Párolgás szabadenergiájának meghatározása súlyozott mintavételezéssel számítógépes szimulációban

Szakdolgozat
Kémia Alapszak

KOHUT GERGELY

Témavezető: Dr. Tóth Gergely

Kémiai Informatikai Laboratórium

Eötvös Loránd Tudományegyetem, Budapest
Természettudományi Kar
Kémiai Intézet
A védés helye: Fizikai Kémiai Tanszék
2014

Köszönetnyilvánítás

Szeretném megköszönni témavezetőmnek, Dr. Tóth Gergelynek a segítségét és hasznos tanácsait a dolgozat megírásában, valamint köszönöm Kiss Péternek, hogy az eredeti BK3 modell molekuláris dinamikai programját rendelkezésünkre bocsátotta. A számításokat a NIIF számítóközpont pécsi szuperszámítógépén végeztük, a lehetőséget ezúton is köszönjük.

Szeretnék köszönetet mondani barátnőmnek, Nagy Flóra Boglárkának, hogy mellettem állt, türelmével és odaadásával segítette munkámat, továbbá bátyámnak, Kohut Attilának, hogy idejét nem kímélve, szakmai tanácsaival és szerkesztési észrevételeivel hozzájárult a dolgozat elkészüléséhez.

Tartalom

1. Bevezetés és célkitűzések 4
2. Irodalmi áttekintés 6
2.1 Vízmodellekről röviden 6
2.2 Víz párolgási szabadenergiájának számítása az irodalomban 7
2.3 Umbrella Sampling 11
2.4 Szabadenergia-számító módszerek 14
2.5 Felszín- és erőszámítás. A Chandler-féle mechanika 17
3. A szimuláció kivitelezése 21
4. Eredmények 27
4.1 WHAM-mal számolt profilok 27
4.2 VFEP-pel számolt profilok 29
4.3 MBAR-ral számolt profilok 30
4.4 Az egyes módszerekkel kapott eredmények összehasonlítása 31
Szakdolgozati összefoglaló 34
Summary 35
Irodalomjegyzék 36
Nyilatkozat 38

1. Bevezetés és célkitűzések

A víz már ősidőktől foglalkoztatja az embereket, hiszen életünk szempontjából alapvető fontosságú vegyület. A víz párolgása az egyik legalapvetőbb folyamat, ami a mindennapi életünk része és folyamatosan körül vesz minket, úgy a természeti körforgása, mint a mindennapi főzés során. A folyamat megértéséhez alapvető fontosságú megérteni a molekulák energetikai változását, éppen ezért azt tűztük ki célul, hogy molekuladinamikai szimulációk segítségével megpróbáljuk meghatározni a víz kondenzációja során történő szabadenergia-változást.

A feladat elvégzéséhez a Baranyi András és Kiss Péter által fejlesztett viszonylag új BK3[1] modellt használtuk, ugyanis ezzel a modellel még nem végeztek szabadenergia számításokat. Ez lehetővé teszi, hogy a csoportjuk által eddig a modellen végzett számításokat kiegészítsük. A téma aktualitását ezen kívül az is adta, hogy a Kiss Dóra Judit és Tóth Gergely által végzett, a párolgás kinetikájára vonatkozó kutatásokat kibővítsük.[2] Az irodalomban ebben a témában született cikkeket áttekintve úgy döntöttünk, hogy a számításainkhoz a témában eddig megjelent legújabb Varilly és Chandler[3] (A következőekben csak Chandler nevével hivatkozuk rá) által közölt cikket fogjuk alapul venni.

A dolgozat második fejezetében, az irodalmi áttekintésben röviden összefoglalom, hogy milyen modelleken végeztek hasonló számításokat és néhány további vízmodell segítségével jellemzem a modellek legfontosabb tulajdonságait. Ezt követően bemutatom az eddig a témában született számunkra releváns eredményeket, a megfelelő szabadenergia-profilokat és a párolgás során a két fázis közötti szabadenergia-különbséget, valamint a szabadenergia-számolás lehetséges módszereit. Mivel munkánkhoz a Chandler-féle számításokat vettük alapul, ezért a fejezet következő részében az általa is használt Umbrella Sampling módszer elméleti hátterét valamint a felszín számításához használt általa alkalmazott modell elméleti hátterét írom le.

A harmadik fejezetben tömören összefoglalom a gyakorlati munka során felmerült problémákat, az általunk írt programrész néhány jellegzetes tulajdonságát, valamint a szimulációk paramétereit.

A dolgozat végén összefoglalom az általunk kapott, különböző módszerekkel számolt eredményeket, és összehasonlítom az eddigi irodalomban tárgyaltakkal, valamint következtetéseket vonok le a különböző módszerek hatásosságáról, számítási pontosságáról.

A munkámhoz a használt modell Fortranban írt forráskódját Kiss Péter bocsátotta rendelkezésemre. A modellt kiegészítő Umbrella Samplinget megvalósító programot, valamint a kiértékeléshez és a párhuzamos futtatáshoz szükséges scripteket én írtam a témavezetőm segítségével. A kiértékelést az interneten szabadon elérhető módszerek programjain keresztül én végeztem el.

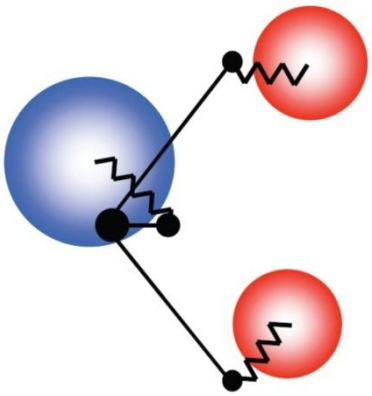
2. Irodalmi áttekintés

2.1 Vízmodellekről röviden

Mivel munkánk során azt tűztük ki célul, hogy egy molekula párolgásának szabadenergia változását meghatározzuk, az irodalmi áttekintésnél arra koncentráltunk, hogy feltérképezzük ki milyen vízmodellel és milyen módszert alkalmazva jutott el a víz párolgása (kondenzációja) során történő szabadenergia-profil meghatározásához, valamint, hogy milyen egyéb szabadenergia számítására szolgáló módszerek vannak még. Ennélfogva a következőkben röviden összefoglalom a témában született eddigi eredményeket valamint a számításokra vonatkozó egyéb új alternatívákat.

A leggyakrabban használt vízmodelleket két különböző osztályba lehet sorolni: merev nem polarizálható és merev polarizálható vízmodellek. Az eddigi legtöbb párolgással kapcsolatos szimulációt merev nem polarizálható vízmodelleken végezték, mint például a TIP3P[6], SPC/E[7], TIP4P/2005[8] vagy az TIP5P[9], az utóbbi évtizedekben azonban egyre inkább utat törnek maguknak a polarizálható modellek, amik sok esetben jobban közelítik a víz tulajdonságainak kísérleti eredményeit.

A molekuladinamikai szimulációk során a legnagyobb processzorigényt a molekulák közötti kölcsönhatások kiszámítása jelenti, ezért azt a vízmodellek fejlesztésének kezdetén a molekulák közötti párkölcsönhatásokra bontották és ennek megfelelően a molekulán kölcsönható helyeket (angolul site-okat) helyeztek el. Ennek megfelelően a TIP3P és az SPC/E három, a TIP4P/2005 négy, míg a TIP5P öt kölcsönható helyet tartalmaz. Egyik modell sem tudja azonban reprodukálni a víz összes tulajdonságát. Az adott modell általában mindig úgy van beállítva, hogy a víz bizonyos tulajdonságait jól visszaadja. A sok site-tal rendelkező molekulák valamelyest lassították a szimulációt a sok kölcsönható hely miatt. A számítókapacitás növekedésével azonban nőni kezdett az igény a modellek finomítására ezért megjelentek a polarizálható modellek, amelyek a polarizáció figyelembevétele miatt nagyobb számításigényűek. Ma az egyik legújabb polarizálható modell a Baranyai András és Kiss Péter által fejlesztett BK3 vízmodell. Ez egy ún. „charge on spring” modell, ami azt jelenti, hogy az atomokon lévő töltések vagy töltéseloszlások harmonikus rugóval vannak rögzítve, ennek az erőssége határozza meg a polarizálhatóságot.



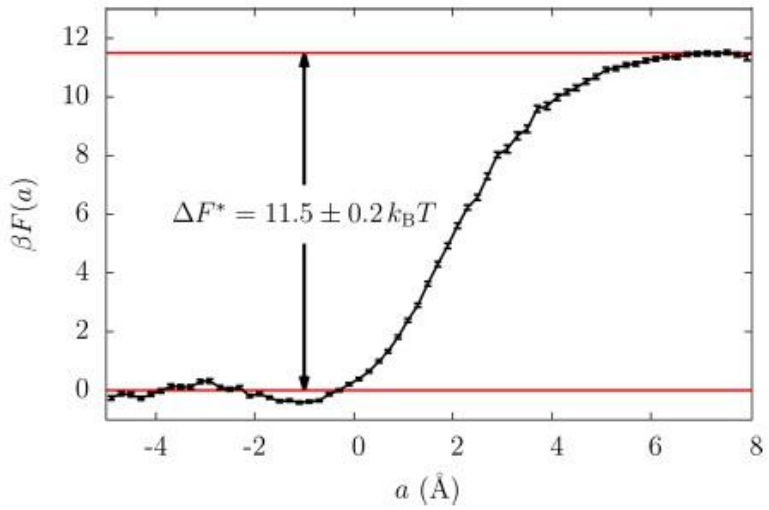
1. ábra: A BK3 modell sematikus felépítése[1]

Munkánk során azért választottuk a BK3 modellt a szabadenergia-számítások elvégzéséhez, mert a jelenlegi modellekhez képest jó közelítéssel képes követni a víz legtöbb tulajdonságának tapasztalati értékeit. Emellett képes kezelni a víz kis koordinációs számú állapotát, ezáltal alkalmas folyadékból a gázfázisba való átmenet vizsgálatára.

2.2 Vízpárolgás szabadenergiájának számítása az irodalomban

A szakirodalomban fellelhető szabadenergia-számítások közül szinte kivétel nélkül azonos eredményre jutnak egy adott molekula tömbfázisból gázfázisba való kiszakadását illetően.

