elő, amely azt mutatja, hogy egy fragmenspár alkalmas-e az összekötésre és finomításra (lásd a következő szakaszt). Az  $f_{CN}$  értéke 1, ha  $t_{CN}$  és  $c_{CN}$  egyenlők, egyébként nulla.

A távolságmátrix elemeit  $d_{CN}$  szerint növekvő sorrendbe rendezzük, majd a fő ciklus kiválasztja az első (következő) még nem ellenőrzött fragmenspárt a listából a legkisebb  $d_{CN}$  értékkel és ellenőrzi a szűrési mátrix megfelelő elemét. Ha  $f_{CN} = 0$ , akkor a következő fragmenspárt ellenőrzi, egyébként a fragmenspárt menti a lehetséges fragmenspárok listájában.

Miután példánkban az AR és TK első fragmenspárjának jelölt szerkezeteinek egy csoportját előállítottuk, ugyanezt az eljárást ismételjük meg a következő TK és QT fragmenspár esetében is, amennyiben az ARTKQT példa szerinti hexapeptidet tekintjük. Miután befejeztük az összes fragmenspár vizsgálatát (példánkban két pár), folytatjuk a feldolgozást, hogy összekapcsoljuk a fragmenspárokat triádokká. Ennek megfelelően a legkisebb  $d_{CN}$  értékkel rendelkező első AR-TK fragmenspárt választjuk ki. Ha ugyanaz a TK-fragmens a TK-QT-párok egyikében előfordul, akkor létrejön az ARTKQT hexapeptid szerkezete (**13. ábra**,  $AR_2-TK_3-QT_1$  vagy  $AR_2-TK_3-QT_4$  stb.). Ha nem ez az eset fordul elő, akkor a következő AR-TK-párnak TK-QT-párokkal közös TK-ját ellenőrzi és az algoritmus addig folytatódik, amíg az összes AR-TK-páron végig nem megy. Az összekapcsoló algoritmus a fenti párosítási folyamat minden szintjén egy szerkezeti gyűjteményt állít elő. A folyamat tetszőleges hosszú peptidláncokon működik, ami azt jelenti, hogy a peptid hosszúságától függően fragmenspárok, triádok és tetrádok állíthatók elő. A gyűjtemények statisztikáit a program egy riportba írja ki (**Függelék D. táblázat**).

### 5.2.3. Összefűzés és finomítás

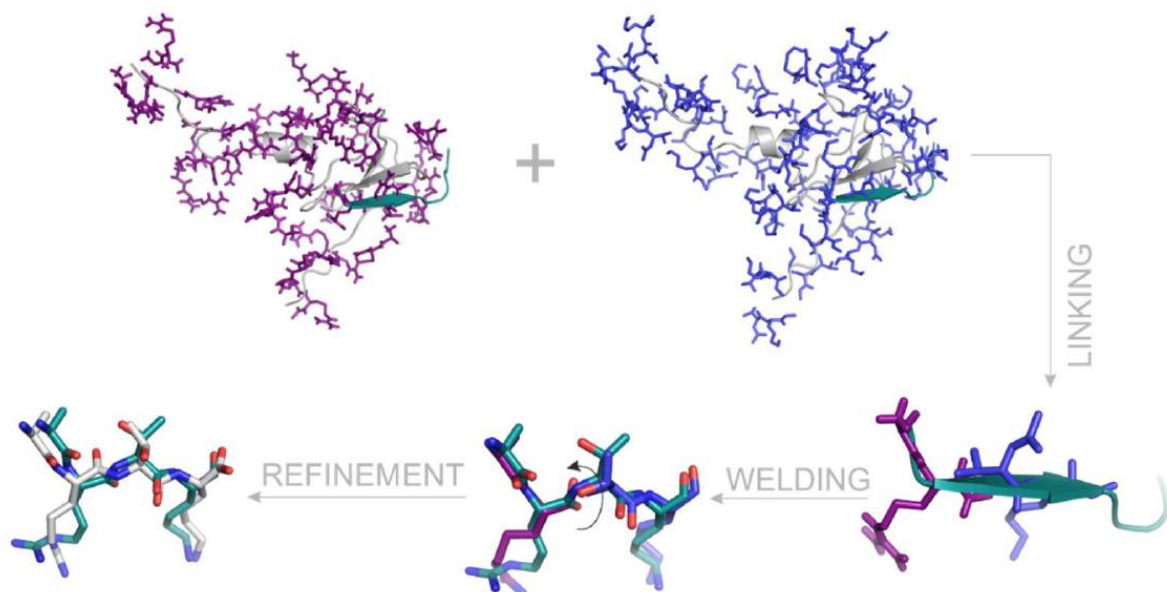
Ezután az összekapcsolt jelölt szerkezetek gyűjteményét további feldolgozásnak vetjük alá a  $C_T$  és  $N_T$  atomok közötti kovalens kötések újrakötéséhez az AR és TK fragmensek között (**14. ábra**). Az illesztés során az AR-t a  $C_\alpha-C_T$  ( $\Psi$  szög), a TK-t az  $N_T-C_\alpha$  ( $\Phi$  szög) kötések mentén forgatjuk el szisztematikusan, 1 fokos lépésközzel.  $\Psi$  szögben egy forgatási lépést követően  $\Phi$  mentén forgatunk egy teljes fordulatig, azaz 360 fokig. Minden egyes forgatási lépés után számoljuk  $d_{CN}$ -t és a megfelelő szögekkel együtt tároljuk. A tárolt  $d_{CN}$  értékek közül a legkisebbet választjuk ki és a megfelelő szerkezetet illesztéssel kapjuk. Ugyanezt az illesztési eljárást kell követni az ARTKQTARKS fennmaradó fragmenseiben, amelyek az összekötött jelöltek listájában szerepelnek.

Az összekapcsolási és illesztési folyamatokat követően a párosított fragmensek szerkezeti finomítását is ajánljuk egy általános MM energia-minimalizálással, lehetőleg explicit vízmodellt is alkalmazva. A 2ke1 rendszer esetében a dokkolás az első két dipeptid fragmenst 1.9 (AR-NHMe) és 2.8 (Ac-TK-NHMe) Å-ös RMSD-értéken találta meg (**11. táblázat**), ezt a két fragmenst a **Függelék D. táblázatában** AR04 és TK01 néven találjuk meg. A dokkolásból nyert fragmenseket összekapcsoltuk, ahogyan azt a **17. ábra** mutatja és a finomítás után az amidkötés  $d_{CN}$  értéke 1.5-ről 1.3 Å-re változott. Az ARTK optimalizált szerkezete megfelel a röntgenszerkezetnek 1.3 Å RMSD-értéken. Ez figyelemre méltó javulás, figyelembe véve a fenti dipeptid fragmensek előbb említett RMSD-értékeit.

**11. Táblázat** Dokkolási eredmények, #sorszám

Rendszer	Fragmenttípus	Fragmens szekvencia	#Ciklus	#Rank	RMSD (Å)
2ke1	N2=C2	AR-NHMe	1	4	1.9
	N2=C2	Ac-TK-NHMe	1	1	2.8
	N3	Ac-KQT-NHMe	2	1	2.4
2pvc	N2	AR-NHMe	1	5	3.2
	N2	Ac-TK-NHMe	1	3	2.8
	N3	Ac-KQTA	1	1	3.0
	C2	Ac-KQ-NHMe	1	2	1.7
3qla	N2=C2	Ac-QT-NHMe	1	12	2.0
	N2=C2	Ac-K(Me <sub>3</sub> )S	1	2	3.7
	N4	ARTK-NHMe	1	1	3.3
4lk9	N2	AR-NHMe	1	1	2.9
	N2	Ac-TK-NHMe	1	4	4.0
5tdw	N2	AR-NHMe	1	1	1.7





**17. ábra** Az egyes lépések folyamata a 2ke1 rendszer H3 hisztonpeptid ligandumából származó AR-NHMe és Ac-TK-NHMe fragmenspárookra. Az összehasonlítás kedvéért a kristallografikus ligandum szerkezet teal cartoon formában látszik az első három képen és teal stick-ben az utolsó kettőn. Az AIRE PHD célpontot szürke cartoon-nal, a számolt kötőmódokat lila és kék pálcikákkal jelöli az ábra. Forrás: Bálint et al. 2019.

### 5.3. Vízhálózatok számítása

A harmadik kéziratban (Jeszenői et al. 2018) vízhálózatok hőmérsékletfüggését vizsgáltuk fehérjefelületen és komplexek interfész-régióiban. Egy standard validációs protokollt alkalmaztunk [25, 167], amelynek segítségével egy automatikus minősítéssel hasonlíthatjuk össze a referencia- és az előrejelzett hidratációs szerkezeteket (**VI. tézispont**). Korábban ezt a validációs protokollt a toleranciaértékek kalibrálásához és a legjobb klaszterezési séma kiválasztásához alkalmaztuk a MobyWat program számára [167], azaz a validációs protokoll a kísérleti és a jósolt vízpozíciók közötti egyezések azonosításán alapult.

$$SR = 100 \frac{\text{Találatok száma}}{\text{Referencia vízpozíciók száma}} \% \quad (5)$$

**12. táblázat** A vízhálózatok szempontjából tárgyalt rendszerek.

PDB kód	Rendszer (ligandum)	Felbontás (Å)	Kristályosodás T (K)	Adatgyűjtés T (K)	Adatgyűjtés p (MPa) <sup>a</sup>	Vízek száma <sup>a b</sup>
Felszín hidratálása						
2o9s	SH3 domain from ponsin	0.83	na <sup>d</sup>	100	na	108
2vb1 <sup>c</sup>	Hen egg white lysozyme	0.65	292	100	na	144
3nir	Crambin	0.48	na	100	na	65
Interfész hidratálása						
3ro3	G-protein-signaling modulator 2 (QVDSVQRWMEDLKLMTE)	1.10	288	100	na	12
4h9o	death domain-associated protein 6 (histone H3.3 G90 M mutant/H4)	2.05	277	na	na	35

<sup>a</sup> Atmoszférikus nyomást feltételezve, ahol nem volt nyomás megadva

<sup>b</sup> A kristályszerkezeti vízmolekulák száma  $b_{\max} = 100 \text{ Å}^2$  és  $d_{\max} = 3.5 \text{ Å}$

<sup>c</sup> Az adatgyűjtési hőmérséklet szimulációjának hatékonyságát vizsgáló szimulációkban felhasznált HEWL (Hen egg white lysozyme)

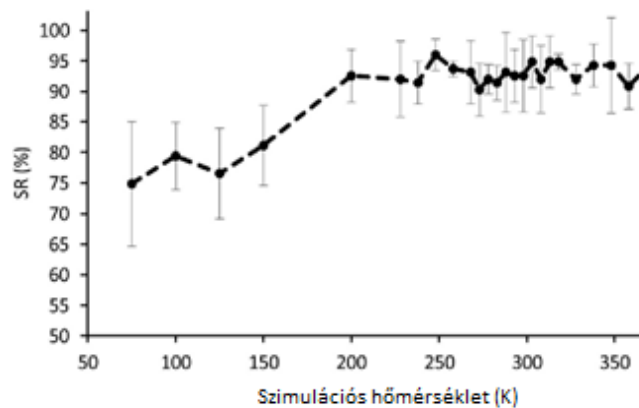
<sup>d</sup> Nem elérhető

Akkor beszélhetünk egyezésről, ha a várható és a referencia víz-oxigénatomok közötti távolság kisebb egy előre meghatározott értékénél. A találatokból egy sikerességi arányt (success rate, SR) számolunk (5. egyenlet), ahol a nevező a kristályban levő referencia vízpozíciókat tartalmazza, amelyek a maximális távolságlimiten ( $d_{\max}$ ) belül vannak. A 100%-os SR-érték teljes egyezést és optimális MD-feltételeket jelent. A MobyWat-előrejelzéseken kívül az SR-értékeket az egyes MD-frame-ekre is kiszámítottuk, ezekben az esetekben az  $SR_n$  szimbólumot használjuk a szövegben,

ahol  $n$  jelöli a MD-frame sorozatszámát. Így az  $SR_n$ -értékek a MD-frame-ekre utaló nyers adatokon alapulnak, és csak összehasonlítás céljából használjuk őket.

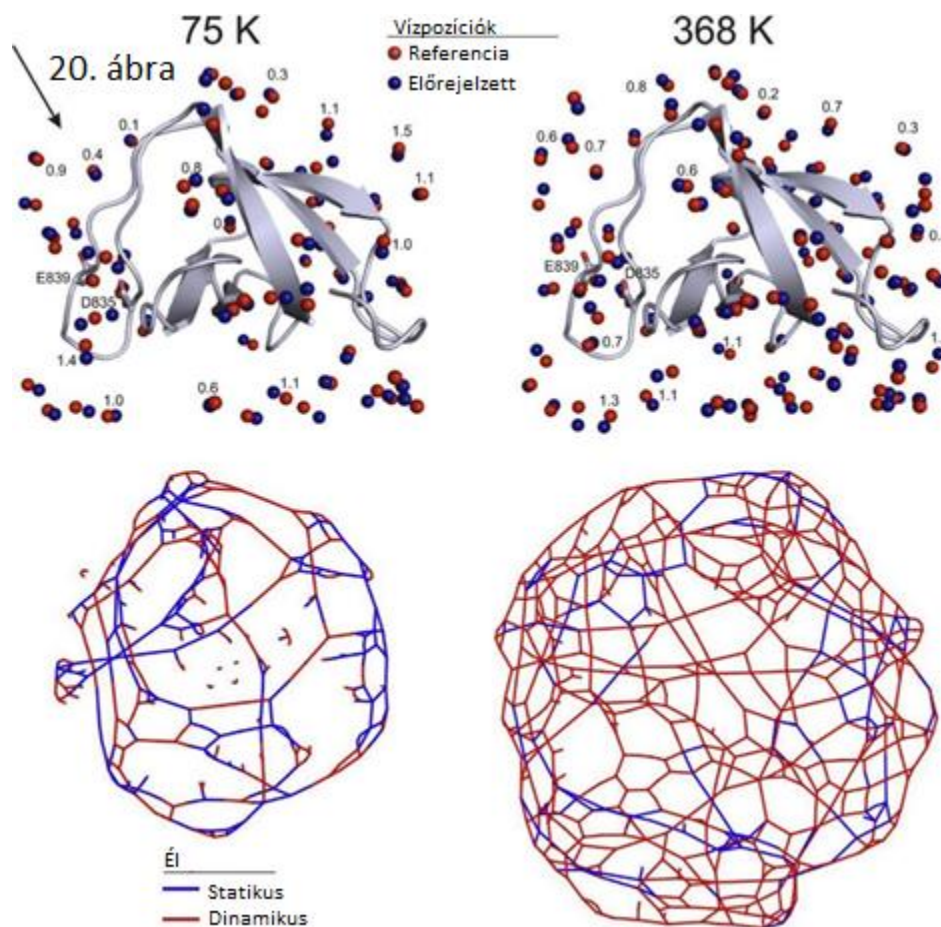
### 5.3.1. Felületi hidratáció

A **12. táblázatban** található meg a vizsgált rendszerek. Magas hőmérsékleten (368 K) több vízpozíció jósolható meg, mint alacsony hőmérsékleten (75 K) ugyanazon fehérje esetén (**18. ábra**), ami az SR növekedésének egyik oka.



