

Matematikai módszerek a tömegspektrometriában, a modellezés különböző szintjei

Doktori értekezés

Peltz Csaba

Eötvös Loránd Tudományegyetem, Kémia Doktori Iskola
Elméleti és fizikai kémia, anyagszerkezet-kutatás doktori program

A doktori iskola vezetője: *Prof. Dr. Inzelt György*

A doktori program vezetője: *Prof. Dr. Surján Péter*

Témavezető: *Dr. Drahos László*



Budapest, 2008

Köszönetnyilvánítás

- Köszönöm Dr. Inzelt Györgynek, a doktori iskola vezetőjének és Dr. Surján Péternek, a doktori program vezetőjének, hogy lehetővé tették doktori értekezésem elkészítését.
- Köszönettel tartozom Dr. Drahos Lászlónak, témavezetőmnek mindenkori segítségét és a doktori munkám során nyújtott támogatását.
- Ezúton mondok köszönetet Újszászy Kálmánnak a tömegspektrometria megismertetéséért és útmutatásáért.
- Dr. Blaskó Gábornak, Dr. Simig Gyulának, Dr. Kövesdi Istvánnak és a Szerkezetkutatási osztály dolgozóinak köszönöm, hogy támogatták doktori munkám.
- Köszönettel tartozom Dr. Vékey Károlynak, amiért lehetővé tette, hogy kutatási témáim a MTA Kémiai Kutatóközpontjának Tömegspektrometria osztályán végezhettem.
- Sztáray Juditnak köszönöm a kvantumkémiai számításokat.
- Dr. Dombi Józsefnek köszönöm a mesterséges intelligenciával való megismerkedés lehetőségét.
- Köszönöm családomnak türelmüket és segítségüket.

Tartalomjegyzék

Bevezetés	5
Célkitűzések	6
1. Irodalmi összefoglalás	7
1.1. Méréstechnikák a tömegspektrometriában	7
1.1.1. Ionizációs módszerek	7
Elektronütöközéses ionizáció.....	7
Elektroporlasztásos ionizáció.....	8
1.1.2. Fourier-transzformációs tömegspektrometria	10
SORI-CID	12
Infravörös hűlés	13
1.1.3. Termokémiai számítások.....	15
A kinetikus módszer	16
A kiterjesztett kinetikus módszer	17
1.1.4. Szimulációs módszerek.....	19
A MassKinetics program	20
1.1.5. Automatizálási lehetőségek.....	21
Tömegspektrum-értékelést támogató szoftverek	22
1.2. A felhasznált matematikai módszerek	23
1.2.1. Paraméterbecslések	23
Lineáris regresszió	24
ODR – ortogonális távolság regresszió.....	25
1.2.2. Numerikus szélsőérték-keresési technikák.....	26
Többdimenziós szélsőérték-keresés, Powell módszere.....	27
Globális optimalizálások, szimulált fagyasztás módszere	29
1.2.3. A gépi tanulás.....	31
Szakértői rendszerek	32
Valószínűségi szabályok indukciója, intervallum-valószínűségek	34
2. A kísérletek és számítások részletei	35
2.1. SORI-CID fragmentáció modellezése	35
2.2. A kinetikus módszer hibájának vizsgálata.....	41
2.2.1. Virtuális kísérleti adatkészletek generálása.....	41
2.3. Valószínűségi fragmentációs szabályok gépi tanulása	44
2.3.1. A tömegspektrometriás mérési körülmények.....	44

2.3.2.	A kémiai szerkezetek reprezentációja és kezelése	44
2.3.3.	A kidolgozott és alkalmazott algoritmus.....	45
2.3.4.	A spektrumok tárolása.....	46
2.3.5.	A fragmentáció tárolása	48
2.3.6.	A fragmentációs szabályok	48
2.3.7.	Adatformátumok	49
3.	Eredmények és értékelésük	50
3.1.	SORI-CID fragmentáció modellezése	50
3.1.1.	A SORI gerjesztési folyamat elméleti leírása.....	50
3.1.2.	A SORI kísérletek numerikus modellezése.....	54
3.2.	A kinetikus módszer hibájának vizsgálata.....	61
3.2.1.	A meghatározható entalpia és entrópia pontossága.....	61
3.3.	Valószínűségi fragmentációs szabályok gépi tanulása.....	71
3.3.1.	A tanult szabályok szerkezete, finomítása és szűrése	71
3.3.2.	A szerkezetek összehasonlítása.....	73
3.3.3.	A megjelenési intenzitások számítása	73
3.3.4.	A megvalósított program.....	74
3.3.5.	A meghatározott szabályok.....	75
4.	Összefoglalás	78
5.	Summary	79
	Függelék. A kinetikus energia eloszlásának meghatározása.....	80
	Irodalomjegyzék.....	82
	Közlemények.....	92
	Az értekezés alapját képező közlemények	92
	Az értekezés témaköréhez nem kapcsolódó közlemények.....	92
	Az értekezés témaköréhez kapcsolódó előadások, poszterek.....	92
	Az értekezés témaköréhez nem kapcsolódó előadások, poszterek.....	93

Bevezetés

Az elmúlt években a tömegspektrometria területén – a nagyműszeres analitika más területeihez hasonlóan – a számítástechnika és az automatizálás nagymértékű térhódítása volt megfigyelhető. A műszervezérlési, üzemeltetési, hangolási, mérési feladatok szükség esetén teljesen automatikussá tehetők, lehetővé téve a nagy áteresztőképességű (HT, high throughput) metodikák kifejlesztését. A felhasználói oldalon túl, a tömegspektrometria elméleti, alapkutatási területein is hasonló tendenciát figyelhetünk meg. Joggal alkothatunk már manapság a CAD (Computer Aided Design – számítógéppel segített tervezés), CADD (Computer Aided Drug Design – számítógéppel segített gyógyszertervezés) és egyéb hasonló betűszavak analógiájára egy CAMS (Computer Aided Mass Spectrometry – számítógéppel segített tömegspektrometria) rövidítést. Érdemes megjegyezni ugyanakkor, hogy nem-CAMS, tehát számítástechnikát mellőző tömegspektrometria már lényegében nem is létezik.

A tapasztalt tendencia azonban több kérdést is felvet. Vajon a különböző feladatok megoldására a megfelelő algoritmusokat alkalmazzuk-e? Értem itt a megfelelő jelző alatt azt, hogy olyan módszereket és eljárásokat használunk, amelyek elegendően pontosak és helyesek az adott feladat megoldására, ugyanakkor nem igényelnek indokolatlanul nagy erőforrásokat az eredmények elérésére.

Jelen dolgozatban a tömegspektrometria számítógéppel segített elméleti modellezésének három különböző szintjét mutatom be, a legprecízebbtől indulva az általánosabb, egyre több közelítést és egyre kevesebb fizikai-kémiai elméleti alapot tartalmazó megoldások felé, bemutatva a különböző matematikai eszközöknek a modellezésben betöltött szerepét és korlátait.

Az első bemutatott modellezési szint célja a tömegspektrométerben lezajló fizikai-kémiai folyamatok minél pontosabb leírása, ismeretlen molekuláris paraméterek meghatározása, illetve az eredményként kapott tömegspektrum lehető legpontosabb becslése. A második részben egyszerű, analitikus formában megadható, termodinamikából vett összefüggések nem-termodinamikai rendszerben való alkalmazását mutatom be, alkalmazhatóságának határaival és a kapott eredmények hibáival. Végül a harmadik megközelítés a mesterséges intelligencia és a tömegspektrometria határterületén az emberi spektrumfejtés modellezését mutatja be.

Célkitűzések

Az általánosan elterjedt modellezési módszerek, közelítő kifejezések és becslések mindegyike meghatározott alkalmazhatósági feltételekkel rendelkezik. Doktori munkám során mindenekelőtt célul tűztük ki annak vizsgálatát, hogy az egyes fizikai-kémiai, matematikai illetve számítástechnikai módszereknek milyen hatása van a modellezés eredményeinek megbízhatóságára, pontosságára. A különböző szimulációk és tömegspektrometriás mérés technikák esetén a közelítések alkalmazhatóságának feltételeit vizsgáltam.

Célunk volt továbbá annak vizsgálata, hogy a tömegspektrometriás folyamatok modellezése esetén milyen matematikai módszerek és algoritmusok alkalmazása a legcélszerűbb és leghatékonyabb, milyen pontosságú és milyen mélységű fizikai-kémiai ismereteket kell felhasználnunk az adott feladat megoldásához.

Szükségesnek tartottuk megvizsgálni azt, hogy a maximális pontosságra törekvő számítási módszerek esetén milyen fizikai hatások figyelembevétele szükséges a fragmentáció mértékének helyes becsléséhez, ezért célul tűztük ki SORI-CID kísérletek modellezését, majd ebből az infravörös hűlés és az energia-átadás hatékonyságának meghatározását és részletes vizsgálatát.

Egy napi gyakorlatban széles körben elterjedt mérés technika, az ún. kinetikus módszer esetén vizsgáltam azt, hogy az entalpia- illetve entrópia-paraméterek meghatározásakor az elméleti elhanyagolások, a számítási pontatlanságok és a kísérleti körülmények bizonytalanságai milyen mértékben tehetők felelőssé a végeredmény hibájáért.

A modellezési lehetőségek legkevesebb fizikai-kémiai háttért igénylő képviselőjeként végül a mesterséges intelligencia módszereinek használhatóságát vizsgáltam a tömegspektrometriás gyakorlatban. Céloom az emberi spektrumértékelési folyamat algoritmikus modellezése volt.

1. Irodalmi összefoglalás

1.1. Méréstechnikák a tömegspektrometriában

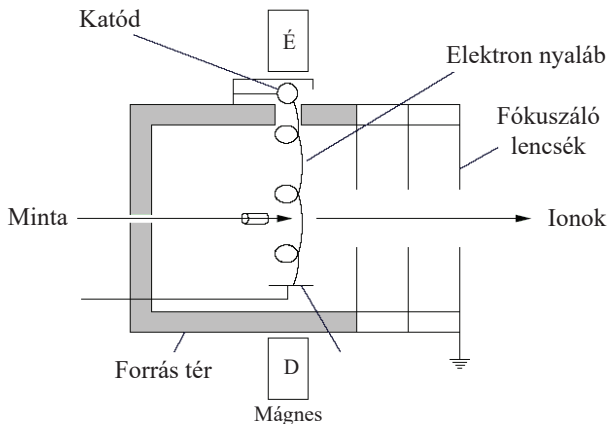
1.1.1. Ionizációs módszerek

A tömegspektrometria töltött részecskék tömegének – illetve tömeg/töltés arányának – meghatározására alkalmas analitikai módszer. A mérési folyamat első lépése az adott minta ionizálása, töltött részecskék létrehozása a gáz- illetve kondenzált fázisú mintából. A bevitt minta tulajdonságaitól (polaritás, molekulaméret, protonaffinitás) függően különböző ionizációs módszerek alkalmazhatók, ilyenek az elektronütközéses ionizáció (EI), a kémiai ionizáció (CI), az elektroporlasztásos ionizáció (ESI), az atmoszférikus nyomású kémiai ionizáció (APCI), a mátrixszal segített lézer deszorpciós ionizáció (MALDI). Ezek közül a továbbiakban a dolgozat későbbi részeiben érintett két ionizációs technikát ismertetem részletesebben: az elektronütközéses és az elektroporlasztásos ionizációt. [1]

Elektronütközéses ionizáció

A hagyományos elektronütközéses ionizáció (electron impact, EI) során a vizsgálni kívánt mintát – nagyvákuum-térben elpárologtatott formában – ionizációs kamrába juttatjuk. A kamrában nagysebességű elektronok a minta semleges molekuláiból egy-egy elektront ütnek ki, majd az így kapott egyszerűen pozitív töltésű gyökiont a pozitív katasztó-potenciál az analizátor felé tereli. (1. ábra)

Az ionizációs folyamat eredményeként nyílthéjú, ún. gyökionok keletkeznek. A bombázó gyors elektronok energiája általában standard módon 70 eV, így a molekula ionizáció során nagy fölös energiára tesz szert, ami általában kiterjedt fragmentációhoz vezet, így szerkezeti információt szolgáltat.



1. ábra. Elektronütközéses ionforrás felépítése.

A 70 eV-os standardizáció és a gerjesztés jellege miatt a fragmentáció – azaz a tömegspektrum – jól reprodukálható, emiatt állhatnak rendelkezésre kiterjedt spektrumkönyvtárak. További előnye a reprodukálható fragmentációnak, hogy a kapott fragmensek kapcsolhatók a molekulában található adott szerkezeti részletekhez, ezzel segítve az ismeretlen szerkezetek meghatározását. Az ionizáció módjából adódóan az EI forrással rendelkező tömegspektrométer könnyen kapcsolható gázkromatográfval.

Elektroporlasztásos ionizáció

Az elmúlt két évtizedben az elektroporlasztásos ionizáció (electrospray ionization, ESI) a tömegspektrometria domináns ionizációs technikájává vált. Köszönhette ezt annak, hogy lényegéből adódóan kiválóan alkalmas folyadékfázisú minták ionizációjára, s így folyadékkromatográfhoz közvetlenül kapcsolható. Lágýionizációs technikaként a hagyományos elektronütközéses ionizációval ellentétben egy molekulaiont (pontosabban ún. kvázimolekulaiont) szolgáltat, amely megkönnyíti a molekulatömeg meghatározását.

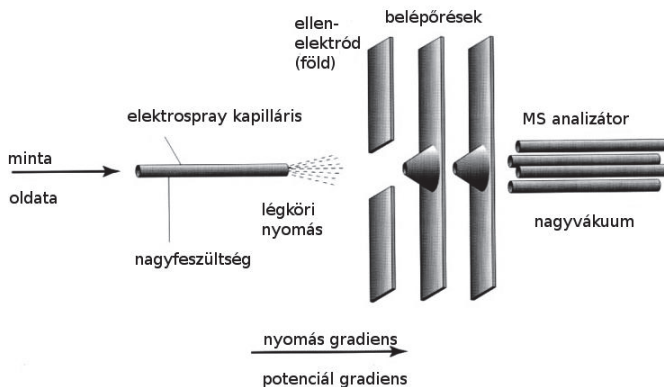
Az elektroporlasztásos ionizáció során a vizsgálandó minta oldatát egy kapillárison keresztül juttatjuk a tömegspektrométerhez. Ennek a kapillárisnak a végére nagyfeszültséget kapcsolnak (készüléktípustól függően ez 3-6 kV), amelynek hatására porlasztás történik.

Néhány $\mu\text{l}/\text{perc}$ áramlási sebesség fölött az elektroporlasztást a kapilláris mellett áramoltatott nagysebességű N_2 -gáz segíti.

Az elektroporlasztásos ionizáció mechanizmusa kétféle modellel magyarázható.

- A *töltésmaradvány modell* (charge residue model) szerint a többszörösen töltött cseppek párolognak és egy kritikus méretet elérve a töltések tasztításának hatására aprózódnak. Az aprózódás során egyre kisebb méretű cseppek jönnek létre, míg a folyamat végén a minta szolvatált és általában protonált (MH^+) ionjai jelennek meg.
- Az *ion-evaporációs modellben* (ion evaporation model) a töltések tasztítása miatt a csepp egy töltött részecskét lök ki magából.

Ezek a folyamatok a tömegspektrométer atmoszférikus interfészében játszódnak le, ahol N_2 -gáz ellenáramoltatás vagy magasabb hőmérséklet alkalmazásával segítik az egyedi ionok kialakulását. (2. ábra)



2. ábra. Elektroporlasztásos ionforrás vázlatos felépítése.

Az egyszerű HPLC-MS kapcsolat kialakíthatósága miatt az elektroporlasztásos ionizáció széles körben elterjedt. A módszer – illetve annak különböző fajtái – a néhány nl/perc

áramlási sebességektől a néhány ml/perc tartományig megfelelően alkalmazható. Az ionizáció kéméletes (ún. lágyonizációs technika), általában valamilyen protonáló ágens segítségével történik meg. A kapott tömegspektrumban fragmensek általában nem, de különböző klaszter ionok nagy valószínűséggel előfordulhatnak. A spektrumban látható ionok néhány speciális esettől eltekintve zárthéjú, ún. kvázi-molekulaionok. A molekulatömeg információ mellett a tandem tömegspektrometriás [2] technikák segítségével szerkezeti információhoz is juthatunk. [3]

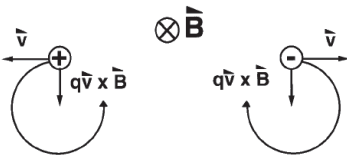
1.1.2. Fourier-transzformációs tömegspektrometria

A tömegspektrométerekben található analízator típusok mindegyike a bejuttatott töltött részecskének elektromágneses térrel való kölcsönhatásán alapul. A különböző analízator-típusok közül a későbbiekben használt Fourier-transzformációs ion-ciklotron-rezonanciát (FT-ICR) mutatom be. [4; 5; 6]

Töltött részecskéknak külső mágneses térben való mozgását a következő összefüggés adja meg.

$$\mathbf{F} = m\mathbf{a} = m \frac{d\mathbf{v}}{dt} = q\mathbf{v} \times \mathbf{B} \quad (1)$$

Az egyenletben \mathbf{F} az adott részecskére ható erő, m a tömege, \mathbf{a} a gyorsulása, \mathbf{v} a sebessége, q a töltése, \mathbf{B} pedig a mágneses térerő. A fenti egyenletből adódóan az ionok röppályájára merőleges külső mágneses tér esetén az adott ionoknak mindenkor a sebességük irányára merőleges gyorsulása lesz, tehát körpályán mozognak. (3. ábra)



3. ábra. Pozitív és negatív ionok mozgása sebességükre merőleges mágneses térben

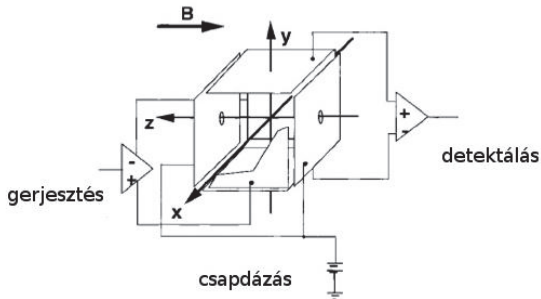
Körmozgás esetén a fenti összefüggés skalár alakja a következő:

$$m \frac{v^2}{r} = qvB_0. \quad (2)$$

B_0 a mágneses térerősség nagysága, $\frac{v^2}{r}$ a sugárirányú gyorsulás. Szögsebességre (ω), majd abból frekvenciára (ν) áttérve az alábbi egyszerű egyenletet kapjuk.

$$\nu_c = \frac{\omega}{2\pi} = \frac{qB_0}{2\pi m} \quad (3)$$

Tehát a ν_c ciklotron frekvencia az m/q tömeg/töltés hányadossal fordítottan, a B_0 mágneses térerősséggel pedig egyenesen arányos, az ionok sebességétől viszont független, így m/q pontos meghatározásához nincs szükség translációs energia-fókuszálásra. Egy általánosnak mondható 7 teslás mágneses térben az egyes ionokra kapott frekvenciák a néhány kHz és néhány MHz közötti tartományba esnek, az egyes ionok röppályáinak sugara pedig a néhány tízed mm-es nagyságrendbe. A 4. ábra egy ICR cella sematikus rajzát mutatja.

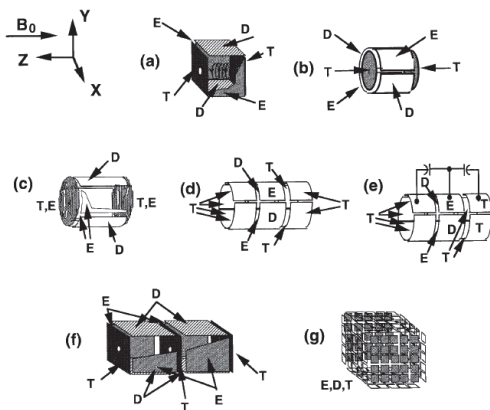


4. ábra. ICR cella vázlatos rajza. (Jobb és bal oldalon a csapdázó elektródok, főt és lent az indukált áram detektálásáért, míg elöl és hátul az esetleges gerjesztésekért felelős elektródok.)