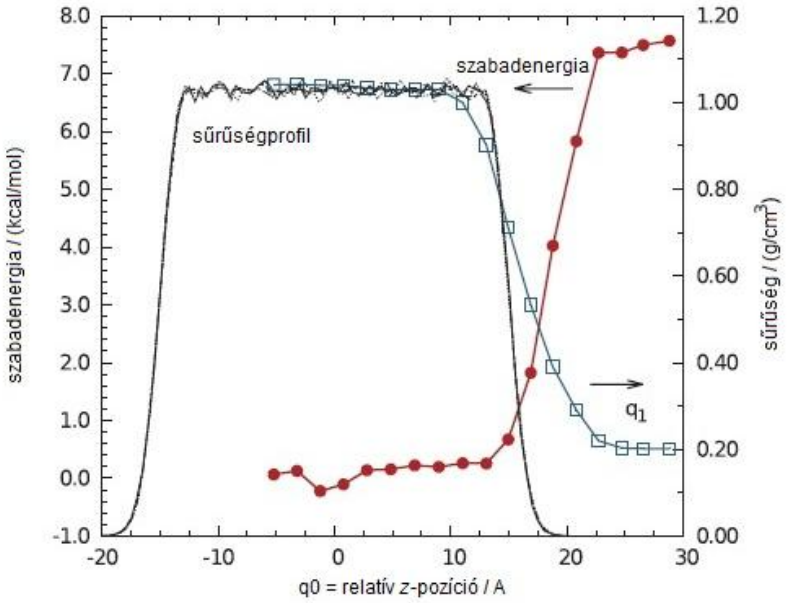
A Chandler és társai által végzett vizsgálatok szerint egy vízmolekula párolgása az 1. ábrán látható szabadenergia-profilt követi:



2. ábra: Vízmolekula kondenzációjának szabadenergia-változása Chandler szerint[3]

Az ábra egy egyedi molekula szabadenergiájának változását mutatja a felszíntől vett a távolságban, ahol a piros egyenesek a folyadék tömbfázisában és a gázfázisban lévő molekula szabadenergiáját mutatják. A görbe alakjából ítélve az első és egyik legalapvetőbb következtetés, hogy a párolgásnak nincs szabadenergia gátja, a reakciókoordináta előrehaladtával folyamatosan növekszik a molekula szabadenergiája. A reakció gátján jelen esetben a profilban megjelenő lokális maximumhelyek kialakulását értjük. Eltekintve a szabadenergia kis fluktuációtól jól látható, hogy a tömbfázis molekulái mély szabadenergia-állapotban vannak. A szimulációt Chandler SPC/E vízmodellen 300 K hőmérsékleten és 1 atm nyomáson végezte az egyes szimulációkhoz tartozó szabadenergiákat pedig MBAR[10,,18,19] módszerrel határozta meg. Átszámítva a kT mértékegységet kJ/mol-ra Chandler szerint a kiszakadás szabadenergia-különbsége $\Delta F = 28{,}68 \pm 0{,}5\ \mathrm{kJ/mol}$.[3]

N. Musolino és társai a következő szabadenergia-profilt határozták meg:

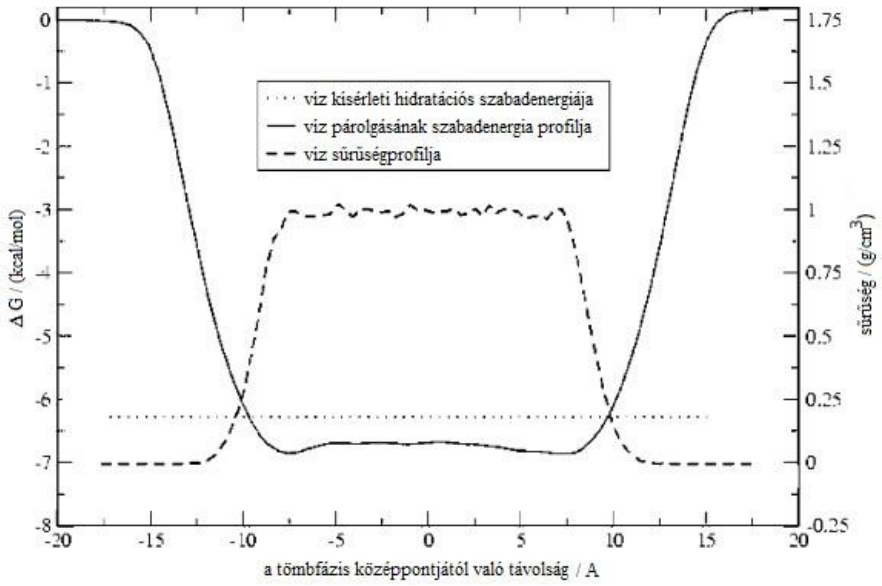


3. ábra.: Szabadenergia-profil Musolino szerint[11]

Számításaik alapján látszik, hogy a molekula szabadenergiájának változása a tömbfázisból a gázfázisba való kiszakadást követően, gyakorlatilag teljesen ugyanazt a profilt adja vissza, mint a Chandleréké. Ebben szerepet játszhat az a tény is, hogy szintén SPC/E vízmodellel dolgoztak. A konkrét szabadenergia számításokat string módszerrel végezték. A felszínt ők is a tömbfázis sűrűségén keresztül definiálják, azonban Chandlerékkel ellentétben, nem egy lépésenként dinamikusan változó felszínt határoznak meg, hanem egy négy szimulációból kiátlagolt sűrűségprofil azon részét veszik alapul, ahol a sűrűség értéke fele a tömbfázis sűrűségének. Ez lényegesen megkönnyíti a számolást, hiszen nem kell minden lépésben

felszínmeghatározással foglalkozni, azonban csak pontatlanabb számolást tesz lehetővé. A mért szabadenergia-változás az általuk végzett vizsgálatok során 30,96 kJ/mol volt. [11]

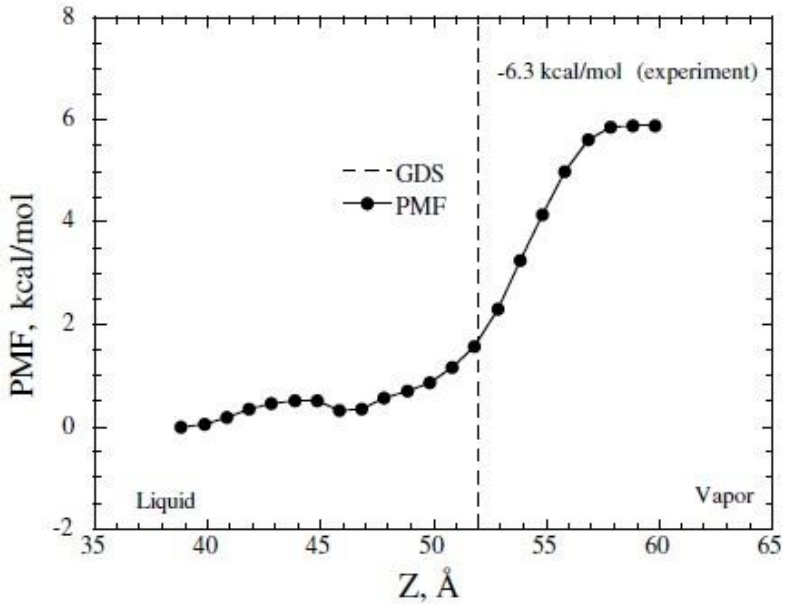
R. Vacha 2004-ben közölt cikkében szintén foglalkoznak a vízmolekulák kiszakadásának szabadenergia profiljával. Az általuk számolt profil a 4. ábrán látható.



4. ábra. Szabadenergia-profil Vacha szerint[12]

A görbe alakja és meredeksége gyakorlatilag teljesen megegyezik az eddig vizsgáltakkal. Szintén SPC/E modellt alkalmaztak a számításokhoz, viszont a teljes szabadenergia különbség némileg kisebb az eddigiekben tárgyaltakhoz képest, ugyanis a számításaik alapján ez megközelítőleg $\Delta F = 28{,}9\ \mathrm{kJ/mol}$. A görbe alakjában a periodicitás azzal magyarázható, hogy a szimulációs cella közepén helyezkedik el a tömbfázis, alatta és felette is vákuum van, tehát mindkét irányba megvalósulhat a kiszakadás.[12]

Dang és Garrett szintén közöltek cikket arra vonatkozólag, hogy milyen szabadenergia-változást követ egy vízmolekula kiszakadása, az általuk meghatározott modell az 5. ábrán látható.



5. ábra. Dang által meghatározott szabadenergia-profil[13]

A profil abban különbözik az eddigiektől, hogy egy általuk fejlesztett 4 site-tal rendelkező merev polarizálható modellen végezték a számításokat. A görbe alakjában az eddig tárgyaltakhoz képest az a változás, hogy már a felszín előtt – melyet a függőleges szaggatott vonal jelez – elkezd nőni a szabadenergia a tömbfázishoz képest. Az általuk meghatározott szabadenergia-változás a tömbfázisból való teljes kiszakadást követően $\Delta F = 24{,}69 \pm 0{,}21\,\mathrm{kJ/mol}$ volt, ami a fentiek közül a legközelebb van a kísérletileg meghatározott 26,36 kJ/mol értékhez. A kísérlettel való jó egyezés a többi munkában is várható volt, ugyanis az SPC/E modell megalkotásakor potenciális célként tűzték ki a víz kísérleti párolgáshőjének reprodukálását.[13]

Az áttekintést követően összefoglalásként elmondhatjuk, hogy minden általunk vizsgált cikkben ugyanarra a következtetésekre jutnak a víz párolgásának szabadenergia-profilját illetően. Vannak különbségek ugyan a számítási módszerek, valamint a felszínmeghatározások között, de a görbe alakja valamint a szabadenergia kis eltéréseit leszámítva gyakorlatilag teljesen megegyezik. Az alábbi táblázat az összes általunk vizsgált profil összehasonlítását tartalmazza.

1. táblázat: A irodalomban fellelt szabadenergia számítások összevetése

	Alkalmazott vízmodell	Szabadenergia-számító módszer	Felszínszámítás	Körülmények T/K, P/atm	ΔA $/\left(\frac{\mathrm{kJ}}{\mathrm{mol}}\right)$	$\Delta z/\text{Å}$
Chandler	SPC/E	Umbrella Sampling	Dinamikus, minden lépésben számol	300, 1	28,68	6,5-7,0
Musolino	SPC/E	String módszer	Statikus, 4 szimulációból átlagolt	298,1	30,96	7,0-8,0
Vacha	SPC/E	Átlagerő integrálás	Statikus, átlagolt	300,1	28,9	7,0-7,5
Dang	Dang-Chang 4-site polarizable model	Átlagerő integrálás	Statikus, átlagolt	298,1	24,69	10-10,5

A fentebb megvizsgált eredmények közül a legújabb a Chandler által végzett munka, ezért úgy gondoltuk, hogy mi is az ő módszerét próbáljuk meg átültetni a BK3 modellre. Ezt az elképzelést erősítette az a tény is, hogy az általunk talált eredmények közül az ő munkájában találhatunk a többitől lényegesen eltérő felszín- és szabadenergia számítást. A következőkben tehát kicsit részletesebben fogom tárgyalni a felszínszámítási módszerét, valamint az Umbrella Samplinget.

2.3 Umbrella Sampling[4,5,22]

Molekuladinamikai szimuláció során legtöbbször egyáltalán nem lehet, vagy nagyon komplikált meghatározni egy reakciókoordináta szabadenergia-függvényét (angol terminológiában Potential of Mean Force, rövidítve PMF). Például a reakciókoordináta mentén esetlegesen megjelenő energiagátak gyakorlatilag lehetetlenné teszik a konfigurációs térben bizonyos állapotok mintavételezését a szimulációs időtartam alatt. Ez a gyakorlatban egy átmeneti állapot esetén valósul meg, amiben a vizsgált molekula olyan ritkán tartózkodik, hogy nem gyűjthető elég adat a szimulációs időtartam alatt ahhoz, hogy a szabadenergiát meghatározzuk. De ez a probléma lép fel akkor is, ha egy bizonyos térkoordináta mentén szeretnénk a szabadenergia függvényt meghatározni, ugyanis a szimuláció időtartama alatt gyakorlatilag nulla a valószínűsége, hogy a kiszemelt részecskénk bejárja azt az utat. A fentebb említett problémák elkerülése végett különböző mintavételezési technikákat

fejlesztettek ki, hogy kiszámíthassuk egy reakciókoordináta mentén a szabadenergiát. Ezen problémák megoldásának egyik lehetősége a Torrie és Valleau[5] által közölt ún. Umbrella Sampling technika. A módszer működésének megértéséhez először definiálnunk kell a szabadenergia-változást egy bizonyos reakciókoordináta mentén:

$$A(\xi) = A^*(\xi) - k \cdot T \cdot \ln\left(\frac{\langle p(\xi)\rangle}{\langle p^*(\xi)\rangle}\right) \tag{1}$$

Ahol $A^*(\xi)$ egy referencia szabadenergia, $\langle p(\xi)\rangle$ pedig a kanonikus sokaság állapotösszege egy reakciókoordináta mentén. A reakciókoordináta függő állapotösszeg egy Boltzmann-faktorral súlyozott átlag, mely a következőképpen néz ki:

$$\langle p(\xi)\rangle = \frac{\int \delta(\xi'(\boldsymbol{R}) - \xi) \cdot e^{-U(\boldsymbol{R})/kt} d\boldsymbol{R}}{\int e^{-U(\boldsymbol{R})/kt} d\boldsymbol{R}} \tag{2}$$

$U(\boldsymbol{R})$ a rendszer teljes energiája az $\boldsymbol{R}$ koordináták függvényében, $\xi'(\boldsymbol{R})$ pedig egy olyan függvény amely a fázistér néhány változójától függ (pl.: egy szög vagy egy távolság), míg a Dirac-delta tulajdonképpen kiválasztja azokat a helyeket és hozzájuk tartozó súlyokat a fázistérben amit a $\xi'(\boldsymbol{R})$ kijelöl. A fentebb említett esetekben azonban bonyolult, néha megvalósíthatatlan a PMF valamint a reakciókoordináta függő állapotösszeg direkt meghatározása a molekuláris dinamikai szimulációban, ezért alkalmazhatjuk többek között az Umbrella Sampling módszert.