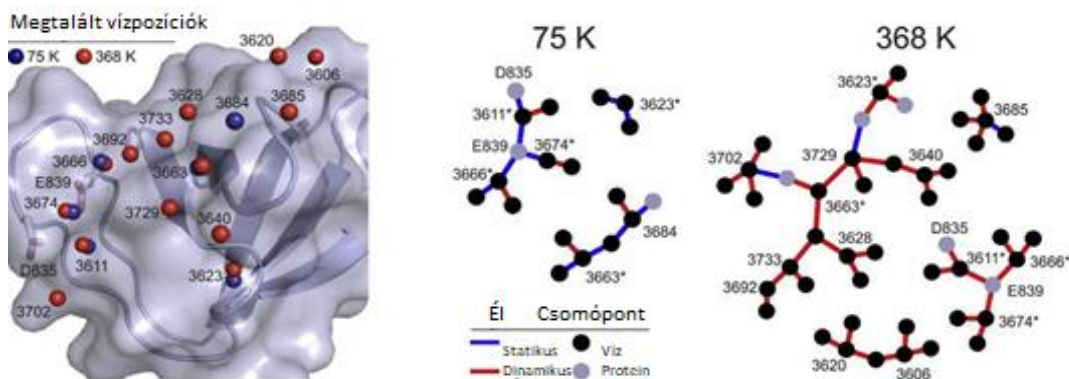
**18. ábra** A szimulációs hőmérséklet hatása a komplex interfész (4h9o rendszer) hidratációs szerkezete előrejelzésének sikerességi rátáira. Minden adatpont egy öt sikeres szimulációból kiszámított átlagos sikerességi rátát mutat, különböző kezdeti sebességeloszlásokkal. A hibaszávok a standard eltéréseket jelölik. Forrás: Jeszenői et al. 2018.

További magyarázatként a várható hidratációs szerkezetek hálózati gráfjait a MobyWat NetDraw módjával készítettük el, a gráfinformáció alapján a csúcsoakat és az éleket statikus és dinamikus kategóriákba soroltuk. A dinamikus csúcsoak arányát az összes csúcs közül több mint 15%-kal magasabbnak találtuk 368 K-en, mint 75 K-en. Ennek eredményeként a kiterjedt dinamikus alhálózati régiók összekapcsolódtak, a statikus alhálózatok vízpozíciói stabilizálódtak 368 K-en (a **19. ábrán** a piros élek) [25] és növekedett az SR.



**19. ábra** Felül: a ponsinból (2o9s) származó SH3 domén felszínén a becsült és a kísérleti vízpozíciók illeszkedése 75, illetve 368 K hőmérsékleten. A találati távolságok Å-ben vannak megadva a jobb összehasonlíthatóság miatt. Alul: a megfelelő hidratációs hálózati gráfok az előrejelzett vízpozíciókból generálva. A nyíl a negatív töltésű D835 és E839 aminosavak körüli régióra mutat, amelyet részletesebben a **20. ábrán** tárgyalunk. Forrás: Jeszenői et al. 2018.

A dinamikus hálózatépítés ilyen stabilizáló hatása nem működik 75 K hőmérsékleten, ahol a gyengén összekapcsolt statikus (kék) régiók dominálnak egy viszonylag kis gráfban. Egy jellegzetes szerkezet részleteit mutatja a **20. ábra**, ahol a statikus 3611, 3666 és a 3674-es vízpozíciók a D835 és E839-hez csatlakoznak és mindkét hőmérsékleten mély zsebekben tudtuk őket reprodukálni; a dinamikus helyzetet csak 368K-en lehetett reprodukálni (**VI. tézispont**).



**20. ábra** A megjósolt vízpozíciók részletei 75 és 368 K hőmérsékleten, amely összhangban áll a 2o9s rendszer kristallografikus referenciapozícióival. Az ábra a fehérje negatív töltésű régiójának felel meg, amint azt a **19. ábrán** mutató nyíl jelzi. A számok a 2o9s-ben lévő vizek számozásának felelnek meg. Az algráfok az előrejelzett vízmolekulák első kapcsolódási szféráját tartalmazzák. A mindkét hőmérsékleten megtalált pozíciókat csillaggal jelöljük a 75 K-es alábbián. Forrás: Jeszenői et al. 2018.

A 75 K hőmérsékleten előállított hidratációs gráfban kisebb, egymással kapcsolatban nem álló algráfok figyelhetők meg. 368 K-en megnövekedett az algráfok száma és sűrűn összekapcsolódtak a felszíneken található dinamikus vízcsúcsok (3663 és 3729) mentén a teljes felületen.

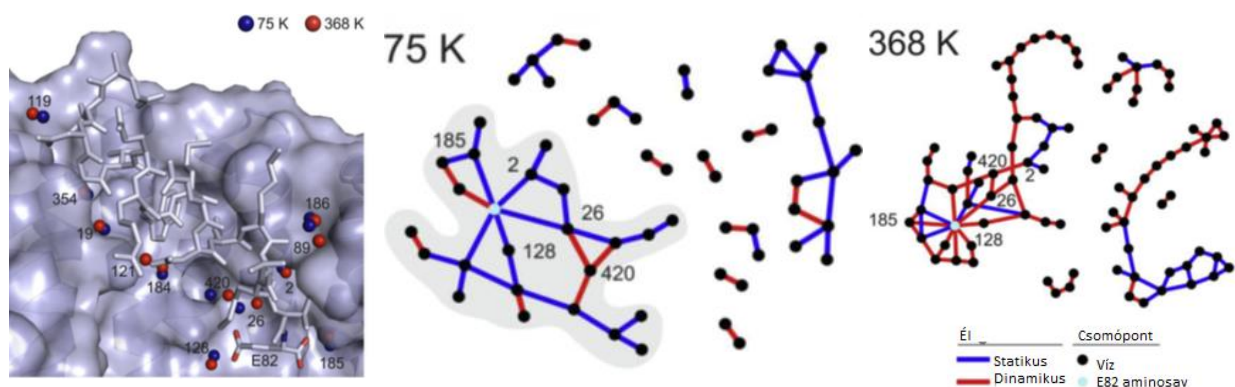
Míg a rendelkezésre álló kötőhelyek hatékony feltárása megköveteli a hőmérséklet emelését, a nagyon magas hőmérséklet ellentétes hatást gyakorolhat az SR-re, ami nagyobb mobilitást, alacsony klaszter-kihasználtságot és végül az SR-görbék telítettségét eredményezi (**18. ábra**), amint a hőmérséklet meghaladja a 200 K-t. Az előrejelzések trajektóriánkénti SR-értékeit, mint pl. a hőmérséklet különbségeit a MobyWat elemzési módjában kiszámított perframe  $SR_n$ -értékekkel lehet szemléltetni. Mindegyik rendszerhez és a kiindulási sebességeloszlási kombinációhoz (MD-trajektória) egyetlen SR-érték és 1001  $SR_n$  tartozik. Így az  $SR_n$  a MD-szimuláció során a hidratációs réteg ingadozását (evolúcióját) tükrözi, az  $SR_n$  statisztikája azt mutatja, hogy az ingadozása a hőmérséklettől függ. A 2vb1 HEWL-fehérje példája azt szemlélteti, hogy a frame-enkénti  $SR_n$ -értékek ingadozása szignifikánsan nagyobb 368 K-en, mint 75 K-en a vízmobilitás fent említett különbségei miatt (Bálint et al. 2019).

Az átlagos  $SR_n$ -értékek ugyancsak magasabbak 368 K-en, mint 75 K-en, az IDa-klaszterezés során elért átlagos SR (84%) alatt maradnak.

### 5.3.2. Interfész hidratáció

Az interfész vízmolekuláinak általános mobilitása topológiaiilag korlátozott, magas hőmérsékleteken a nagy vízmobilitás (**5. egyenlet**) felülmúlja ezeket a korlátozásokat, az interfész hozzáférhetővé válik, így a vízpozíciók nagy csoportja biztosított mind az öt MobyWat-klaszterezés számára. Azaz a magas hőmérséklet alkalmazása megnöveli a vízmolekulák mobilitását, ami javítja a MobyWat foglaltságon alapuló algoritmusának hatékonyságát. Így a kapott SR-értékek és szórásuk közötti különbség csökken. Felületi hidratáció esetében, ahol ilyen topológiai korlátozások nincsenek, a vízmolekulák szabadon elfoglalhatják a hidratációs helyeket mind az öt pályán és a hőmérséklet növekedése nem korrelál az SR szórásával.

Az interfészeknél magasabb SR-értékek figyelhetők meg 75 K-en, mint a felületeknél (**18. ábra**), valamint a felületek legnagyobb átlagos SR-értéke 60% (3nir, Bálint et al. 2019), ami 15 százalékponttal alacsonyabb, mint az interfészek ugyanazon értékénél (75%, 4h9o). Még a referenciákkal való teljes egyezést is sikerült elérni (3ro3, T = 100 K, 5. seed). Az összes interfészvíz megtalálása nem volt ritka, míg a felületi előrejelzések legjobb SR-értéke 91% volt (2vb1, T = 318 K, 3. seed, Bálint et al. 2019). Alacsony és magas hőmérsékleten egyaránt találtunk magas SR-értékeket (3ro3, 75 K-en 92%, 368 K hőmérsékleten 100%), különbségek az E82 aminosav környéki hidratációs gráfban figyelhetők meg (**21. ábra**).



**21. ábra** Előrejelzett vízpozíciók 75 és 368 K hőmérsékleten, amelyek megegyeznek a 3ro3 rendszer kristallografikus referenciapozícióival; a számok megfelelnek a vizek azonosítóinak a 3ro3 rendszerben. Az E82 ligandum aminosav körül 75 K-en statikus alhálózat helyezkedik el (szürke terület), ez a központi régió 368 K-en dinamikussá válik, ahol a dinamikus vizek a fragmentált statikus alhálózatot veszik körül. Forrás: Jeszenői et al. 2018.

Kiterjedt statikus alhálózatok alakulnak ki 75 K-en az E82 középpontjából kiinduló, az összesen hatból öt statikus éllel. A hálózat több alhálózatra oszlik szét és hiányoznak a statikus régiókat egymással összekötő dinamikus vizek. 368 K-en ez megváltozik, statikus helyett többnyire dinamikus kapcsolatok képződnek az E82 aminosav körül és a hidratációs hálózat nagy része dinamikussá válik (**VI. tézispont**). Az interfész-előrejelzések magas SR-értékei és azok hőmérsékleti változásokkal szembeni relatív immunitása a ligandum molekula és a párhuzamos lépések (**21. ábra**) alkalmazásának következményei az előrejelzési sémában.

## 6. Összefoglalás

A jelen PhD-disszertáció a célpont-ligandum komplexeken elvégzett számításaink eredményeit mutatja be a molekulatervezés fragmens alapú módszereinek alkalmazásával. A fragmens alapú megközelítést a komplexnek mind a célpont, mind a ligandum részein alkalmaztuk és a hidratációs hálózatok hatásait is megvizsgáltuk. A tanulmány olyan farmakológiai jelentőségű rendszereket is magában foglal, mint a hiszton-komplexek, amelyek az epigenetikus gyógyszertervezés kulcsszereplői. Kihívást jelentő nagy rugalmasságú és kiterjedt hidratációval rendelkező peptideken teszteltük a szerkezetmeghatározó és számoló módszereinket.

Egy új fragmentáló eljárást (Fragmenter) dolgoztunk ki a célpontokra. A Fragmenter a célpont-ligandum komplexek felületei szerkezetének automatizált és szabványosított kivágását hajtja végre kvantummechanikai számítások számára. Kivágott interfész-szerkezetek vizsgálatával szisztematikus vizsgálatunk megmutatta, hogy egy hibrid vízmodell alkalmazásával kapjuk a legjobb korrelációt szemiempirikus kvantumkémiai szinten a számított reakcióhők és a kísérleti kötési entalpiaváltozások között. Ez az új modell kombinálja a COSMO implicit vízmodellt és a korábban publikált MobyWat módszerrel előrejelzett explicit vízmolekulákat. Kiszámítottunk egy skálázó faktort is, amely egyszerű, lineáris kapcsolatot hoz létre a szemiempirikus kvantumkémiai szinten számolt reakcióhők és a kísérleti kötési entalpiaváltozások között tengelymetszet használata nélkül. Ily módon egy új, közvetlen kapcsolatot tártunk fel a molekulaszerkezet és a kötési termodinamika között. Kidolgoztuk és teszteltük a fragment blind docking (FBD) eljárást, amely új módszerként kombinálja a blind docking módszert és a fragmens dokkolás technikáját. Az FBD általános algoritmusával kidolgoztunk, elkészítettünk és kalibráltunk egy eljárást, amelynek segítségével a célpont csomagolási lépése után a peptid fragmenseket összekapcsoltuk és összehegesztettük. Kiszámoltuk hidratáló vízmolekulák hálózatának szerkezetét és kimutattuk hőmérsékletfüggésüket interfészek és teljes fehérjefelületek esetén. Hálózati megfontolásaink segítségével magyarázatot adtunk a vízszerkezet molekuladinamikán alapuló számításainak sikerére.