A bejuttatott ionok a cellában csapdázódnak és a mágneses tér irányára merőleges körpályán mozognak. Az adott frekvenciákkal keringő ionok a detektáló elektródparban áramokat indukálnak. A különböző ionok által indukált áramok egymásra szuperponálódnak. Az indukált áram idő függvényében való méréséből Fourier transzformációval [7; 8] megkapható a különböző frekvencia-tagok súlya az összáram előállításával, s így különböző tömeg/töltés hányadossal rendelkező ionok relatív mennyisége. A valós készülékekben alkalmazott cella-geometriák igen változatosak lehetnek, ahogyan azt a 5. ábra is mutatja.

Az ioncsapdás technikák és a tandem tömegspektrometria napjaink elterjedt, szerves molekulák szerkezetvizsgálatára széleskörűen alkalmazott módszereivé váltak. Az egyre

inkább terjedő proteomikai alkalmazások [9; 10; 11] mellett az FT-ICR spektrométerek legjelentősebb alkalmazási területe a gázfázisú ionkémia. [12; 13; 14; 15] Az extrém nagy tömegfelbontás és pontosság mellett az FT-ICR előnyei között mindenképpen megemlítendő a kontrollált, ugyanakkor flexibilis gerjesztési módszerek alkalmazásának lehetősége: például a rezonancia gerjesztés, infravörös sugárzással való gerjesztés (BIRD) [16], hosszantartó nem rezonáns gerjesztés (SORI). [15] A következő részben a doktori munkám során vizsgált SORI gerjesztési módszert mutatom be.



5. ábra. Néhány ICR cellageometria. (E: gerjesztő, D: detektáló, T: csapdázó elektródák) (a) köbös, (b) hengeres, (c) "virtuális" végtelen hengeres, (d,e) nyitott hengeres, (f) duális, (g) mátrix típusú cella.

SORI-CID

A hosszantartó nem rezonáns gerjesztésű ütközéses aktiválás (sustained off-resonance irradiation collision induced dissociation, SORI-CID) az FT-ICR tömegspektrométerek egy kedvelt gerjesztési módja. [17; 18] A technika az ICR cellában tárolt ionok periodikus gyorsításával és lassításával, viszonylag „magas” (10^{-5} – 10^{-6} torr) cella-nyomással kombinálva éri el az ionok belső energiájának a fragmentációhoz szükséges szintjét. A rezonáns gerjesztésekkel ellentétben SORI esetén lehetőségünk van arra, hogy kontrolláltan, fokozatosan emeljük a belső energiát, hosszabb aktiválási időt, ugyanakkor alacsony kinetikus energiát alkalmazva. [14; 19; 20]

Az FT-ICR cellában a külső mágneses tér és a gerjesztő oszcilláló elektromos tér hatását együttesen figyelembe véve a következő mozgásegyenlet írható fel:

$$m \frac{d^2 \mathbf{r}}{dt^2} = q\mathbf{E} + q \left(\frac{d\mathbf{r}}{dt} \times \mathbf{B} \right), \quad (4)$$

ahol $\mathbf{r} = (x, y, z)$ a helyvektor, $\mathbf{B} = (0, 0, B_z)$ a homogén külső mágneses tér, $\mathbf{E} = (E_0 \cos \omega_{\text{exc}} t, -E_0 \sin \omega_{\text{exc}} t, 0)$ a gerjesztő elektromos tér ω_{exc} frekvencián, m és q pedig rendre a részecske tömege és töltése. A cella geometriája függvényében $E_0 = \frac{\beta V_{\text{p-p}}}{2d}$ alakban adható meg, ahol $V_{\text{p-p}}$ a gerjesztő potenciál, d az ICR cella átmérője, β pedig a cella geometriai paramétere [21], az előző részben bemutatott – az ideális végtelen hengeres elrendezéstől eltérő – cellageometriák hatását figyelembe vevő korrekciós tag.

A fenti egyenlet megoldásaként az ionok pályájának a sugarára az

$$r = \frac{E_0}{2B(\omega_{\text{exc}} - \omega_0)} \sin \frac{(\omega_{\text{exc}} - \omega_0)t}{2} \quad (5)$$

kifejezés adódik.

Ebből már $B = \frac{\omega_0 m}{q}$, $E_0 = \frac{\beta V_{\text{p-p}}}{2d}$ és $\omega_{\text{exc}} - \omega_0 = 2\pi\Delta\nu$ helyettesítésekkel megadható a laboratóriumi vonatkoztatási rendszerben érvényes kinetikus energia:

$$E^{\text{kin}} = \frac{m\omega_0^2 r^2}{2} = \frac{q^2 \beta^2 V_{\text{p-p}}^2}{64m d^2 \pi^2 \Delta\nu^2} (1 - \cos 2\pi\Delta\nu t). \quad (6)$$

A SORI kísérleteknek több elméleti modellezési módja is ismert. [22] Ezek a SORI gerjesztést általában egy periodikus gyorsítási és lassítási folyamat és véletlenszerű ütközések kombinációjaként írják le. A véletlen és az ütközések statisztikus kezelése nyitott utat a különböző Monte-Carlo típusú megoldások előtt. [23] A tanulmányok egy része behatóan vizsgálja az ütközőgáz nyomásának hatását, amely a SORI gerjesztés során történt ütközések számát befolyásolja. [18]

Infravörös hűlés

A gerjesztési és ütközési folyamatok mellett az ionok belsőenergia-eloszlását legjelentősebben befolyásoló fizikai folyamat a sugárzással való energiaátadás. Ahhoz, hogy a sugárzásos folyamatok jelentős effektust okozzanak, az ionok tartózkodási idejének megfelelően hosszúnak kell lennie. A legtöbb tömegspektrometriás méréstípus esetén ezek

elhanyagolhatók, ugyanakkor az FT-ICR mérések esetén a hosszú, akár másodperces tartózkodási idők következtében mindenképpen figyelembe veendő. A sugárzásos folyamatok tömegspektrometriás mérés technikák esetén elsősorban hűlést jelentenek. Gerjesztési folyamatra kiváló példa a feketetest infravörös sugárzásos disszociáció (BIRD: blackbody infrared irradiation dissociation), ahol a gerjesztés infravörös „melegítés” révén érhető el. [24; 25; 26; 27]

A sugárzás intenzitását alapvetően a Planck összefüggéssel írhatjuk le, mely szerint

$$\rho(\nu) = \frac{8\pi h \nu^3}{c^3 \exp\left(\frac{h\nu}{k_B T_{MS}}\right)}, \quad (7)$$

ahol T_{MS} a tömegspektrométer hőmérséklete, $\rho(\nu)$ pedig sugárzás intenzitása egy adott ν frekvencián.

A fotonok emisszióját és abszorpcióját az Einstein koeficienssekkel jellemezhetjük.

$$A(\nu, n, I_0) = \frac{2,88 \times 10^{-9}}{c^2} \nu^2 (n+1) I_0, \quad (8)$$

$$B(\nu, n, I_0) = \frac{c^3 A(\nu, n, I_0)}{8\pi h \nu^3}.$$

Itt I_0 az adott fény intenzitása, ν pedig a frekvencia. Az így kapott A emisszióhoz és B abszorpcióhoz tartozó koeficienssek segítségével kifejezhető annak a valószínűsége, hogy a molekula sugárzásos folyamat(ok) hatására E_1 belsőenergia állapotból E_2 -be jut. Attól függően, hogy vibrációs vagy elektron-állapotokat vizsgálunk, analóg egyenletekkel leírható az infravörös és az UV/látható energiatranszfer is.

Az így kapott modell fizikailag megalapozott, ugyanakkor igen számításgényes. Alternatívaként alkalmazható a Dunbar által kidolgozott szénhidrogén-modell. [24] Ennek alapfeltételezése az, hogy hasonló szerkezetű – egyszerű szerves – molekulák esetén a sugárzási folyamatok paraméterei alapvetően hasonlóak, amennyiben bizonyos molekuláris paraméterekkel – például a szabadsági fokok számával – megfelelően korrigálunk. Dunbar modelljét eredendően infravörös hűlés esetére dolgozta ki. Kísérleti eredményeiből adódott, hogy a molekula fölös (excess) energiája, tehát az aktuális környezet termikus energiája fölötti rész az időben exponenciálisan csökken. A csökkenés sebességét a szabadsági fokokkal arányosnak feltételezte. A sebességi állandót meghatározta egy kisméretű szénhidrogén

esetén, így a közelítés a standard szénhidrogén modell nevet kapta. Az előbb leírtak a következő egyenletben foglalhatók össze:

$$E_{\text{excess}}(t) = E_{\text{excess}}(0) \times \exp(-k_{\text{cool}}t), \quad (9)$$

ahol E_{excess} jelöli a termikus fölötti energiát, k_{cool} a sebességi állandó, t pedig az idő.

1.1.3. Termokémiai számítások

A termokémiai paraméterek a tömegspektrometria és a gázfázisú ion-kémia területén a kémia más területeihez hasonlóan fontos szerepet töltenek be. A meghatározásukra alkalmas módszerek a következőképpen osztályozhatók:

1. kalorimetria,
2. spektroszkópia,
3. elméleti számítások,
 - a. reakciósebesség számítások,
 - b. egyensúlyi állandó számítása,
4. termokinetikus módszerek,
 - a. *kinetikus módszer*,
 - b. ion-molekula reakciók (bracketing),
 - c. küszöb analízis.

Abszolútnak tekinthető módszerek ritkán alkalmazhatók, ezért a kísérleti meghatározások többsége valamilyen relatív módszer segítségével történik. Ezek közé a relatív kísérleti módszerek közé tartoznak az utolsó pontban említett termokinetikus módszerek, amelyekben a kísérleti úton mért reakciósebességek alapján határozhatók meg a keresett termokémiai paraméterek.

A termokinetikus módszerek egyik csoportját az úgynevezett küszöb analízis módszerek alkotják. Ezek közé tartozik az ionizációs energiák illetve adott fragmens-ionok képződéshőinek mérése a megjelenési energiák alapján. Minthogy a különböző megjelenési energiaértékek valamilyen idealizált körülményekhez tartoznak, amelyek kísérletileg elérhetetlenek, minden küszöb mérési módszer az energia függvényében elvégzett kísérletsorozatból extrapolál a meghatározni kívánt energiaértékre. A termokinetikus módszerek másik csoportja (bracketing) ion-molekula reaktivitás mérésén alapul. Proton- és

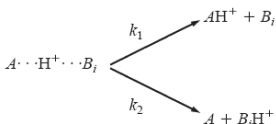
elektron-transzfer reakciók segítségével proton- és elektron-affinitások határozhatók meg. A kinetikus módszer a fenti termokinetikus meghatározások harmadik csoportját alkotja. Egyszerűsége miatt – elsősorban a tömegspektrometrián belül – széles körben elterjedt. [28; 29]

A kinetikus módszer

A kinetikus módszer alkalmazásakor a termokémiai paraméterek meghatározása az alábbi egyenlet szerint referencia-anyagok használatával történik.

$$\ln\left(\frac{I_A}{I_B}\right) \approx \ln\left(\frac{[AH^+]}{[BH^+]}\right) = \ln\left(\frac{k_1}{k_2}\right) \approx \ln\left(\frac{Q_1^*}{Q_2^*}\right) + \frac{\Delta\epsilon_0}{RT_{\text{eff}}} \approx \frac{\Delta(\Delta G)}{RT_{\text{eff}}} \approx \frac{\Delta(\Delta H)}{RT_{\text{eff}}}, \quad (10)$$

amennyiben a következő kompetitív reakcióegyenleteket tekintjük. [28; 30]



6. ábra. Protonnal összetartott komplex lehetséges bomlási útjai.

Az összefüggésben I_X X csúcshintenzitása, $[X]$ X koncentrációja, k_i a megfelelő sebességi állandó, Q_i^* az adott állapotösszeg, $\Delta\epsilon_0$ a kritikus energiák különbsége, $\Delta(\Delta G)$ a szabadentalpia-, $\Delta(\Delta H)$ pedig az entalpiaváltozások különbsége. T_{eff} definíció szerint az a hőmérséklet, amelyen a Boltzmann-eloszlású aktivált komplex ionok azonos termékarányokat adnának az adott kísérleti körülmények között mérttel. [31; 32] A fenti közelítés-sorozatnak leglényegesebb pontja az utolsó lépés, amely az entrópia-effektusok elhanyagolását jelenti. [33] Érdemes rámutatni arra is, hogy a fenti összefüggés bizonyos határok között érvényes akkor is, ha nem termikus energia-eloszlással állunk szemben – mint például egy tömegspektrométer esetén. Meg kell jegyeznünk, hogy a módszer alkalmazásainál általában eltekintenek a számított termokémiai paraméterek (ΔH és ΔS) hőmérséklet-függésétől, tehát az $\ln\left(\frac{k_1}{k_2}\right) \leftrightarrow \frac{1}{T}$ függvény nem-linearitásától.

A módszer elterjedésének oka egyrészt az, hogy meglepően pontos eredményeket szolgáltat, illetve a mért kompetitív reakciósebességek igen érzékenyen változnak a termokémiai paraméterek kis különbségei esetén is, másrészt vitathatatlan a mérés

gyorsasága, egyszerűsége és az a tény, hogy egyéb tandem tömegspektrometriás mérésekhez hasonlóan a minták szennyezettsége nem okoz problémát. További előny, hogy biológiai szempontból érdekes molekulák, amelyek más mérésekkel nem tanulmányozhatók, könnyen vizsgálhatók a kinetikus módszer segítségével. Felhasználási területe kiterjed a proton-affinitások, gázfázisú bázicitás és savasság, protonalódás esetén bekövetkezett entrópiaváltozások, ionizációs energiák, elektron-affinitások becsülésének körére. [28; 34; 35; 36; 37; 38]

A kinetikus módszer alkalmazása során a kísérleti körülményektől függően (referencia-anyagok száma, a vizsgált ütközési energiák száma és az entrópia kezelésének módja) különböző pontosság érhető el. A tipikusan alkalmazott módszereket a következő táblázatban foglaltam össze. (1. táblázat)

1. táblázat. A kinetikus módszer különböző típusai, pontosságuk szerint rendezve.

Név	Referencia-anyagok száma	Ütközési energiák száma	Entrópia elhanyagolása
Közrefogásos	3	1	azonos
Standard	sok	1	azonos
Izoentropikus	sok	1	minden dimerre azonos entrópia-különbség
Autentikált	sok	legalább 2	elhanyagolhatóság kísérleti úton ellenőrizve
Kiterjesztett	sok	legalább 3	nincs, pontosan figyelembe véve

A kiterjesztett kinetikus módszer

A kinetikus módszer alkalmazásának nagy előrelépése volt, amikor kiderült, hogy az entrópia-effektusok nemcsak elfogadhatók és számításoknál figyelembe vehetők, hanem ugyanezzel a technikával becsülhetők is. A korai megoldásokkal ellentétben, ahol az entrópia-effektusok kiküszöbölése érdekében szerkezetileg hasonló referencia anyagokat használtak, a kiterjesztett módszernél lehetőség van ennek az elhanyagolt effektusnak a kísérleti becsülésére, meghatározására. [39; 40; 33]

Amennyiben a B mintától szerkezetileg különböző, de egymáshoz hasonló B_i referenciasorozatot használunk, akkor az $\ln\left(\frac{Q_i^*}{Q^*}\right)$ tagok nem hanyagolhatók el, de referenciák egymás közötti hasonlósága miatt minden B_i -re hasonló, így $\ln\left(\frac{k_i}{k}\right)$ értékét ábrázolva $\Delta H(B_i)$ függvényében egyeneseket kaphatunk. Az $\ln\left(\frac{Q_i^*}{Q^*}\right)$ kifejezés ekkor a két disszociációs út entrópia-változásának különbségét adja meg, tehát $\ln\left(\frac{Q_i^*}{Q^*}\right) = \frac{\Delta(\Delta S)}{R} = \frac{(\Delta S_i^0 - \Delta S^0)}{R}$. Ekkor a 16. oldal 10. egyenlete a következő alakot ölti

$$\ln\left(\frac{k_1}{k_2}\right) = \ln\left(\frac{[AH^+]}{[BH^+]}\right) = \frac{\Delta H(B_i)}{RT_{\text{eff}}} - \frac{\Delta H(B) - T_{\text{eff}}\Delta(\Delta S)}{RT_{\text{eff}}}, \quad (11)$$

ahol a $\Delta H(B) - T_{\text{eff}}\Delta(\Delta S)$ tagra, mint a látszólagos szabadentalpiára hivatkozhatunk ($\Delta G(B)^{\text{app}}$). Figyeljünk arra, hogy az entrópia-változások különbségének szerepeltetése miatt ez nem egy szabadentalpia jellegű tag, csupán formailag hozható kapcsolatba vele.

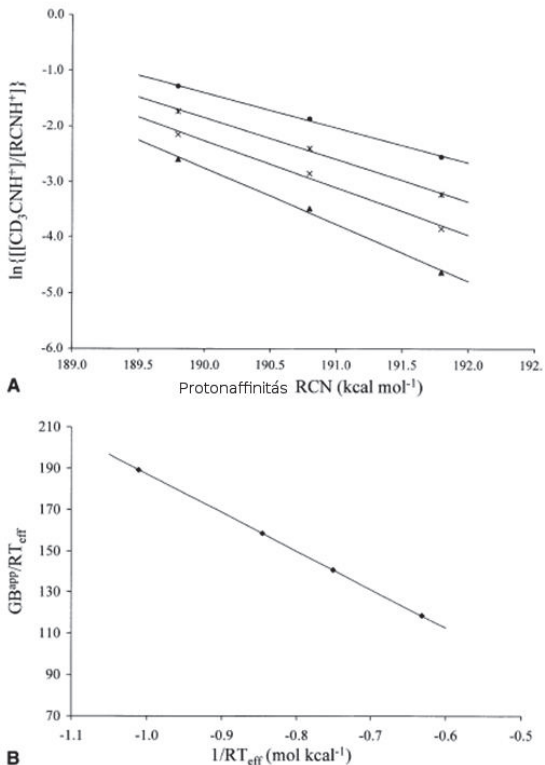
Feltételezve, hogy az átmeneti állapot hasonló a termékekhez, a $\Delta(\Delta S) \approx \Delta S_{X^+}(B_i) - \Delta S_{X^+}(B)$ közelítést tehetjük. $\ln\left(\frac{k_i}{k}\right)$ értékét ábrázolva $\Delta H(B_i)$ függvényében az y tengelymetszet $-\frac{\Delta G(B)^{\text{app}}}{RT_{\text{eff}}}$ értékét adja, a meredekség pedig $\frac{1}{RT_{\text{eff}}}$ -et. Különböző ütközési energiákon – amelyek különböző effektív hőmérsékleteket jelentenek – elvégezve a méréseket, a $\frac{\Delta G(B)^{\text{app}}}{RT_{\text{eff}}}$ kifejezés $\frac{1}{RT_{\text{eff}}}$ függvényében $\Delta H(B)$ -t adja meredekségképpen és $-\Delta(\Delta S)/R$ -t tengelymetszetként. (7. ábra)

A kiterjesztett kinetikus módszer pontosabb eredményeket szolgáltat azokban az esetekben, ahol intramolekuláris kötések léphetnek fel. A kapott $\Delta(\Delta S)$ értékének megbízhatósága azonban sok szempontból megkérdőjelezhető. Egyrészt T_{eff} nem egy valós hőmérséklet-paraméter, másrészt pedig az előbbi levezetésben feltételeztük, hogy $\Delta(\Delta S)$ a hőmérséklettől független. A 7. ábra a standard és a kiterjesztett kinetikus módszer alkalmazását mutatja deuterio-acetonitril termokémiai paramétereinek meghatározására.

A fenti problémák ellenére a kinetikus módszer – bizonyos esetektől eltekintve – jól használható, de az irodalomban régóta vitatottak a módszer használhatóságának korlátai.