Az eljárás során egy bizonyos $V(\xi)$ torzító potenciált használunk, hogy ezáltal az általunk vizsgált rendszerben elősegítsük a minél jobb mintavételezést. A torzító potenciál arra szolgál, hogy a ξ koordinátát egy olyan kisebb intervallumra szűkítsük, ami az adott helyen egy hatékonyabb mintavételezést tesz lehetővé, vagyis növeli egy adott helyen az állapot megvalósulási valószínűségét. Ezt például egy harmonikus potenciál segítségével valósítjuk meg, mely a következő alakot veszi fel:

$$V(\xi) = \frac{\kappa}{2} \cdot (\xi - \xi_0)^2 \tag{3}$$

Ahol ξ_0 a potenciálgödör koordinátája, κ pedig a potenciál meredekségét meghatározó konstans. Egy szimuláció alatt ezzel a módszerrel azonban csak a ξ_0 koordináta egy szűk környezetében tudjuk meghatározni a relatív értékhez viszonyított szabadenergiát, ezért ahhoz, hogy az általunk vizsgált rendszerben a teljes reakciókoordináta mentén meghatározhassuk, bizonyos számú szimulációt kell (ún. umbrella ablakokban) elvégeznünk a ξ_0 és a κ megfelelő megválasztásával, hogy a ξ_0 körüli eloszlások átfedjenek egymással. Ez

történhet a ξ_0 ekvidisztáns változtatásával úgy, hogy a kappát végig azonos értéken tartjuk, de megvalósítható mindkettő változtatásával is az egyes szimulációk között.

Egy szimulációban a torzított potenciál hatására a reakciókoordináta menti állapotösszeg a következőképpen változik meg:

$$\langle p(\xi)\rangle_i = \frac{e^{-V(\xi)/kt}.\langle p(\xi)\rangle}{\langle e^{-V(\xi)/kt}\rangle} \tag{4}$$

Ebből kifejezve az eredeti reakciókoordináta menti állapotösszeget kapjuk, hogy:

$$\langle p(\xi)\rangle = \frac{\langle p(\xi)\rangle_i \langle e^{-V(\xi)/kt}\rangle}{e^{-V(\xi)/kt}} \tag{5}$$

amit behelyettesítve PMF függvénybe a következő egyenletre vezet:

$$A(\xi) = A^*(\xi) - k \cdot T \cdot \ln\left(\frac{\langle \rho(\xi)\rangle_i}{\langle \rho^*(\xi)\rangle}\right) - V(\xi)/kt + F_i \tag{6}$$

Ahol F_i egy ismeretlen szabadenergia konstans, amit a következőképpen definiálunk:

$$e^{-F_i/kt} = \langle e^{-V_i(\xi)/kt}\rangle \tag{7}$$

Abban az esetben, ha sikerült megfelelő átfedést biztosítanunk az ablakok között, meghatározhatjuk az egyes F_i konstansokat és az egyes ablakokhoz tartozó eloszlásokból a felesleges adatokat elhagyva kaphatjuk meg a teljes reakciókoordináta menti szabadenergia-függvényt. Ez azonban nem a legmegfelelőbb módszer az F_i konstansok és a PMF meghatározására, ugyanis a statisztikus hibák elkerülése végett meglehetősen nagy átfedést kell biztosítani az egyes ablakok között, ami nagyszámú adatból álló sort kíván meg, ezzel hosszú szimulációt eredményezve. További problémák merülnek fel akkor, ha a szabadenergia nem egyetlen reakciókoordináta függvénye, hiszen két különböző változó mentén eltérő helyen lehetnek az átfedések, amik más és más konstansok számításához vezethetnek. Az egyes eloszlásokból a szabadenergia függvény meghatározására a fentebb említett problémák miatt különböző módszereket fejlesztettek ki. A következőekben ezek közül tekintem át a WHAM-ot, VFEP-et és MBAR-t.

2.4 Szabadenergia-számító módszerek

Weighted Histogram Analysis Method (WHAM)[14,15]

A WHAM a használt módszerek közül a mai napig is a legelterjedtebb a szabadenergia-függvények kiszámítására több különálló adatsorból. Mint minden módszernél, mely torzító potenciál alkalmazásával előállított eloszlásokból számítja ki a szabadenergiát, a leglényegesebb lépés itt is az, hogy hogyan számoljuk ki torzított eloszlásokból a torzítatlanokat.

Hogy a módszer lényegét megértsük, tekintsük úgy, hogy S darab szimulációt végeztünk, melyben az i-edik szimulációban N_i pontot vettünk fel. (Esetünkben a molekula z koordinátája, de lehetne akár pl. egy szög is molekulakonformáció változás esetében). Ezt követően képezzünk a pontjainkból egy M osztályból álló hisztogramot, ahol minden pontot az M osztály szélességétől függően besorolunk valahová. Legyen p_{ij} az i-edik szimuláció során a j-edik osztály valószínűsége, melyet úgy kapunk meg, hogy az eredeti j-edik osztály valószínűségét megszorozzuk egy súlyfaktorral, ami az umbrella potenciáltól függ:

$$p_{ij} = f_i \cdot c_{ij} \cdot p_j^0 \tag{8}$$

Az egyenletben f_i az i-edik szimulációra vonatkozó normálási konstans, melyre igaz, hogy $f_i^{-1} = \sum_j c_{ij} \cdot p_j^0$, c_{ij} pedig a torzító faktor, amelyre ha $\beta = 1/kT$ fenn áll, hogy

$$c_{ij} = e^{-\beta \cdot V_i(\xi - \xi_0)} \tag{9}$$

A (8) egyenletben tehát tulajdonképpen azt írjuk le, hogy az umbrella potenciál hatása hogyan változtatja meg egy szimulációban, az egy osztályhoz tartozó valószínűséget az eredeti előfordulási valószínűséghez képest és tulajdonképpen megegyezik a (4) egyenlettel. Látható azonban, hogy a normálási konstans függ az i-edik szimulációban az osztályok eredeti megvalósulási valószínűségétől. Ez elsőre problémát jelenthet, de kvalitatívan megvizsgálva logikusnak tűnik, hogy az f_i függ az eredeti valószínűségtől, hiszen a torzító potenciál nélkül is volt valamilyen megvalósulási valószínűség, ami hozzáadódik a torzított potenciállal kapott szimulációból nyert valószínűséghez. A probléma tehát az, hogy a torzító potenciállal kapott eloszlásunkból, hogyan számoljunk vissza, az eredeti diszkretizált torzítatlan eloszlásra, amiből majd a szabadenergia változását kiszámolhatjuk. Innen elindulhatunk úgy, hogy a szimulációból először kapott p_{ij} valószínűségekből kiszámoljuk az egyes p_j^0

valószínűségeket, ebből új f_i -ket számolunk, és ezt folytatjuk iteratívan egy bizonyos konvergenciáig. Ekkor azonban az i-edik szimulációból nyert p_j^0-k csak egy szűk környezetre vonatkoznak, hiszen csak az umbrella ablak szűk környezetéből van adatunk és nem tudjuk, hogy egymáshoz képest hogyan helyezkednek el, milyen súllyal számítanak a teljes eloszlásban. Ezzel szemben, ha az egész reakciókoordináta mentén az S szimuláció összes pontját egyszerre rendezzük j osztályba, valamint a súlyfüggvényeket összeadjuk, akkor a teljes eredeti eloszlásra tudunk következtetni az alábbi egyenletek segítségével:

$$p_j^0 = \frac{\sum_i^S n_{ij}}{\sum_i^S N_i \cdot f_i \cdot c_{ij}} \tag{10}$$

$$f_i^{-1} = \sum_j^M c_{ij} \cdot p_j^0 \tag{11}$$

Ekkor a különböző szimulációból származó eloszlások különböző mértékben fogják meghatározni a teljes eloszlásfüggvényt ezzel pontosabb eloszlást eredményezve. A (10), (11) egyenleteket nevezzük a WHAM egyenleteinek. Ez tulajdonképpen $M + S$ nem lineáris egyenlet, melyet iteratív módszerrel f_i önkényes megválasztásával egy bizonyos konvergenciakritérium megvalósulásáig számolunk.

Ha vesszük az f_i^{-1} természetes alapú logaritmusát a következők szerint:

$$F_i = -\beta^{-1} \cdot \ln(f_i^{-1}) \tag{12}$$

és azt a (10) egyenletbe behelyettesítjük, akkor a következő egyenlethez jutunk:

$$p_j^0 = \frac{\sum_i^S n_{ij}}{\sum_i^S N_i \cdot e^{\beta(F_i - V_i(x))}} \tag{13}$$

Mellyel tulajdonképpen összefüggésbe hozhatjuk a (6) egyenletben tárgyalt F_i konstansokat a WHAM által számolt konstansokhoz. A WHAM tehát ezen F_i konstansok valamint az általuk definiált p_j^0 valószínűségek segítségével határozza meg a teljes szabadenergia profilt.

Variational Free Energy Profile (VFEP)[16]

A VFEP maximum likelihood módszert alkalmaz a teljes szabadenergiafüggvény meghatározására. A maximum likelihood módszer célja, hogy adott mérési mintához egy függvény paraméterezését úgy optimalizáljuk, hogy az a legnagyobb valószínűséggel adja vissza a mért értékeket. A likelihood függvény egy olyan valószínűségi függvény, amely

megadja, hogy az adott valószínűségi sűrűség függvény egy választott paraméterezése mellett mekkora a valószínűsége annak, hogy ez a mért adatsort fogja visszaadni. Esetünkben vegyünk egy olyan eloszlás-függvényt ami megadja, hogy mekkora valószínűséggel található bizonyos x állapotokban a rendszer:

$$p(x) = \frac{e^{-\beta F(x)}}{\int e^{-\beta F(x\prime)} dx\prime} \tag{14}$$

Most vegyük az $F(x|\theta_m)$ függvény egy paraméterezett modelljét, ahol a θ_m az F függvény m-edik paramétere. A valószínűségi eloszlás F függése miatt szintén függeni fog a θ_m paramétertől. Ha vesszük a mintánk során kapott $\{x_n\}$ adatot és a mintavételezett pontok függetlenek egymástól, akkor felírhatjuk, hogy:

$$p(\{x_n\}|\theta_m) = p(x_1|\theta_m) \cdot\ p(x_2|\theta_m) \cdots) \cdot\ p(x_n|\theta_m) \tag{15}$$

Az $F(x|\theta_m)$ függvény likelihood függvénye tehát:

$$L(F\{\theta_m\}|x_1 \ldots . x_n) = \prod_i^n p(x_i|\{\theta_m\}) \tag{16}$$

Mivel szorzat helyett összeggel számolni praktikusan jobb a numerikus stabilitás megőrzése végett, ezért a legtöbbször loglikelihood függvénnyel számolnak, ami a likelihood függvény természetes alapú logaritmusa:

$$l(F\{\theta_m\}|x_1 \ldots . x_n) = \frac{1}{N}\sum_i^N ln\ p(x_i|\{\theta_m\}) \tag{17}$$

A feladat tehát az, hogy megtaláljuk az F azon paraméterezését, amely maximalizálja a likelihood függvényt. Ha egy torzító $V_i(x)$ potenciált használunk az i-edik szimulációban, akkor a valószínűség, hogy egy x pontban találjuk a rendszert:

$$p_i(x) = \frac{e^{-(F(x)+V_i(x))}}{\int e^{-(F(x)+V_i(x))} dx} \tag{18}$$

Ezek szerint az F szabadenergia függvény likelihood-ja S szimuláció esetén:

$$l(F) = \sum_i^S c_i \cdot l_i(\{\theta_m\}|\{x_n^i\}) \tag{19}$$

Ahol c_i a különböző umbrella ablakok súlyfaktorai. Említést érdemel még az a tény, hogy az optimális szabadenergia-függvénynél az összes ablaknak külön-külön is optimuma van. A módszer tehát ezt a függvényt maximalizálja az F paramétereinek változtatásával, ezzel megkapva az optimális szabadenergia-függvényt.