A dolgozat eredményeit három kutatási publikációban és egy, a tudományos közösség számára szabadon elérhető weboldalon tettük közzé.



## 7. Summary

This thesis presents our results on calculation of target-ligand complexes using fragment-based methods of molecular design. The fragmenting approach was adopted on both the target and ligand parts of the complex and the effects of hydration networks were investigated, as well. The study involves systems of pharmacological relevance such as histone complexes, key participants of epigenetic drug discovery. Peptides with high flexibility and extensive hydration were also involved as challenging test systems of current structure determination and calculation methods. A new fragmenting procedure (Fragmenter) was elaborated on the target side. Fragmenter performs an automatized and standardized extraction of structures of interfaces of target-ligand complexes for quantum mechanical calculations. Using a series of extracted interface structures, our systematic investigation showed that a hybrid water model yields the best correlation between reaction heats calculated at semi-empirical quantum chemical level and experimental binding enthalpy changes. This new model combines the COSMO implicit water model and explicit water molecules calculated by our previously published MobyWat method. A scaling factor was derived which establish a simple, linear relationship between reaction heats calculated at semi-empirical quantum chemical level and experimental binding enthalpy changes without the involvement of an intercept. In this way, a new, direct relationship was established between molecular structure and binding thermodynamics. The fragment blind docking (FBD) algorithm was elaborated and tested which is a new method combining our blind docking approach and the fragment docking technique. Within the general algorithm of FBD, a procedure was elaborated, coded and calibrated for linking and welding of peptide fragments covering the target after the wrapping step. The structure of networks of hydrating water molecules was calculated and their temperature dependence was uncovered in the cases of interfaces and whole protein surfaces. With the help of our network considerations, an explanation was given for the success of molecular dynamics-based calculation of water structure.

The results of the present thesis were disseminated in three research articles and one web page freely available for the scientific community.

## Köszönetnyilvánítás

A dolgozat alapját képező kutatómunka anyagi támogatásáért köszönetet mondok a Nemzeti Kutatási, Fejlesztési és Innovációs Hivatalnak (NKFIH K 123836). A számítógépidő oroszlánrészét a Kormányzati Informatikai Fejlesztési Ügynökség (KIFÜ) szuperszámítógépes rendszere biztosította.

Elsősorban köszönettel tartozom témavezetőmnek, **Dr. Habil. Hetényi Csaba** egyetemi docensnek. Köszönöm végtelen türelmét, nélküle ez a dolgozat nem jöhetett volna létre.

Köszönöm közvetlen szerzőtársaimnak, **Dr. Bálint Mónikának, Dr. Jeszenői Norbertnek** és **Dr. Paragi Gábornak**, hogy szakmailag és emberileg is támogattak, kérdések esetén azonnal rendelkezésemre álltak és mindenben segítették munkámat.

Köszönöm oktatóimnak, **Prof. Dr. Penke Botondnak, Dr. Bogár Ferencnek,** és **Prof. Dr. Galbács Gábornak**, hogy a szakterülethez közvetlenül és nem közvetlenül szükséges tudást, iránymutatást megadták, valamint **Dr. Bartha Ferencnek**, hogy a KIFÜ oldaláról szoftveres háttértámogatást biztosított.

Hálás vagyok **Dr. Habil. Rajkó Róbertnek és Dr. Habil. Czakó Gábornak**, az MTA Szegedi Akadémiai Bizottság Kemometriai és Molekulamodellezési Munkabizottsága társelnökeinek, hogy lehetőséget adtak a jelen disszertáció anyagának a bizottság előtti konstruktív megvitatására.

Köszönöm továbbá a doktori szigorlati bizottság tagjainak, **Prof. Dr. Janáky Tamásnak** és **Dr. Ötvös Ferencnek**, hogy a szakterületeiken felvetett kérdéseikkel ismereteim további elemzésére inspiráltak.

Végül, de nem utolsó sorban köszönöm az egész **családomnak, szüleimnek**, valamint **barátaimnak**, hogy mellettem álltak, támogattak és hittek bennem. Nélkületek nem sikerült volna.

## Irodalomjegyzék

1. Desjarlais, R. and P.J. Tummino, *The role of histone-modifying enzymes and their complexes in regulation of chromatin biology*. Biochem., 2016. **55**: p. 1584–1599.
2. Portela, A. and M. Esteller, *Epigenetic modifications and human disease*. Nat. Biotechnol., 2010. **28**: p. 1057–1068.
3. Rodenhiser, D. and M. Mann, *Epigenetics and human disease: Translating basic biology into clinical applications*. Can. Med. Assoc. J., 2006. **174**: p. 341–348.
4. Taverna, S.D., et al., *How chromatin-binding modules interpret histone modifications: Lessons from professional pocket pickers*. Nat. Struct. Mol. Biol., 2007. **14**: p. 1025–1040.
5. Musselman, C.A., et al., *Perceiving the epigenetic landscape through histone Catherine*. Nat. Struct. Mol. Biol., 2012. **19**: p. 1218–1227.
6. Strahl, B.D. and C.D. Allis, *The language of covalent histone modifications*. Nature, 2000. **403**: p. 41–45.
7. Jenuwein, T. and C.D. Allis, *Translating the histone code*. Science, 2001. **293**(1074–1081).
8. Davis, A.M., S.J. Teague, and G.J. Kleywegt, *Application and limitations of X-ray crystallographic data in structure-based ligand and drug design*. Angew. Chem., 2003. **42**: p. 2718–2736.
9. Blundell, T.L., H. Jhoti, and C. Abell, *High-Throughput crystallography for lead discovery in drug design*. Nat. Rev. Drug Discov., 2002. **1**: p. 45–54.
10. Leach, A.R., *Molecular Modelling: Principles and applications*. 2001.
11. Andrews, D.H., *The Relation Between the Raman Spectra and the Structure of Organic Molecules*. Physical Review, 1930. **36**(3): p. 544–554.
12. Case, D., et al., *AmberTools, 15*. 2015: Amber.
13. Brooks, B.R., et al., *CHARMM: A program for macromolecular energy, minimization, and dynamics calculations*. J. Comput. Chem., 1983. **4**(2): p. 187–217.
14. Jorgensen, W.L., D.S. Maxwell, and J. Tirado-Rives, *Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids*. J. Am. Chem. Soc., 1996. **118**: p. 11225–11236.
15. Daura, X., A.E. Mark, and W.F. Van Gunsteren, *Parametrization of aliphatic CH<sub>n</sub> united atoms of GROMOS96 force field*. J. Comp. Chem., 1998. **19**(5): p. 535–547.
16. Poór, M., et al., *Interaction of citrinin with human serum albumin*. Toxins, 2015. **7**(12): p. 5155–5166.
17. Allingham, J., R. Smith, and I. Rayment, *The structural basis of blebbistatin inhibition and specificity for myosin II*. Nat. Struct. Mol. Biol., 2005. **12**: p. 378–9.
18. Shan, Y., et al., *How does a drug molecule find its target binding site?* Am. Chem. Soc., 2011. **133**(24): p. 9181–9183.
19. Prakash, P., J. F Hancock, and A. Gorfe, *Binding hotspots on K-ras: Consensus ligand binding sites and other reactive regions from probe-based molecular dynamics analysis: Consensus Ligand Binding Sites on Ras*. Proteins, 2015. **83**.
20. Kukol, A., *Methods in Molecular Biology*, vol. 443, *Molecular Modeling of Proteins: Preface*. 2008. **443**: p. v–vi.
21. Hetényi, C. and D. van der Spoel, *Efficient docking of peptides to proteins without prior knowledge of the binding site*. Protein Sci., 2002. **11**: p. 1729–1737.
22. Hetényi, C. and D. Van Der Spoel, *Toward prediction of functional protein pockets using blind docking and pocket search algorithms*. Protein Sci., 2011. **20**: p. 880–893.

23. Trott, O. and A.J. Olson, *AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading*. J. Comput. Chem., 2010. **31**(2): p. 455-61.
24. Morris, G.M., et al., *AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility*. J. Comp. Chem., 2009. **28**: p. 73–86.
25. Jeszenői, N., et al., *Exploration of interfacial hydration networks of target–ligand complexes*. J. Chem. Inf. Model., 2016. **56**: p. 148–158.
26. Serrano-Cinca, C., Y. Fuertes Callén, and C. Molinero, *Measuring DEA efficiency in Internet companies*. Decis. Support Syst., 2005. **38**: p. 557-573.
27. Fischer, M., et al., *Incorporation of protein flexibility and conformational energy penalties in docking screens to improve ligand discovery*. Nat Chem, 2014. **6**(7): p. 575-83.
28. Bálint, M., et al., *Systematic exploration of multiple drug binding sites*. J. Cheminform., 2017. **9**: p. 65.
29. Bellissent-Funel, M.C., et al., *Water Determines the Structure and Dynamics of Proteins*. Chem Rev, 2016. **116**(13): p. 7673-97.
30. Rastelli, G., et al., *Discovery of new inhibitors of aldose reductase from molecular docking and database screening*. Bioorg. Med. Chem., 2002. **10**(5): p. 1437-50.
31. Hou, X., et al., *Protein Flexibility in Docking-Based Virtual Screening: Discovery of Novel Lymphoid-Specific Tyrosine Phosphatase Inhibitors Using Multiple Crystal Structures*. J. Chem. Inf. Model., 2015. **55**(9): p. 1973-83.
32. Pan, A.C., et al., *Molecular determinants of drug-receptor binding kinetics*. Drug. Discov. Today., 2013. **18**(13-14): p. 667-73.
33. Halperin, I., et al., *Principles of docking: An overview of search algorithms and a guide to scoring functions*. Proteins., 2002. **47**(4): p. 409-43.
34. Brooijmans, N. and I.D. Kuntz, *Molecular recognition and docking algorithms*. Annu. Rev. Biophys. Biomol. Struct., 2003. **32**: p. 335-73.
35. Hetényi, C. and D. van Der Spoel, *Blind docking of drug-sized compounds to proteins with up to a thousand residues*. FEBS Lett., 2006. **580**: p. 1447–1450.
36. Grinter, S.Z. and X. Zou, *Challenges, applications, and recent advances of protein-ligand docking in structure-based drug design*. Molecules, 2014. **19**(7): p. 10150-76.
37. Yuriev, E., M. Agostino, and P.A. Ramsland, *Challenges and advances in computational docking: 2009 in review*. J Mol Recognit, 2011. **24**(2): p. 149-64.
38. Yuriev, E., J. Holien, and P.A. Ramsland, *Improvements, trends, and new ideas in molecular docking: 2012-2013 in review*. J Mol Recognit, 2015. **28**(10): p. 581-604.
39. Hocker, H.J., N. Rambahal, and A.A. Gorfe, *LIBSA--a method for the determination of ligand-binding preference to allosteric sites on receptor ensembles*. J. Chem. Inf. Model., 2014. **54**(2): p. 530-8.
40. Whalen, K.L., et al., *Nature of allosteric inhibition in glutamate racemase: discovery and characterization of a cryptic inhibitory pocket using atomistic MD simulations and pKa calculations*. J Phys Chem B, 2011. **115**(13): p. 3416-24.
41. Schindler, C.E.M., S.J. de Vries, and M. Zacharias, *Fully Blind Peptide-Protein Docking with pepATTRACT*. Structure, 2015. **23**(8): p. 1507-1515.
42. Bugatti, A., et al., *Molecular interaction studies of HIV-1 matrix protein p17 and heparin: identification of the heparin-binding motif of p17 as a target for the development of multitarget antagonists*. J. Biol. Chem., 2013. **288**(2): p. 1150-1161.
43. Garcia-Sosa, A.T., S. Sild, and U. Maran, *Design of multi-binding-site inhibitors, ligand efficiency, and consensus screening of avian influenza H5N1 wild-type neuraminidase and of the oseltamivir-resistant H274Y variant*. J. Chem. Inf. Model., 2008. **48**(10): p. 2074-80.

44. Todorova-Balvay, D., et al., *An efficient two step purification and molecular characterization of beta-galactosidases from Aspergillus oryzae*. J. Mol. Recognit., 2006. **19**(4): p. 299-304.
45. Kovacs, M., et al., *Mechanism of blebbistatin inhibition of myosin II*. J. Biol. Chem., 2004. **279**(34): p. 35557-63.
46. Agarwal, T., et al., *Molecular docking and dynamic simulation evaluation of Rohinitib - Cantharidin based novel HSF1 inhibitors for cancer therapy*. Mol. Graph. Mod., 2015. **61**: p. 141-149.
47. Gutierrez, L.J., R.D. Enriz, and H.A. Baldoni, *Structural and thermodynamic characteristics of the exosite binding pocket on the human BACE1: a molecular modeling approach*. J. Phys. Chem. A., 2010. **114**(37): p. 10261-9.
48. Pagadala, N.S., K. Syed, and J. Tuszynski, *Software for molecular docking: a review*. Biophys. Rev., 2017. **9**(2): p. 91-102.
49. Ghersi, D. and R. Sanchez, *Improving accuracy and efficiency of blind protein-ligand docking by focusing on predicted binding sites*. Proteins, 2009. **74**(2): p. 417-24.
50. Gianti, E., et al., *Computational analysis of EBNA1 "druggability" suggests novel insights for Epstein-Barr virus inhibitor design*. Journal of computer-aided molecular design, 2016. **30**(4): p. 285-303.
51. Alonso, H., A.A. Bliznyuk, and J.E. Gready, *Combining docking and molecular dynamic simulations in drug design*. Med. Res. Rev., 2006. **26**(5): p. 531-68.
52. Durrant, J.D. and J.A. McCammon, *Molecular dynamics simulations and drug discovery*. BMC. Biol., 2011. **9**: p. 71.
53. Ganesan, A., M.L. Coote, and K. Barakat, *Molecular dynamics-driven drug discovery: leaping forward with confidence*. Drug. Discov. Today., 2017. **22**(2): p. 249-269.
54. Dror, R.O., et al., *Biomolecular simulation: a computational microscope for molecular biology*. Annu. Rev. Biophys., 2012. **41**: p. 429-52.
55. Shan, Y., et al., *Molecular basis for pseudokinase-dependent autoinhibition of JAK2 tyrosine kinase*. Nat Struct Mol Biol, 2014. **21**(7): p. 579-84.
56. Buch, I., T. Giorgino, and G. De Fabritiis, *Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations*. Proc Natl Acad Sci U S A, 2011. **108**(25): p. 10184-9.
57. Limongelli, V., M. Bonomi, and M. Parrinello, *Funnel metadynamics as accurate binding free-energy method*. Proc. Natl. Acad. Sci. USA, 2013. **110**(16): p. 6358-63.
58. Kuzmanic, A., et al., *Changes in the free-energy landscape of p38alpha MAP kinase through its canonical activation and binding events as studied by enhanced molecular dynamics simulations*. Elife, 2017. **6**.
59. Niu, Y., et al., *Revealing inhibition difference between PFI-2 enantiomers against SETD7 by molecular dynamics simulations, binding free energy calculations and unbinding pathway analysis*. Sci. Rep., 2017. **7**: p. 46547.
60. Pabon, N.A. and C.J. Camacho, *Probing protein flexibility reveals a mechanism for selective promiscuity*. Elife., 2017. **6**.
61. Celik, L., et al., *Exploring interactions of endocrine-disrupting compounds with different conformations of the human estrogen receptor alpha ligand binding domain: a molecular docking study*. Chem. Res. Toxicol., 2008. **21**(11): p. 2195-206.
62. Amaral, M., et al., *Protein conformational flexibility modulates kinetics and thermodynamics of drug binding*. Nat. Commun., 2017. **8**(1): p. 2276.
63. Borhani, D.W. and D.E. Shaw, *The future of molecular dynamics simulations in drug discovery*. J Comput Aided Mol Des, 2012. **26**(1): p. 15-26.
64. Casasnovas, R., et al., *Unbinding Kinetics of a p38 MAP Kinase Type II Inhibitor from Metadynamics Simulations*. J. Am. Chem. Soc., 2017. **139**(13): p. 4780-4788.