A vélemények abban térnek el, hogy mekkora az így meghatározott ΔH és ΔS paraméterek szisztematikus és véletlen hibája. A statisztikai becslésekhez hasonlóan itt is a torzítatlanság kérdése merül fel. Szintén teljes az egyetértés abban, hogy ΔH energia-

(effektív hőmérséklet) függését a legriválisabb esetekben is ellenőrizni kell, és szükség esetén a kiterjesztett kinetikus módszert kell használni. [40; 39; 41]



7. ábra. A standard (A) és a kiterjesztett (B) kinetikus módszer alkalmazása CD₃CN termokémiai paramétereinek meghatározására. A: a koncentrációk logaritmusai a referencia-anyagok (RCN) protonaffinitása függvényében, B: a látszólagos GB^{app}/RT_{eff} ábrázolása 1/RT_{eff} függvényében [36].

1.1.4. Szimulációs módszerek

A tömegspektrometrumok elméleti és numerikus modellezésére több különböző megközelítés létezik. Kiemelt szerepet foglalnak el a modellek között azok, amelyek célja –

minél pontosabb fizikai-kémiai háttér figyelembevételével – egy tetszőleges molekula tömegspektrumának kialakulását „pontosan” becslő modell, majd algoritmus megalkotása.

A tömegspektrométerekben lejátszódó folyamatok a makroszkópikus és a mikroszkópikus folyamatok határterületén kezelhetők a legnagyobb sikerrel, az ilyen területen születő modellek többsége az egyedi részecskemozgás szimulációk és a reakciókinetikai megközelítések között helyezkedik el. Az előbbit elsősorban a különböző analízátor típusok fejlesztésekor ion-röppályák számítására, stabilitásának vizsgálatára használják, míg utóbbit a fragmentációs folyamatok „statisztikus” kezelésére, de teljesen nem választhatjuk szét ezt a két csoportot. A SIMION program[42] például figyelembe vesz ion-molekula ütközéseket, de belsőenergia-változások és fragmentációs reakciók csak nagyon durva közelítésekkel szerepelnek az algoritmusában. Hatékonyabb, de specializáltabb algoritmusok születtek ionsapdás készülékek modellezésére. [43] Az ITSIM [44] nevű ionsapda modellező programot Plass és Cooks kiterjesztették fragmentáció és ütközéses hűlés figyelembevételére. Goeringer és McLuckey pedig ütközéses aktiválás és dezaktiválás numerikus modellezését tanulmányozta ionsapdás körülmények között. [45; 46]

A MassKinetics program

A MassKinetics program célja az összes molekuláris paraméter és fizikai folyamat figyelembevételével a teljes tömegspektrometriás kísérlet modellezése az ionok képződésétől a detektálásukig, az egyes résztvevő ionok és semleges részecskék az idő függvényében változó energia-eloszlásának és koncentrációjának meghatározása. [47; 48]

A MassKinetics algoritmus a következő fizikai-kémiai folyamatokat veszi figyelembe.

- *Ionizáció.* kezdő energia-eloszlások (ionforrás hőmérséklete) [49], az energia-eloszlások változása ionizáció során.
- *Gyorsítás.* az ionok gyorsítása elektrosztatikus térben.
- *Reakciók.* Megadhatók az egyes reaktánsok, azok molekuláris paraméterei (tömeg, töltés, szabadsági fok), a reakciók molekuláris paraméterei (kritikus energiák, frekvencia paraméterek). A számítások a Rice-Ramsperger-Kassel-Marcus (RRKM) elmélet [50; 51] alapján történnek.
- *Ütközések.* Az ütközések kezelése két lépésben szemléltethető: kezelni kell az ütközések gyakoriságát, illetve az ütköző gázzal való egyedi energiacserekek hatását. A

gyakoriságok kezelése ütközési hatáskeresztmetszetek és frekvenciák formájában történik. Az egyedi ütközések leírására a hosszú életű ütközési komplex modell és a részlegesen rugalmatlan ütközési modell használható.

- *Sugárzási átmenetek.* Az irodalmi részben már említett módszerekkel lehetőség van a számítások során infravörös hűlés figyelembevételére.

Az így definiált fizikai-kémiai folyamatok egy – általában csatolt – differenciál egyenlet-rendszer adnak, amelynek az adott kezdeti feltételekkel az idő szerinti integrálása [52] a feladat. A MassKinetics az egyetlen olyan szoftver az irodalomban, amely az összes fontos folyamatot egyidejűleg veszi figyelembe.

1.1.5. Automatizálási lehetőségek

A tömegspektrometria területén – a műszeres analitikai technikák többségéhez hasonlóan – a mesterséges intelligencia módszerek a mérési folyamat három különböző szakaszában alkalmazhatók. Mindenekelőtt az egyre bonyolultabb és automatizáltabb készülékek mérés előtti állapotellenőrzésének és hangolási, kalibrálási fázisainak feladatait látják el. Második lépésben a tényleges mérési adatgyűjtés vezérlését végzik, végül pedig a mérési adatok több-kevesebb emberi közreműködést igénylő feldolgozását és kiértékelését. [9; 53; 54]

Az előbb említett mérési illetve készülékhasználati részfolyamatok közül a hardver vezérlése és hangolása, folyamatos kontrollja a mai nagyműszerek és ezen belül a tömegspektrométerek esetén már teljesen általános. Ezen funkciók legtöbbjéhez a felhasználónak már nincs is lehetősége manuálisan hozzáférni. A második rész, a tényleges mérési fázis szintén egyre inkább automatizált, a vizsgálandó minta előállításától kezdve emberi közreműködést egyre kevésbé igényel. A napjainkban kapható, kromatográfiai technikákkal kapcsolt tömegspektrométerek többsége már tartalmaz automata mintaváltót.

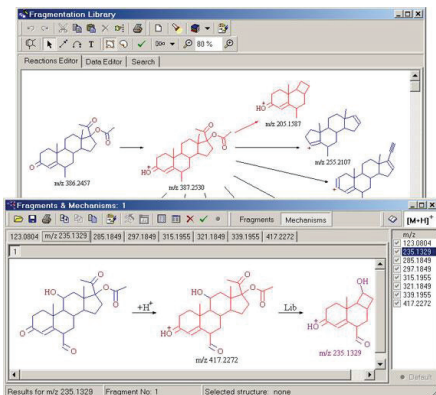
A legelhanyagoltabb fázis az automatizálás terén a kiértékelés, annak ellenére, hogy az automatikus mérésekkel előálló óriási mennyiségű adat kézi feldolgozása egyre lehetetlenebbé válik. Különböző adatfeldolgozási algoritmusokkal találkozhatunk az egyes alkalmazási területekre specializáltan, amelyek többsége valamilyen adatbányászati megoldással kapcsolatos. [55; 56; 57; 58; 59; 60; 61; 62; 63]

Tömegspektrum-értékelést támogató szoftverek

Napjainkban több kereskedelmi forgalomban kapható tömegspektrum-értékelést segítő szoftver létezik. Ezek nem elsősorban az automatikus értékelést, sokkal inkább a manuális kiértékelésnek a grafikus, kényelmi funkciókkal gazdagon ellátott támogatását tűzik ki célul. Felépítésüket tekintve három fő részre bonthatók.

1. Tartalmazzak egy *kémiai szerkezeteket rajzoló illetve megjelenítő modult*, általában valamilyen kémiai adatbázissal való kapcsolatot is biztosítva. A kémiai szerkezetek tárolása, visszakeresése és rajzolása természetesen nemcsak a tömegspektrometria területén fontos. Számos kémiai adatbáziskezelő, elektronikus „laborfüzet” kapható kereskedelmi forgalomban. (A teljesség igénye nélkül néhány kiragadott példa: CambridgeSoft: ChemOffice, MDL: Isis, Chemaxon: Jchem.)
2. Hasonló stílusú rajzoló, adatfeldolgozó és adatbázis-keresési funkciókat biztosítanak a vizsgálni kívánt *tömegspektrummal* kapcsolatban is. Ez a funkció egy másik szoftverkategóriához vezet, az ún. spektroszkópiai adatbázis szoftverekhez, illetve a LIMS (Laboratory Information Management System) rendszerekhez.
3. Végül pedig legfontosabb részük egy *fragmentációs utakat tartalmazó adatbázis*. Ez az adott szoftver minőségétől függően néhány egyszerű szabálytól a több tízezres szabály-adatbázisig terjedhet. Ezen szabályoknak a következetes alkalmazásával végzik el a kapott spektrum csúcsainak asszignációját, magyarázatát.

A harmadik funkciót is tartalmazó, tehát a tényleges spektrum-asszignációs munka szempontjából valóban használható eszközt már kevesebbet találhatunk. Kiemelkedő képviselőik a Thermo cég MassFrontier szoftvere, a ACD Labs MS Processor modulja, illetve újabban a Waters cég MassFragment programja. Ezek mindegyike egy több ezer fragmentációs szabályt találhatók könyvtár elemeit alkalmazza az adott molekula–spektrum pár esetén.



8. ábra. A Mass Frontier nevű spektrumkiértékelést támogató programrendszer fragmentációs könyvtárának egy darabja.

1.2. A felhasznált matematikai módszerek

1.2.1. Paraméterbecslések

A statisztikai becslést a következőképpen definiálhatjuk. Tekintsünk egy $f(x_1, x_2, \dots, x_n)$ n -változós függvényt. Ekkor egy $(\xi_1, \xi_2, \dots, \xi_n)$ statisztikai minta esetén az $S_n = f(\xi_1, \xi_2, \dots, \xi_n)$ kifejezést statisztikának nevezzük. Amennyiben θ egy ismeretlen paramétere a minta eloszlásának, célunk θ becslése S_n segítségével. [64; 7; 65]

Statisztikai becsléseket többféleképpen készíthetünk. Egyrészt megsejtés illetve kitalálás útján, másrészt valamilyen konstruktív módszerrel. A konstruktív módszerek közül a leggyakoribbak az úgynevezett maximum-likelihood (legnagyobb valószínűség) elvén működő módszerek, melyek lényege, hogy egy $f(x, \theta) \approx P_\theta(\xi = x)$ függvényt konstruálunk, ahol θ becslendő paramétere a ξ eloszlásnak, x pedig az adott minta. Ekkor ξ értékét n -szer függetlenül megfigyelve kapjuk az x_1, x_2, \dots, x_n konkrét mintát. Legyen

$$L(\theta) = f(x_1, \theta) \cdot f(x_2, \theta) \cdot \dots \cdot f(x_n, \theta). \quad (12)$$

$L(\theta)$ -t likelihood-függvénynek nevezzük, mivel annak a valószínűségét mutatja meg, hogy éppen az adott x_1, x_2, \dots, x_n mintát kapjuk. Legyen ezek után θ becslése az a paraméterérték, amelyre $L(\theta)$ maximális, ezt maximum-likelihood becslésnek nevezzük.

Lineáris regresszió

Adott x bemenő értékekhez egy $f(x)$ pontatlanul mért kimenő érték függvény, $y = f(x) + \epsilon$ mért értékkel, ahol ϵ a mérési hiba. A feladat x_1, x_2, \dots, x_n és Y_1, Y_2, \dots, Y_n esetén meghatározni $f(x)$ -et. Lineáris regresszióról beszélünk, ha a keresett függvény $f(x) = ax + b$ alakú. [65; 64; 7]

A statisztikai feladat a következő alakban írható le:

$$Y_i = ax_i + b + \epsilon_i, \text{ ahol } a \text{ és } b \text{ becsülendő paraméterek} \quad (13)$$

$$\epsilon_i \in N(0, \sigma^2), \quad i = 1, 2, \dots, n$$

A hibákra normálist eloszlást feltételezünk, ebből következően az Y_i értékek is normális eloszlásúak és Y_i sűrűségfüggvénye a következő alakú

$$f_{a,b}(Y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(Y-ax_i-b)^2}, \quad (14)$$

ebből következően az $L(a, b)$ likelihood-függvény

$$L(a, b) = f_{a,b}(Y_1) \cdot \dots \cdot f_{a,b}(Y_n) = \frac{1}{(\sqrt{2\pi})^n \sigma^n} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (Y_i - ax_i - b)^2} \quad (15)$$

A keresett a és b paraméterek meghatározásához $L(a, b)$ -t kell maximalizálni.

$$L(a, b) = \max \Leftrightarrow Q(a, b) = \sum_{i=1}^n (Y_i - ax_i - b)^2 = \min \quad (16)$$

Tehát a $Q(a, b)$ négyzetösszeget kell minimalizálnunk a és b függvényében. A $Q(a, b)$ kifejezés szemléletes jelentése ugyanakkor a mért Y_i adatok függőleges távolsága az $ax + b$ regressziós egyenestől.

Deriválással és átrendezésekkel adódik a és b becslése:

$$a = r \cdot \sqrt{\frac{V_n(Y)}{V_n(x)}}, \quad (17)$$

ahol $r = \frac{C_n(x, Y)}{\sqrt{V_n(Y)}\sqrt{V_n(x)}}$ az empirikus korreláció,

$$V_n(x) = \frac{1}{n} \sum_{i=1}^n x_i^2 - \left(\frac{1}{n} \sum_{i=1}^n x_i \right)^2, \quad V_n(Y) = \frac{1}{n} \sum_{i=1}^n Y_i^2 - \left(\frac{1}{n} \sum_{i=1}^n Y_i \right)^2, \quad (18)$$

$$C_n(x, Y) = \frac{1}{n} \sum_{i=1}^n x_i Y_i - \left(\frac{1}{n} \sum_{i=1}^n x_i \right) \left(\frac{1}{n} \sum_{i=1}^n Y_i \right), \quad (19)$$

$$b = \frac{1}{n} \sum_{i=1}^n Y_i - a \frac{1}{n} \sum_{i=1}^n x_i. \quad (20)$$

Megfigyelhető, hogy a levezetés során feltételeztük, hogy csak az Y értékeknek van mérési hibája, tehát az x értékek rögzítettek. A fenti lineáris regressziós formulákat használhatjuk különböző – általában az Y értékek hibája szerint – súlyozott formákban is.

ODR – ortogonális távolság regresszió

A fentiekkel ellentétben legyen most a kiinduló feltételezésünk a következő:

$$(X_i, Y_i), \quad i = 1, 2, \dots, n \quad (21)$$

a megfigyelt valószínűségi változók

$$(x_i, y_i), \quad i = 1, 2, \dots, n \quad (22)$$

valós értékekkel, tehát

$$X_i = x_i - \delta_i, \quad (23)$$

$$Y_i = y_i - \epsilon_i.$$

Továbbá feltételezzük még, hogy y megadható x és valamely $\boldsymbol{\beta}$ paraméter vektor függvényében, vagyis

$$y_i = f(x_i, \boldsymbol{\beta}), \quad (24)$$

$$Y_i = f(X_i + \delta_i, \boldsymbol{\beta}) - \epsilon_i.$$

Mivel mind az X , mind pedig az Y mért értékeknek van hibája, ezért bármely korrekten becsülnek ezek mindegyikét figyelembe kell vennie. Ennek a problémának a megoldása érdekében az (X_i, Y_i) mért adatpontnak definiáljuk az $f(x, \boldsymbol{\beta})$ görbétől való *ortogonális* távolságát (r_i) a következőképpen:

$$r_i^2 = \min\{\epsilon^2 + \delta^2\}, \quad i = 1, 2, \dots, n. \quad (25)$$

Amennyiben a δ és ϵ hibák 0 körüli normális eloszlásúak, akkor a lineáris regressziós részben ismertetett levezetéshez hasonlóan, a $\boldsymbol{\beta}$ paraméter vektor maximum likelihood

becslése a definiált r_i^2 ortogonális távolságok négyzetösszegének minimalizálásával egyenértékű.

A lineáris regressziós módszer és az ortogonális lineáris regresszió egyetlen, de lényeges, elvi különbsége, hogy ez utóbbi az x értékek hibáját is figyelembe veszi. Levezetésük módja és statisztikai hátterük azonos, mindkettő az adott probléma maximum likelihood becsléseket adódik. A hagyományos lineáris regresszió alkalmazásakor ugyanakkor egyszerűségének és az x értékek általában elhanyagolható hibájának köszönhetően lényegesen elterjedtebb. [66; 67]

1.2.2. Numerikus szélsőérték-keresési technikák

A legegyszerűbb megfogalmazásban az általános feladat a következő. Adott egy f függvény, egy vagy több független változóval. A kérdés a független változóknak az a vektora, ahol az f függvény – egy adott változó-tartományon belül – a maximális vagy minimális értékét felveszi. A maximum- illetve minimum-keresés nyilvánvalóan ekvivalens feladat, a függvény negálásával egyik visszavezethető a másikra. A numerikus szélsőérték-keresési technikák létrejöttét az indokolta, hogy azokra a függvényekre, ahol a szélsőérték-keresési feladat analitikus formában nem adható meg – például azért mert már maga a függvény sem adható meg egy zárt analitikus alakban –, a számítógépek fejlődésével ott is lehetőség volt ezeknek az extrémumoknak gyors és pontos kiszámítására. [68; 7; 8; 69]

Az előbb felvázolt feladat tovább bontható: beszélhetünk globális és lokális szélsőértékekről. Első esetben a teljes, a feladat szempontjából lényeges tartományon keressük a függvény szélsőértékét, míg a második esetben csak egy lokális maximumra vagy minimumra vagyunk kíváncsiak, vagyis egy olyan pontra, amelynek létezik olyan lokális környezete, amin belül f -nek szélsőértéke van.

A globális szélsőérték-keresésnek alapvetően két heurisztikus módja van.

1. Változó pontokból indulva keressünk lokális extrémumokat, majd válasszuk ezek közül a legkisebbet / legnagyobbat. A kiindulási pontok megválasztása ilyenkor általában valamilyen kvázi-véletlenszerű algoritmussal történik.

2. Egy adott lokális extrémum helyétől távolodjunk el egy adott véges lépéssel és vizsgáljuk meg, hogy a lokális szélsőérték-keresés ugyanoda jut-e vissza, vagy más eredményre vezet.

Egyik szélsőérték-keresési algoritmus sem nevezhető ki jobbnak, mint a többi. Minden esetben az adott probléma függvényében érdemes választanunk köztük. A következő néhány pont összefoglalja a legfontosabb választási szempontokat.

1. Érdemes figyelembe venni, hogy van-e lehetőségünk a függvényérték számításán túl a deriváltak használatára is. A deriváltakat is figyelembe vevő algoritmusok általában hatékonyabbak, ugyanakkor csak megfelelő óvatossággal használhatók, ha a derivált-függvényeknek szakadási helyeik vannak. Ekkor egy egyszerű aranymetszéses módszer is jobb megoldás lehet.
2. Többdimenziós esetben fontos lehet az alkalmazott algoritmus tárigénye ($\sigma(N)$ vagy $\sigma(N^2)$ ahol N a dimenziók száma). [70]
3. Speciális függvényalak (pl. lineáris) esetén, illetve bonyolult függvényalak globális optimum-keresése esetén az adott célra fejlesztett algoritmusokat érdemes alkalmazni.

A következő részben rövid összefoglalását adom a munkám során használt optimumkeresési módszereknek: Powell többdimenziós szélsőérték-keresési algoritmusának és a sokdimenziós esetben globális optimum keresésére alkalmas szimulált fagyasztás (simulated annealing) módszerének.

Többdimenziós szélsőérték-keresés, Powell módszere

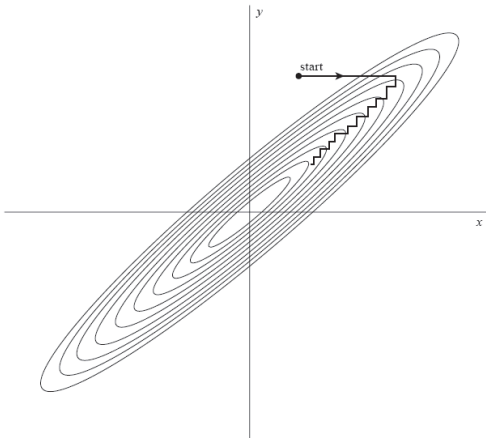
Induljunk ki abból az alapfeltételezésből, hogy létezik hatékonyan működő egydimenziós optimumkereső módszerünk. Ha egy N -dimenziós térben egy adott \mathbf{P} pontból elindulunk valamely \mathbf{n} vektor irányába, akkor egy tetszőleges $f(\mathbf{P})$ függvény optimalizálható egydimenziós módszerekkel \mathbf{n} egyenes mentén.