Multistate Bennett Acceptance Ratio (MBAR)

A módszer elmélete összetett, pár oldalban nem vezethető le. A részletes levezetések C. H. Bennett[10] és Shirts M. R., Chodera J. D.[18] munkáiban találhatóak meg. Az elmúlt évben ez a módszer is egy standard kiértékelő eljárássá vált, amit számításaik során Chandlerék is használtak.

2.5 Felszín- és erőszámítás. A Chandler-féle mechanika[3,17]

A fejezetben összefoglalom a Chandler által alkalmazott felszín- és erőszámítást, amit első körben mi is alkalmaztunk a számításaink során.

A Chandler által használt kvadratikus umbrella potenciál, amivel egy adott helyre kényszerítjük a vizsgált részecskét:

$$V_{\text{umbr}}(\boldsymbol{r}^N) = \frac{\kappa}{2} \cdot (z_n - h(x_n, y_n, \boldsymbol{r}^N) - \tilde{\alpha})^2 \tag{20}$$

Ahol, κ konstans, mely a potenciál meredekségét határozza meg, x_n, y_n, z_n a vizsgált részecske koordinátái a térben, $\tilde{\alpha}$ a potenciálminimum koordinátája a térben, $h(x_n, y_n, \boldsymbol{r}^N)$ a felszín z koordinátája a próbamolekulával megegyező x_n és y_n koordinátákon.

Ahhoz, hogy meghatározzuk a potenciál hatására a részekékre ható erőt, a potenciál negatív gradiensét kell vennünk. Mivel a sűrűségfüggvény és a felszín is egyszerre függ a molekulák koordinátájától a sűrűségfüggvényből explicit módon nem kapható meg a felszín koordinátája, ezért a felszín koordinátáját csak implicit alakban tudjuk kifejezni, mely a következőképpen írható fel:

$$\tilde{\rho}(x_n, y_n, h(x_n, y_n; \boldsymbol{r}^N); \boldsymbol{r}^N) = \frac{\rho_l}{2} \tag{21}$$

A felszínt tehát ott definiáljuk, ahol a sűrűségfüggvény értéke megegyezik a tömbfázis ρ_l sűrűségének felével. A következő felmerülő kérdés azonban az, hogy a sűrűségfüggvényt milyen alakban definiáljuk. Ennek egyik legegyszerűbb interpretációja, mellyel a mező egy $\boldsymbol{r}$ pontjának egy adott *t* időpillanatban az értékét meghatározhatjuk, a következő:

$$\rho(\boldsymbol{r}, t) = \sum_i \ \delta(\boldsymbol{r} - \boldsymbol{r}_i(t)) \tag{22}$$

Ez azonban egy komplikált és hirtelen változó mezőhöz és felszínhez vezetne, ezért egy sokkal kényelmesebb módja a sűrűségmező definiálásának, ha a molekulákra Gauss-függvényeket helyezünk és így egy bizonyos távolságban lecsengő Gauss-függvények

értékeinek összegét vesszük egy bizonyos pontban. A Gauss-függvényeket a következőképpen írtuk fel:

$$\varphi = (2\pi\xi^2)^{-d/2} \cdot \exp(-r^2/2\xi^2) \tag{23}$$

Ahol d a Gauss-függvény dimenziója, ξ a szórása. A sűrűségfüggvény tehát egy adott koordinátán és pillanatban a következő összegfüggvénnyel írható fel:

$$\rho(\boldsymbol{r}, t) = \sum_i \ \varphi_i(|\boldsymbol{r} - \boldsymbol{r_i}(t)|, \xi) \tag{24}$$

Ahol φ az i-edik részecske Gauss-függvényének az értéke a számolt **r** koordinátán egy t időpillanatban.

Most deriváljuk le a potenciált az $\boldsymbol{r_i}$ koordináta szerint. Ehhez két részre bontjuk a műveletet. Először a próbamolekulán kívüli összes molekula koordinátája szerint deriváljuk, majd külön az adott molekula koordinátája szerint.

$$-\frac{\partial V}{\partial \boldsymbol{r_i}} = F_i = \frac{\partial V}{\partial h} \cdot \frac{\partial h}{\partial \boldsymbol{r_i}} = -\kappa \cdot \frac{\partial h}{\partial \boldsymbol{r_i}} \cdot (z_n - h - \tilde{a}) \tag{25}$$

Jól látható, hogy a derivált függvény tartalmazza a felszín deriváltját, amit viszont csak az implicit sűrűségfüggvényből lehet kifejezni. Implicit függvény esetén a deriváltfüggvény általánosan így néz ki:

$$y' = \frac{dy}{dx} = -\frac{\frac{\partial f(x,y)}{\partial x}}{\frac{\partial f(x,y)}{\partial y}} \tag{26}$$

Ezek alapján a felszín térkoordináták szerinti deriváltjai:

$$\frac{\partial h}{\partial \boldsymbol{r_i}} = -\frac{\partial \rho}{\partial \boldsymbol{r_i}} \Big/ \frac{\partial \rho}{\partial h} \tag{27}$$

Mivel a h egy t időpillanatban vett molekulaelrendeződésnél a felszín z koordinátáját jelöli, ezért az egyenlet a következőképpen írható át:

$$\frac{\partial h}{\partial \boldsymbol{r_i}} = -\frac{\partial \rho}{\partial \boldsymbol{r_i}} \Big/ \frac{\partial \rho}{\partial z} \tag{28}$$

Az egyenletet kifejtve a különböző részecskék koordinátáira külön-külön:

$$\frac{\partial h}{\partial x_i} = -\frac{\partial \rho}{\partial x_i} \Big/ \frac{\partial \rho}{\partial z} \tag{29}$$

$$\frac{\partial h}{\partial y_i} = -\frac{\partial \rho}{\partial y_i} \Big/ \frac{\partial \rho}{\partial z} \tag{30}$$

$$\frac{\partial h}{\partial z_i} = -\frac{\partial \rho}{\partial z_i} \Big/ \frac{\partial \rho}{\partial z} \tag{31}$$

Míg az n-edik részecskére vonatkoztatva az egyenletek:

$$\frac{\partial h}{\partial x_n} = -\frac{\partial \rho}{\partial x_n} \Big/ \frac{\partial \rho}{\partial z} \tag{32}$$

$$\frac{\partial h}{\partial y_n} = -\frac{\partial \rho}{\partial y_n} \Big/ \frac{\partial \rho}{\partial z} \tag{33}$$

$$\frac{\partial h}{\partial z_n} = -\frac{\partial \rho}{\partial z_n} \Big/ \frac{\partial \rho}{\partial z} = 0 \tag{34}$$

A h független a z_n változásától, ugyanis: $z_n = z$, mivel a felszínt az n-dik részecske $x_n,\ y_n$ pontjában keressük. Persze az utolsó 3 egyenletnek csak akkor van jelentősége, ha a felszín meghatározásakor a próbamolekula térkoordinátáit is belevesszük a számításba.

Most már tehát csak a sűrűségfüggvényt kell deriválnunk a részecskék koordinátái, valamint a felszín z koordinátája szerint.

A tömbfázisban lévő vízmolekulák térkoordinátái szerinti derivált:

$$\frac{\partial \rho}{\partial z_i} = \frac{(2\pi\xi^2)^{-d/2} \cdot exp\left(-\frac{(x-x_i)^2+(y-y_i)^2+(z-z_i)^2}{2\xi^2}\right)}{\partial z_i} =$$
$$-(2\pi\xi^2)^{-d/2} \cdot exp\left(-\frac{(x-x_i)^2+(y-y_i)^2+(z-z_i)^2}{2\xi^2}\right) \cdot \frac{(z-z_i)}{\xi^2} \tag{35}$$

$$\frac{\partial \rho}{\partial x_i} = \frac{(2\pi\xi^2)^{-d/2} \cdot exp\left(-\frac{(x-x_i)^2+(y-y_i)^2+(z-z_i)^2}{2\xi^2}\right)}{\partial x_i} =$$
$$-(2\pi\xi^2)^{-d/2} \cdot exp\left(-\frac{(x-x_i)^2+(y-y_i)^2+(z-z_i)^2}{2\xi^2}\right) \cdot \frac{(x-x_i)}{\xi^2} \tag{36}$$

$$\frac{\partial \rho}{\partial y_i} = \frac{(2\pi\xi^2)^{-d/2} \cdot exp\left(-\frac{(x-x_i)^2+(y-y_i)^2+(z-z_i)^2}{2\xi^2}\right)}{\partial y_i} =$$
$$-(2\pi\xi^2)^{-d/2} \cdot exp\left(-\frac{(x-x_i)^2+(y-y_i)^2+(z-z_i)^2}{2\xi^2}\right) \cdot \frac{(y-y_i)}{\xi^2} \tag{37}$$

Míg a felszín z koordinátája szerinti derivált:

$$\frac{\partial \rho}{\partial z} = \sum_{i \neq n} \frac{(2\pi\xi^2)^{-\frac{d}{2}} \cdot exp\left(-\frac{(x-x_i)^2+(y-y_i)^2+(z-z_i)^2}{2\xi^2}\right)}{\partial z} = \sum_{i \neq n} (2\pi\xi^2)^{-d/2} \cdot exp\left(-\frac{(x-x_i)^2+(y-y_i)^2+(z-z_i)^2}{2\xi^2}\right) \cdot \frac{(z-z_i)}{\xi^2} \quad (38)$$

Ezzel a levezetéssel gyakorlatilag megkaptuk, hogy a sűrűségfüggvény implicit volta miatt milyen erők hatnak a tömbfázist alkotó molekulákra. Abban az esetben, ha a felszínszámításba a próbamolekulát nem vesszük bele, a molekulára a következő z irányú erő hat:

$$-\frac{\partial V}{\partial z_n} = F_{z_n} = -\kappa \cdot (z_n - h - \tilde{a}) \quad (39)$$

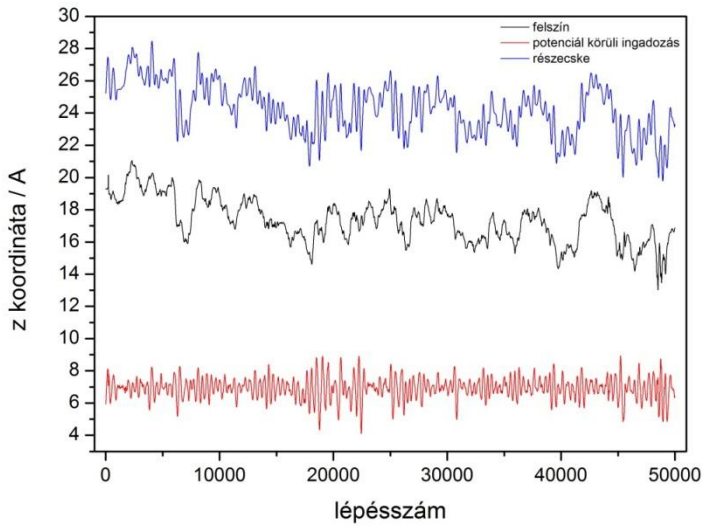
Ez tehát lehetővé teszi, hogy kiszámoljuk az n-edik részecskénkre a potenciál által kifejtett erő hatását, valamint a potenciál által a tömbfázis többi részecskéjére ható erőket. Felmerülhet azonban a kérdés, hogy vajon érdemes-e beleszámolni a próbarészecskét a felszínszámításba. Amíg a részecske a felszíntől távol helyezkedik el ez probléma elhanyagolható, itt ugyanis nem fogja befolyásolni a felszín helyzetét, mivel a rá illesztett Gauss-függvény ebben a távolságban már lecseng. A felszínhez közel azonban a hatása egyre nagyobb lesz, ami erősen befolyásolhatja a felszín aktuális helyzetét így nem biztos, hogy célszerű elhanyagolni a felszínszámításnál.