65. Tiwary, P., et al., *Kinetics of protein-ligand unbinding: Predicting pathways, rates, and rate-limiting steps*. Proc. Natl. Acad. Sci. USA, 2015. **112**(5): p. E386-91.
66. Tiwary, P., et al., *Role of water and steric constraints in the kinetics of cavity-ligand unbinding*. Proc. Natl. Acad. Sci. USA, 2015. **112**(39): p. 12015-9.
67. De Vivo, M. and A. Cavalli, *Recent advances in dynamic docking for drug discovery: Dynamic docking for drug discovery*. Wiley Interdiscip. Rev. Comp. Mol. Sci., 2017. **7**: p. e1320.
68. Balint, M., et al., *Dynamic changes in binding interaction networks of sex steroids establish their non-classical effects*. Sci Rep, 2017. **7**(1): p. 14847.
69. Lotz, S.D. and A. Dickson, *Unbiased Molecular Dynamics of 11 min Timescale Drug Unbinding Reveals Transition State Stabilizing Interactions*. J. Am. Chem. Soc., 2018. **140**(2): p. 618-628.
70. Poole, P.L. and J.L. Finney, *Hydration-induced conformational and flexibility changes in lysozyme at low water content*. Int. J. Biol. Macromol., 1983. **5**: p. 308-310.
71. Kunstmann, S., et al., *Solvent networks tune thermodynamics of oligosaccharide complex formation in an extended protein binding site*. J. Am. Chem. Soc., 2018. **140**: p. 10447-10455.
72. Brysbaert, G., R. Blossey, and M.F. Lensink, *The inclusion of water molecules in residue interaction networks identifies additional central residues*. Front Mol. Biosci., 2018. **5**.
73. Choi, H., H. Kang, and H. Park, *New solvation free energy function comprising intermolecular solvation and intramolecular self-solvation terms*. J. Cheminform., 2013. **5**: p. 5-8.
74. Cui, G., J.M. JasonSwails, and E.S. Manas, *SPAM: A simple approach for profiling bound water molecules*. J. Chem. Theory Comput., 2013. **9**: p. 5539-5549.
75. Garcia-Sosa, A.T., S. Firth-Clark, and R.L. Mancera, *Including tightly-bound water molecules in de novo drug design. Exemplification through the in silico generation of poly(ADP-ribose)polymerase ligands*. J. Chem. Inf. Model., 2005. **45**: p. 624-633.
76. Huggins, D.J. and B. Tidor, *Systematic placement of structural water molecules for improved scoring of protein-ligand interactions*. Protein Eng. Des. Sel., 2011. **24**: p. 777-789.
77. Ladbury, J.E., *Just add water! The effect of water on the specificity of proteinligand binding sites And its potential application to drug design*. Cell Chem. Biol., 1996. **3**: p. 973-980.
78. Lloyd, D.G., et al., *The effect of tightly bound water molecules on the structural interpretation of ligand-derived pharmacophore models*. J. Comp. Aided Mol. Des., 2004. **18**: p. 89-100.
79. Michel, J., J. Tirado-Rives, and W.L. Jorgensen, *Energetics of displacing water molecules from protein binding sites: consequences for ligand optimization*. J. Am. Chem. Soc., 2009. **131**: p. 15403-15411.
80. Zheng, Z., *Computational modeling of solvent effects on protein-ligand interactions Using fully polarizable continuum model and rational drug design*. Commun. Comput. Phys., 2011. **13**: p. 31-60.
81. Quioco, F., D. Wilson, and N. Vyas, *Substrate specificity and affinity of a protein modulated by bound water molecules*. Nature, 1989. **340**: p. 404-407.
82. Afonine, P.V., et al., *Bulk-solvent and overall scaling revisited: faster calculations, improved results*. Acta Cryst., 2013. **69**: p. 625-634.
83. Badger, J., *Modeling and refinement of water molecules and disordered solvent*. Meth. Enzymol., 1997. **277**: p. 344-352.
84. Halle, B., *Protein hydration dynamics in solution: a critical survey*. Phil. Trans. R. Soc. Lond. B, 2004. **359**: p. 1207-1224.
85. Elsässer, S.J., et al., *DAXX envelops a histone H3.3-H4 dimer for H3.3-specific recognition*. Nature, 2012. **491**(7425): p. 560-565.
86. Pearlstein, R., et al., *New hypotheses about the structure-function of proprotein convertase subtilisin/kexin type 9: Analysis of the epidermal growth factor-like repeat A docking site using WaterMap*. Proteins, 2010. **78**(12): p. 2571-2586.

87. Vukovic, S., P.E. Brennan, and D. Huggins, *Exploring the role of water in molecular recognition: predicting protein ligandability using a combinatorial search of surface hydration sites*. J. Phys. Condens. Matter., 2016 **28**(34).
88. Hylsová, M., et al., *Explicit treatment of active-site waters enhances quantum mechanical/implicit solvent scoring: Inhibition of CDK2 by new pyrazolo[1,5-a]pyrimidines*. Eur. J. Med. Chem., 2016. **126**: p. 1118-1128.
89. García-Sosa, A.T. and R.L. Mancera, *Free Energy Calculations of Mutations Involving a Tightly Bound Water Molecule and Ligand Substitutions in a Ligand-Protein Complex*. Mol. Inf., 2010. **29**(8-9): p. 589-600.
90. Lee, S.H. and P.J. Rossky, *A comparison of the structure and dynamics of liquid water at hydrophobic and hydrophilic surfaces—a molecular dynamics simulation study*. J. Chem. Phys., 1994. **100**(4): p. 3334-3345.
91. Watanabe, G., et al., *Analysis of water channels by molecular dynamics simulation of heterotetrameric sarcosine oxidase*. Biophys. Physicobiol., 2015. **12**: p. 131-137.
92. Patodia, S., A. Bagaria, and D. Chopra, *Molecular Dynamics Simulation of Proteins: A Brief Overview*. J. Phys. Chem. Bioph., 2014. **4**(6).
93. Abel, R., et al., *Role of the active-site solvent in the thermodynamics of factor Xa ligand binding*. J. Am. Chem. Soc., 2008. **130**: p. 2817–2831.
94. Skyner, R.E., et al., *A review of methods for the calculation of solution free energies and the modelling of systems in solution*. Phys. Chem. Chem. Phys., 2015. **17**(9): p. 6174-6191.
95. Cramer, C.J. and D.G. Truhlar, *Implicit solvation models: equilibria, structure, spectra, and dynamics*. Chemical Reviews, 1999. **99**(8): p. 2161-2200.
96. Persico, M. and G. Granucci, *Continuum solvation models in chemical physics: from theory to application*. 2007.
97. Cramer, C.J., *Essentials of computational chemistry, theories and models*. J. Chem. Inf. Comput. Sci., 2003. **43**(5): p. 1720-1720.
98. Jensen, F., *Introduction to computational chemistry*. 2007: John Wiley and Sons.
99. Palmer, D.S., et al., *First-principles calculation of the intrinsic aqueous solubility of crystalline druglike molecules*. J. Chem. Theory Comput., 2012. **8**(9): p. 3322-3337.
100. Klamt, A. and G. Schuurmann, *COSMO: A new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient*. J. Chem. Soc., Perkin Trans. 2, 1993. **2**: p. 799-805.
101. Klamt, A. and V. Jonas, *Treatment of the outlying charge in continuum solvation models*. J. Chem. Phys., 1996. **105**(22): p. 9972-9981.
102. Klamt, A., C. Moya, and J. Palomar, *A comprehensive comparison of the IEFPCM and SS(V)PE continuum solvation methods with the COSMO approach*. J. Chem. Theory Comput., 2015. **11**(9): p. 4220-4225.
103. Price, D.J. and C.L. Brooks, *A modified TIP3P water potential for simulation with Ewald summation*. J. Chem. Phys., 2004. **121**(20): p. 10096-10103.
104. Ponder, J.W., et al., *Current status of the AMOEBA polarizable force field*. J. Phys. Chem. B., 2010. **114**(8): p. 2549-2564.
105. Tomasi, J., B. Mennucci, and R. Cammi, *Quantum mechanical continuum solvation models*. Chem. Rev., 2005. **105**(8): p. 2999-3094.
106. Kamerlin, S.C.L., M. Haranczyk, and A. Warshel, *Are mixed explicit/implicit solvation models reliable for studying phosphate hydrolysis? A comparative study of continuum, explicit and mixed solvation models*. ChemPhysChem, 2009. **10**(7): p. 1125-1134.
107. Nisius, L. and S. Grzesiek, *Key stabilizing elements of protein structure identified through pressure and temperature perturbation of its hydrogen bond network*. Nat Chem, 2012. **4**(9): p. 711-7.

108. Zhao, L., W. Li, and P. Tian, *Reconciling mediating and slaving roles of water in protein conformational dynamics*. PLoS One, 2013. **8**(4): p. e60553.
109. Cheung, M.S., A.E. Garcia, and J.N. Onuchic, *Protein folding mediated by solvation: water expulsion and formation of the hydrophobic core occur after the structural collapse*. Proc Natl Acad Sci U S A, 2002. **99**(2): p. 685-90.
110. Levy, Y. and J.N. Onuchic, *Water mediation in protein folding and molecular recognition*. Annu Rev Biophys Biomol Struct, 2006. **35**: p. 389-415.
111. Frauenfelder, H., et al., *Protein folding is slaved to solvent motions*. Proc Natl Acad Sci U S A, 2006. **103**(42): p. 15469-72.
112. Thirumalai, D., G. Reddy, and J.E. Straub, *Role of water in protein aggregation and amyloid polymorphism*. Acc Chem Res, 2012. **45**(1): p. 83-92.
113. Chong, S.H. and S. Ham, *Distinct role of hydration water in protein misfolding and aggregation revealed by fluctuating thermodynamics analysis*. Acc Chem Res, 2015. **48**(4): p. 956-65.
114. Grisanti, L., et al., *A computational study on how structure influences the optical properties in model crystal structures of amyloid fibrils*. Physical Chemistry Chemical Physics, 2017. **19**(5): p. 4030-4040.
115. Jong, K., L. Grisanti, and A. Hassanali, *Hydrogen Bond Networks and Hydrophobic Effects in the Amyloid beta30-35 Chain in Water: A Molecular Dynamics Study*. J Chem Inf Model, 2017. **57**(7): p. 1548-1562.
116. Li, Y., et al., *Modeling of the water network at protein-RNA interfaces*. J Chem Inf Model, 2011. **51**(6): p. 1347-52.
117. Kao, Y.T., et al., *Ultrafast dynamics of nonequilibrium electron transfer in photoinduced redox cycle: solvent mediation and conformation flexibility*. J Phys Chem B, 2012. **116**(30): p. 9130-40.
118. Sappati, S., et al., *Nuclear quantum effects in a HIV/cancer inhibitor: The case of ellipticine*. J Chem Phys, 2016. **145**(20): p. 205102.
119. Law, Y.K. and A.A. Hassanali, *Role of Quantum Vibrations on the Structural, Electronic, and Optical Properties of 9-Methylguanine*. J Phys Chem A, 2015. **119**(44): p. 10816-27.
120. Giberti, F. and A.A. Hassanali, *The excess proton at the air-water interface: The role of instantaneous liquid interfaces*. J Chem Phys, 2017. **146**(24): p. 244703.
121. Daly, C.A., Jr., et al., *Decomposition of the Experimental Raman and Infrared Spectra of Acidic Water into Proton, Special Pair, and Counterion Contributions*. J Phys Chem Lett, 2017. **8**(21): p. 5246-5252.
122. Chen, M., et al., *Hydroxide diffuses slower than hydronium in water because its solvated structure inhibits correlated proton transfer*. Nat Chem, 2018. **10**(4): p. 413-419.
123. Cuny, J. and A.A. Hassanali, *Ab initio molecular dynamics study of the mechanism of proton recombination with a weak base*. J Phys Chem B, 2014. **118**(48): p. 13903-12.
124. Zundel, G., *Hydrogen Bonds with Large Proton Polarizability and Proton Transfer Processes in Electrochemistry and Biology*. Advances in Chemical Physics, 1999: p. 1-217.
125. Ishikita, H. and K. Saito, *Proton transfer reactions and hydrogen-bond networks in protein environments*. J R Soc Interface, 2014. **11**(91): p. 20130518.
126. Williams, M.A., J.M. Goodfellow, and J.M. Thornton, *Buried waters and internal cavities in monomeric proteins*. Protein science : a publication of the Protein Society, 1994. **3**(8): p. 1224-1235.
127. Carugo, O., *Structure and function of water molecules buried in the protein core*. Curr Protein Pept Sci, 2015. **16**(3): p. 259-65.
128. Garcia-Sosa, A.T., S. Firth-Clark, and R.L. Mancera, *Including tightly-bound water molecules in de novo drug design. Exemplification through the in silico generation of poly(ADP-ribose)polymerase ligands*. J Chem Inf Model, 2005. **45**(3): p. 624-33.