A többdimenziós optimumkeresési módszerek egy jelentős része egydimenziós optimalizálás egymásutánjából áll, az irányok megfelelő egymás utáni megválasztásával. Ezeket a módszereket osztályozhatjuk annak függvényében, hogy használnak-e gradiens számítást az irányok kiválasztásában vagy sem. Nyilvánvaló, hogy ha az egydimenziós részalgoritmusunkban szükség van gradiens számítására, lehet azt használni az irányok megválasztásánál is. Amennyiben azonban az egydimenziós algoritmus nem ilyen és a

függvény deriváltjának kiszámítása költséges, akkor az irányok megválasztásánál is más módszert kell választanunk.

Legegyszerűbb esetben mondhatjuk azt, hogy amennyiben az N -dimenziós tér egységvektorai $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_N$, akkor válasszuk ezeket a megfelelő irányoknak, tehát végezzünk rendre egydimenziós optimalizálást az \mathbf{e}_i egységvektorok mentén, majd kezdjük újra, mindaddig, amíg el nem értünk az optimumba. A módszer egyszerűsége ellenére meglepően jól működik, kivéve ha az f függvény keskeny völgyeket tartalmaz valamely az egységvektorokkal nem megegyező irányba. Ekkor ugyanis a völgyben lefelé csak nagyon apró lépések sorozatával haladhatunk, tehát N növekedésével arányosan nő a – fölösleges – lépések száma is, ahogyan azt a 9. ábra mutatja.

Megfelelőbb irányok választásához juthatunk, ha sikerül olyan irányt kiválasztani, amely mentén a minimumba eljutva az f függvény gradiense ott merőleges erre az irányra.



9. ábra. Sorozatos egydimenziós minimumkeresés x és y irányokban egy nem-tengelyirányú keskeny völgyben. [8]

Becsüljük f -et a Taylor-sorával:

$$f(\mathbf{x}) = f(\mathbf{P}) + \sum_i \frac{\partial f}{\partial x_i} x_i + \frac{1}{2} \sum_{i,j} \frac{\partial^2 f}{\partial x_i \partial x_j} x_i x_j + \dots \approx c - \mathbf{b}\mathbf{x} + \frac{1}{2} \mathbf{x}\mathbf{A}\mathbf{x}, \quad (26)$$

ahol $c = f(\mathbf{P})$, $\mathbf{b} = -\nabla f|_{\mathbf{P}}$, $[A]_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j} \Big|_{\mathbf{P}}$.

\mathbf{u} és \mathbf{v} két megfelelő irány ekkor, amennyiben

$$0 = \mathbf{u} \cdot \delta(\nabla f) = \mathbf{u} \cdot \mathbf{A} \cdot \mathbf{v}. \quad (27)$$

Ebben az esetben konjugált irányoknak hívjuk őket. Feladatunk N dimenzióban megtalálni azokat az irányokat, amelyek páronként konjugáltak. Ekkor ezek mentén N lépés egydimenziós optimalizálást végrehajtva kvadratikusan függvény esetén a pontos minimumot kapjuk. Természetesen nem minden f függvény kvadratikusan, ekkor újabb iterációra lehet szükség, ám mindenképpen elmondhatjuk, hogy kvadratikusan közelítjük az optimumot.

Powell módszere a következőképpen adható meg. Tekintsük a következő eljárást.

1. Vegyük az optimalizálás \mathbf{u}_i irányvektorainak az N -dimenziós bázisvektorokat, kezdőpontnak pedig \mathbf{P}_0 -t.
2. Rendre keressük meg az \mathbf{u}_i irányokban a minimumokat, lépünk oda és nevezzük \mathbf{P}_i -nek.
3. 1-től $N - 1$ -ig legyen $\mathbf{u}_i = \mathbf{u}_{i+1}$, végül $\mathbf{u}_N = \mathbf{P}_N - \mathbf{P}_0$.
4. \mathbf{u}_N mentén vigyük \mathbf{P}_N -et a minimumba, legyen ez az új \mathbf{P}_0 .

Bizonyítható, hogy k egymást követő végrehajtása az előbbi eljárásnak egy olyan \mathbf{u}_i irányosorozatot ad meg, amelynek utolsó k tagja páronként konjugált. Tehát N -szeri végrehajtása – összesen $N(N + 1)$ egydimenziós minimumkeresés – után pontosan megadja egy kvadratikusan függvény minimumát.

A módszer hibája, hogy az \mathbf{u}_1 értékek sorozatos eldobása után a kapott irányosorozat lineárisan összefüggő, tehát a minimumot nem a teljes térben, csak annak egy ezek által kifeszített alterében adja meg. Ennek kiküszöbölésére több módszer is létezik, ilyen például, ha minden ciklus elején újra a bázisvektor irányokkal kezdünk. [8]

Globális optimalizálások, szimulált fagyasztás módszere

A szimulált fagyasztás módszere igen hatékonynak bizonyult olyan sokdimenziós problémák kezelésében, ahol olyan globális optimum keresése a feladat, amely számos egyéb lokális optimum közül választandó ki. A módszer elsőként adott hatékony megoldást az „utazó ügynök problémára”, amelyben egy N szögpontú súlyozott gráfban keressük a

legrövidebb minden szögpontot érintő kört. Az utazó ügynök probléma tipikus képviselője a kombinatorikus optimalizálás feladatkörének.

A szimulált fagyasztás módszere elsősorban diszkrét, de igen nagy konfigurációs terű feladatok esetén bizonyul hatékonynak. Az ügynökprobléma esetén a vizsgálandó konfigurációs tér $N!$ nagyságrendű [70]. Az ilyen típusú problémák esetén a többi általános sokdimenziós algoritmus nem használható, a legmeredekebb csökkenés iránya és a hasonló keresési feltételek értelmüket veszítik. A szimulált fagyasztás módszere ugyanakkor átülthető folytonos paraméterteret esetére is.

Nevéhez híven a módszer a termodinamikai szimulációkkal hozható közeli kapcsolatba. Szemléletes fizikai képet mutatva, ahogyan a folyadékok egy folytonos fázisteret szabadon bejárnak, majd a hőmérséklet lassú, fokozatos csökkenésével befagynak egy adott konfigurációba, úgy a szélsőérték-kereső módszer is kezdetben könnyen bejárja a paraméterteret, majd egyre inkább a minimum környékére korlátozódik, míg végül egy pontra „fagy be”. A módszer leglényegesebb eleme tehát a lassú „hőmérséklet” csökkentés.

Legyen adott egy minimalizálandó E függvény.

1. Ekkor az adott pontban kiszámítjuk az értékét, majd a paraméterter valamely másik – szomszédos – pontjára léptetve, ott is megismételjük a függvényérték számítását.
2. Az új pontot elfogadjuk, ha kisebb függvényértéket kapunk az előzőnél.
3. p valószínűséggel elfogadjuk továbbá akkor is, ha az új függvényérték nagyobb a réginél.

Az elfogadási kritérium a 3. pont esetén

$$p = \exp \left[-\frac{(E_2 - E_1)}{kT} \right], \quad (28)$$

ahol E_1 és E_2 rendre a régi és az új függvényérték, a kT tag pedig a Boltzmann eloszlással analóg „hőmérséklet” jellegű paramétert képviseli. (Megjegyzés: $E_2 < E_1$ esetén a p érték nyilvánvalóan nagyobb 1-nél, tehát az előbbi 2. pont tulajdonképpen nem képvisel külön feltételt.)

Ezt az algoritmust először 1953-ban Metropolis és munkatársai implementálták numerikus számításaikban, a későbbiekben pedig mint Metropolis Monte Carlo módszer terjedt el.

A módszer nem-termodinamikai szimulációkban való alkalmazhatóságához a következő feltételeket kell teljesítenünk.

1. Meg kell adnunk a lehetséges konfigurációk általános leírását.
2. Hatékonyan megoldani a véletlenszerű új lépések kiválasztását.
3. Kiszámítani egy – az energiával analóg – célfüggvényt.
4. Létrehozni egy – hőmérséklet jellegű – T paramétert és annak egy az adott feladathoz alkalmazkodó csökkenési ütemét.

Az előbbi négy feltételből a másodiknál lényeges szerepet játszik az, hogy a véletlenszerű kiválasztások valóban bejárják a lehetséges konfigurációs teret, a negyedik esetén pedig a feladatban szereplő függvény „alakjához” illeszkedő befagyasztási sebesség. Természetesen ez utóbbira – teljesen ismeretlen függvényalakok esetén – használhatunk valamilyen előre definiált lassú csökkentést, ám ekkor az optimumkeresés igen lassan fog konvergálni.

Összefoglalva elmondható, hogy a kevés dimenziós esetekben, kevés lokális szélsőérték esetén az egydimenziós optimalításokra visszavezethető módszerek előnyt élveznek, míg sokdimenziós – kombinatorikus típusú – problémák esetén, globális minimum keresésére a szimulált fagyasztás módszerét érdemes alkalmazni. [8]

1.2.3. A gépi tanulás

A gépi tanulás a mesterséges intelligenciának egy kiterjedt részterületét képviseli, olyan algoritmusok és módszerek megalkotását tűzve ki célul, amelyek felruházzák a számítógépeket a tanulás képességével. Tanulási folyamatuk a legtöbb esetben induktív, tehát a feldolgozandó adattömegből különböző szabályok és mintázatok automatikus felismerését végzik változatos számítástechnikai és statisztikai módszerek alkalmazásával. A gépi tanulás emiatt szoros kapcsolatban áll az adatbányászat és a statisztika vizsgálati módszereivel. [71; 72]

Alkalmazási területei közül kiemelendők a természetes nyelvek feldolgozása, beszédfelismerés, orvosi diagnosztika, bioinformatika, kézírás- és általában képfelismerés és nem utolsósorban a számítógépes játékok és robotok vezérlése. [73; 74; 75; 76; 77]

A leggyakrabban alkalmazott típusokat elsősorban a tanulás módja szerint szokták osztályokba sorolni.

- *Felügyelt tanulás.* Az algoritmus olyan „függvényt” keres, amely a bemenő adatokat legjobb közelítéssel képezi a tanuló példák elvárt eredményeire.
- *Felügyelet nélküli tanulás.* Lényegében a bemenő adatok osztályozását végzi el, hasonlóságuk szerint. Az algoritmus tanulópéldáihoz nincsenek elvárt kimeneti értékek.
- *Félig felügyelt tanulás.* A tanulópéldák vegyesen tartalmaznak várt eredményekkel rendelkező és anélküli bemeneteket.
- *Megerősítéses tanulás.* Az algoritmus egy adott környezetben való „viselkedést” tanul. Minden lehetséges reakciója valamilyen hatása bír a környezetére, a környezet egy folyamatos visszacsatolást ad a tanuló algoritmusnak.
- *Transzdukció.* A felügyelt tanulásnak egy változata, annyi eltéréssel, hogy az algoritmus nem egy adott leképezést tanul, csupán a tanulópéldák alapján további eredményeket próbál becsülni.
- *Tanulni tanulás.* Az algoritmus az indukcióhoz szükséges mintáknak a keresését maga végzi előzetes tapasztalatai alapján.

Szakértői rendszerek

A szakértői rendszerek, vagy más néven tudás alapú rendszerek, olyan számítógépes programok, amelyek valamilyen területen analitikus képességekkel, szaktudással rendelkeznek. Elsődleges célja az emberi döntések támogatása adott munkafolyamatokban. Tipikus alkalmazási területük az orvosi diagnosztika, gazdasági elemzések és gyártásközi ellenőrzések.

A szakértői rendszerekkel kapcsolatos legfontosabb fogalom a következtetési szabály. Ez a legegyszerűbb esetben egy „*ha X akkor Y*” típusú következtetési szabály. Működés közben a következtető motor, ilyen illetve ehhez hasonló szabályok segítségével keres választ a feltett kérdésekre, illetve tanítás esetén ilyen formájú szabályok indukciója történik.

A következtető motor jellegzetessége, hogy általában független az adott szakterületől, bár néha szükség lehet specifikus következtetési módokra. Ez a tulajdonság egyébként a döntő különbség a szakértői rendszerek és a hagyományos probléma-megoldó rendszerek között. Míg ez utóbbiakban az adott szakmai ismeretek, tudásanyag többé-kevésbé beágyazódik az algoritmusokba, a programkódba, addig a szakértői rendszerek esetén a következtető

algoritmus az esetek döntő többségében ettől független, a szaktudást a következtetési szabályok hordozzák.

A szakértői rendszerek két hagyományos képviselője a Dendral és a Mycin. [78; 79; 80] A Dendral az 1960-as években éppen tömegspektrumok elemzésére kidolgozott rendszer volt. Feladata ismeretlen szerves molekulák azonosítása volt tömegspektrumaik és bizonyos kémiai tudás ismeretében. A belőle származtatott rendszerek egyike volt a Mycin, amelyet az 1970-es években orvosi diagnosztikai feladatra, vérmérgezés diagnosztizálására és megfelelő antibiotikum kezelés kiválasztására dolgoztak ki. A későbbiekben általam kidolgozott algoritmus is a Dendral-lal mutat rokonságot.



10. ábra. A tömegspektrumok értékelésére kidolgozott HEURISTIC DENDRAL szakértői rendszer folyamatábrája. [79]

Valószínűségi szabályok indukciója, intervallum-valószínűségek

Az orvosi diagnosztikában néhány évvel (évtizeddel) ezelőtt terjedtek el a különböző betegségek diagnosztizálására és javasolt terápia megadására megalkotott szakértői rendszerek (MYCIN, INFERNO...). Ezek közös tulajdonsága, hogy valamilyen – az addigi esetek statisztikáiból leszűrt – valószínűségi szabályok alkalmazásával az adott tünetekhez képesek megbecsülni a felállított diagnózisok helyességének valószínűségét. A számítások általában Bayes tételét alkalmazzák. [81]

A fix szabálykészlettel dolgozó módszerek mellett léteznek olyan algoritmusok is, amelyek ilyen valószínűségi szabályok tanulását és további finomítását végzik. Ezekben a szabályok képzését megadó módszeren túl jellemző a szabályok finomítását és az ellentmondó illetve redundáns szabályok kiszűrését végző, ezáltal a szabályadatbázis méretét csökkentő algoritmusrészlet jelenléte is. [82]

A bizonytalan következtetések kezelésének és reprezentálásának másik lehetősége az egzaktul megadott valószínűség-értékek helyett valószínűségi intervallumok használata. Az intervallum-használat egyik motivációjának azt tartják, hogy az emberi ítéletek jellegéhez és bizonytalanságához közelebb áll, mint a hagyományos egyértékű valószínűségi modell. I. J. Good az egyértékű valószínűségi elmélet mintájára megalkotott egy olyan axiómahalmazt az intervallumvalószínűségek elméletének felépítésére. [83; 82; 84]

2. A kísérletek és számítások részletei

2.1. SORI-CID fragmentáció modellezése

A hőmérsékletfüggő SORI kísérleteket Guo és társai végezték egy módosított Bruker APEX 7.0e FT-ICR tömegspektrométeren. A kísérlet részletes leírása megtalálható Guo és társai cikkében, a következőkben csak a modellezés szempontjából fontos részletekre térek ki. [85; 86]

A vizsgált modellrendszer leucin-enkefalin (YGGFL), ütköző gázként argont használva. A mérés során a leucin-enkefalin protonált molekulaionját (m/z 556) izolálták és néhány másodpercnyi termikus ekvilibráció után SORI gerjesztésnek vetették alá különböző cellahőmérsékletek esetén a 143-296 K hőmérséklet-tartományban. Off-rezonáns RF gerjesztést alkalmaztak $\Delta\nu = -1000$ Hz frekvencia-eltéréssel $V_{p-p} = 2,25, 2,75, 3,00, 3,25$ V gerjesztő potenciállal. A felsorolt értékek a 6. egyenlet alapján 3,9, 5,9, 7,0 és 8,2 eV kinetikus energiáknak felelnek meg laboratóriumi vonatkoztatási rendszerben. A mérés során a molekulaion %-os intenzitását követték az idő, azaz a SORI ciklusok függvényében. Minden mérés során Ar ütközőgázt használtak 5 másodpercig, 5×10^{-6} mbar maximális nyomással. A következő táblázatban a három egymástól függetlenül változó paraméter – idő, hőmérséklet és kezdő kinetikus energia – terében felvett kísérleti adatokat foglaltam össze. Fontos megjegyezni, hogy a későbbiekben tárgyalt számítások és illesztések minden esetben a teljes adatkészletre történtek, így biztosítva azt, hogy a modellezés más kísérleti körülmények esetén is megfelelő eredményt szolgáltatson.

A hőmérsékletfüggő SORI-CID mérések eredményeit – a mért molekulaion-intenzitát (%) az idő, kinetikus energia és hőmérséklet függvényében – a MassKinetics programmal [47; 48] modelleztem. Érdemes megjegyezni, hogy az alkalmazott – a SORI-CID esetén tipikus – kísérleti körülmények az alacsony-nyomású tartományba esnek, tehát SORI RF-ciklusonként átlagosan egynél kevesebb ütközés történt. [87] Így a modellezés során használt, később részletesen tárgyalt összefüggésekben is az alacsony-nyomású közelítést használtam. [18]

2. táblázat. A hőmérséklet, kinetikus energia és időfüggő SORI kísérletek mérési eredményei. [85]

Hőmérséklet (K)	Várható kinetikus energia (eV)	Gerjesztési idő (s)	Bomlatlan molekulaionok aránya
296	2,60	0,01	0,987
296	2,60	0,05	0,986
296	2,60	0,10	0,985
296	2,60	0,15	0,976
296	2,60	0,20	0,965
296	2,60	0,25	0,951
296	2,60	0,30	0,945
296	2,60	0,35	0,941
296	2,60	0,40	0,939
296	2,60	0,45	0,932
296	2,60	0,50	0,924
296	2,60	0,60	0,910
296	2,60	0,80	0,872
296	2,60	1,20	0,793
296	2,60	1,60	0,724
296	2,60	2,00	0,650
296	2,60	3,00	0,513
296	2,60	4,00	0,394
296	2,60	5,00	0,293
296	2,60	6,00	0,163
296	3,93	0,06	0,968
296	3,93	0,09	0,951
296	3,93	0,12	0,945
296	3,93	0,15	0,930
296	3,93	0,18	0,919
296	3,93	0,21	0,900
296	3,93	0,24	0,860
296	3,93	0,27	0,826
296	3,93	0,30	0,773
296	3,93	0,33	0,724
296	3,93	0,36	0,656
296	3,93	0,38	0,634
296	3,93	0,40	0,598
296	3,93	0,41	0,594
296	3,93	0,42	0,563
296	3,93	0,43	0,564
296	3,93	0,44	0,554
296	3,93	0,46	0,508
296	3,93	0,48	0,488
296	3,93	0,50	0,455
296	3,93	0,53	0,407
296	3,93	0,56	0,364
296	3,93	0,60	0,325
296	3,93	0,65	0,276
296	3,93	0,70	0,222
296	3,93	0,75	0,204