3. A szimuláció kivitelezése

Munkám során az első feladatom tehát az volt, hogy a BK3 modell Fortranban megírt forráskódjához külön rutinban hozzáírjam az Umbrella Sampling módszert, beleértve a felületdefiníciót és a fellépő erőket a Chandler-féle modellben. Ehhez legelőször definiálni kell azt a molekulát, melyet egy szimuláció során követni akarunk. Ezt úgy oldottuk meg, hogy a torzító potenciál és a felszín z koordinátájának összegét felülről leginkább megközelítő z koordinátájú pontot keressük meg, minden egyes szimuláció elején.

A molekula megtalálása után a következő feladat a molekulával azonos x és y koordinátájú helyen a felszín z koordinátájának megtalálása, hiszen a potenciált és ezáltal a részecskékre ható erőket a felszín koordinátáján keresztül definiáltuk. Mivel Chandler munkáját követtük, az általa használt felszíndefiníciót vettük mi is alapul és úgy döntöttünk, hogy számításaink során figyelembe fogjuk venni a felszínszámításnál a próbarészecskénk helyzetét is. A konkrét számításhoz először meg kellett keresni a sűrűségmező azon pontját, ami pontosan a fele a tömbfázis sűrűségének. Vagyis tulajdonképpen az a pont, ahol a próbamolekulán átmenő, a felszín síkjára merőleges egyenes metszi a felszínt. A felszínkereséshez több különböző numerikus gyökkereső algoritmus használható. Chandler munkájában a Newton-Raphson módszert használta, mi azonban a biztosabb konvergencia és a függvények bonyolultsága miatt egyszerűbben beprogramozható intervallumfelezésen alapuló numerikus gyökkereső algoritmus mellett döntöttünk. Az intervallum két szélét úgy választottuk meg, hogy az egyik a gázfázis, míg a másik a tömbfázis közepén helyezkedjen el, ezzel biztosítva, hogy a függvény áthaladjon a nulla ponton. Azonban az elméletet a gyakorlatba ültetve, a beprogramozást követően a felszínszámolás nem ment olyan gördülékenyen, mint azt feltételezni lehetett. A hosszabb futtatások során a felszínszámításnál egy idő után leállt a program. Először a numerikus gyökkereső módszerben láttuk a hibát, melyet úgy próbáltunk meg orvosolni, hogy szelőmódszerrel is kiszámítottuk a felszíneket, azonban így is leállt a program, ezért úgy gondoltuk, hogy a próbamolekula alatti felszínkoordináta helyett felszínátlagot fogunk számolni. Ehhez a kiválasztott molekula második koordinációs szféráján, vagyis 5,6 Å sugarú körön belül, száz véletlenszerű x és y koordinátájú pontban számoltunk ki felszínkoordinátákat és ennek az átlagával számoltunk tovább, viszont a program sok esetben így is leállt. Ennek okát a Chandler-féle számítási módszer mechanikájában találtuk. Mivel a potenciál definíciójában a felszín dinamikus megváltozása miatt minden részecske szerepet játszik, az erő kiszámolásában megjelenő $\frac{\partial h}{\partial \boldsymbol{r_i}}$ tag miatt

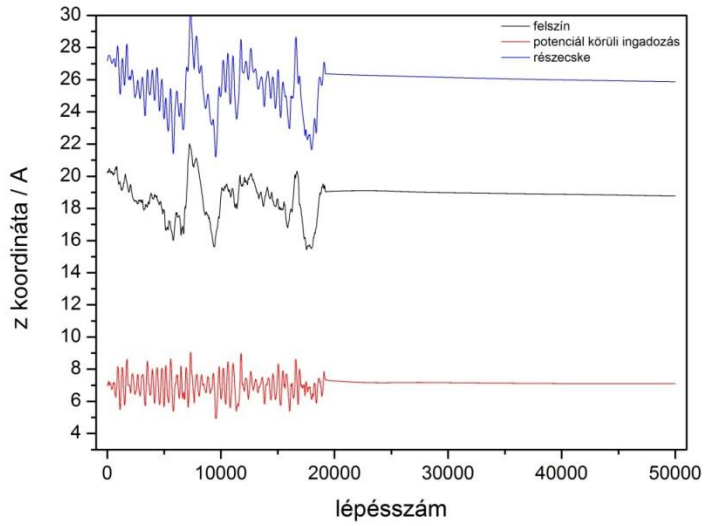
kvalitatív megfontolások alapján kijelenthető, hogy az egyes molekulákra ható erő arányos az ezen molekula elmozdulása esetén fellépő felszínváltozással. Ebből viszont az következik, hogy a próbarészecskénkre ható erő ellenereje nem a teljes fázison hat, hanem a próbarészecske alatti felszínt alkotó néhány részecskén, ami tulajdonképpen egy „lyukat" okoz a tömbfázisban, melynek következében a szimuláció megbízhatatlanná válik (6. ábra).



6. ábra: A Chandler-féle számítási módszer szerinti felszín- és részecske z koordináta valamint a potenciál körüli mozgás a szimuláció során.

Jól látható, hogy a szimuláció előrehaladtával a felszín koordinátája csökken a próbarészecske környezetében, és mivel a potenciál a felszínhez képest van definiálva a részecske z koordinátájának a csökkenését is magával vonja.

A tömbfázisban kialakult „lyuk" másik következménye, hogy így a részecskénk x, y irányú elmozdulásakor $\frac{\partial h}{\partial x_{próba}}$ és $\frac{\partial h}{\partial y_{próba}}$ értékek nagyon megnőnek, ezzel felgyorsítva a próbarészecske x és y irányú mozgását. A részecske nagy sebességének hatására megnövekszik a rendszer hőmérséklete, aminek következtében a termosztálás során a tömbfázist alkotó többi molekula sebessége gyakorlatilag nullára csökken ezzel „megfagyasztva" a rendszert. Ezt a 7. ábra szemlélteti.



7. ábra. Az x, y irányban felgyorsult próbarészecske miatti „befagyás”

A fentebb említett problémák tehát sok esetben a szimuláció leállását eredményezték, ezért a Chandler-féle mechanika helyett más módszerhez folyamodtunk. A felszínszámítás során meghagytuk az általa használt Gauss-függvényekkel definiált, sűrűségszámításon alapuló meghatározást, de a részecskékre ható erők számítását elvetettük. Helyette a newtoni mechanika következtében a próbarészecskére ható erő ellenerejét a tömbfázis összes részecskéje között egyenletesen osztottuk szét, ezzel kiküszöbölve a futás közbeni „lefagyást”.

Kanonikus sokaságon végzett szimuláció során a hőmérséklet állandó értéken való tartásához termosztálásra van szükség. A BK3 modell programja ezt Nose-Hoover termosztáttal valósítja meg, ami a részecskék sebességeinek skálázásán keresztül állítja be a hőmérsékletet. A tömbfázisban lévő molekuláink ütköznek egymással, ezáltal sebesség és energiaátadást valósítva meg egymás között, így gyakorlatilag termosztálódnak. Viszont a kint tartott részecske nem ütközik, ezért nincs energiaátmenete, vagyis annak hőmérsékletét önállóan kell szabályozni. Ezért a tér azon részén, ahol még nem ütközik a felszíni részecskékkel a próbamolekulánk, azt az egyszerű megoldást választottuk, hogy mintavételezést végzünk a megfelelő kanonikus sokaságból. Ezt úgy valósítottuk meg, hogy bizonyos lépésközönként a tömbfázisból véletlenszerűen kiválasztott molekula egyik rotációs vagy transzlációs szabadsági fokának megfelelő sebességet adtunk a próbamolekulánknak. Ez átlagosan 250-400 K között tartotta a részecskénket a tömbfázistól távol futtatott szimulációk során.

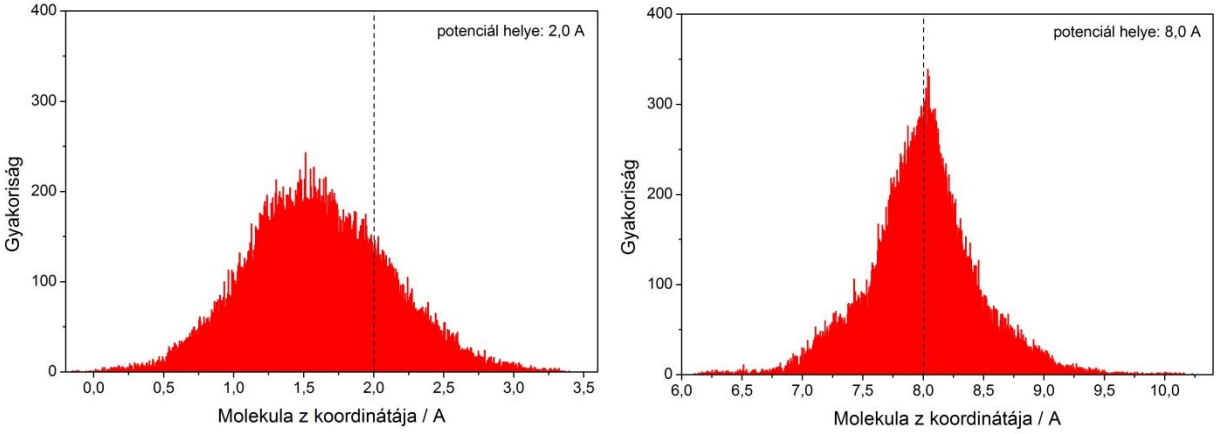
A szimulációk során 1000 db molekulát tartalmazó $28{,}0 \times 28{,}0 \times 120{,}5\ \text{Å}^3$ térfogatú szimulációs cellát használtunk és periodikus peremfeltételt alkalmaztunk, tehát a cellát saját eltoltjaival vettük körbe. A cella a 8. ábrán látható:

8. ábra: A szimulációs cella

A szimulációt NVT sokaságon végeztük, 1 atm nyomáson és 298 K hőmérsékleten. Minden szimuláció elején a mintagyűjtés kezdete előtt egy 5000 időlépésből álló termosztálást végeztünk. Egy időlépés 2 fs-nak felel meg, tehát a termosztálás teljes hossza 10 ps volt. Egy számítást 50000 időlépésben maximalizáltunk, ami 0,1 ns-nak felel meg. A potenciál meredekségét beállító κ konstans értékét 10 kJ/mol-ra állítottuk az egyes futtatások kezdetén.