129. Garcia-Sosa, A.T. and R.L. Mancera, *The effect of a tightly bound water molecule on scaffold diversity in the computer-aided de novo ligand design of CDK2 inhibitors*. J Mol Model, 2006. **12**(4): p. 422-31.
130. Garcia-Sosa, A.T., R.L. Mancera, and P.M. Dean, *WaterScore: a novel method for distinguishing between bound and displaceable water molecules in the crystal structure of the binding site of protein-ligand complexes*. J Mol Model, 2003. **9**(3): p. 172-82.
131. Mancera, R.L., *Molecular modeling of hydration in drug design*. Curr Opin Drug Discov Devel, 2007. **10**(3): p. 275-80.
132. Roberts, B.C. and R.L. Mancera, *Ligand-protein docking with water molecules*. J Chem Inf Model, 2008. **48**(2): p. 397-408.
133. Wong, S.E. and F.C. Lightstone, *Accounting for water molecules in drug design*. Expert Opin Drug Discov, 2011. **6**(1): p. 65-74.
134. De Lucca, G.V., S. Erickson-Viitanen, and P.Y.S. Lam, *Cyclic HIV protease inhibitors capable of displacing the active site structural water molecule*. Drug Discovery Today, 1997. **2**(1): p. 6-18.
135. de Beer, S.B.A., N.P.E. Vermeulen, and C. Oostenbrink, *The Role of Water Molecules in Computational Drug Design*. Current Topics in Medicinal Chemistry, 2010. **10**(1): p. 55-66.
136. Garcia-Sosa, A.T., *Hydration properties of ligands and drugs in protein binding sites: tightly-bound, bridging water molecules and their effects and consequences on molecular design strategies*. J Chem Inf Model, 2013. **53**(6): p. 1388-405.
137. Cavasotto, C.N. and A.T. García-Sosa, *Role of Water Molecules and Hydration proPerties in Modeling LigandeProtein Interaction and Drug Design, in Silico Drug Discovery and Design: Theory, Methods, Challenges, and Applications*. 2015: CRC Pres.
138. Pitt, W.R., J. Murray-Rust, and J.M. Goodfellow, *AQUARIUS2: Knowledge-based modeling of solvent sites around proteins*. Journal of Computational Chemistry, 1993. **14**(9): p. 1007-1018.
139. Pitt, W.R. and J.M. Goodfellow, *Modelling of solvent positions around polar groups in proteins*. Protein Eng, 1991. **4**(5): p. 531-7.
140. Vedani, A. and D.W. Huhta, *Algorithm for the systematic solvation of proteins based on the directionality of hydrogen bonds*. Journal of the American Chemical Society, 1991. **113**(15): p. 5860-5862.
141. Rossato, G., et al., *AcquaAlta: a directional approach to the solvation of ligand-protein complexes*. J Chem Inf Model, 2011. **51**(8): p. 1867-81.
142. Ross, G.A., G.M. Morris, and P.C. Biggin, *Rapid and accurate prediction and scoring of water molecules in protein binding sites*. PLoS One, 2012. **7**(3): p. e32036.
143. Mason, J.S., et al., *High end GPCR design: crafted ligand design and druggability analysis using protein structure, lipophilic hotspots and explicit water networks*. In Silico Pharmacology, 2013. **1**: p. 23.
144. Pettitt, M.B. and M. Karplus, *The structure of water surrounding a peptide: a theoretical approach*. Chemical Physics Letters, 1987. **136**(5): p. 383-386.
145. Rossky, P.J. and M. Karplus, *Solvation. A molecular dynamics study of a dipeptide in water*. Journal of the American Chemical Society, 1979. **101**(8): p. 1913-1937.
146. van Gunsteren, W.F., et al., *Computer simulation of the dynamics of hydrated protein crystals and its comparison with x-ray data*. Proceedings of the National Academy of Sciences of the United States of America, 1983. **80**(14): p. 4315-4319.
147. Dror, R.O., et al., *Pathway and mechanism of drug binding to G-protein-coupled receptors*. Proc Natl Acad Sci U S A, 2011. **108**(32): p. 13118-23.
148. Dror, R.O., et al., *Structural basis for modulation of a G-protein-coupled receptor by allosteric drugs*. Nature, 2013. **503**(7475): p. 295-9.

149. Abraham, M.J., et al., *GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers*. SoftwareX, 2015. **1-2**: p. 19-25.
150. Salomon-Ferrer, R., et al., *Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald*. J Chem Theory Comput, 2013. **9**(9): p. 3878-88.
151. Gotz, A.W., et al., *Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born*. J Chem Theory Comput, 2012. **8**(5): p. 1542-1555.
152. Pan, A.C., et al., *Demonstrating an Order-of-Magnitude Sampling Enhancement in Molecular Dynamics Simulations of Complex Protein Systems*. J Chem Theory Comput, 2016. **12**(3): p. 1360-7.
153. Shaw, D.E., et al., *Anton, a special-purpose machine for molecular dynamics simulation*. SIGARCH Comput. Archit. News, 2007. **35**(2): p. 1-12.
154. Virtanen, J.J., et al., *Modeling the hydration layer around proteins: HyPred*. Biophys J, 2010. **99**(5): p. 1611-9.
155. Lounnas, V., B.M. Pettitt, and G.N. Phillips, Jr., *A global model of the protein-solvent interface*. Biophys J, 1994. **66**(3 Pt 1): p. 601-14.
156. Makarov, V., B.M. Pettitt, and M. Feig, *Solvation and hydration of proteins and nucleic acids: a theoretical view of simulation and experiment*. Acc Chem Res, 2002. **35**(6): p. 376-84.
157. Makarov, V.A., B.K. Andrews, and B.M. Pettitt, *Reconstructing the protein-water interface*. Biopolymers, 1998. **45**(7): p. 469-78.
158. Makarov, V.A., et al., *Residence times of water molecules in the hydration sites of myoglobin*. Biophysical journal, 2000. **79**(6): p. 2966-2974.
159. Pettitt, B.M., V.A. Makarov, and B.K. Andrews, *Protein hydration density: theory, simulations and crystallography*. Curr Opin Struct Biol, 1998. **8**(2): p. 218-21.
160. Schoenborn, B.P., A. Garcia, and R. Knott, *Hydration in protein crystallography*. Prog Biophys Mol Biol, 1995. **64**(2-3): p. 105-19.
161. Lensink, M.F., et al., *Blind prediction of interfacial water positions in CAPRI*. Proteins: Structure, Function, and Bioinformatics, 2014. **82**(4): p. 620-632.
162. Henchman, R.H. and J.A. McCammon, *Structural and dynamic properties of water around acetylcholinesterase*. Protein Sci, 2002. **11**(9): p. 2080-90.
163. Huang, H.C., et al., *Cluster analysis of hydration waters around the active sites of bacterial alanine racemase using a 2-ns MD simulation*. Biopolymers, 2008. **89**(3): p. 210-9.
164. Madhusudhan, M.S. and S. Vishveshwara, *Deducing hydration sites of a protein from molecular dynamics simulations*. J Biomol Struct Dyn, 2001. **19**(1): p. 105-14.
165. Lazaridis, T., *Inhomogeneous Fluid Approach to Solvation Thermodynamics. 1. Theory*. The Journal of Physical Chemistry B, 1998. **102**(18): p. 3531-3541.
166. Lazaridis, T., *Inhomogeneous Fluid Approach to Solvation Thermodynamics. 2. Applications to Simple Fluids*. The Journal of Physical Chemistry B, 1998. **102**(18): p. 3542-3550.
167. Jeszenői, N., et al., *Mobility-based prediction of hydration structures of protein surfaces*. Bioinf., 2015. **31**: p. 1959-1965.
168. Vega, C., et al., *What ice can teach us about water interactions: a critical comparison of the performance of different water models*. Faraday Discussions, 2009. **141**(0): p. 251-276.
169. Nayar, D. and C. Chakravarty, *Sensitivity of local hydration behaviour and conformational preferences of peptides to choice of water model*. Phys Chem Chem Phys, 2014. **16**(21): p. 10199-213.
170. Nerenberg, P.S. and T. Head-Gordon, *Optimizing Protein-Solvent Force Fields to Reproduce Intrinsic Conformational Preferences of Model Peptides*. J Chem Theory Comput, 2011. **7**(4): p. 1220-30.

171. Florova, P., et al., *Explicit Water Models Affect the Specific Solvation and Dynamics of Unfolded Peptides While the Conformational Behavior and Flexibility of Folded Peptides Remain Intact*. J Chem Theory Comput, 2010. **6**(11): p. 3569-79.
172. Matthes, D. and B.L. de Groot, *Secondary structure propensities in peptide folding simulations: a systematic comparison of molecular mechanics interaction schemes*. Biophys J, 2009. **97**(2): p. 599-608.
173. Best, R.B. and J. Mittal, *Protein Simulations with an Optimized Water Model: Cooperative Helix Formation and Temperature-Induced Unfolded State Collapse*. The Journal of Physical Chemistry B, 2010. **114**(46): p. 14916-14923.
174. Fadda, E. and R.J. Woods, *On the Role of Water Models in Quantifying the Binding Free Energy of Highly Conserved Water Molecules in Proteins: The Case of Concanavalin A*. J Chem Theory Comput, 2011. **7**(10): p. 3391-8.
175. Rahaman, O., et al., *Configurational Disorder of Water Hydrogen-Bond Network at the Protein Dynamical Transition*. J Phys Chem B, 2017. **121**(28): p. 6792-6798.
176. Field, M.J., P.A. Bash, and M. Karplus, *A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations*. J. Comp. Chem., 1990. **11**: p. 700-733.
177. Hayik, S.A., R. Dunbrack, and K.M. Merz, *A mixed QM/MM scoring function to predict protein-ligand binding affinity*. J. Chem. Theory Comp., 2010. **6**: p. 3079–3091.
178. Menikarachchi, L.C. and J.A. Gascón, *QM/MM approaches in medicinal chemistry research*. Curr. Topics in Med. Chem., 2010. **10**: p. 46-54.
179. van der Kamp, M.W. and A.J. Mulholland, *Combined quantum mechanics/molecular mechanics (QM/MM) methods in computational enzymology*. Biochem., 2013. **52**: p. 2708–2728.
180. Warshel, A. and M. Levitt, *Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme*. J. Mol. Biol., 1976. **103**: p. 227-249.
181. Vidossich, P. and A. Magistrato, *QM/MM molecular dynamics studies of metal binding proteins*. Biomolecules, 2014. **4**: p. 616-645.
182. Zhang, D.W. and J.Z.H. Zhang, *Molecular fractionation with conjugate caps for full quantum mechanical calculation of protein–molecule interaction energy*. J. Chem. Phys., 2003. **119**: p. 3599-3605.
183. Nikitina, E., et al., *Semiempirical calculations of binding enthalpy for protein–ligand complexes*. Int. J. Quantum Chem., 2003. **97**: p. 747-763.
184. Nikitina, E., et al., *Mixed implicit/explicit Solvation models in quantum mechanical calculations of binding enthalpy for protein–ligand complexes*. Int. J. Quantum Chem., 2004. **106**: p. 1943-1963.
185. Dobeš, P., et al., *Interaction energies for the purine inhibitor roscovitine with cyclin-dependent kinase 2: correlated ab Initio quantum-chemical, DFT and empirical calculations*. Chemistry 2006. **12**: p. 4297-4304.
186. Freire, E., *Do enthalpy and entropy distinguish first in class from best in class?* Drug Discov. Today, 2008. **13**: p. 869-874.
187. Ferenczy, G.G. and G.M. Keserü, *Thermodynamics guided lead discovery and optimization*. Drug Discov. Today, 2012. **15**: p. 919-932.
188. Hann, M.M. and G.M. Keserü, *Finding the sweet spot: the role of nature and nurture in medicinal chemistry*. Nat. Rev. Drug. Discov., 2012. **11**: p. 355-365.
189. Berman, H.M., et al., *The Protein Data Bank*. Nucleic Acids Res., 2000. **28**: p. 235–242.
190. Stewart, J.J.P., *Openmopac online manual* - <http://www.openmopac.net>. 2016.
191. Stewart, J.J.P., *Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters*. J. Mol. Model., 2013. **19**: p. 1-32.