Hőmérséklet (K)	Várható kinetikus energia (eV)	Gerjesztési idő (s)	Bomlatlan molekulaionok aránya
296	3,93	0,80	0,187
296	3,93	0,90	0,124
296	3,93	1,00	0,089
296	3,93	1,10	0,060
296	4,67	0,05	0,962
296	4,67	0,07	0,952
296	4,67	0,09	0,940
296	4,67	0,11	0,933
296	4,67	0,13	0,912
296	4,67	0,15	0,879
296	4,67	0,17	0,841
296	4,67	0,19	0,783
296	4,67	0,21	0,738
296	4,67	0,23	0,658
296	4,67	0,25	0,598
296	4,67	0,27	0,543
296	4,67	0,28	0,504
296	4,67	0,29	0,467
296	4,67	0,30	0,448
296	4,67	0,31	0,432
296	4,67	0,32	0,388
296	4,67	0,33	0,362
296	4,67	0,34	0,354
296	4,67	0,35	0,315
296	4,67	0,36	0,293
296	4,67	0,38	0,252
296	4,67	0,40	0,218
296	4,67	0,43	0,169
296	4,67	0,46	0,128
296	4,67	0,50	0,090
296	4,67	0,55	0,055
296	4,67	0,60	0,036
296	4,67	0,65	0,024
296	4,67	0,70	0,014
296	5,47	0,03	0,969
296	5,47	0,05	0,947
296	5,47	0,07	0,938
296	5,47	0,09	0,909
296	5,47	0,11	0,876
296	5,47	0,12	0,836
296	5,47	0,13	0,803
296	5,47	0,14	0,775
296	5,47	0,15	0,727
296	5,47	0,16	0,674
296	5,47	0,17	0,649
296	5,47	0,18	0,587
296	5,47	0,19	0,573
296	5,47	0,19	0,521

Hőmérséklet (K)	Várható kinetikus energia (eV)	Gerjesztési idő (s)	Bomlatlan molekulaionok aránya
296	5,47	0,20	0,513
296	5,47	0,20	0,497
296	5,47	0,21	0,455
296	5,47	0,21	0,448
296	5,47	0,22	0,416
296	5,47	0,22	0,391
296	5,47	0,23	0,378
296	5,47	0,23	0,356
296	5,47	0,24	0,323
296	5,47	0,25	0,277
296	5,47	0,26	0,238
296	5,47	0,28	0,190
296	5,47	0,30	0,150
296	5,47	0,32	0,108
296	5,47	0,34	0,084
296	5,47	0,36	0,059
183	4,67	0,05	0,974
183	4,67	0,10	0,945
183	4,67	0,20	0,909
183	4,67	0,30	0,740
183	4,67	0,35	0,632
183	4,67	0,40	0,509
183	4,67	0,45	0,415
183	4,67	0,50	0,329
183	4,67	0,55	0,264
183	4,67	0,60	0,217
183	4,67	0,65	0,174
183	4,67	0,70	0,138
183	4,67	0,80	0,085
183	4,67	0,90	0,055
183	4,67	1,00	0,035
183	4,67	1,20	0,012
143	4,67	0,05	0,971
143	4,67	0,10	0,953
143	4,67	0,20	0,932
143	4,67	0,30	0,810
143	4,67	0,35	0,736
143	4,67	0,40	0,639
143	4,67	0,45	0,553
143	4,67	0,50	0,454
143	4,67	0,55	0,364
143	4,67	0,60	0,305
143	4,67	0,65	0,233
143	4,67	0,70	0,186
143	4,67	0,80	0,124
143	4,67	0,90	0,084
143	4,67	1,00	0,055
143	4,67	1,20	0,025

A számításokhoz használt kritikus energia és pre-exponenciális faktor értékeket Williams és csoportja munkájából vettem, míg a szükséges rezgési frekvenciákat kvantumkémiai számításokból (B3LYP szint, 6-31G*(d) bázisfüggvény-készlet). [88; 89; 90] A felhasznált rezgési frekvenciákat a 3. táblázat mutatja.

A mért és számított molekulaion-arány illesztésével az ütközéses energiáttranszfer (η) és az infravörös hűlési sebesség (k_{cool}) értékeire adtam becslést. Az első modellezéseket a globális optimum keresése miatt a szimulált fagyasztás módszerével végeztem, de a későbbiekben kipróbált Powell módszer gyorsabbnak bizonyult amellet, hogy megbízható eredményeket adott. [8] Fontos megjegyezni, hogy az ütközési energia-transzfer nem egy átlagos értéket jelent, hanem egy eloszlást az ütközési energia függvényében. [91]

Háromféle energia szerinti eloszlás hatását vizsgáltam, az Armentrout által kísérletileg meghatározott [92], a trajektória számításokon alapuló Hase-félét [93] és a legtöbb esetben használt egyszerű exponenciális eloszlást. A számítások többségében Hase függvényalakját használtam. A használt ütközési modell feltételezi továbbá, hogy az ütközési energia-transzfer arányos az ütközési energiával.

Az infravörös hűlést a MassKinetics algoritmusában leírt módon vettem figyelembe, Dunbar modelljével dolgozva, a következő összefüggéssel:

$$E_{\text{excess}}(t) = E_{\text{excess}}(0) \times \exp(-k_{\text{cool}}t), \quad (29)$$

ahol $E_{\text{excess}} = E_{\text{int}} - E_{\text{therm}}$ az ionok többlet-energiája (az adott hőmérséklethez tartozó termikus energia feletti belső energia), k_{cool} a hűlés sebességét leíró empirikus konstans, t pedig az idő. (Érdemes megjegyezni, hogy a fenti egyenlet egyaránt alkalmas a sugárzásos hűlés és gerjesztés leírására E_{excess} előjelétől függően.) Dunbar kísérleteivel összhangban a k_{cool} paraméter értéke függ a vizsgált molekula szerkezetétől, de a mérettől nem (a molekula szabadsági fokával normált). Így hasonló szerkezetek egyúttal hasonló k_{cool} értékeket is jelentenek.

3. táblázat. A modellezés során használt számított leucin-enkefalin rezgési frekvenciák. (B3LYP, 6-31G*(d))

Leucin-enkefalin számított rezgési frekvenciái (cm ⁻¹)					
5,10	359,63	796,00	1140,11	1479,58	3102,52
10,10	376,65	808,30	1167,16	1484,50	3108,13
19,44	378,92	824,25	1175,34	1497,02	3113,52
30,07	386,07	829,90	1189,54	1508,30	3117,78
31,92	399,69	841,09	1207,10	1510,40	3120,29
38,04	404,98	853,41	1218,08	1531,58	3124,39
44,60	409,85	855,48	1221,13	1537,12	3144,98
52,46	420,70	857,24	1225,99	1541,03	3151,13
59,75	424,26	858,97	1228,83	1544,84	3156,71
68,93	428,41	876,09	1232,42	1547,89	3157,45
76,15	436,28	900,11	1234,81	1553,12	3180,70
82,16	438,84	906,14	1238,89	1555,91	3182,21
84,86	440,88	931,21	1239,80	1559,96	3184,08
91,60	451,40	938,28	1244,94	1561,22	3195,24
98,45	487,26	940,84	1246,93	1568,52	3197,44
110,76	505,63	962,94	1260,26	1570,06	3200,55
115,52	525,11	965,29	1267,88	1572,00	3203,54
134,89	541,03	973,52	1273,38	1585,86	3210,54
138,13	549,79	980,44	1287,56	1608,92	3218,42
146,00	565,67	991,40	1297,09	1610,68	3226,12
161,42	571,88	996,61	1310,84	1616,92	3239,04
168,23	602,41	999,15	1315,89	1627,00	3279,22
175,45	604,38	1002,99	1318,87	1651,34	3366,84
180,45	608,61	1010,32	1330,99	1653,26	3382,64
188,86	624,80	1018,99	1340,74	1674,56	3524,36
201,47	655,71	1027,48	1345,60	1700,87	3524,46
218,15	659,04	1029,82	1360,75	1712,59	3547,70
233,24	660,05	1033,13	1362,83	1727,22	3555,84
234,48	663,59	1038,67	1366,25	1756,29	
240,57	674,80	1046,34	1368,26	1792,30	
254,88	678,44	1049,03	1369,77	2949,36	
260,42	689,73	1049,58	1370,44	3038,59	
271,57	703,37	1053,16	1372,38	3040,72	
285,97	750,87	1060,92	1394,43	3045,88	
292,26	757,12	1066,10	1400,50	3052,98	
303,71	757,65	1076,15	1401,83	3054,34	
308,66	762,49	1100,09	1412,03	3084,60	
315,48	767,69	1114,06	1413,69	3087,26	
321,41	788,75	1131,20	1433,47	3100,01	
338,20	793,45	1135,66	1460,17	3101,24	

2.2. A kinetikus módszer hibájának vizsgálata

2.2.1. Virtuális kísérleti adatkészletek generálása

A kinetikus módszerrel kapcsolatos vizsgálatokhoz a bemenő „kísérleti” adatokat MassKinetics szimulációk segítségével generáltam. Az eredmények statisztikai értékeléséhez több ezer különböző mérési körülmények között elvégzett kísérlet kiértékelése szükséges, pontosan ismert/definiált entalpia és entrópia különbségekkel. A MassKinetics algoritmusával alkalmas a tömegspektrometriás kísérletek megfelelő pontosságú reprodukálására [47], ezért a tényleges kísérleti adatok helyett kiértékelhető az így kapott mesterséges tömegspektrum is. Munkám során módszert dolgoztam ki arra, hogy automatikusan generált kísérleti körülmények esetén lehetőség legyen a MassKinetics algoritmusával becsült tömegspektrumokból elvégezni a kinetikus módszer szerinti kiértékelést, majd a kapott eredményeket statisztikusan értékelni.

A kiterjesztett kinetikus módszer segítségével meghatározva a $\Delta H_{\text{kinetic}, 298\text{K}}$ és $\Delta S_{\text{kinetic}, 298\text{K}}$ értékeit, a becsült paraméterek és az elméleti – jelen esetben egzaktul ismert – $\Delta H_{\text{theor}, 298\text{K}}$ és $\Delta S_{\text{theor}, 298\text{K}}$ közötti összefüggéseket vizsgáltuk. A kezdeti feltételek véletlen hibájának figyelembevétele az ion-intenzitások mérési hibájának és a termokémiai paraméterek hibájának mesterséges előállításával történt. A kiterjesztett kinetikus módszer kiértékelésekor mind az Ervin és Armentrout által felvetett ortogonális lineáris regressziós, mind pedig a hagyományosan használt lineáris regressziós kiértékeléseket elvégeztem.

Többféle ion-molekula rendszert tanulmányoztunk, a komplexek mérete 70 és 300 szabadsági fok között változott, ennek 100-500 Da tömegtartomány felelt meg. Az aktiválási energiák az 1,0-1,3 eV tartományban változtak, míg az entrópia változások a $-80..+100 \text{ J mol}^{-1}\text{K}^{-1}$ tartományban. Többféle átmeneti állapotot használtunk, közelítőleg pre-exponenciális faktorokkal jellemezve a 10^{10} - 10^{20} tartományban. A generált mesterséges adatkészletekben olyan extrém molekuláris paraméterek is szerepeltek, amelyeket kísérleti úton nehéz lenne vagy egyáltalán nem lehet vizsgálni. Összességében elmondható, hogy a kinetikus módszer jelen tesztelésekor lényegesen tágabb kezdeti paraméter-tartományokat vizsgáltunk, mint ami a módszer napi használatakor előfordul.

Változó kísérleti körülményeket és készülék-paramétereiket vizsgáltunk, a kezdeti belsőenergia eloszlást a 300-500 K termikus eloszlással jellemezve, az ütközési cella hosszát 0,1 és 1 m között változtatva. A vizsgált ütközési energia tartomány 5-150 eV volt (laboratóriumi koordináta-rendszerben), az ütközési cellában a nyomást közben úgy változtatva, hogy egyszeres és többszörös ütközések egyaránt vizsgálhatók legyenek. Az átlagos ütközési szám 1 és 10 között változott.

4. táblázat. A változtatott molekuláris és készülék-paraméterek vizsgált tartományai (minimális érték, maximális érték).

Paraméter neve	Minimum	Maximum
Szabadsági fokok száma	70	300
Molekulatömeg (Da)	100	500
Aktiválási energia (eV)	1,0	1,3
Entrópiaváltozás ($\text{J mol}^{-1} \text{K}^{-1}$)	-80	100
Pre-exponenciális faktor	10^{10}	10^{20}
Kezdeti energia-eloszlásnak megfelelő hőmérséklet (K)	300	500
Ütközési cella hossza (m)	0,1	1,0
Ütközési energia (eV)	5	150
Ütközések átlagos száma	1	10

A realizisztikus mérési adatok generálása érdekében a mért ion-intenzitásokra is megszorításokat tettünk. A 150:1 és 1:150 közötti intenzitás arányokat tekintettük mérhetőnek, ez $-5 < \ln\left(\frac{I_1}{I_2}\right) < 5$ tartománynak felel meg. Továbbá nem tartottuk értékelhetőnek azokat a spektrumokat, amelyekben az fragmensionok intenzitása nem érte el az anyaion 1%-át. A valós mérésekhez hasonlóan a referencia-anyagok száma a 3-8 tartományba esett, és 3-6 különböző ütközési energián „mértünk”. A mérési hibák szimulációjára 2,5, 5 és 10% RSD intenzitás-meghatározási hibákat generáltunk.

A referens anyagok termokémiai paraméterei is hibával terheltek, ahogyan arra Ervin és Armentrout is rámutatott. Ennek figyelembevételére egy $\pm 2 \text{ kJ mol}^{-1}$ illetve $\pm 3 \text{ kJ mol}^{-1} \text{K}^{-1}$ hibával terheltük őket. Összességében a felsorolt pontatlanságok a kapott eredményeknek egy $\pm 2..2,5 \text{ kJ mol}^{-1}$ mértékű hibájaként jelentkeztek, ahogyan azt Ervin és Armentrout javasolta.

A mérési adatkészletek generálását és az automatikus kiértékelést egy saját fejlesztésű – a MassKinetics programmal együttműködő – C++ [94] nyelven írt szoftver segítségével végeztem. Az ODR típusú [66; 67] statisztikai kiértékelést egy az előbbivel együttműködő Fortran77 nyelven írt modullal végeztem.

Az általam kidolgozott mérésadatokat generáló és kiértékelő algoritmus minden tárgyalt esetben 1000 rendszer értékelését végezte el, az érdekesebb kérdések vizsgálatakor 3000 különböző molekuláris rendszerét. Összességében mintegy 500 000 tömegspektrum kinetikus módszerrel való automatikus kiértékelése történt meg.

2.3. Valószínűségi fragmentációs szabályok gépi tanulása

2.3.1. A tömegspektrometriás mérési körülmények

A vizsgálatokhoz leírt méréseket az Egis Gyógyszergyár Nyrt. Szerkezetkutatási osztályán végeztem egy Waters-Micromass GCT és egy Fisons Trio 1000 típusú direkt adagolási lehetőséggel ellátott GC-MS készüléken. A vizsgálatok során az adott minta néhány mikrogrammját mértem kvarc adagolóval, szobahőmérsékletről 10 °C/perc sebességgel 220 °C-ra fűtve, 70 eV elektron-energián, 220 °C-os ionforrás hőmérséklettel.

2.3.2. A kémiai szerkezetek reprezentációja és kezelése

A kémiai szerkezetek kezelésére a molekula szerkezeti gráfja konnektivitási táblájának tárolását választottam, mint legegyszerűbb ismert megoldást. Minthogy a fejlesztés során nincs szükség több ezres darabszámban szerkezetek tárolására, így ez az átlátható módszer megfelelően hatékony. Egy tömegspektrum adatsorának digitalizálására, szintén a legegyszerűbb megoldást, a tömeg-intenzitás adatpárok vektorának tárolását választottam.

A vizsgált molekulát – az általánosan elterjedt módszerrel – egy olyan irányítatlan gráffal reprezentáltam, melyben a csúcsok címkézése a megfelelő atomokat jelölte (C_{sp1} , C_{sp2} , O, N, S...), az éleké pedig az egyes kötésrendeket (a legegyszerűbb esetben 1, 2, 3...). A reprezentációs gráfban a hidrogének ábrázolását elhanyagoltam, a szerves molekulának csak az úgynevezett nehézatom vázát használtam. (A megfelelő nem-hidrogén atomokból és azok kötésrendjéből a hidrogének száma egyértelműen megadható.)

A kémiai szerkezetek kezeléséhez, vagyis a lehetséges fragmenseknek a szerkezetekből való generálásához, továbbá a különböző fragmensek szerkezeteinek későbbi összehasonlításához a következő gráf-algoritmusokat használtam. [95; 96]

Fragmensek generálása. A vizsgált szerkezeti gráfok nem feltétlenül körmentesek (tartalmazhatnak gyűrűket). Ezért nemcsak az élek egyenkénti elhagyását, hanem élpárok törlését is vizsgáltam. Egy élpár törlése akkor tekinthető „jónak”, ha a szerkezeti gráf pontosan két darabra esik szét. Tehát a kapott szerkezetvágó algoritmus a fentiek fényében a következőképpen írható le.

1. Válasszuk ki rendre a gráf éleit.
2. Vizsgáljuk meg, hogy a megfelelő él elhagyásával megszűnik-e az összefüggőség.
 - a. Amennyiben igen, úgy megkaptunk egy lehetséges fragmenspárt.
 - b. Ha nem, akkor a maradék gráf éleit rendre elhagyva vizsgáljuk újra az összefüggőség megszűnését.

Egy-egy él elhagyásakor azt kell meghatározni, hogy az elhagyott él (élek) két végpontja elérhető-e egymásból a maradék gráfban, tehát a gráf pontosan az adott él elhagyásakor esett-e szét két komponensre. Ezt az elhagyott él végpontjából induló szélességi kereséssel határoztam meg. [96; 95]

Kémiai szerkezetek összehasonlítása, részszerkezetek keresése. A részszerkezet keresés a gráfok egymásra való leképezésének és a részgráf-kereséseknek speciális esetei, ebből következően NP-teljes [70] problémák. A létező algoritmusok a gyorsítás érdekében különböző heurisztikákat alkalmaznak, kihasználva azt, hogy a kémiai szerkezeti gráfok – az esetek legnagyobb részében – síkgráfok, továbbá egyéb speciális kémiai szerkezetekre jellemző tulajdonságokat. [97; 98; 99; 100; 101; 102]

Jelen esetben, minthogy elég kis szerkezeti gráfok összehasonlítását kellett elvégezni egy sztochasztikus megoldást használtam, amely polinomiális időben [70] számítható, kisméretű gráfok esetén jó közelítéssel megadja az egzakt eredményt, nagyobb méretűeknél pedig a méret függvényében elfogadhatóan közelíti azt. A módszer részletei az „Eredmények és értékelésük” fejezetben található.

2.3.3. A kidolgozott és alkalmazott algoritmus

A szerkezet-spektrum párojaink feldolgozása során háromféle listát generáltam: a szabályok, az aktuális spektrum csúcainak és az aktuális szerkezet lehetséges fragmenseinek listáját. A későbbiekben ezek egymáshoz való viszonyát definiálom.

A szabálytanuló algoritmus ezek után a következőképpen épül fel.

1. Amennyiben szükséges, olvassuk be a meglévő (esetlegesen üres) szabálylistát.
2. Képezzük a beolvasott szerkezetből a hasadáslistát.
3. Készítsük el a beolvasott tömegspektrumból a csúcslistát.