A szimuláció során csak a futás elején bizonyos módszer szerint kiválasztott molekula paramétereit kellett tárolni, ezért legegyszerűbbnek a fájlba való kiíratás tűnt. A szabadenergia-profil pontjait a felszíntől való távolságuk alapján akartuk felvenni, ezért minden lépésben kiírattuk a $z_n - h$ távolságot, ahol a h a felszín aktuális z koordinátáját jelöli. A szimuláció követéséhez és esetleges ellenőrzéshez jónak láttuk néhány további érték tárolását, mint a molekula z koordinátája és a felszín koordinátája külön-külön valamint a részecske hőmérséklete és sebessége minden időlépésben. A molekula gyorsabb potenciálgödörbe való kényszerítése végett a következő szimulációt az előző szimuláció utolsó lépésének molekulakonfigurációjából indítottuk, ehhez minden szimuláció végén elmentettük a konfigurációt és egy script segítségével írtuk át inputkonfigurációvá. A szimuláció két paraméterét, a potenciálgödör helyzetét valamint a κ értékét szintén script segítségével változtattuk meg.

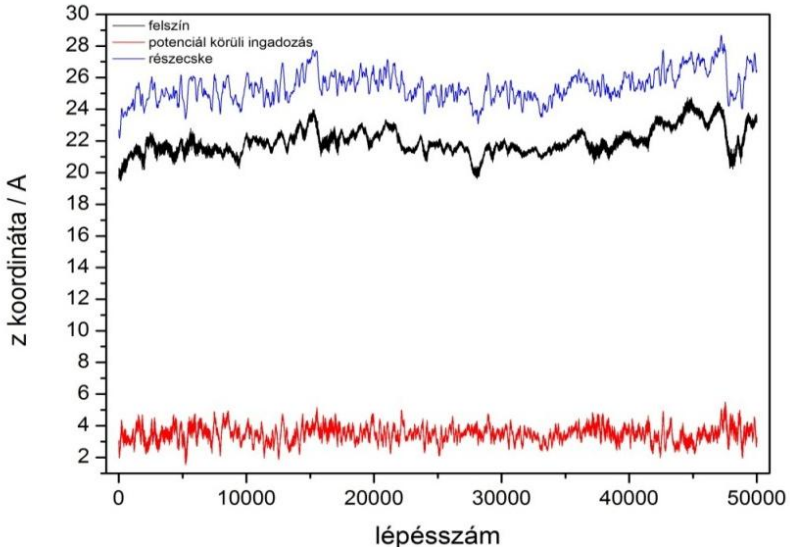
A kiértékeléshez a próbamolekula potenciál körüli mozgására van szükségünk. A felszíntől távolabb, ahol gyakorlatilag még nincs kölcsönhatás a mi molekulánk és a tömbfázis többi molekulája között, a potenciál körüli eloszlásnak közel szimmetrikusnak kell lennie, ami összhangban van az általunk kapott eredménnyel is (9-10. ábra).



9-10. ábra: A kiválasztott molekula *z* koordinátájának eloszlása a felszíntől 2,0 Å-re és 8,0 Å-re elhelyezett potenciál körül

A jobb oldali ábra alapján megállapítható, hogy a felszíntől távoli potenciálkoordinátán szimmetrikus eloszlást követ a próbarészecske *z* koordinátájának eloszlása. A bal oldali ábrán, mely a 2,0 Å-re elhelyezett potenciál körüli eloszlást ábrázolja jól látható, hogy a felszínhez közelebbi állapotok sokkal gyakoribbak, ami a tömbfázist alkotó részecskékkel való kölcsönhatásnak a bizonyítéka.

A szimulációk során az általunk bevezetett erő-, valamint a Chandler-féle felszínszámítás alkalmazásával kapott részecskemozgást a következő ábra szemlélteti:



11. ábra: A kijelölt részecskénk z koordinátája, a potenciál körüli váltakozása valamint a felszín z koordinátája

A kijelölt részecskére ható erő egyenletes elosztása az egész tömbfázison egy egyenletesebb felszínmozgást eredményezett, ezáltal stabilabbá téve a felszínszámítást is. Azt a problémát, hogy a kijelölt részecskét használjuk-e a felszínszámításhoz, azzal az egyszerű megoldással orvosoltuk, hogy a felszínhez közeli koordinátákon a részecske *z* irányban felszín vetített helyzetétől egy 2,5 Å sugarú körön belül egyáltalán nem vettünk ki felszínkoordinátát. Ez azt eredményezi, hogy ezekben az eseteken a felszín meghatározásához szükséges 100 pontot a részecske koordinátájától 2,5-5,6 sugarú tartományból mintavételezzük, ezzel tárgytalanná téve a részecske ráhatásának kérdését.

4. Eredmények

A szimuláció során előállított adatsorokból három, az elméleti bevezetőben tárgyalt módszerrel számítottuk ki a szabadenergia-profilt, következőkben ezen három módszerrel kapott eredményeket értékeljük ki. A szakdolgozat témáján túl mutat az egyes módszerek saját kódolása ezért általában a módszerek szerzői által az interneten ingyenesen közzétett programokat használtuk a kiértékeléshez[19,20,21]. Az egyes szimulációkat a következő potenciálhelyekről vettük fel:

-9,5 – 7,0 Å intervallumban 0,5 Å lépésekkel,

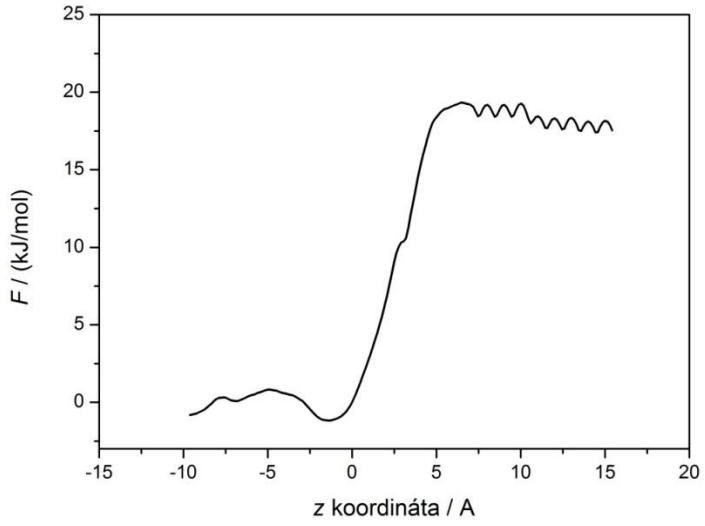
7,0 – 15,0 Å tartományban 1,0 Å lépésekkel.

Az eredmények kiértékelése során a szabadenergia-profil alakjának, valamint a tömb- és gázfázis szabadenergia különbségének meghatározásán túl célul tűztük ki az egyes szabadenergia-számító módszerek összehasonlítását is. Mivel az irodalomban az egyes módszerek tárgyalásánál az egyik legfontosabb különbségnek azt veszik, hogy mekkora adatmennyiség és átfedés szükséges a pontos számításhoz, ezért az összehasonlításnál erre is kitérünk.

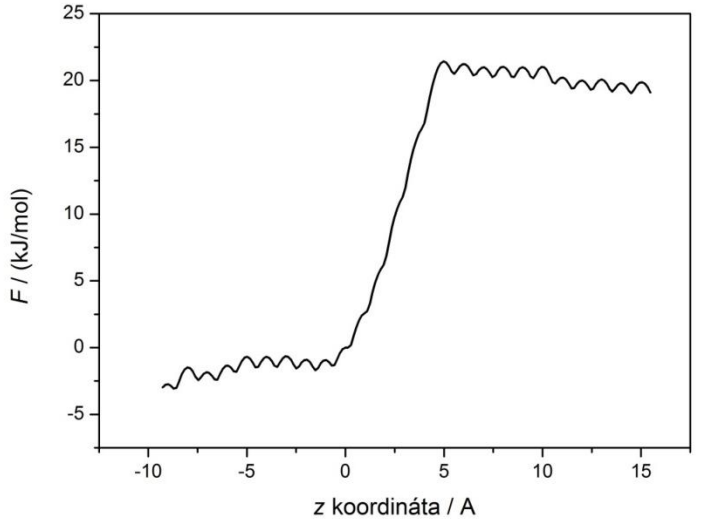
4.1 WHAM-mal számolt profilok

A WHAM a legrégebbi a használt módszerek közül, a legtöbb szabadenergia-számítást ma is ezzel a megoldással végzik. Mivel az adatok előállítása során még nem tudtuk, hogy mely módszerek programjaihoz sikerül hozzájutnunk, nem tudtuk, hogy az adatokat milyen formátumban tároljuk, ezért a legegyszerűbb megoldást az ASCII-fájlba való soronkénti tárolást választottuk. A program inputjához egy metaadat-fájl szükséges, mely soronként 3 adatot tartalmaz, az egyes szimulációk adatsorának elérési útját, a potenciál helyét és a potenciál konstansának értékét, amin a szimulációkat végeztük. Az egyes szimulációk ASCII fájljainak azonban két oszlopot kell tartalmaznia. Az első oszlopban a szimulációs lépés a másodikban a részecskénk adott időpillanatban vett z koordinátája, melynek mentén a szabadenergiát számoljuk ki. Ezeknek a formázott fájloknak az előállításához külön scriptet írtunk. A formázott fájlok és a metaadat fájl előállítása, valamint a program lefutása után egy eredmény fájl jön létre, melyben paraméterként változtatható számú pontban tárolódnak a részecske z koordinátái és a hozzá tartozó szabadenergiák. A WHAM által kapott eredményeket a következő két grafikon szemlélteti, melyek közül az elsőn az összes adatból

számolt profilt láthatjuk, míg a másodikon az egész Angström helyzetű potenciálokkal rendelkező szimulációk adataival számoltunk.



12. ábra: WHAM-mal számolt szabadenergia-profil minden adatsor felhasználása esetén



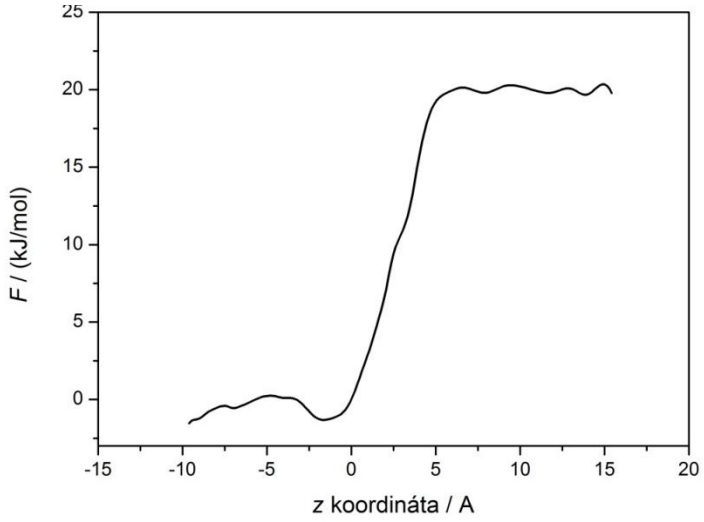
13. ábra: 1,0 Å lépésközzel rendelkező szimulációkból számolt WHAM szabadenergia profil

A szabadenergia pontok számítása során a 0 Å-höz tartozó pontot vettük a számítás referenciapontjaként. A görbén (12. ábra) a fenti platón látható hullámzás valószínűleg azt a bizonytalanságot jelzi, ami a nem megfelelő átfedés következtében valósulhatott meg, ugyanis ebben az esetben már 1,0 Å a távolság a két potenciál között. Látható, hogy már a 0,5 Å helyetti 1,0 Å-ös potenciálváltoztatás is nagy bizonytalanságot visz a számolásba, melyet az

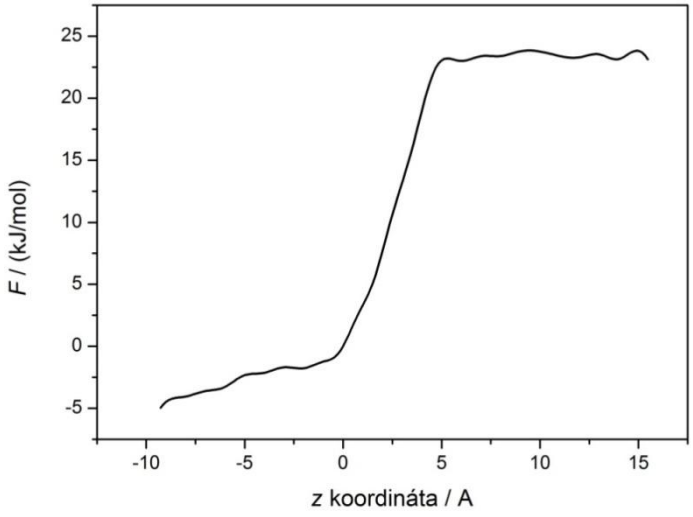
alsó plató hullámzása bizonyít (13. ábra). A meredeken emelkedő részben is megjelennek az egyenetlenségek, ami szintén a ritkításnak köszönhető.