192. Stewart, J.J.P., *Application of localized molecular orbitals to the solution of semiempirical self-consistent field equations*. Int. Journ. Quant. Chem., 1996. **58**(133-146).
193. Baker, J., *An algorithm for the location of transition states*. J. Comp. Chem., 1986. **7**: p. 385.
194. Mehler, E.L. and T. Solmajer, *Electrostatic effects in proteins: Comparison of dielectric and charge models*. Prot. Eng., 1991. **4**: p. 903–910.
195. Lindorff-Larsen, K., et al., *Improved side-chain torsion potentials for the Amber ff99SB protein force field*. Proteins Struct. Funct. Bioinform., 2010. **78**: p. 1950–1958.
196. Boyle, N.M.O., et al., *Open Babel: An open chemical toolbox*. J. ChemInf., 2011. **3**: p. 1-14.
197. Bastian, M., S. Heymann, and M. Jacomy, *Gephi: An Open Source Software for Exploring and Manipulating Networks*. 2009.
198. Bajusz, D., G. G Ferenczy, and G. Keserű, *Structure-based Virtual screening approaches in kinase-directed drug discovery*. Curr. Top. Med. Chem., 2017. **17**(20): p. 2235-2259.
199. Talhout, R. and J.B. Engberts, *Thermodynamic analysis of binding of p-substituted benzamidines to trypsin*. Eur. J. Biochem., 2001. **268** (6): p. 1554-1560.
200. Sleight, S.H., et al., *Crystallographic and calorimetric analysis of peptide binding to OppA protein*. J. Mol. Biol., 1999. **291** (2): p. 393-415.
201. Dullweber, F., et al., *Factorising ligand affinity: A Combined thermodynamic and crystallographic study of trypsin and thrombin inhibition*. J. Mol. Biol., 2001. **313** (3): p. 593-614.
202. Palencia, A., et al., *Thermodynamic dissection of the binding energetics of proline-rich peptides to the Abl-SH3 domain: implications for rational ligand design*. J. Mol. Biol., 2004. **336** (2): p. 527-537.
203. McNemar, C., et al., *Thermodynamic and structural analysis of phosphotyrosine polypeptide binding to Grb2-SH2*. BioChem., 1997. **36**: p. 10006-10014.
204. Wang, C., N.H. Pawley, and L.K. Nicholson, *The role of backbone motions in ligand binding to the c-Src SH3 domain*. J. Mol. Biol., 2001. **313**: p. 873-887.
205. Org, T., et al., *The autoimmune regulator PHD finger binds to non-methylated histone H3K4 to activate gene expression*. EMBO Rep., 2008. **9**: p. 370-376.
206. Chrencik, J.E., et al., *Structure and thermodynamic characterization of the EphB4/Ephrin-B2 antagonist peptide complex reveals the determinants for receptor specificity*. Struct., 2006. **14**: p. 321-330.
207. Kozlov, G., et al., *Structural basis of ligand recognition by PABC, a highly specific peptide-binding domain found in poly(A)-binding protein and a HECT ubiquitin ligase*. EMBO J., 2004. **23**: p. 272-281.
208. Day, C.L., et al., *Structure of the BH3 domains from the p53-inducible BH3-only proteins Noxa and Puma in complex with Mcl-1*. J. Mol. Biol., 2008. **380**: p. 958-571.
209. Lu, Z. and Y. Zhang, *Interfacing ab initio quantum mechanical method with classical drude oscillator polarizable model for molecular dynamics simulation of chemical reactions*. J. Chem. Theory Comput., 2008. **4**: p. 1237-1248.
210. Vanommeslaeghe, K., O. Guvench, and A.D. MacKerell, *Molecular mechanics*. Curr. Pharm. Des., 2014. **20**: p. 3281-3292.
211. Ganoth, A., et al., *A molecular dynamics study and free energy analysis of complexes between the Mlc1p protein and two IQ motif peptides*. Biophys. J., 2006. **91**: p. 2436-2450.
212. Fenley, A.T., H.S. Muddana, and M.K. Gilson, *Entropy–enthalpy transduction caused by conformational shifts can obscure the forces driving protein–ligand binding*. Proc. Natl. Acad. Sci., 2012. **109**: p. 20006-20011.
213. Huggins, D.J., *Quantifying the entropy of binding for water molecules in protein cavities by computing correlations*. Biophys. J., 2015. **108**: p. 928-936.

214. Dougherty, D.A., *Cation- $\pi$  interactions involving aromatic amino acids*. J. Nutr., 2007. **137**: p. 1504-1508.
215. Majumdar, S., S. Maiti, and S. Ghosh Dastidar, *Dynamic and static water molecules complement the TN16 conformational heterogeneity inside the tubulin cavity*. BioChem., 2016. **55**: p. 335-347.
216. Zhang, D.W., Y. Xiang, and J.Z.H. Zhang, *New advance in computational chemistry: Full quantum mechanical ab Initio computation of streptavidin-biotin interaction energy*. J. Phys. Chem. B., 2003. **107**: p. 12039-12041.
217. Zhang, D.W., et al., *Quantum mechanical map for protein-ligand binding with application to  $\beta$ -trypsin/benzamidine complex*. J. Chem. Phys., 2004. **120**: p. 1145-1148.
218. Brown, S., M. Shirts, and D. Mobley, *Free-energy calculations in structure-based drug design*. 2010.
219. Weichenberger, C.X., et al., *The solvent component of macromolecular crystals*. Acta Cryst., 2015. **71**: p. 1023-1038.
220. Halle, B., *Biomolecular cryocrystallography: Structural changes during flash-cooling*. Proc. Natl. Acad. Sci. USA, 2004. **101**: p. 4793-4798.
221. Schmidtke, P., et al., *Shielded hydrogen bonds as structural determinants of binding kinetics: application in drug design*. J. Am. Chem. Soc., 2011. **133**: p. 18903-18910.
222. Hanson, R.M., et al., *JSmol and the next-generation web-based representation of 3D molecular structure as applied to proteopedia*. 2013.
223. Rentzsch, R. and B.Y. Renard, *Docking small peptides remains a great challenge: An assessment using AutoDock Vina*. Brief Bioinform., 2015. **16**: p. 1045-1056.
224. Saladin, A., et al., *PEP-SiteFinder: A tool for the blind identification of peptide binding sites on protein surfaces*. Nucleic Acids Res., 2014. **42**: p. W221-W226.
225. Yan, C., X. Xu, and X. Zou, *Fully blind docking at the atomic level for protein-peptide complex structure prediction*. Structure, 2016. **24**: p. 1842-1853.
226. Joseph-McCarthy, D., et al., *Fragment-based lead discovery and design*. J. Chem. Inf. Model., 2014. **54**: p. 693-704.
227. Liao, J., Y.-T. Wang, and C.S. Lin, *A fragment-based docking simulation for investigating peptide-protein bindings*. Phys. Chem. Chem. Phys., 2017. **19**: p. 10436-10442.
228. Taylor, R.D., P.J. Jewsbury, and J.W. Essex, *A review of protein-small molecule docking methods*. J. Comput. Aided Mol. Des., 2002. **16**: p. 151-166.
229. Chignola, F., et al., *The solution structure of the first PHD finger of autoimmune regulator in complex with non-modified histone H3 tail reveals the antagonistic role of H3R2 methylation*. Nucleic Acids Res., 2009. **37**: p. 2951-2961.
230. Li, H., et al., *PaFlexPepDock: parallel ab-initio docking of peptides onto their receptors with full flexibility based on Rosetta*. PLoS ONE 2014. **9**: p. e105715.
231. Verschueren, E., et al., *Protein-peptide complex prediction through fragment interaction patterns*. Structure 2013. **21**: p. 789-797.

# Függelék

A táblázat A 6. táblázat nyers adatai

## A 3. egyenlet paramétere

Rendszer	$\alpha$	$\Delta\alpha^*$	$\beta$	$\Delta\beta^*$
Vákuum/Száraz	$4.3975 \times 10^{-3}$	$4.8897 \times 10^{-3}$	-7.8349	1.4104
Vákuum /Shell 1	$6.0855 \times 10^{-3}$	$3.6583 \times 10^{-3}$	-7.0296	1.3944
Vákuum /Shell 2	$4.0942 \times 10^{-3}$	$2.3105 \times 10^{-3}$	-6.7765	1.4465
Vákuum /Shell 3	$6.7440 \times 10^{-3}$	$2.1121 \times 10^{-3}$	-5.2456	1.3445
COSMO/Száraz	$7.9586 \times 10^{-2}$	$2.1696 \times 10^{-2}$	-3.4980	1.6027
COSMO/Shell 1	$3.7379 \times 10^{-2}$	$7.5864 \times 10^{-3}$	-4.3391	1.0864
COSMO/Shell 2	$1.1187 \times 10^{-2}$	$4.4373 \times 10^{-3}$	-5.9395	1.4008
COSMO/Shell 3	$2.1972 \times 10^{-2}$	$3.7381 \times 10^{-3}$	-3.1009	1.1038

\*Standard error of regression coefficient.

## Vákuum/Száraz\*

Rendszer	$\Delta_f H(\text{Ligand})$	$\Delta_f H(\text{Target})$	$\Delta_f H(\text{Target:Ligand}[\text{H}_2\text{O}]_s)$	$\Delta_f H$	$\Delta H_b(\text{calc})$	$ \varepsilon $
3ptb_ben	178.10	-1490.53	-1396.08	-83.65	-8.20	3.70
3ptb_pme	162.59	-1632.88	-1510.70	-40.41	-8.01	3.60
3ptb_pam	167.30	-1497.83	-1403.59	-73.06	-8.16	1.74
3ptb_pmo	130.65	-1718.48	-1669.42	-81.59	-8.19	4.45
3ptb_pad	136.51	-1672.10	-1705.79	-170.20	-8.58	5.65
1k1l	262.49	-1928.22	-1798.53	-132.81	-8.42	0.56
1k1m	19.38	-1688.03	-1720.73	-52.08	-8.06	0.16
1k1i	-117.59	-1957.77	-2164.68	-89.31	-8.23	2.67
1k1j	-33.07	-1617.02	-1695.03	-44.94	-8.03	1.43
1jyr	-1038.98	-1744.62	-2984.92	-201.32	-8.72	0.78
1rlq	-222.49	-1576.62	-2185.57	-386.47	-9.53	0.67
2ke1	-71.83	-2361.52	-3322.49	-889.14	-11.74	2.54
2bba	-740.01	-2916.44	-3749.75	-93.29	-8.25	7.25
1jgn	-1455.38	-2702.38	-4405.34	-247.57	-8.92	5.88
2roc	-2618.04	-2389.65	-5350.57	-342.88	-9.34	4.96

\*(kcal/mol)

## Vákuum/Shell 1\*

Rendszer	$\Delta_f H(\text{Ligand})$	$\Delta_f H(\text{Target})$	$\Delta_f H(\text{Target:Ligand}[\text{H}_2\text{O}]_s)$	$\Delta_f H$	$\Delta H_b(\text{calc})$	$ \varepsilon $
3ptb_ben	178.10	-1490.53	-1478.31	-108.08	-7.69	3.18
3ptb_pme	162.59	-1632.88	-1605.48	-77.39	-7.50	3.09
3ptb_pam	167.30	-1497.83	-1615.82	-111.89	-7.71	1.29
3ptb_pmo	130.65	-1718.48	-1747.77	-102.14	-7.65	3.91
3ptb_pad	136.51	-1672.10	-1818.83	-167.64	-8.05	5.11
1k1l	262.49	-1928.22	-2155.93	-201.20	-8.25	0.39
1k1m	19.38	-1688.03	-2004.46	-104.61	-7.67	0.56
1k1i	-117.59	-1957.77	-2302.42	-111.46	-7.71	3.19
1k1j	-33.07	-1617.02	-1926.14	-102.66	-7.65	1.81
1jyr	-1038.98	-1744.62	-3036.58	-195.19	-8.22	0.28
1rlq	-222.49	-1576.62	-2330.39	-415.69	-9.56	0.64
2ke1	-71.83	-2361.52	-4251.26	-1124.31	-13.87	4.67
2bba	-740.01	-2916.44	-4692.84	-342.79	-9.12	6.38
1jgn	-1455.38	-2702.38	-5383.05	-416.08	-9.56	5.24
2roc	-2618.04	-2389.65	-6337.01	-520.12	-10.19	4.11

\*(kcal/mol)

**Vákuum/Shell 2\***

Rendszer	$\Delta_f H(\text{Ligand})$	$\Delta_f H(\text{Target})$	$\Delta_f H(\text{Target:Ligand}[\text{H}_2\text{O}]_s)$	$\Delta_f H$	$\Delta H_b(\text{calc})$	$ \varepsilon $
3ptb_ben	178.10	-1490.53	-1814.24	-155.01	-7.41	2.90
3ptb_pme	162.59	-1632.88	-1926.73	-167.43	-7.46	3.05
3ptb_pam	167.30	-1497.83	-1689.75	-128.02	-7.30	0.88
3ptb_pmo	130.65	-1718.48	-2100.65	-166.02	-7.46	3.71
3ptb_pad	136.51	-1672.10	-2290.93	-292.94	-7.98	5.04
1k1l	262.49	-1928.22	-2533.46	-289.74	-7.96	0.10
1k1m	19.38	-1688.03	-2627.09	-264.84	-7.86	0.36
1k1i	-117.59	-1957.77	-3113.13	-286.36	-7.95	2.95
1k1j	-33.07	-1617.02	-2666.43	-264.95	-7.86	1.60
1jyr	-1038.98	-1744.62	-4006.52	-413.72	-8.47	0.53
1rlq	-222.49	-1576.62	-4014.08	-769.98	-9.93	0.27
2ke1	-71.83	-2361.52	-5031.56	-1731.21	-13.86	4.66
2bba	-740.01	-2916.44	-4892.75	-369.30	-8.29	7.21
1jgn	-1455.38	-2702.38	-6921.86	-798.90	-10.05	4.75
2roc	-2618.04	-2389.65	-8129.03	-924.94	-10.56	3.74

\*(kcal/mol)

**Vákuum/Shell 3\***

Rendszer	$\Delta_f H(\text{Ligand})$	$\Delta_f H(\text{Target})$	$\Delta_f H(\text{Target:Ligand}[\text{H}_2\text{O}]_s)$	$\Delta_f H$	$\Delta H_b(\text{calc})$	$ \varepsilon $
3ptb_ben	178.10	-1490.53	-1894.35	-177.32	-6.44	1.93
3ptb_pme	162.59	-1632.88	-1994.60	-177.50	-6.44	2.03
3ptb_pam	167.30	-1497.83	-1904.57	-169.43	-6.39	0.03
3ptb_pmo	130.65	-1718.48	-2161.13	-168.70	-6.38	2.64
3ptb_pad	136.51	-1672.10	-2400.94	-287.35	-7.18	4.25
1k1l	262.49	-1928.22	-2906.69	-373.96	-7.77	0.10
1k1m	19.38	-1688.03	-2918.14	-324.68	-7.44	0.79
1k1i	-117.59	-1957.77	-3260.66	-318.30	-7.39	3.51
1k1j	-33.07	-1617.02	-2889.86	-314.97	-7.37	2.10
1jyr	-1038.98	-1744.62	-4063.15	-412.56	-8.03	0.09
1rlq	-222.49	-1576.62	-4143.39	-783.69	-10.53	0.33
2ke1	-71.83	-2361.52	-5363.26	-1369.31	-14.48	5.28
2bba	-740.01	-2916.44	-5814.11	-597.05	-9.27	6.23
1jgn	-1455.38	-2702.38	-7969.57	-1037.41	-12.24	2.56
2roc	-2618.04	-2389.65	-9169.78	-1156.49	-13.04	1.26