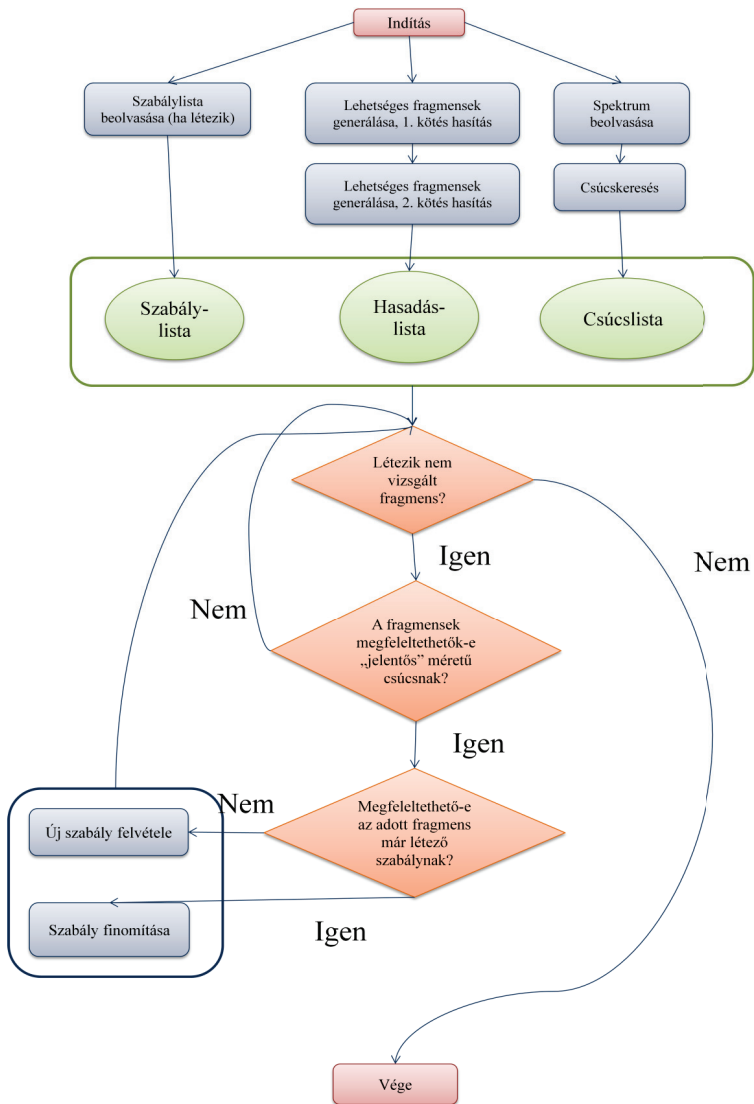
4. Minden hasadás esetén vizsgáljuk meg, hogy megjelenik-e hozzárendelhető csúcs a csúcslistában.
 - a. Amennyiben a megjelenés „jelentős”, keressük meg, hogy a szabálylistában melyik (melyek) az adott hasadáshoz leginkább hasonló szabály (szabályok).
 - i. Ha ennek a hasonlóságnak a mértéke meghalad egy előre megszabott küszöböt, növeljük a szabály előfordulási gyakoriságát és finomítjuk a megjelenési intenzitást az újnak a függvényében.
 - ii. Amennyiben nem éri el a hasonlóság a meghatározott küszöbértéket, úgy vegyünk fel egy új szabályt az aktuális hasadás függvényében.
5. Rendszeresen végezzük el a szabálylista karbantartását, a ritkán előforduló illetve nem szignifikáns szabályok kiszűrését. (Ennek módját a későbbiekben tárgyalom.)

A 11. ábra a kifejlesztett tanuló algoritmus felépítését mutatja.

2.3.4. A spektrumok tárolása

Az emberi spektrumolvasás modellezésére a legmagasabb csúcs intenzitása (BPI) szerint normáltam a beolvasott spektrumokat.

A csúcsok abszolút intenzitása helyett az egyes tömegcsúcsoknak a környezetükben tapasztalható dominanciáját használtam jellemző adatként. Ez gyakorlati szempontból azt jelenti, hogy a tömegcsúcs adott (10 tömegegységnyi) környezetében jel/zaj viszonyt számoltam. Az így kapott intenzitásmérték közel áll az emberi kiértékelésnél használthoz, amikor is észreveszünk egy tömegcsúcsot, ha nincs a környezetében másik, de nem számít jelentősnek, ha egy intenzív csúcs-köteg közepén van – bármekkora is abszolút intenzitásértéke.



11. ábra. A megvalósított tanuló algoritmus vázlatos diagramja.

2.3.5. A fragmentáció tárolása

A hasadások listájának elemei fragmenspárok. Egy hasadás a molekula reprezentációs gráfjának (illetve egy másik fragmens gráfjának) elvágásával kapott két részgráfot jelenti. Mindkét részgráfnak a szükséges vágások számával megegyező darabszámú végpontja lesz, amely jellemző a fragmenspárra.

A vizsgált átlagos méretű gyógyszermolekulák esetén – kb. 30 nem-hidrogén atomot tartalmazó molekulában – a gyűrűk számától függően mintegy 50-100 elsődleges (molekulaionból képzett) hasadás szerepelt és további néhány ezer másodlagos.

2.3.6. A fragmentációs szabályok

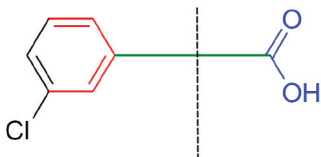
A szabálylistán fragmentációs szabályoknak a halmazát értjük, ahol a szabály a következőképpen definiált összetett adatszerkezetet jelenti:

$$G_1 \wedge G_2 \Rightarrow J_1 \wedge J_2, \quad (30)$$

ahol G_1 és G_2 két részgráfot jelöl, J_1 és J_2 pedig két „megjelenési intenzitást”.

Egy fragmentálódó ion egy hasadás során valahogyan két részre esik szét. Amennyiben a két rész megfelel valamely szabály két részgráfjának, akkor a hasadás eredményeképpen valamelyik részlet megjelenik a spektrumban.

A 12. ábra például klórbenzoesav karboxil-vesztéséhez tartozó fragmentációs szabályt demonstrálja.



12. ábra. Klórbenzoesav karboxil-vesztéséhez rendelhető fragmentációs szabály ábrázolása. A szabály bal oldalának két részgráfja jelen esetben például a pirossal illetve kékkel kiemelt két szerkezeti részlet.

J_1 és J_2 az intenzitások értéktartományát jelentik. A megjelenési intenzitásokat mint intenzitás-intervallumokat értelmeztem. Így megfelelően kezelhető egy-egy csúcs

intenzitásának a bizonytalansága. A nem-specifikus szabályok – tehát azok, amelyek vagy alkalmazhatók, vagy nem – egy széles intervallumot fognak adni, így ezek könnyen kiszűrhetők a megjelenési intervallum szélessége alapján.

2.3.7. Adatformátumok

A szerkezetek tárolására – elterjedtsége miatt – az MDL (Molecular Design Laboratories) nevéhez fűződő MOL formátumot választottam.

A nagyobb szerkezet-adatbázisok feldolgozása érdekében az elkészült program képes SD (structure definition) formátum beolvasására is. Ez egy olyan kötött szerkezetű szövegfájl, amelyben beágyazott MOL részletek szerepelnek. Az SD fájl alkalmas több rekord egyetlen fájlba foglalására, ahol egy rekord egy MOL-formátumú szerkezeti rész és egy tömegspektrum-lista együttese. SDF formátumba a legtöbb kémiai adatbázis kezelő képes exportálni adatait, illetve képes abból importálni is rekordokat.

3. Eredmények és értékelésük

3.1. SORI-CID fragmentáció modellezése

A dolgozatban tárgyalt három módszer közül a „legpontosabbnak” tekinthető MassKinetics szimulációt alkalmaztam a leucin-enkefalin ütközéses aktiválós disszociáció (SORI-CID) során tapasztalt fragmentációjának vizsgálatára.

A SORI kísérletek modellezése a MassKinetics programmal elvileg lehetséges, azonban az egy kísérlethez tartozó nagyszámú (~1000) gerjesztési ciklus követése a ma elérhető számítási kapacitások esetén több hónapos szimuláció lenne. Ezért kidolgoztam egy olyan megközelítést, amelyben az időfüggés problémája megszüntethető.

3.1.1. A SORI gerjesztési folyamat elméleti leírása

SORI kísérletekben a gerjesztendő ionok sebessége illetve ezzel összefüggésben a (laboratóriumi vonatkoztatású) kinetikus energiája periodikusan változik (13. ábra). Egy hengeres ICR cellát tekintve a kinetikus energia időfüggését az alábbi egyenlet írja le.

$$E^{\text{kin}}(t) = \frac{\beta^2 q^2 V_{\text{p-p}}^2}{64\pi^2 m d^2 \Delta v^2} (1 - \cos 2\pi \Delta v t) \quad (31)$$

β az ICR cella geometriai faktora [21], q és m rendre az ion töltése és tömege, $V_{\text{p-p}}$ a gerjesztő feszültség, d az ICR cella átmérője és Δv a ciklotron és RF gerjesztési frekvenciák közti különbség. Ebből következően a maximális kinetikus energia

$$E_{\text{max}}^{\text{kin}} = \frac{\beta^2 q^2 V_{\text{p-p}}^2}{32\pi^2 m d^2 \Delta v^2}. \quad (32)$$

Az irodalomban tipikusan ezt az $E_{\text{max}}^{\text{kin}}$ értéket használják a SORI kísérletek kinetikus energiájának jellemzésére. [85; 91] Érdemes megjegyeznünk, hogy ez maximális kinetikus energia, tehát a vizsgált ionok kinetikus energiája az idő nagy részében ennél lényegesen kisebb.

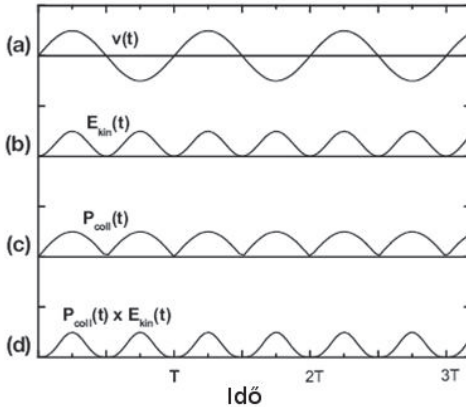
E^{kin} kinetikus energiára az ütközési gyakoriság a

$$P^{\text{coll}}(E^{\text{kin}}) = \sqrt{E^{\text{kin}}}/N \quad (33)$$

egyenlettel adható meg, ahol N egy alkalmasan választott normáló faktor. Míthogy a gyors ionok a fentiek értelmében nagyobb valószínűséggel ütköznek, mint a lassúbbak, az ütközések valószínűsége szintén periodikusan változik az időben. Az előbbi kifejezésbe helyettesítve a kinetikus energia időfüggését:

$$P^{\text{coll}}(t) = \sqrt{\frac{\beta^2 q^2 V_{p-p}^2}{64\pi^2 m d^2 \Delta v^2} (1 - \cos 2\pi \Delta v t)} / N. \quad (34)$$

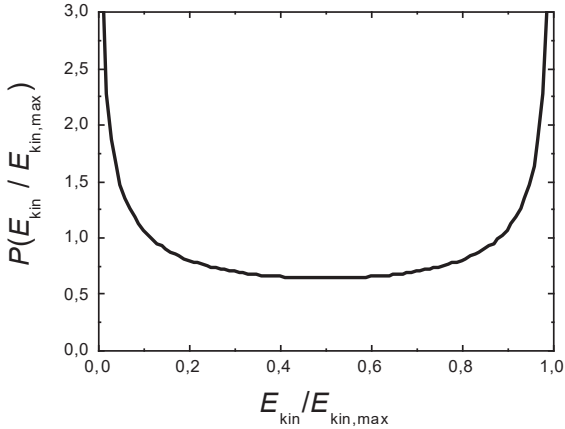
Az energiaátadás az ütközési energiától és az ütközések gyakoriságától függ, közelítőleg az $E^{\text{kin}}(t)$ és $P^{\text{coll}}(t)$ szorzatával arányos (13. ábra, d).



13. ábra. Sebesség (a), kinetikus energia (b), ütközési valószínűség (c) és az ütközés hatásossága (d) az idő függvényében ábrázolva SORI kísérlet esetén.

Mivel a SORI-CID folyamat során az ütközések véletlenszerűen következnek be a számítások egyszerűsítése érdekében a kinetikus energia időfüggése helyett a sűrűségfüggvényét használhatjuk. Ez a közelítés analóg a termodinamikában használatos ergodicitási feltétellel. Az áttérés matematikai levezetése a függelékben található. A leírt módszerrel E^{kin} sűrűségfüggvényét meghatározva az adott SORI periodikus időfüggvény esetén a következő $P(E^{\text{kin}})$ kifejezést kapjuk:

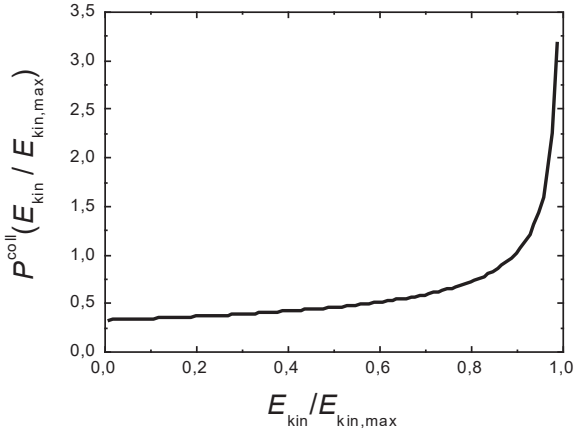
$$P(E^{\text{kin}}) = \frac{1}{\pi} \frac{E_{\text{max}}^{\text{kin}}}{\sqrt{(E_{\text{max}}^{\text{kin}} - E^{\text{kin}})E^{\text{kin}}}} \quad (35)$$



14. ábra. Kinetikus energia-eloszlás sűrűségfüggvénye. $P(E^{\text{kin}}/E_{\text{max}}^{\text{kin}})$ ábrázolva $E^{\text{kin}}/E_{\text{max}}^{\text{kin}}$ függvényében.

A 14. ábra az energia-intervallum két végén (0-nál és 1-nél) a végtelenbe tart, azt jelezve, hogy az ionok döntő többsége vagy nagyon alacsony, vagy nagyon magas kinetikus energiával rendelkezik. Hasonlóan megadható a ténylegesen ütköző ionoknak az energia szerinti eloszlása is (15. ábra):

$$P^{\text{coll}}(E^{\text{kin}}) = \frac{\sqrt{E^{\text{kin}}} P(E^{\text{kin}})}{N} = \frac{\sqrt{E^{\text{kin}}}}{N} \frac{1}{\pi} \frac{E_{\text{max}}^{\text{kin}}}{\sqrt{(E_{\text{max}}^{\text{kin}} - E^{\text{kin}})E^{\text{kin}}}} \quad (36)$$



15. ábra. Ütközési energia-eloszlás sűrűségfüggvénye. $P^{\text{coll}}(E^{\text{kin}}/E^{\text{kin,max}})$ ábrázolva $E^{\text{kin}}/E^{\text{kin,max}}$ függvényében.

Megfigyelhető, hogy a kinetikus és az ütközési energia-eloszlások jelentősen különböznek. Alacsony ütközési energiánál nem látunk maximumot, ugyanis kis kinetikus energián ütközések ritkábban következnek be.

Az átlagos (laboratóriumi vonatkoztatású) ütközési energia mindezek után már egyszerűen megadható

$$E_{\text{lab,avg}}^{\text{coll}} = \int_0^{\infty} E P^{\text{coll}}(E) dE = \frac{2}{3} E_{\text{max}}^{\text{kin}}. \quad (37)$$

A fenti levezetésnek az eredménye, hogy az átlagos ütközési energia a SORI kísérlet során $E_{\text{lab,avg}}^{\text{coll}} = \frac{2}{3} E_{\text{max}}^{\text{kin}}$. Feltételezhetjük, hogy – legalábbis kis energiatartományon belül – az ütközési energiátaszfer arányos az ütközési energiával, illetve az ütközési hatáskeresztmetszetek nem függenek attól. Ekkor kijelenthető, hogy a SORI kísérlet eredménye praktikusán megfeleltethető annak a kísérletnek, ahol minden ion kinetikus energiája egységesen $\frac{2}{3} E_{\text{max}}^{\text{kin}}$. Az egy ütközéssel átadott energia (ΔE_{int}) ekkor már egyszerűen megadható a következő kifejezésekkel:

$$\Delta E_{\text{int}} = \eta_{\text{single,com}} E_{\text{com}}, \quad (38)$$

$$E_{\text{com}} = E_{\text{coll}} \frac{m_g}{m_g + m}, \quad (39)$$

ahol E_{com} a tömegközépponti ütközési energia, m_g és m rendre az ütközőgáz és a vizsgált ion tömege, $\eta_{\text{single,com}}$ pedig az energia-átadás hatékonysága. A fenti egyszerűsítések lehetővé teszik a SORI kísérletek modellezését, minthogy a kinetikus energia időbeli változását a számítások során nem kell figyelembe vennünk. Az eredmény azt is mutatja, hogy a SORI-CID kísérletnek az – irodalomban szokásos – $E_{\text{max}}^{\text{kin}}$ értékével való jellemzése jogos, megfelelően visszaadja az ütközések energetikai vonatkozásait.

3.1.2. A SORI kísérletek numerikus modellezése

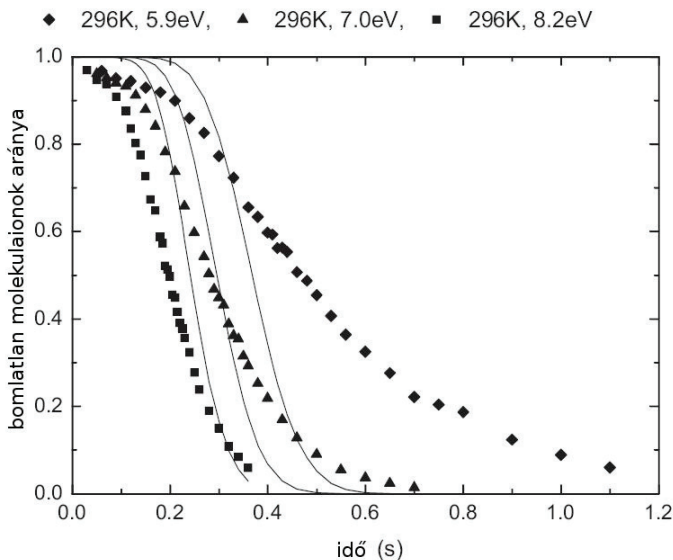
Az előző részben tárgyalt eredményeket a MassKinetics keretei között alkalmazva leucin-kefalin SORI fragmentációjának modellezését végeztük el. A kísérleti részben leírt módszerrel ütközési energia, készülék-hőmérséklet és gerjesztési idő függvényében felvett molekulaion intenzitásadatokat modelleztem. A meghatározandó két paramétert, az ütközési energiáttranszfer hatékonyságát $\eta_{\text{single,com}}$ és az infravörös hűlés sebességét k_{cool} , párhuzamosan a 122 adatpontot tartalmazó adatkészletre illeszttem.

Az első vizsgálatok az infravörös hűlés figyelembe vétele nélkül történtek. Ennek az illesztésnek az eredményeképpen egy 6,7 %-os energiaátadás ($\eta_{\text{single,com}}$) érték adódott. A mért és számított görbék illeszkedését a következő két ábra mutatja (16. ábra, 17. ábra), a bomlatlan molekulaion arányát ábrázolva a gerjesztési idő függvényében. Jól látható, hogy a sugárzásos energiavesztés figyelembevétele nélkül a számított görbék alakja jelentősen eltér a mért adatokétól. A görbék lefutása sokkal meredekebb, a különböző hőmérsékleteken illetve különböző ütközési energiákkal számított görbék lényegesen kisebb mértékben térnek el egymástól, mint a megfelelő mért adatok. A 17. ábra esetén például a három számított görbe között a különbség az eltérő kezdeti termikus energia-eloszlásban van. Ez az egy paraméter láthatóan nem elég a mért görbék közötti sokkal szembetűnőbb eltérések megmagyarázásához.

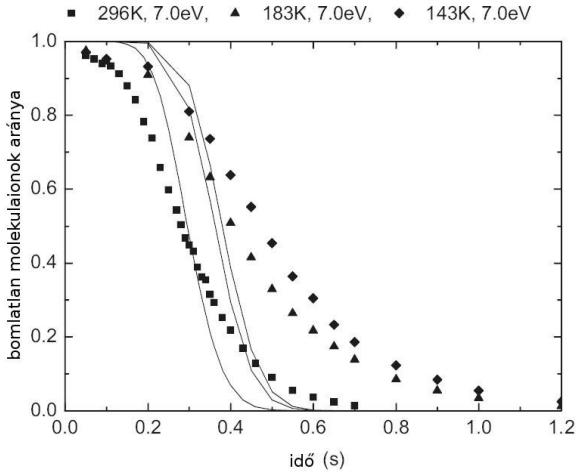
Az ütközéses energiáttranszfer vizsgálatok felmerül a kérdés, hogy a használt függvényalak mennyiben befolyásolja a szimuláció eredményét. Nemrégiben két különböző energia-transzfer eloszlásfüggvény jelent meg az irodalomban: a Hase és csoportja által – egy

kisméretű peptid trajektória-számításai alapján – kidolgozott [93], és az Armentrout által kísérleti úton $\text{Cr}(\text{CO})_6^+$ -ra meghatározott [92].