4.2 VFEP-el számolt profilok

A WHAM-hoz hasonlóan ebben az esetben is a szerzők által az interneten közzétett, ingyenesen elérhető szoftvert használtuk a szabadenergia-számításhoz. Mivel ugyanazt az inputot kellett megadni, mint a WHAM esetében is, ezért további változtatásokat nem végeztünk, a kiértékelés a korábbiakban ismertetett szempontok szerint zajlott, az eredményt a 14. és 15. ábra szemlélteti.



14. ábra: VFEP módszerrel számolt szabadenergia-profil minden adatsort felhasználva

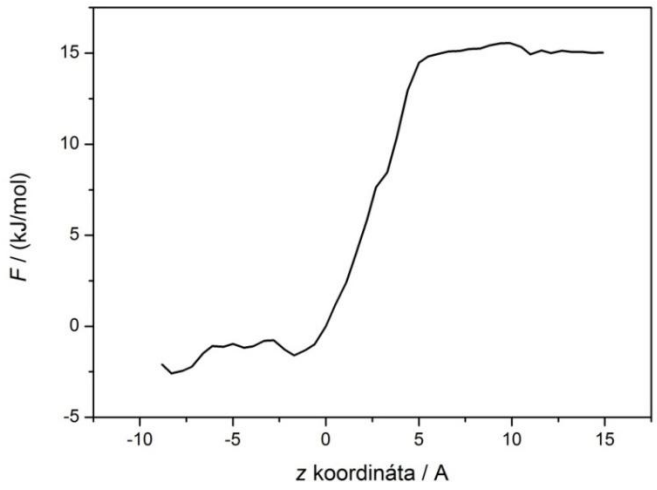


15. ábra: 1,0 Å lépésközzel rendelkező szimulációkból számolt VFEP szabadenergia-profil

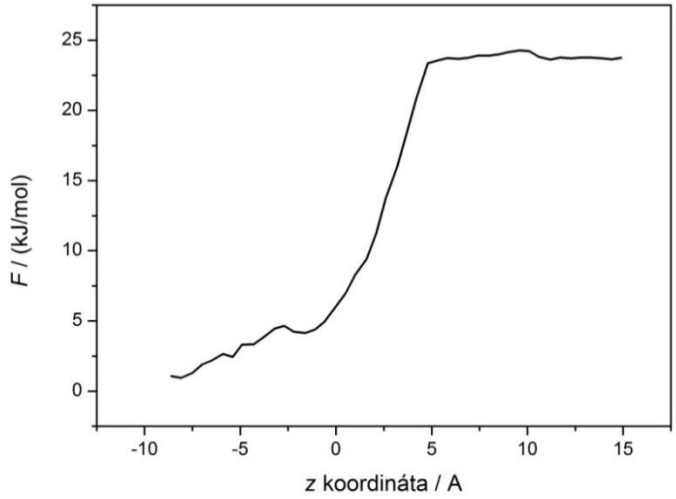
A WHAM-mal ellentétben itt nem változtat annyit a profilon a potenciálok számának ritkítása a z koordináta mentén. Az egyedüli változás az összes adatsorral számolt profilhoz képest az, hogy nem figyelhető meg alsó plató, ehelyett egy kisebb meredekségű csökkenést tapasztalhatunk a tömbfázisban haladva.

4.3 MBAR-ral számolt profilok

Ez esetben egy Pythonban írt forrásfájl állt csak rendelkezésünkre az előző módszerekkel szembeni kész programok helyett. Így a program telepítése valamint futtatása nagyobb problémát okozott, ugyanis a számítási módszer miatt a program inputja mást vár, mint az előző két esetben. A probléma megoldásában az segített, hogy egy szabadon elérhető mások által írt Python scriptet használva a WHAM-ra készített input segítségével tudtuk a számításokat elvégezni. A következő ábrán az ezen módszer segítségével számolt szabadenergia-profilt láthatjuk az összes szimuláció adatsorából számolva:



16. ábra MBAR-ral számolt szabadenergia profil

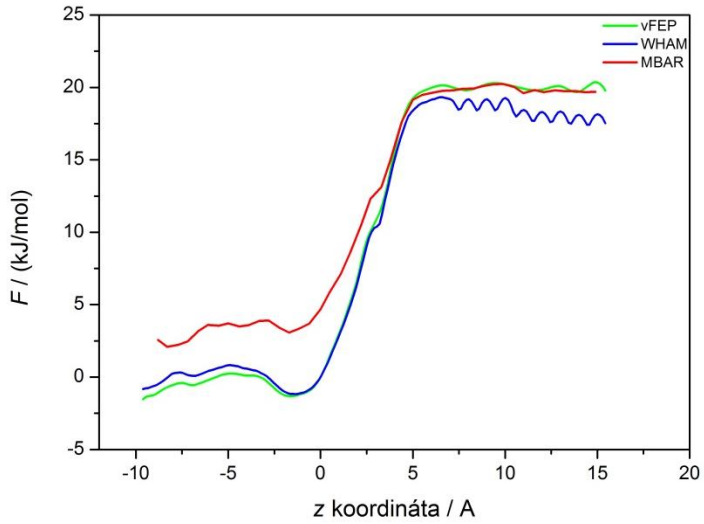


17. ábra MBAR-ral számolt szabadenergia profil ritkított potenciállal.

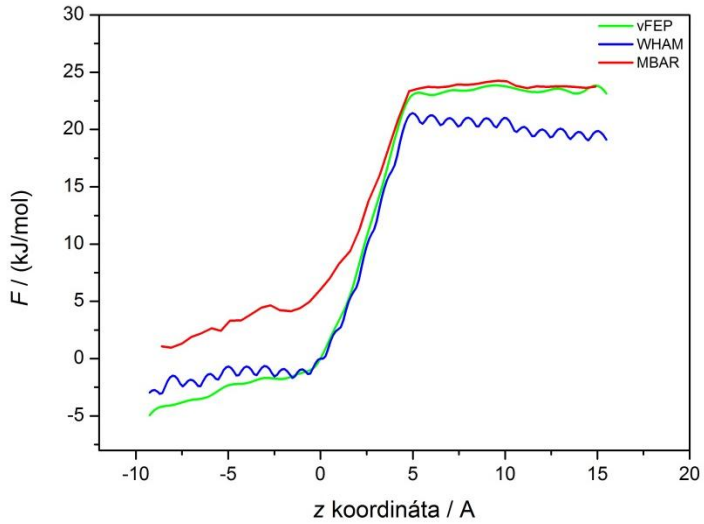
A VFEP módszerrel számított profilhoz hasonlóan itt sem figyelhető meg a WHAM-nál látott, a nem megfelelő átfedésre utaló hullámzás. Az egyedüli változás itt is az alsó plató helyetti csökkenő rész. Azonban ez némileg meredekebb, mint a VFEP esetében.

4.4 Az egyes módszerekkel kapott eredmények összehasonlítása

Az összes módszerrel számolt energiaprofilokat a minden adatsorral, valamint a csak 1,0 Å-önként elhelyezett potenciálok esetén számolt profilok a következő két ábrán láthatóak:



18. ábra: A három különböző módszerrel számolt profil összehasonlítása



19. ábra: A három különböző módszerrel, ritkított adatsorból számolt profil összehasonlítása

A profilokat összehasonlítva az első szembetűnő különbség a három különböző módszer között, hogy az MBAR-ral számolt profilban a két fázis közötti szabadenergia-különbség kisebb, mint a másik két módszer esetén. A VFEP és a WHAM meglehetősen hasonló profilt ad, az egyedüli eltérés a fenti plató bizonytalansága a WHAM esetén. Minden módszerben közös, hogy az alsó platón a szabadenergiának a felszín környékén minimuma van, majd a tömbfázisban beljebb haladva egy magasabb állapotba kerül, bár az MBAR esetében ez egy jóval kisebb minimum, mint ami a másik kettőben látható. A profil negatív irányú csökkenését a tömbfázisban az eredményezi, hogy azokban a pontokban már nem történik szimuláció, ezért a módszernek nem áll rendelkezésére megfelelő számú adat a szabadenergia-számítás elvégzéséhez. A szimuláció számbeli eredményeit a 2. táblázat foglalja össze.

2. táblázat: összefoglaló táblázat az általunk, különböző módszerekkel számolt, valamint az irodalomban talált szabadenergia-különbségekről

	$F_{\text{alsó}}$ / (kJ/mol)	$F_{\text{felső}}$ / (kJ/mol)	ΔF / (kJ/mol)	Δz/Å
VFEP	-0,52	19,98	20,50	6,4
WHAM	-0,03	18,47	18,50	6,4
MBAR	2,57	19,83	15,15	6,7
Chandler	–	–	28,68	6,5-7,0
Musolino	–	–	30,96	7,0-8,0
Vacha	–	–	28,90	7,0-7,5
Dang	–	–	24,69	10-10,5
Kísérlet	–	–	26,36	–

A táblázatban az $F_{\text{alsó}}$ és $F_{\text{felső}}$ a két plató szabadenergiájának átlagát fejezi ki, míg ΔF a köztük lévő szabadenergia-különbség, Δz pedig a meredeken növekvő, rész szélessége. A ritkított potenciállal számolt profilokból – mivel nem tartalmaznak alsó platót – nem számoltunk szabadenergia különbséget, bár érdemes megjegyezni, hogy a VFEP esetében például jobban közelít a tapasztalati értékhez a különbség, hogyha a plató helyett a töréspontot vesszük referenciának.