\*(kcal/mol)

**COSMO/Száraz\***

Rendszer	$\Delta_f H(\text{Ligand})$	$\Delta_f H(\text{Target})$	$\Delta_f H(\text{Target:Ligand}[\text{H}_2\text{O}]_s)$	$\Delta_f H$	$\Delta H_b(\text{calc})$	$ \varepsilon $
3ptb_ben	121.51	-1586.93	-1524.91	-59.49	-8.23	3.73
3ptb_pme	110.35	-1747.18	-1655.03	-18.20	-4.95	0.53
3ptb_pam	111.93	-1588.82	-1513.72	-36.83	-6.43	0.01
3ptb_pmo	75.10	-1835.89	-1808.97	-48.18	-7.33	3.59
3ptb_pad	68.28	-1843.55	-1806.26	-31.00	-5.97	3.03
1k1l	105.03	-2055.90	-1973.22	-22.35	-5.28	2.59
1k1m	-45.84	-1848.02	-1919.22	-25.37	-5.52	2.71
1k1i	-187.90	-2072.55	-2318.30	-57.86	-8.10	2.80
1k1j	-92.20	-1757.78	-1917.79	-67.81	-8.89	0.57
1jyr	-1234.57	-1878.39	-3190.48	-77.52	-9.67	1.73
1rlq	-425.59	-1856.53	-2371.00	-88.88	-10.57	0.37
2ke1	-503.91	-2988.96	-3619.50	-126.63	-13.58	4.38
2bba	-919.44	-3156.63	-4172.38	-96.30	-11.16	4.34
1jgn	-1612.06	-2918.85	-4652.64	-121.72	-13.19	1.61
2roc	-2889.82	-2613.33	-5604.22	-101.07	-11.54	2.76

\*(kcal/mol)

**COSMO/Shell 1\***

Rendszer	$\Delta_f H(\text{Ligand})$	$\Delta_f H(\text{Target})$	$\Delta_f H(\text{Target:Ligand}[\text{H}_2\text{O}]_s)$	$\Delta_f H$	$\Delta H_b(\text{calc})$	$ \varepsilon $
3ptb_ben	121.51	-1586.93	-1597.54	-66.91	-6.84	2.33
3ptb_pme	110.35	-1747.18	-1733.93	-31.89	-5.53	1.12
3ptb_pam	111.93	-1588.82	-1728.52	-56.00	-6.43	0.02
3ptb_pmo	75.10	-1835.89	-1896.88	-70.87	-6.99	3.25
3ptb_pad	68.28	-1843.55	-1951.67	-45.99	-6.06	3.12
1k1l	105.03	-2055.90	-2343.04	-66.11	-6.81	1.05
1k1m	-45.84	-1848.02	-2227.18	-72.48	-7.05	1.17
1k1i	-187.90	-2072.55	-2469.51	-78.64	-7.28	3.62
1k1j	-92.20	-1757.78	-2113.17	-67.55	-6.86	2.60
1jyr	-1234.57	-1878.39	-3264.81	-86.64	-7.58	0.36
1rlq	-425.59	-1856.53	-2512.44	-99.90	-8.07	2.13
2ke1	-503.91	-2988.96	-4517.93	-242.54	-13.41	4.21
2bba	-919.44	-3156.63	-5100.12	-241.52	-13.37	2.13
1jgn	-1612.06	-2918.85	-5730.98	-287.11	-15.07	0.27
2roc	-2889.82	-2613.33	-6649.32	-233.22	-13.06	1.24

\*(kcal/mol)

**COSMO/Shell 2\***

Rendszer	$\Delta_f H(\text{Ligand})$	$\Delta_f H(\text{Target})$	$\Delta_f H(\text{Target:Ligand}[\text{H}_2\text{O}]_s)$	$\Delta_f H$	$\Delta H_b(\text{calc})$	$ \varepsilon $
3ptb_ben	121.51	-1586.93	-1947.84	-91.16	-6.96	2.45
3ptb_pme	110.35	-1747.18	-2047.97	-85.09	-6.89	2.48
3ptb_pam	111.93	-1588.82	-1819.98	-82.25	-6.86	0.44
3ptb_pmo	75.10	-1835.89	-2252.35	-100.29	-7.06	3.32
3ptb_pad	68.28	-1843.55	-2414.92	-117.97	-7.26	4.32
1k1l	105.03	-2055.90	-2725.01	-122.03	-7.30	0.56
1k1m	-45.84	-1848.02	-2813.28	-136.90	-7.47	0.75
1k1i	-187.90	-2072.55	-3293.99	-185.81	-8.02	2.88
1k1j	-92.20	-1757.78	-2881.42	-183.71	-7.99	1.47
1jyr	-1234.57	-1878.39	-4240.22	-214.31	-8.34	0.40
1rlq	-425.59	-1856.53	-4238.84	-326.46	-9.59	0.61
2ke1	-503.91	-2988.96	-5274.97	-803.95	-14.93	5.73
2bba	-919.44	-3156.63	-5321.09	-266.86	-8.92	6.58
1jgn	-1612.06	-2918.85	-7249.62	-501.55	-11.55	3.25
2roc	-2889.82	-2613.33	-8455.49	-474.33	-11.25	3.05

\*(kcal/mol)

**COSMO/Shell 3\***

Rendszer	$\Delta_f H(\text{Ligand})$	$\Delta_f H(\text{Target})$	$\Delta_f H(\text{Target:Ligand}[\text{H}_2\text{O}]_s)$	$\Delta_f H$	$\Delta H_b(\text{calc})$	$ \varepsilon $
3ptb_ben	121.51	-1586.93	-2039.70	-117.81	-5.69	1.18
3ptb_pme	110.35	-1747.18	-2130.87	-102.78	-5.36	0.95
3ptb_pam	111.93	-1588.82	-2042.38	-109.02	-5.50	0.92
3ptb_pmo	75.10	-1835.89	-2348.58	-131.31	-5.99	2.24
3ptb_pad	68.28	-1843.55	-2566.22	-138.85	-6.15	3.22
1k1l	105.03	-2055.90	-3126.41	-197.37	-7.44	0.43
1k1m	-45.84	-1848.02	-3101.83	-164.61	-6.72	1.50
1k1i	-187.90	-2072.55	-3441.60	-203.00	-7.56	3.34
1k1j	-92.20	-1757.78	-3077.59	-184.24	-7.15	2.32
1jyr	-1234.57	-1878.39	-4335.35	-244.23	-8.47	0.53
1rlq	-425.59	-1856.53	-4376.79	-333.99	-10.44	0.24
2ke1	-503.91	-2988.96	-5643.13	-389.58	-11.66	2.46
2bba	-919.44	-3156.63	-6229.64	-392.88	-11.73	3.77
1jgn	-1612.06	-2918.85	-8181.94	-520.91	-14.55	0.25
2roc	-2889.82	-2613.33	-9481.52	-587.42	-16.01	1.71

\*(kcal/mol)



**Vákuum\***

Rendszer	$\Delta_f H$
Víz	-57.80

\*(kcal/mol)

**COSMO\***

Rendszer	$\Delta_f H$
Víz	-65.21

\*(kcal/mol)

**Három modell részletes eredményei** $\beta \neq 0$ **3. Egyenlet paraméterei**

Modell	$\alpha$	$\Delta\alpha^*$	$\beta$	$\Delta\beta^*$
Vákuum/Száraz	$1.6204 \times 10^{-2}$	$9.6984 \times 10^{-3}$	-6.2967	1.7530
COSMO/Száraz	$1.0029 \times 10^{-1}$	$2.2342 \times 10^{-2}$	-2.5495	1.5319
COSMO/Shell 3	$2.3097 \times 10^{-2}$	$3.7761 \times 10^{-3}$	-3.0011	1.0851

\*A regressziós egyenlet standard hibája.

**Vákuum/Száraz\***

Rendszer	$\Delta H_b(\text{calc})$	$ \varepsilon $
3ptb_ben	-7.65	3.15
3ptb_pme	-6.95	2.54
3ptb_pam	-7.48	1.06
3ptb_pmo	-7.62	3.88
3ptb_pad	-9.05	6.12
1k1l	-8.45	0.59
1k1m	-7.14	1.08
1k1i	-7.74	3.16
1k1j	-7.02	2.44
1jyr	-9.56	1.62
1rlq	-12.56	2.36
2bba	-7.81	7.69
1jgn	-10.31	4.49
2roc	-11.85	2.45
<b>R<sup>2</sup></b>		<b>0.19</b>
<b>F</b>		<b>2.79</b>
<b>RMSE</b>		<b>3.88</b>

\*(kcal/mol)

**COSMO/Száraz\***

Rendszer	$\Delta H_b(\text{calc})$	$ \varepsilon $
3ptb_ben	-8.52	4.01
3ptb_pme	-4.37	0.04
3ptb_pam	-6.24	0.17
3ptb_pmo	-7.38	3.64
3ptb_pad	-5.66	2.72
1k1l	-4.79	3.07
1k1m	-5.09	3.13
1k1i	-8.35	2.55
1k1j	-9.35	0.11
1jyr	-10.32	2.38
1rlq	-11.46	1.26

Rendszer	$\Delta H_b(\text{calc})$	$ \varepsilon $
2bba	-12.21	3.29
1jgn	-14.76	0.04
2roc	-12.69	1.61
<b>R<sup>2</sup></b>		<b>0.63</b>
<b>F</b>		<b>20.15</b>
<b>RMSE</b>		<b>2.63</b>

\*(kcal/mol)

### COSMO/Shell 3\*

Rendszer	$\Delta H_b(\text{calc})$	$ \varepsilon $
3ptb_ben	-5.72	1.22
3ptb_pme	-5.38	0.96
3ptb_pam	-5.52	0.90
3ptb_pmo	-6.03	2.29
3ptb_pad	-6.21	3.27
1k1l	-7.56	0.30
1k1m	-6.80	1.42
1k1i	-7.69	3.21
1k1j	-7.26	2.21
1jyr	-8.64	0.70
1rlq	-10.72	0.52
2bba	-12.08	3.42
1jgn	-15.03	0.23
2roc	-16.57	2.27
<b>R<sup>2</sup></b>		<b>0.76</b>
<b>F</b>		<b>37.41</b>
<b>RMSE</b>		<b>2.12</b>

\*(kcal/mol)

$\beta=0$

### 3. Egyenlet paramétere

Modell	$\alpha$	$\Delta\alpha^*$	$\beta$	$\Delta\beta^*$
Vákuum/Száraz	$4.4281 \times 10^{-2}$	$7.9453 \times 10^{-3}$	0	-
COSMO/Száraz	$1.3332 \times 10^{-1}$	$1.0941 \times 10^{-2}$	0	-
COSMO/Shell 3	$1.1997 \times 10^{-2}$	$2.4289 \times 10^{-3}$	0	-

\*A regressziós koefficiens standard hibája.

### Vákuum/Száraz\*

Rendszer	$\Delta H_b(\text{calc})$	$ \varepsilon $
3ptb_ben	-3.70	0.80
3ptb_pme	-1.79	2.62
3ptb_pam	-3.24	3.18
3ptb_pmo	-3.61	0.13
3ptb_pad	-7.54	4.60
1k1l	-5.88	1.98
1k1m	-2.31	5.92
1k1i	-3.95	6.94
1k1j	-1.99	7.47
1jyr	-8.91	0.97
1rlq	-17.11	6.91
2bba	-4.13	11.37
1jgn	-10.96	3.84

Rendszer	$\Delta H_b(\text{calc})$	$ \varepsilon $
2roc	-15.18	0.88
<b>R<sup>2</sup></b>		<b>0.70</b>
<b>R<sup>2</sup>(cv)</b>		<b>0.71</b>
<b>F</b>		<b>31.06</b>
<b>RMSE</b>		<b>5.37</b>

\*(kcal/mol)

#### COSMO/Száraz\*

Rendszer	$\Delta H_b(\text{calc})$	$ \varepsilon $
3ptb_ben	-7.93	3.42
3ptb_pme	-2.43	1.99
3ptb_pam	-4.91	1.51
3ptb_pmo	-6.42	2.68
3ptb_pad	-4.13	1.20
1k1l	-2.98	4.88
1k1m	-3.38	4.84
1k1i	-7.71	3.18
1k1j	-9.04	0.42
1jyr	-10.34	2.40
1rlq	-11.85	1.65
2bba	-12.84	2.66
1jgn	-16.23	1.43
2roc	-13.48	0.82
<b>R<sup>2</sup></b>		<b>0.92</b>
<b>R<sup>2</sup>(cv)</b>		<b>0.91</b>
<b>F</b>		<b>148.48</b>
<b>RMSE</b>		<b>2.81</b>

\*(kcal/mol)

#### COSMO/Shell 3\*

Rendszer	$\Delta H_b(\text{calc})$	$ \varepsilon $
3ptb_ben	-3.77	0.74
3ptb_pme	-3.29	1.12
3ptb_pam	-3.49	2.93
3ptb_pmo	-4.20	0.46
3ptb_pad	-4.44	1.51
1k1l	-6.32	1.55
1k1m	-5.27	2.95
1k1i	-6.50	4.40
1k1j	-5.90	3.57
1jyr	-7.81	0.13
1rlq	-10.69	0.49
2bba	-12.57	2.93
1jgn	-16.67	1.87
2roc	-18.80	4.50
<b>R<sup>2</sup></b>		<b>0.93</b>
<b>R<sup>2</sup>(cv)</b>		<b>0.91</b>
<b>F</b>		<b>173.54</b>
<b>RMSE</b>		<b>2.61</b>

\*(kcal/mol)

**B táblázat** Lineáris regresszió három vízmodellre,  $\beta=0$

### 3. Egyenlet paraméterei

Modell	$\alpha$	$\Delta\alpha^*$	$\beta$	$\Delta\beta^*$
Vákuum/Száraz	$2.2784 \times 10^{-2}$	$6.3704 \times 10^{-3}$	0	-
COSMO/Száraz	$1.2143 \times 10^{-1}$	$1.1437 \times 10^{-2}$	0	-
COSMO/Shell 3	$3.1025 \times 10^{-2}$	$2.3146 \times 10^{-3}$	0	-

\*A regressziós koefficiens standard hibája.