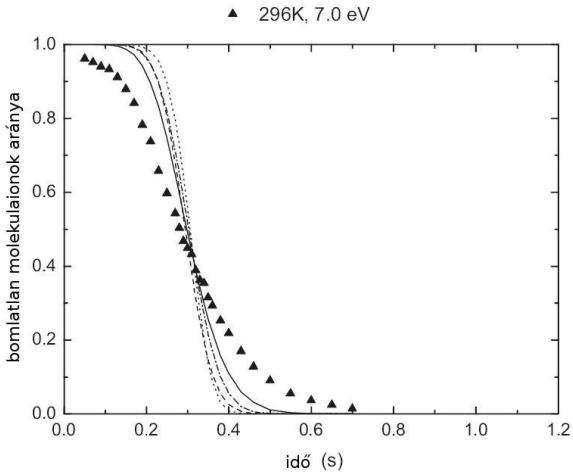
A 18. ábra a különböző energia-átadás függvényalakok hatását mutatja. Az energia-transzfer függvényalakok közötti különbség szintén nem magyarázza a kísérleti adatok és a szimulált eredmények eltérését. A legdurvább közelítésnek számító fix η értékű energia-átadás áll ugyan a legmesszebb a kísérleti görbétől, de a szimulált görbealakok meredeksége még a legjobb esetben (Hase) sem hasonló a kísérleti adatokéhoz.



16. ábra. Molekulaion intenzitás adatok az idő függvényében 296K hőmérsékleten különböző kinetikus energiák esetén. Folytonos vonal: az optimált paraméterekkel modellezett értékek, pontok: kísérleti adatok. Az illesztések az infravörös hűlés figyelembe vétele nélkül történtek.



17. ábra. Molekulaion intenzitás adatok az idő függvényében különböző hőmérsékleteken 4,67 eV kinetikus energia esetén. Folytonos vonal: az optimált paraméterekkel modellezett értékek, pontok: kísérleti adatok. Az illesztések az infravörös hűlés figyelembe vétele nélkül történtek.

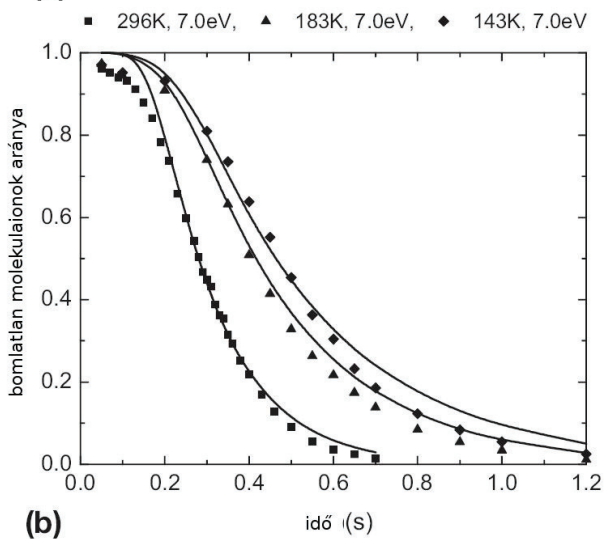
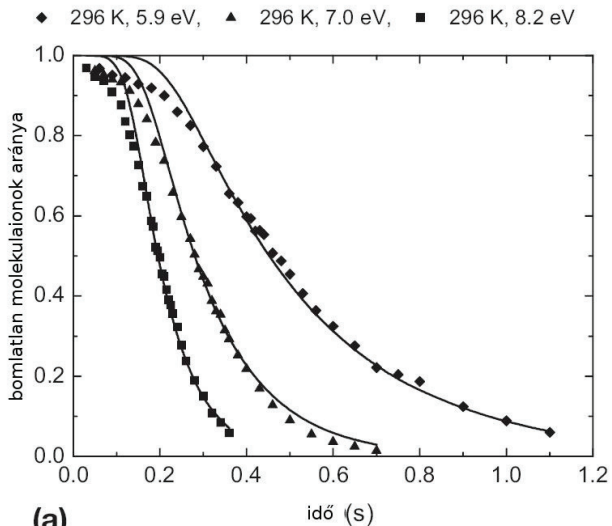


18. ábra. Molekulaion intenzitása az idő függvényében 296 K hőmérsékleten, 7,0 eV energia esetén különböző energiatranszfer függvényalakok használatával. Folytonos vonal: Hase, szaggatott: exponenciális, szaggatott-pontozott: Armentrout, pontozott: fix érték.

A különböző függvényalakok illesztésekor η értéke hasonlóan adódott: 6,5%, 6,7%, 6,6% és 6,5% rendre fix, exponenciális, Hase és Armentrout esetén. Összefoglalva elmondható, hogy bár a SORI görbék lefutása függ a használt függvényalaktól, de a kísérleti értékeket *egyik esetben sem közelíti jól*.

Az infravörös hűlést figyelembe véve a két paraméter párhuzamos illesztésével $\eta_{\text{single,com}} = 12,8\%$ és $k_{\text{cool}} = 7,5 \text{ s}^{-1}$ értékek adódtak. A mért és számított adatok illeszkedése jónak tekinthető, egy 0,9 fölötti regressziós együtthatóval jellemezve. Az előző modellel ellentétben a számított görbék alakja nagy energián is megfelelően illeszkedik a mérési eredményekhez. Érdemes kiemelni, hogy a számítások során csupán ez a két paraméter változott a három független kísérleti körülmény (hőmérséklet, kinetikus energia és idő) terében felvett 122 adatpontra való illesztéskor, a modellezés összes többi paramétere állandó volt. Ez megerősíti azt a felvetést, hogy a használt szimulációs modell megfelelően kezeli a fragmentációs és gerjesztési folyamatokat. A SORI görbék leírása szempontjából ez egyúttal azt is jelenti, hogy az infravörös hűlés nem hanyagolható el SORI-CID kísérletek esetén, hiszen nélküle a modell csak gyengén tudta reprodukálni a mérési adatokat.

Az itt kapott eredmények jól illeszkednek a korábban az irodalomban rezonancia gerjesztés esetén, infravörös hűlés figyelembevételével már meghatározott energiatranszfer értékhez. Rezonancia gerjesztés alkalmazásakor az infravörös hűlés hatása kisebb, köszönhetően a rövidebb időskálának. Az ott meghatározott 9,6%-os (rezonancia-gerjesztés, IR hűlés nélkül) érték így érthetően a most megadott 6,6% (SORI, IR hűlés nélkül) és 12,8% (SORI, IR hűléssel) között helyezkedik el.



19. ábra. Molekulaion intenzitása az idő függvényében az infravörös hűlést figyelembe véve. (a) különböző kinetikus energiák, azonos hőmérséklet, (b) különböző hőmérsékletek, azonos kinetikus energia. Folytonos vonal: szimulált értékek, pontok: mérési eredmények.

Az eddig tárgyalt két paraméter mellett az ütközéses hűlés kaphat még szerepet a modellezésekben. Ugyanakkor ez csak abban az esetben válhat jelentőssé, ha az ütközések egy jelentős hányada a termikus energia alatt történik. SORI esetén ez a feltétel nem teljesül, ugyanis a 7 eV kinetikus energia egy 0,5 eV-os tömegközépponti ütközési energiának felel meg. A SORI folyamat során a leucin-enkefalin valóban gerjesztődik, belső energiája 3 eV környékéig emelkedik, ez durván 500 K hőmérsékletnek feleltethető meg. 500 K-en a termikus tömegközépponti ütközési energia viszont mindössze 0,02 eV, ami csupán 4 %-a a tapasztaltnak. A 15. ábra azt mutatja, hogy az ütközéseknek mindössze 1-2 %-a történik ilyen alacsony energián.

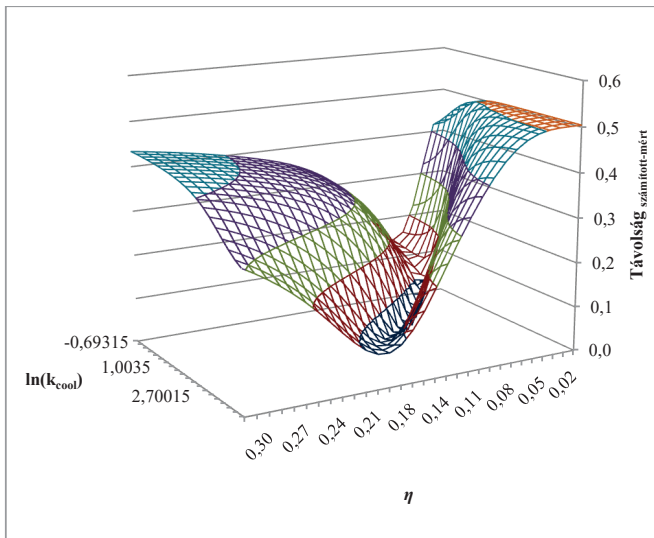
A modellezés megbízhatósága nehezen becsülhető meg. A becült paraméterek hibáját a kiindulási adatok egy reális szórását feltételezve határoztam meg. Ezek 0,05 eV a kritikus energia esetén, 10 cm^{-1} a rezgési frekvenciákra, 10 % a nyomásértékekre és 1 K a mért hőmérséklet értékekre. Mindezek mellett az ütközéses energia-átadás eloszlásfüggvényének alakját a már tárgyalt módon figyelembe véve az eredmények $\eta_{\text{single,com}} = 12,8 \% \pm 2,1 \%$ és $k_{\text{cool}} = 7,5 \pm 0,5 \text{ s}^{-1}$ hibákkal adódtak.

Az ütközéses gerjesztés és az infravörös hűlés a belsőenergia változása szempontjából ellentétes effektusok. Megfigyelhető, hogy a hűlési folyamat figyelembe vétele nélkül az illesztés erősen alulbecsülte az energia-átadás hatékonyságát. Ennek a jelenségnek a tanulmányozására az illeszkedés „jószágát” – tehát a mért és számított értékek eltérését – vizsgáltam $\eta_{\text{single,com}}$ és k_{cool} értékeit változtatva. A 20. ábra az illesztés hibáját ábrázolja a $\eta_{\text{single,com}}$ és $\ln(k_{\text{cool}})$ függvényében. Szembetűnő, hogy sokféle $\eta_{\text{single,com}}$ és k_{cool} kombináció ad jó illeszkedést. A háromdimenziós felszín egy átlós völgyet mutat ezen a féllogaritmikus ábrán, azt jelezve, hogy $\eta_{\text{single,com}}$ értékének alulbecslése könnyen kompenzálható k_{cool} alulbecslésével és fordítva. A számítások másik eredménye, hogy az $\eta_{\text{single,com}}/\ln(k_{\text{cool}})$ hányados lényegesen kisebb hibával határozható meg, mint külön-külön a két paraméter bármelyike.

Az eddigieket röviden összefoglalva a következő eredményeket kaptam.

- SORI kísérletek modellezése esetén meghatároztam az átlagos laboratóriumi vonatkoztatási rendszerben megadott ütközési energiát (hengeres cellageometriára), ez a maximális kinetikus energia 2/3-ának adódott.

- Az elvégzett szimulációk segítségével megállapítottam, hogy az infravörös hűlés nem hanyagolható el a SORI kísérletek modellezése során.
- A vizsgált hőmérsékletfüggő, 122 adatpontból álló mérési adatkészletre illesztettem a MassKinetics program segítségével számított görbéket. Az illesztés eredményeképpen az átlagos energia-átadás és az infravörös hűlés sebességének értékét becsültem. Leucin-enkefalin esetén ezek értékei: $\eta_{\text{single,com}} = 12,8 \% \pm 2,1 \%$ és $k_{\text{cool}} = 7,5 \pm 0,5 \text{ s}^{-1}$.
- A hűlés sebessége és az energia-transzfer optimális értéke között megfigyelt összefüggés nyilvánvalóan mutatja a két folyamat kompetitivitását.



20. ábra. Az illeszkedés „jósága” az energiatranszfer hatékonysága és a hűlési sebességi állandó logaritmusában.

3.2. A kinetikus módszer hibájának vizsgálata

Az előbbi „pontos” szimulációs módszer után egy termodinamikai alapokon levezetett összefüggés nem-termodinamikai rendszerben – tömegspektrométerben – való alkalmazásának feltételeit vizsgálom. A MassKinetics szimulációs módszereit itt a statisztikai vizsgálatokhoz „virtuális kísérleti adatok” generálására használom több ezres nagyságrendben.

3.2.1. A meghatározható entalpia és entrópia pontossága

A kísérleti részben részletesen leírt hibák megfelelnek a kinetikus módszer esetén általános gyakorlatban szereplő értékeknek. A vizsgált paramétereket az előzőekben megadott tartományokban az általam kifejlesztett program egyenletes eloszlással (kivételet képez az ütközési energia, amelyet logaritmikusan) véletlenszerűen választotta. Az így kiválasztott paraméterekkel egy fiktív molekuláris rendszert generált, amelyet átadott számításra a MassKinetics modulnak.

A számítások első részében 3000 különböző molekuláris rendszert értékelt ki az algoritmus. Ekkor a referencia anyagok protonaffinitásait pontosan ismertnek feltételeztem, továbbá az ionintenzitások mérési hibáját sem vettem figyelembe.

Az effektív hőmérsékletek a 280-850 K tartományban mozogtak, 465 K-es átlagos értékkel. Ez megfelel a legtöbb valós kísérlet esetén tapasztaltaknak, az alacsonyabb effektív hőmérsékletek ionscspadás, míg a magasabbak szektor készülékek esetén fordulnak elő.

A ΔH és ΔS meghatározások hibáit mutatja a 21. ábra, vagyis

$$\begin{aligned}\Delta H_{\text{error}} &= \Delta H_{\text{theor}} - \Delta H_{\text{kinetic}}, \\ \Delta S_{\text{error}} &= \Delta S_{\text{theor}} - \Delta S_{\text{kinetic}}\end{aligned}\tag{40}$$

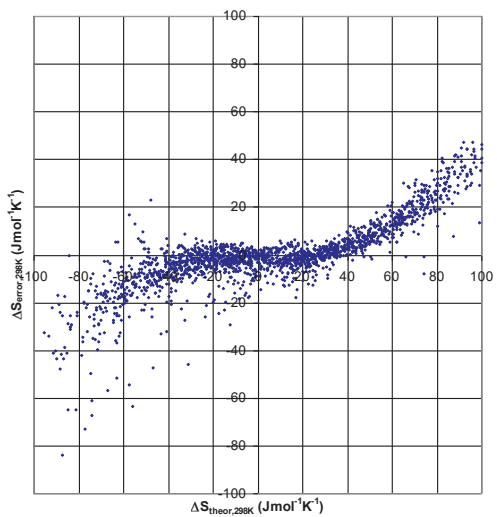
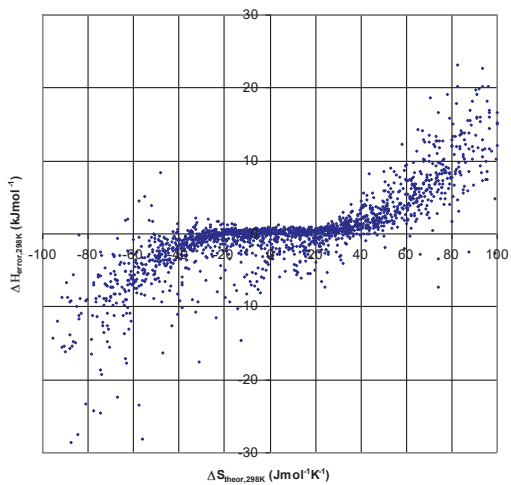
ahol „error” alsóindex jelöli a hibát, „theor” az elméleti (jelen esetben pontosan ismert) értéket, „kinetic” pedig a módszer által meghatározottat. A grafikonok a meghatározott entalpia- illetve entrópiaváltozás hibáját az elméleti entrópiaváltozás függvényében ábrázolják.

Az ábrázolt adatok együttesen mutatják a kinetikus módszer összes belső hibáját, úgymint az „effektív” hőmérséklet, a nem-termikus energia-eloszlások, a kinetikus módszer grafikonjainak a lineáristól való eltérése, a ΔH és ΔS értékek hőmérsékletfüggésének problémáit.

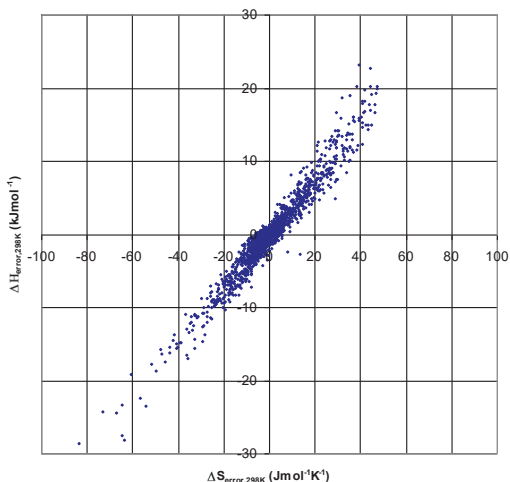
Az így elvégzett számítások legfontosabb eredményeként az emelendő ki, hogy a vizsgált minta és a referencia-anyagok közötti nem túl nagy entrópia-különbségek esetén (vagyis $|\Delta\Delta S| < 35\text{-}40 \text{ J mol}^{-1}\text{K}^{-1}$ esetén) a kiterjesztett kinetikus módszer nem mutat nagy szisztematikus alul- vagy fölülbecslési hibákat ($0,5 \text{ kJ mol}^{-1}$ entalpia és $-1,9 \text{ J mol}^{-1} \text{ K}^{-1}$ entrópia értékek esetén), tehát ΔH_{298} és ΔS_{298} értéke jó közelítéssel meghatározható.

A becslült értékek szórásai $1,9 \text{ kJ mol}^{-1}$ és $4,9 \text{ J mol}^{-1} \text{ K}^{-1}$ rendre entalpia és entrópia meghatározásakor. Ezek tehát a módszerből, az alkalmazott közelítésekből adódó eltérések, a méréstechnikai hibák és a referencia anyagok termokémiai paramétereinek pontatlan ismeretét nem számítva.

ΔH és ΔS becslésének hibáiban erős korreláció fedezhető fel (22. ábra), ahogyan azt az irodalomban is említik már Ervin és munkatársai. [40] A becslült entrópiák pozitív eltéréseit a meghatározott entalpia-értékek pozitív hibái ellensúlyozzák. Szemléletesen megfogalmazva ez azt jelenti, hogy míg a gázfázisú bázicitás, tehát $\Delta G_{\text{app},298}$ még nagy entrópia-különbségek esetén is kielégítő pontossággal határozható meg a kinetikus módszer segítségével, addig ΔG értékének entalpia és entrópia tagokra való szétválasztása számottevő hibát eredményezhet. A 23. ábra a meghatározott szabadentalpia-változást mutatja az elméleti entrópia-változás függvényében.

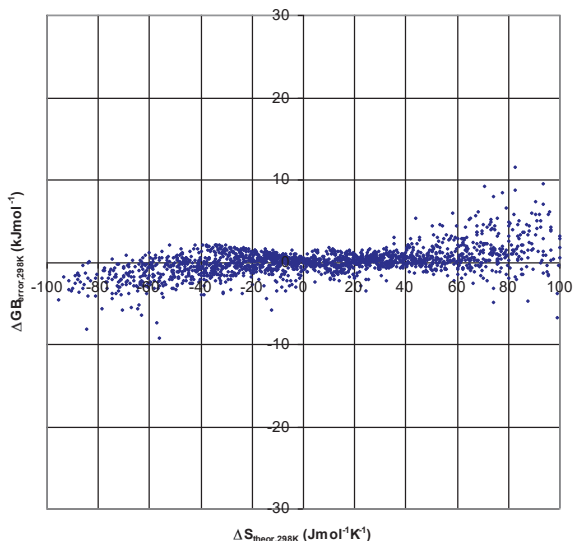


21. ábra. A kinetikus módszerrel meghatározott entalpia- (a) és entrópiaváltozás hibája (b) az elméleti entrópiaváltozás függvényében.