Az eredményeinket figyelembe véve megállapítható, hogy a BK3 modellel végzett számítások is jó közelítéssel visszaadják az irodalomban fellelhető eredményeket. A profilok alakja minden esetben hasonló, a kölcsönhatás hosszát meghatározó meredek rész is hasonló távon cseng le az irodalmi áttekintésben tárgyaltakkal. Az alsó plató fluktuációjának okát a tömbfázis réteges szerkezetében kereshetjük, a rendezett szerkezetben ugyanis a molekula

több hidrogénkötést tud kialakítani, ennek következtében azok a molekulák melyek a rétegek között helyezkednek el magasabb szabadenergia-állapotban vannak. A szabadenergia-különbség a két fázis között azonban kisebb a mások által szimulációval számolttól, valamint a kísérletileg megmérttől is. Ennek lehetséges oka a modellben való eltérésen (a paraméterezésénél ugyanis nem az volt az elsődleges cél, hogy a párolgási szabadenergia-különbséget megfelelően visszaadja, ellentétben az SPC/E modellel, ahol a párolgáshőre is illesztettek) túl az lehet, hogy a részecskénk – különösen a felszíntől távoli koordinátákon – nem a megfelelő hőmérsékleten van a termosztálás hiányosságai miatt. Ezen probléma természetesen az irodalomban tárgyalt mások által végzett számításoknál is felléphetett, de a szimuláció során végzett hőmérsékletkövetésre utaló információt másoknál nem találtunk. Külön kérdés, hogy a részecske termosztálását egységesen, vagy koordinátánként külön valósítsuk meg, ugyanis a kirepülő részecske z koordinátához tartozó hőmérséklete a magasabb sebesség miatt nagyobb. Kísérletileg igazolták ugyanis, hogy a felszín fölötti részecskék magasabb hőmérsékletűek, a szimulációk alapján feltehetően ez a v_z koordinátára vonatkozik.[2,3]

A felület definíciója szintén meghatározhatja a szimulációs eredményeinket. Mi azért használtuk a Chandler által definiált Gauss-függvényekkel számolt sűrűségmezőt, mert egy folytonos, kevéssé változó mezőt biztosít. A Chandler által használt mechanikát azonban el kellett vetnünk. Az egyenletes ellenerő elosztást azért használtuk, mert a vízmodellünket és a tömbfázist ez torzítja a legkevésbé. A felület meghatározásához természetesen más módszer is használható, például az ún. ITIM-módszer.[23]

A szimulációink NVT szimulációban történtek, melyben lehetséges volt implementálni az Umbrella Sampling módszert, úgy, hogy a reakciókoordináta a felülettől számított távolság. Ha a nyomás értékét dinamikusan, pl. NPT szimuláció segítségével próbálnánk beállítani, akkor az adott hőmérsékleten többnyire csak a stabilis folyadék fázis létezne a megfelelő térfogattal, ahol nem lenne hely a szimulációs cellában a reakciókoordináta és a felület definiálásához. Ennek megfelelően nem tudtuk értelmezni az irodalmi előzmények esetén az NPT szimulációra való utalást.

Szakdolgozati összefoglaló

Párolgás szabadenergiájának meghatározása súlyozott mintavételezéssel számítógépes szimulációban

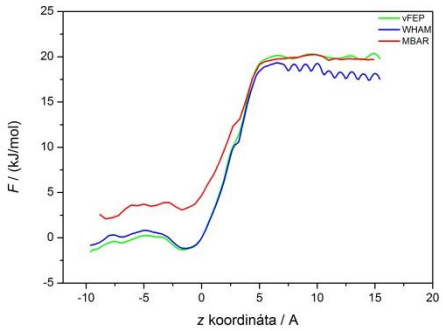
Kohut Gergely, kémia alapszakos hallgató

ELTE TTK Kémiai Intézet, Fizikai Kémiai Tanszék

Témavezető: **Dr. Tóth Gergely** egyetemi docens
ELTE Fizikai Kémiai Tanszék

Munkánk során molekuláris dinamikai szimuláció segítségével BK3[1] vízmodellen egy molekula gázfázisból tömbfázisba történő kondenzációjakor bekövetkező szabadenergia változásának, valamint a szabadenergia profiljának meghatározását tűztük ki célul. A szimulációs adatsorokat Umbrella Sampling módszerrel állítottuk elő, a felszín meghatározásához a Chandler-féle eljárást[2] vettük alapul. Az adatsorokból a kiértékelést három különböző szabadenergia-számítására alkalmas módszerrel végeztük el, melyek a WHAM az MBAR és a VFEP voltak. A molekuláris dinamikai szimulációt 1000 molekulával NVT sokaságot alkalmazva futtattuk. A kutatás kapcsolódik a frissen fejlesztett BK3 modell teljesítőképességének vizsgálatához és háttérinformációkat nyúlt egy folyamatban lévő másik kutatáshoz ahol párolgási trajektóriákat vizsgálunk.

A szimulációs adatsorokat a különböző metódusokkal kiértékelve az eredményül kapott szabadenergia-profilok az alábbi ábrán láthatóak:



1. ábra: A három különböző módsszerrel számolt profil összehasonlítása

A szabadenergia-profil a várakozásoknak megfelelően megegyezett az irodalomban fellelhetőekkel, a párolgásnál fontos kis koordinációjú szerkezeteket jól leíró BK3 modell esetében sincs a profilnak szabadenergia-gátja. A két fázis közötti átmentet energiakülönbsége kisebb, mint az irodalomban mások által számolt, illetve a kísérletileg meghatározott érték. Ennek okát kereshetjük a modell sajátosságaiban is, de két szimulációs részlettel – a hőmérséklet kérdésével, valamint a nyomás értékével, aminek beállítása pl. NPT szimulációban nem ötvözhető az Umbrella Sampling módszerrel – is magyarázható. A molekula hőmérsékletének beállítása külön kérdés, ugyanis a felszín feletti részecske z irányú sebessége kísérleti és szimulációs adatok alapján nagyobb az átlagnál. Ezeknek a figyelembe vétele a szabadenergia számításoknál még további vizsgálatokat igényel.

A módszereket összehasonlítva megállapítható, hogy legjobban a VFEP (20,50 kJ/mol) közelíti a kísérleti eredményeket (26,36 kJ/mol), kevesebb adatsorból való értékelés során is jó közelítéssel visszaadta a teljes adatsorral számolt eredményt.

1. P. T. Kiss, A. Baranyai, *J. Chem. Phys.*, **138**, 204507 (2013)
2. P. Varilly, D. Chandler, *J. Phys. Chem. B*, **117**, 1419-1428 (2013)

Summary

Free energy calculation of evaporation using weighted sampling in computer simulation

Gergely Kohut, BSc in Chemistry student

Place of diploma work: Physical Chemistry Department, Institute of Chemistry, Eötvös University, Budapest

Place of defence: Physical Chemistry Department

Supervisor: **Gergely Tóth, PhD**, associate professor
ELTE, Physical Chemistry Department

Our major aim was to calculate and investigate the free energy profile of the evaporation of a molecule from the liquid phase to the gas phase using the polarizable BK3[1] water model. The simulation data were obtained by the Umbrella Sampling method. We used Chandler's technique[2] to define the proper location of the surface. The free energy profile was evaluated by three methods, namely the WHAM, MBAR and VFEP. The molecular dynamics was performed on 1000 molecules put in an *NVT* slab containing two gas/liquid interfaces. An another aim of the project is to test the evaporation properties of the BK3 model and to supply background information to ongoing evaporation trajectory calculations.

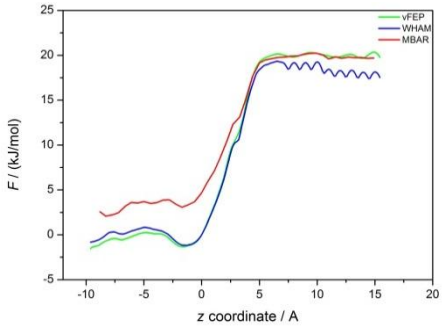


Figure 1: Comparison of the three different method

Considering the free energy profiles shown in Figure 1 we might say that their shapes are nearly the same to those ones in the literature. The profile of the BK3 model seems to be barrierless, as in the literature, despite the ability of the BK3 model to describe low-coordination states superior to other models. The free-energy difference between the phases is less than the experimental value. It could be an attribute of the BK3 water model. The temperature of the evaporating molecule is crucial as well, because experimental and simulation data suggest enhanced z-direction velocity just over the surface. The pressure control of the simulation also can be an important question, because the chosen reaction coordinate in the umbrella sampling (the distance from the local surface) cannot be implemented into NPT ensemble. These questions have to be investigated in careful ongoing studies.

After the comparison of the three different methods, our conclusion is that the VFEP provided better results than the two other methods. It needs less data and overlap to calculate the free energy profile. The free-energy difference of the phases (20,5 kJ/mol) using this method is closer to the experimental value (26,4 kJ/mol) than for the other methods.

1. P. T. Kiss, A. Baranyai, *J. Chem. Phys.*, **138**, 204507 (2013)
2. P. Varilly, D. Chandler, *J. Phys. Chem. B*, **117**, 1419-1428 (2013)

Irodalomjegyzék

1. P. T. Kiss, A. Baranyai, *J. Chem. Phys.*, **138**, 204507 (2013)

2. Kiss Dóra Judit, Szakdolgozat: Párolgás vizsgálata számítógépes szimulációval, ELTE, Fizikai Kémiai Tanszék (2013)

3. P. Varilly, D. Chandler, *J. Phys. Chem.* B, **117**, 1419-1428 (2013)

4. B. Roux, *Comp. Phys. Comm.*, **91**, 275-282 (1995)

5. G. M. Torrie, J. P. Valleau, *J. Comput. Phys*., **23**, 187-199 (1977)

6. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, *J. Chem. Phys.* **79,** 926-935 (1983)

7. H. J. C. Berendsen, J. R. Grigera and T. P. Straatsma, *J. Phys. Chem,*. **91**, 6269-6271 (1987)

8. J. L. F. Abascal, C. Vega, *J. Chem. Phys*. **123**, 234505, (2005)

9. M. W. Mahoney and W. L. Jorgensen, *J. Chem. Phys.* **112,** 8910-8922 (2000)

10. C. H. Bennett, *J. Comput, Phys*., **22**, 245-268 (1976)

11. N. Musolino, B. L. Trout, *J. Chem. Phys.*, **138**, 134707 (2013)

12 R. Vácha, P Slavíček, M. Mucha, B. J. Finlayson-Pitts, P, Jungwirth, *J. Phys. Chem*. A, **108**, 11573-11579 (2004)

13. Dang, L. X.; B. C. Garrett, *Chem. Phys. Lett*., **385**, 309-313 (2004)

14. Kumar, S., Bouzida, D., Swendsen, R. H., Kollman, P. A., and Rosenberg, J.M. *J. Comput. Chem.* **13**, 1011–1021 (1992)

15. M. Andrec, The Weighed Histogram Analysis Method (WHAM), Jegyzet, (2010), http://biomaps.rutgers.edu/wham_derivation.pdf, (letöltve: 2014. április 5.)

16. Lee, T. S., B. K. Radak, A. Pabis, D. M. York, *J. Chem. Theory Comput.*, **9(1)**, 153-164 (2013)

17. A. P. Willard, D. Chandler, J. Phys. Chem. B, **114**, 1954-1958 (2010)

18. Shirts M. R., Chodera J. D., *J. Chem. Phys.* **129**, 124105 (2008)

19. pymbar-2.0beta, https://simtk.org/project/xml/downloads.xml?group_id=264, (letöltve: 2014. április 18.)

20. Grossfield, Alan, "WHAM: the weighted histogram analysis method", version 2.09, http://membrane.urmc.rochester.edu/content/wham, (letöltve: 2014. április 18.)

21. VFEP, version 0.1, https://sites.google.com/site/cancersimulation/software, (letöltve: 2014. április 18.)

22. Néda Zoltán, Tyukodi Botond és Kacsó Ágota – Enikő, A Klasszikus Statisztikus Fizika Alapjai, Jegyzet, http://www.phys.ubbcluj.ro/~zneda/statfiz/stat-fiz.pdf (letöltve: 2014. 04. 02)

23. L. B. Pártay, Gy. Hantal, P. Jedlovszky, Á. Vincze, G. Horvai, J. Comput. Chem., **29**, 945 (2008)

NYILATKOZAT

Név: Kohut Gergely

ELTE, Természettudományi Kar, Szak: Kémia BSc

NEPTUN azonosító: JIYXLQ

Szakdolgozat címe:

Párolgás szabadenergiájának meghatározása súlyozott mintavételezéssel számítógépes szimulációban

A s**zakdolgozat** szerzőjeként fegyelmi felelősségem tudatában kijelentem, hogy a dolgozatom önálló munkám eredménye, saját szellemi termékem, abban a hivatkozások és idézések standard szabályait következetesen alkalmaztam, mások által írt részeket a megfelelő idézés nélkül nem használtam fel.

Budapest, 2014. május 12.

_____

a hallgató aláírása