#### Vákuum/Száraz\*

Rendszer	$\Delta H_b(\text{calc})$	$ \varepsilon $
3ptb_ben	-1.91	2.60
3ptb_pme	-0.92	3.49
3ptb_pam	-1.66	4.75
3ptb_pmo	-1.86	1.88
3ptb_pad	-3.88	0.94
1k1l	-3.03	4.84
1k1m	-1.19	7.04
1k1i	-2.03	8.86
1k1j	-1.02	8.44
1jyr	-4.59	3.35
1rlq	-8.81	1.39
2ke1	-20.26	11.06
2bba	-2.13	13.37
1jgn	-5.64	9.16
2roc	-7.81	6.49
<b>R<sup>2</sup></b>		<b>0.48</b>
<b>R<sup>2</sup>(cv)</b>		<b>0.27</b>
<b>F</b>		<b>12.79</b>
<b>RMSE</b>		<b>7.12</b>

\*(kcal/mol)

#### COSMO/Száraz\*

Rendszer	$\Delta H_b(\text{calc})$	$ \varepsilon $
3ptb_ben	-7.22	2.72
3ptb_pme	-2.21	2.20
3ptb_pam	-4.47	1.94
3ptb_pmo	-5.85	2.11
3ptb_pad	-3.76	0.83
1k1l	-2.71	5.15
1k1m	-3.08	5.14
1k1i	-7.03	3.87
1k1j	-8.23	1.23
1jyr	-9.41	1.47
1rlq	-10.79	0.59
2ke1	-15.38	6.18
2bba	-11.69	3.81
1jgn	-14.78	0.02
2roc	-12.27	2.03
<b>R<sup>2</sup></b>		<b>0.89</b>
<b>R<sup>2</sup>(cv)</b>		<b>0.87</b>
<b>F</b>		<b>112.73</b>
<b>RMSE</b>		<b>3.27</b>

\*(kcal/mol)

### COSMO/Shell 3\*

Rendszer	$\Delta H_b(\text{calc})$	$ \varepsilon $
3ptb_ben	-3.66	0.85
3ptb_pme	-3.19	1.22
3ptb_pam	-3.38	3.03
3ptb_pmo	-4.07	0.33
3ptb_pad	-4.31	1.37
1k1l	-6.12	1.74
1k1m	-5.11	3.12
1k1i	-6.30	4.60
1k1j	-5.72	3.75
1jyr	-7.58	0.36
1rlq	-10.36	0.16
2ke1	-12.09	2.89
2bba	-12.19	3.31
1jgn	-16.16	1.36
2roc	-18.22	3.92
<b>R<sup>2</sup></b>		<b>0.93</b>
<b>R<sup>2</sup>(cv)</b>		<b>0.91</b>
<b>F</b>		<b>179.66</b>
<b>RMSE</b>		<b>2.65</b>

\*(kcal/mol)

**C táblázat** Két részre osztva újraszámoltuk a regressziót

$\beta \neq 0$

### 3. Egyenlet paraméterei

Alrendszer	$\alpha$	$\beta$
1	$5.7500 \times 10^{-2}$	2.1228
2	$1.9629 \times 10^{-2}$	-3.9127

### COSMO/Shell 3\*

Rendszer	$\Delta H_b(\text{calc})$	$ \varepsilon $
3ptb_ben	-4.65	0.14
3ptb_pme	-3.79	0.62
3ptb_pam	-4.15	2.27
3ptb_pmo	-5.43	1.69
3ptb_pad	-5.86	2.93
1k1l	-9.23	1.36
1k1m	-7.34	0.88
1k1i	-9.55	1.35
1k1j	-8.47	0.99
1jyr	-8.71	0.77
1rlq	-10.47	0.27
2ke1	-11.56	72.36
2bba	-11.62	3.88
1jgn	-14.14	0.66
2roc	-15.44	1.14
<b>R<sup>2</sup></b>		<b>0.80</b>
<b>F</b>		<b>50.91</b>
<b>RMSE</b>		<b>1.67</b>

\*(kcal/mol)

$\beta = 0$

### 3. Egyenlet paramétere

Alrendszer	$\alpha$	$\beta$
1	$4.4122 \times 10^{-2}$	-
2	$2.8461 \times 10^{-2}$	-

### COSMO/Shell 3\*

System	$\Delta H_b(\text{calc})$	$ \varepsilon $
3ptb_ben	-5.20	0.69
3ptb_pme	-4.54	0.12
3ptb_pam	-4.81	1.61
3ptb_pmo	-5.79	2.05
3ptb_pad	-6.13	3.19
1k1l	-8.71	0.85
1k1m	-7.26	0.96
1k1i	-8.96	1.94
1k1j	-8.13	1.34
1jyr	-6.95	0.99
1rlq	-9.51	0.69
2ke1	-11.09	1.89
2bba	-11.18	4.32
1jgn	-14.83	0.03
2roc	-16.72	2.42
<b>R<sup>2</sup></b>		<b>0.96</b>
<b>F</b>		<b>336.60</b>
<b>RMSE</b>		<b>1.93</b>

\*(kcal/mol)

### D táblázat A FragmentMerge egy példa riportja

<b>System</b>
2KE1
<b>ID</b>
C2
<b>Remark</b>
\home\user\CBDD\PDB\LINKED\2KE1\C2\LINKING\
<b>Ligand</b>
ARTKQTARKS
<b>Parameters</b>
dCN,max = 6.0
dCN,min = 0.75
alpha = 1.0
dall = 0.75
<b>Path of input pools (fragment copies)</b>
\home\user\CBDD\PDB\LINKED\2KE1\C2\LINKING\INPUT

**List of fragments; count of copies**

AR1 28

TK2 39

QT3 69

AR4 31

KS5 81

**Path of output pools (linked structures)**

\\home\\user\\CBDD\\PDB\\LINKED\\2KE1\\C2\\LINKING\\OUTPUT\\

**List of output pools; count of linked peptides****Pentads 2**

\\home\\user\\CBDD\\PDB\\LINKED\\2KE1\\C2\\LINKING\\OUTPUT\\Pentads\\

AR25 - TK04 - QT07 - AR06 - KS14

AR17 - TK04 - QT07 - AR06 - KS14

**Tetrads 2**

\\home\\user\\CBDD\\PDB\\LINKED\\2KE1\\C2\\LINKING\\OUTPUT\\Tetrads\\

2018.10.13.14.44.16\\part2-5

TK02 - QT51 - AR02 - KS27

TK04 - QT07 - AR06 - KS14

**Triads 20**

\\home\\user\\CBDD\\PDB\\LINKED\\2KE1\\C2\\LINKING\\OUTPUT\\Triads\\

AR22 - TK28 - QT56

AR22 - TK28 - QT18

AR14 - TK38 - QT68

TK17 - QT52 - AR08

TK21 - QT35 - AR14

TK02 - QT51 - AR21

TK07 - QT14 - AR03

QT57 - AR25 - KS42

QT07 - AR06 - KS14

QT50 - AR26 - KS30

QT58 - AR27 - KS52

QT58 - AR27 - KS18

QT03 - AR10 - KS38

QT03 - AR10 - KS56

QT39 - AR15 - KS51

QT39 - AR15 - KS47

QT51 - AR02 - KS27

QT32 - AR15 - KS51

QT32 - AR15 - KS47

QT32 - AR28 - KS25

**Dyads (Pairs) 59**

\home\user\CBDD\PDB\LINKED\2KE1\C2\LINKING\OUTPUT\Dyads\

AR17 - TK24

AR18 - TK39

AR04 - TK01

AR20 - TK24

AR07 - TK39

AR27 - TK24

TK22 - QT41

TK05 - QT67

TK36 - QT17

TK35 - QT55

TK29 - QT36

TK17 - QT16

TK06 - QT30

TK37 - QT38

TK09 - QT49

TK20 - QT60

TK19 - QT27

TK22 - QT25

TK38 - QT68

TK23 - QT30

TK10 - QT25

TK13 - QT37

TK34 - QT19

TK10 - QT41

TK28 - QT56

TK05 - QT20

TK28 - QT18

QT33 - AR30

QT51 - AR21

QT09 - AR17

QT66 - AR23

QT53 - AR19

QT29 - AR18

QT59 - AR04

QT42 - AR18

QT52 - AR08

QT62 - AR05

QT35 - AR14

QT58 - AR23

QT08 - AR16

QT31 - AR08

QT14 - AR03



AR06 - KS14  
 AR26 - KS30  
 AR28 - KS25  
 AR27 - KS52  
 AR13 - KS01  
 AR02 - KS27  
 AR15 - KS51  
 AR10 - KS38  
 AR12 - KS14  
 AR27 - KS18  
 AR10 - KS56  
 AR01 - KS31  
 AR25 - KS42  
 AR12 - KS10  
 AR24 - KS38  
 AR11 - KS28  
 AR15 - KS47

**A ábra** Az input paraméterek template fájlja

```
# comment: #
# optional: $ e.g. $EPS=78.4
# e.g. @RESTRAINT=CALPHA, RESTRAINT=NONE, RESTRAINT=HEAVY, RESTRAINT=BACKBONE
#METHODS=AM1,PM3, PM6 , PM6-DH+, PM7, RM1      # methods, separated with comma
METHODS=PM7                                     # method
WATERDIST=5                                     # water distance
LIGDIST=5                                       # ligand distance
NEIGHBOUR=0                                    # neighbour number
PDBPATH=/home/user/fragment/uploads/           # pdb path
PDBNAME=1k1j.pdb                               # pdb file name
#ALL_IN_VACUUM                                  # create complex, ligand and target in vacuum
#LIGANDCHARGE=-1                               # ligand charge
#FRAGMENTCHARGE=-1                             # fragment charge
TARGETCHAINID=A                                # target chain ID
LIGANDCHAINID=B                                # ligand chain ID
RESTRAINT=NONE                                 # restraint (N=NONE, H=HEAVY, B=BACKBONE, A=CALPHA)
MOPACPARAMS=PREC GRAD EF GNORM=1 MMOK GEO-OK DUMP=30M T=4W LET CYCLES=1000000000 MOZYME
                                                # all other mopac parameters
SCRIPTPATH=$HOME/fragment/bin/createmopacinput.sh # script path
MOPAC_LICENSE=$HOME/bin/mopac2012              # mopac license path
MOPAC=$HOME/bin/mopac2012/MOPAC2012.exe        # mopac path
#@RESTRAINT=CALPHA, RESTRAINT=NONE, RESTRAINT=HEAVY, RESTRAINT=BACKBONE # for list
#@RESTRAINT=CALPHA, RESTRAINT=NONE             # for list
PROCESSNAME=1k1j.pdb                          # this name appears in the qsub statistics
PHPPATH=/usr/bin/php                           # php binary path
```

## Tartalomjegyzék

1. A szerző tudományos közleményei .....	2
2. Az értekezés főbb eredményeinek összefoglalása (tézispontok)* .....	4
3. Bevezetés.....	5
3.1. A hiszton-komplexek szerkezetmeghatározásának nehézségei .....	5
3.2. A dokkolási probléma és eddigi megoldásai .....	7
3.3. A hidratációs probléma és eddigi megoldásai .....	9
3.4. Fragmens alapú számítások a kvantummechanikában.....	15
4. Módszerek .....	17
4.1. A módszerfejlesztés informatikai háttere .....	17
4.2. A Fragmenter eljárás és a kötési entalpiaváltozás számításának módszerei .....	19
4.2.1. A Fragmenter technikai megvalósítása.....	19
4.2.2. A képződéshő számítása .....	20
4.2.3. Célpont–ligandum intermolekuláris kölcsönhatási energia .....	21
4.2.4. Statisztika .....	22
4.3. A fragmens blind docking eljárás módszerei .....	22
4.3.1. A Célpont–ligandum intermolekuláris kölcsönhatási energia (E <sub>inter</sub> ) számítása.....	22
4.3.2. Összekapcsolás és hegesztés.....	22
4.4. A vízhálózatok számításának módszerei .....	23
5. Eredmények és értékelésük .....	25
5.1. Fragmenter eljárás és a kötési entalpiaváltozás számítása .....	25
5.1.1. Végpont-megközelítés.....	26
5.1.2. Fragmenter .....	28
5.1.3. Száraz rendszerek és implicit vízmodell.....	34
5.1.4. Explicit hidratáció és hibrid modell.....	35
5.1.5. A skálázó faktor .....	39
5.1.6. Vízszerkezeti esettanulmányok.....	39
5.1.7. A Fragmenter az interneten .....	41
5.1.8. Input .....	42
5.1.9. Fő algoritmus .....	43
5.1.10. Output.....	44
5.2. Fragmens blind docking.....	46
5.2.1. Fragmentálás .....	48

5.2.2. Összekapcsolás .....	53
5.2.3. Összefűzés és finomítás.....	55
5.3. Vízhálózatok számítása.....	58
5.3.1. Felületi hidratáció.....	59
5.3.2. Interfész hidratáció.....	62
6. Összefoglalás .....	64
7. Summary.....	65
Köszönetnyilvánítás.....	66
Irodalomjegyzék .....	67
Függelék .....	78
Tartalomjegyzék .....	90





Home

Create Account

Help

LIVE CHAT



**Title:** Analysis of the influence of simulation parameters on biomolecule-linked water networks

**Author:** Norbert Jeszenői, Gabriella Schilli, Mónika Bálint, István Horváth, Csaba Hetényi

**Publication:** Journal of Molecular Graphics and Modelling

**Publisher:** Elsevier

**Date:** June 2018

© 2018 Elsevier Inc. All rights reserved.

LOGIN

If you're a [copyright.com](#) user, you can login to RightsLink using your [copyright.com](#) credentials.

Already a [RightsLink](#) user or want to [learn more?](#)

Please note that, as the author of this Elsevier article, you retain the right to include it in a thesis or dissertation, provided it is not published commercially. Permission is not required, but please ensure that you reference the journal as the original source. For more information on this and on your other retained rights, please visit: <https://www.elsevier.com/about/our-business/policies/copyright#Author-rights>

BACK

CLOSE WINDOW