22. ábra. A meghatározott entalpia- és entrópia-értékek hibája (az elméleti és a kinetikus módszerrel becsült érték különbsége) közötti korreláció.

Meg kell jegyezni itt azonban, hogy az entrópia és az entalpia tagok korrelációja nem pusztán a kinetikus módszerre jellemző, hanem a termokémiai mérési technikák egy jelentős részének sajátja. A 23. ábra rámutat arra is, hogy $\Delta G_{app,298}$ a teljes vizsgált entrópia-változás tartományában kielégítően meghatározható, a $35 \text{ J mol}^{-1} \text{ K}^{-1}$ -nél kisebb esetben pedig kiváló pontossággal, $0,9 \text{ kJ mol}^{-1}$ szórással becsülhető.



23. ábra. A számított szabadentalpia-változás hibája az elméleti entrópia-változás függvényében.

A valós kísérleti hibák becslése érdekében az általam kidolgozott automatikus kiértékelő algoritmussal a kísérleti csúcshintenzitás hibájának hatását is vizsgáltam. Véletlenszerű 5 %-os hibát feltételezve a ΔH_{298} és ΔS_{298} becslések szórásai rendre 3,3 kJ mol⁻¹ és 7,5 J mol⁻¹ K⁻¹ értékekre emelkedtek. Hasonló számítások készültek 2,5 illetve 10 % véletlen intenzitás-mérési hibát feltételezve (5. táblázat). Minden esetben az előzőekben már tárgyalt módon, 1000-1000 véletlenszerűen kiválasztott molekuláris rendszer kiértékelését végezte el a program.

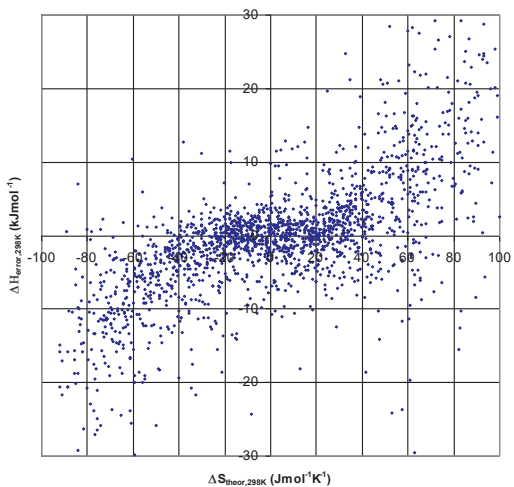
A módszer eredendő hibái és az intenzitás-mérési pontatlanságok után az eredményeket befolyásoló harmadik lényeges tényező a referencia anyagok protonaffinitásainak (és protonálódási entrópiáinak) pontatlan ismerete. Természetesen a referencia anyagok pontatlanságaival összhangban növekszik a kinetikus módszer becslésének hibája is. A protonaffinitások esetén egy ± 2 kJ mol⁻¹ mértékű hibát feltételezve a ΔH_{298} és ΔS_{298} meghatározásának hibája $\pm 4,4$ kJ mol⁻¹ illetve $\pm 9,2$ J mol⁻¹ K⁻¹ értékre növekszik. Az előző számításokhoz hasonlóan ezek az értékek itt is $|\Delta\Delta S| < 35$ J mol⁻¹ K⁻¹ esetén értendők. A hibaszámítást elvégeztük ± 1 kJ mol⁻¹ és ± 4 kJ mol⁻¹ esetén is. A táblázat

adatait vizsgálva azt láthatjuk, hogy a referencia anyagok pontatlan ismerete „hozzáadódik” a módszer eredendő hibáihoz.

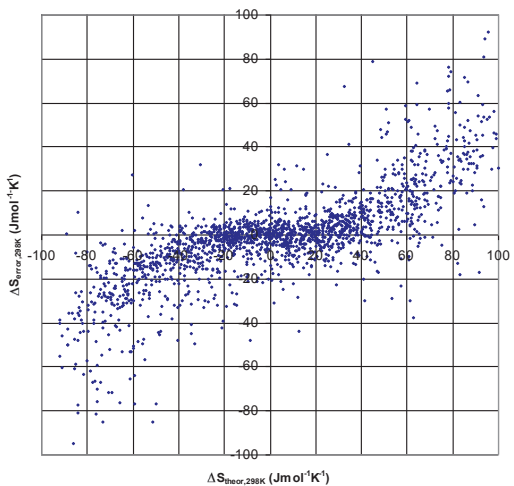
5. táblázat. A kinetikus módszerrel meghatározott entalpia és entrópia értékek hibái. Eredendő hibák (első sor), a kísérleti pontatlanságok (2-4. sor), a referenciák pontatlan ismeretének (5-7. sor) hatása, illetve e hibák összesítése (8. sor).

Vizsgált pontatlanságok		A meghatározott paraméterek hibái		
A csúcsintenzitások mérésének pontatlansága (relatív szórás)	A referencia anyagok proton-affinitásának pontatlansága (kJmol^{-1} , szórás)	ΔH_{298} hibája (kJmol^{-1} , szórás)	ΔS_{298} hibája ($\text{Jmol}^{-1}\text{K}^{-1}$, szórás)	$\Delta G_{\text{app},298}$ hibája (kJmol^{-1} , szórás)
---	---	1.9	4.9	0.9
2.5%	---	2.5	6.0	1.1
5 %	---	3.3	7.5	1.4
10 %	---	4.0	9.5	1.5
---	1.0	3.0	7.0	1.4
---	2.0	4.4	9.2	2.1
---	4.0	6.0	11.2	3.5
5 %	2.0	4.6	10.3	2.1

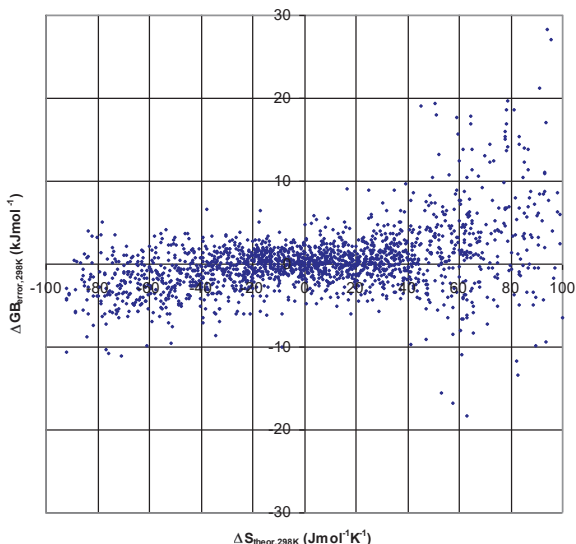
Mindezek után elvégeztük a kinetikus módszer tesztelését az összes eddig felsorolt tényező figyelembe vételével. Tehát a módszer eredendő hibáin kívül a kísérleti mérési pontatlanságokat és a referencia anyagok tökéletlen ismeretét is modellezve 3000 véletlenszerűen generált fiktív molekuláris rendszer esetén ΔH_{298} , ΔS_{298} és $\Delta G_{\text{app},298}$ összesített szórása rendre $4,6 \text{ kJ mol}^{-1}$, $10,3 \text{ J mol}^{-1} \text{ K}^{-1}$ és $2,1 \text{ kJ mol}^{-1}$ -nak adódott, amennyiben az entrópia különbség $35 \text{ J mol}^{-1} \text{ K}^{-1}$ alatt volt. Az összesített számításnál az intenzitások pontatlanságát $\pm 5\%$, a protonaffinitások bizonytalanságát pedig $\pm 2 \text{ kJ mol}^{-1}$ szórásokkal vettem figyelembe. Fontos megemlíteni azt az eredményt, hogy nem látszik számottevő szisztematikus hiba a becslésekben, ΔH_{298} átlagos hibája $\pm 0,3 \text{ kJ mol}^{-1}$, míg ΔS_{298} esetén $\pm 1 \text{ J mol}^{-1} \text{ K}^{-1}$. A becslött entalpia- és entrópia-változás értékek korrelációja itt is domináns, mint a kísérleti hibák nélküli esetben. Kiválóan szemlélteti ezt, hogy a $\Delta G_{\text{app},298}$ szórása mindössze $2,1 \text{ kJ mol}^{-1}$ még ebben az esetben is.



24. ábra. Az entalpiaváltozás hibája az elméleti entrópiaváltozás függvényében minden felsorolt hibalehetőség figyelembevételével (csússzintenzitások pontatlansága, referencia értékek hibás ismerete).



25. ábra. Az entrópiaváltozás hibája az elméleti entrópiaváltozás függvényében minden felsorolt hibalehetőség figyelembevételével (csússzintenzitások pontatlansága, referencia értékek hibás ismerete).



26. ábra. A szabadentalpia-változás hibája az elméleti entrópiaváltozás függvényében minden felsorolt hibalehetőség figyelembevételével (csúcsintenzitások pontatlansága, referencia értékek hibás ismerete).

Amennyiben az entrópia-különbség nagyobb, mint $35 \text{ J mol}^{-1} \text{ K}^{-1}$, a kinetikus módszerrel meghatározott ΔH és ΔS értékek szisztematikus hibája jelentős, 10 kJ mol^{-1} illetve $25 \text{ J mol}^{-1} \text{ K}^{-1}$ fölött van. A számítások eredményei azt mutatták, hogy ebben a tartományban a kinetikus módszer jelentősen alulbecsli az entrópia-különbséget, átlagosan mintegy $16 \text{ J mol}^{-1} \text{ K}^{-1}$ -nel. Az entrópia tagnak hasonló alulbecslését megtalálhatjuk az irodalomban is, valós kísérletek kiértékelésekor. Természetesen az entrópia tagok szisztematikus hibája maga után vonja az entalpiák szisztematikus hibáit is. Ugyanakkor a látszólagos gázfázisú bázicitások hibája – az egyes komponensek szisztematikus hibái ellenére is – meglepően kicsi, $5,5 \text{ kJ mol}^{-1}$ a teljes vizsgált tartományon (26. ábra). Amennyiben szükséges, az alulbecsült tartományban használható egy empirikus skálázó faktor, de feltétlenül szem előtt kell tartani azt, hogy az ilyen korrekciós megoldások mindenképpen egy hamis biztonság látszatát keltik. Az alulbecsült tartományon a kinetikus módszer alapvetően hibás eredményt ad, mivel a módszer alkalmazásának az alapfeltételei nem teljesülnek. Egy empirikus skálázás nem mutat rá erre a problémára, és egy az adott tartományban alkalmatlan módszer használatát segíti elő.

Egy kiterjedt és részletes modellezés eredményeit foglaltam össze az előbbieken. A statisztikai kiértékeléshez adatokat biztosító módszer hatékonyan generált nagy számban „virtuális kísérleti adatokat”, amelyek a későbbi kiterjesztett kinetikus módszer szerinti kiértékelés alapját képezték.

A statisztikai értékelés célja elsősorban az volt, hogy a kinetikus módszer alkalmazhatósági körét felderítsük. Valóban igaz, hogy a módszer, mind standard, mind pedig kiterjesztett változatában, könnyen és kényelmesen használható protonált molekulák, ion-molekula komplexek termokémiai adatainak becslésére. Szintén igaz az ellenérv is, mely szerint a módszer nem kellően körültekintő használatakor – az entrópia tényező indokolatlan elhanyagolásakor, túl nagy entrópia-különbségek esetén, a termokémiai paraméterek esetleges hőmérsékletfüggésekor, nem-linearitási problémák fellépésekor – igen könnyen hamis kvantitatív eredményekre és ennek további mérlegelés nélkül való alkalmazásával akár kvalitatíve is fals eredményekre juthatunk. A két ellentétes érv között azonban érdemes megkeresni azokat a körülményeket, amelyek megszabhatják a módszer alkalmazhatóságának korlátait, továbbá lehetővé teszik a meghatározott paraméterek hibájának a becslését is. A több ezer szimulált mérésből készült statisztika alapján a következő „recept” adható a kinetikus módszer alkalmazására.

1. A legegyszerűbb esetekben is érdemes ellenőrizni a módszer ütközési energiától (tehát effektív hőmérséklettől) való függését. Amennyiben a meghatározott látszólagos bázicitás függ az ütközési energiától, akkor mindenképpen a kiterjesztett kinetikus módszert kell használni. Az irodalomban létező kiértékelési sémák mindegyike lényegében ekvivalens megközelítést képvisel és azonos eredményre vezet. Amennyiben nincs lehetőség a kiterjesztett módszer alkalmazására, az eredmények csak „látszólagos” bázicitásként interpretálhatók.
2. Amennyiben a meghatározott entrópia-különbség $35 \text{ J mol}^{-1} \text{ K}^{-1}$ alatt van, akkor a becslőt esetén ΔH_{298} és ΔS_{298} hibája $\pm 5 \text{ kJ mol}^{-1}$, $\pm 10 \text{ J mol}^{-1} \text{ K}^{-1}$ környékére esik. A megadott számadatok abszolút szórásadatok, mintegy $\pm 5\%$ intenzitásmérési pontatlanságot és $\pm 2 \text{ kJ mol}^{-1}$ referencia protonaffinitás bizonytalanságot feltételeznek.
3. Ha az entrópia-különbség lényegesen $35 \text{ J mol}^{-1} \text{ K}^{-1}$ fölött van, akkor a kinetikus módszer alkalmazásával kapott eredmények hibája nagy, szisztematikusan alulbecslésük akár 30-50 %-os is lehet.

4. Az előbbi hibák nagy része kiküszöbölhető azzal, hogy a látszólagos bázicitást adjuk meg eredményként.

$$GB_{\text{app},298} = \Delta H_{298} - 298 \text{ K} \times \Delta S_{298} = GB_{298} - (\Delta H_{\text{ref}} - GB_{\text{ref}}) \quad (41)$$

$GB_{\text{app},298}$ értéke az előbbiekből tárgyaltak alapján az entalpia értékénél lényegesen nagyobb pontossággal meghatározható, 2 kJ mol⁻¹ körüli hibával, amennyiben az alacsony entrópia-különbség tartományban vagyunk és 5 kJ mol⁻¹ környékén a teljes szimulált tartományon.

5. A kiterjesztett kinetikus módszer alkalmazásakor mindenképpen javasolt legalább három referencia használata, három lényegesen különböző ütközési energián. Amennyiben erre nincs lehetőség, az eredmény becslét hibáját tekintjük kétszer nagyobbobbnak.
6. Néhány esetben, amennyiben egymáshoz nagyon közeli szerkezeteket vizsgálunk (pl. izomereket), a módszer alkalmazásakor az eddig tárgyaltaknál lényegesen kisebb hibákat tapasztalhatunk. Ennek oka az, hogy a kinetikus módszer levezetésénél alkalmazott közelítések nagy része teljesül, s így az eredendő hibája lényegesen kisebb a dolgozatban előbb tárgyaltaknál.
7. A felhasznált statisztikai kiértékelő módszer lehet mind a lineáris regresszió, mind pedig az ortogonális távolság regresszió. Kétségtelen, hogy jelen esetben az utóbbi a korrekt statisztikai megközelítés, ugyanakkor az egyszerűbb és elterjedtebb lineáris regresszió használata nem okoz látható hibát a becslét entrópia és entalpia paraméterekben, szemben a módszer egyéb, már tárgyalt közelítéseivel.

Megfelelő körültekintéssel alkalmazva a fentiek alapján a kinetikus módszer egy használható, általában torzítatlan becslését adja a meghatározni kívánt termokémiai paramétereknek.

3.3. Valószínűségi fragmentációs szabályok gépi tanulása

A kísérleti részben bemutatott megoldások együttes alkalmazásával megvalósítottam és teszteltem egy olyan „valószínűségi” típusú szabályokat tanuló algoritmust, amely alkalmas gyógyszermolekulák és hasonló mérettartományba eső egyéb kismolekulák (illetve általánosítva akár bármilyen molekula) tömegspektrumából valószínűségi fragmentációs szabályok megismerésére, tanulására.

3.3.1. A tanult szabályok szerkezete, finomítása és szűrése

A generált szabályok első része két részszerkezetből áll. A szabály generálásának optimált lépései a következők.

1. Tekintjük egy vizsgált molekula hasadáslistájának egy tagját. Az adott hasadás a molekulaionnak vagy valamelyik előző fragmensnek két részszerkezetét tartalmazza.
2. A hasadás mindkét oldalából létrehozunk egy-egy részszerkezetet, úgy hogy tekintjük a hasadási ponttól legfeljebb 4 kötési távolságra lévő környezetét.
3. Megvizsgáljuk, hogy az így kapott 4-mélységű környezethez létezik-e az adott hasonlósági küszöböt ($2 \times 0,9$) meghaladó meglévő szabály.
 - a. Ha igen, az adott szabály megjelenési oldalát korrigáljuk az új példa alapján, módosítjuk a megjelenési intenzitásokat, továbbá a meglévő és az új részszerkezeteknek vegyük a metszetét.
 - b. Ha nem, akkor a következő két lépést végezzük el.
 - i. Létrehozunk egy új szabályt a meghatározott 4-mélységű részszerkezetek és az adott spektrumban való megjelenési intenzitások alapján.
 - ii. Szűkítjük mindkét oldalt a 3-mélységű környezetre, majd erre a kisebb szerkezetre megismételjük a hasonló szabály keresését és finomítását.

Az előbbi algoritmussal megkapjuk a szabályoknak egy olyan halmazát, amelyben egyrészt szerepel minden a megadott küszöbnél ($0,9$) kevésbé hasonló 4-mélységű (tehát legfeljebb 4 kötés távolságig tekintett) részszerkezet. Másrészt szerepelnek ezeknek 3-mélységű metszetei is, a megjelenési intenzitás-intervallumok uniójával.

A mélységek meghatározásának oka, hogy 4 kötés távolsággal már egy kémiai szerkezetben egy 6-os gyűrű is leírható egy kötés távolságra a hasadás végétől, leírható például egy benzil-hasadás. Három kötésnél erőteljesebb szűkítést nem érdemes alkalmazni, hiszen a szabályok így veszítenek specifikusságukból.

Egy szabály második részét a megjelenési oldal adja. Ez azt mutatja meg, hogy a két részszerkezet közül melyik oldal milyen intenzíven szokott megjelenni a spektrumban. Ennek leírására egy öt tagból álló sorozatot használtam, tehát összesen 10 intenzitás értéket (az intervallum alsó és felső végét számítva). A megjelenés leírásának öt tagja tehát a tényleges szerkezetet (0), az egy (1) illetve két (2) hidrogénnel nagyobbat, és az egy (-1) illetve két (-2) hidrogénnel kisebbet jelenti. A módszer egyik előnye, hogy bizonyos egyszerűbb átrendeződési reakciókat is képes formálisan leírni. Másik előnye, hogy függetlenné teszi az algoritmust az ionizáció módjától. Az algoritmus további módosítás nélkül használható, csupán minden szabályban lesz egy eltolódás pozitív vagy negatív irányba, attól függően, hogy M+H, vagy M-H kvázi-molekulaiont vizsgálunk.

Az implementált algoritmus képes egyszerű és kevésbé egyszerű szabályok generálására a spektrumok és szerkezetek alapján. Az alábbi példa egy kéndioxid (SO₂) heterociklusos gyűrűből való kilépésének kódját írja le. A szabály első részében M BEGIN és M END kifejezések között a két részszerkezet szerepel MOL formátumban, a második részben pedig a két oldal megjelenések 5-ös intenzitás-csoportjai.

```
M BEGIN
...
3 2
0.0000 0.0000 0.0000 S& 0
0.0000 0.0000 0.0000 O 0
0.0000 0.0000 0.0000 O 0
1 2 2
1 3 2
M END
M BEGIN
...
M END
0.113651 0.0177676 0.000982431 1.01588 0.175274
...
```


3.3.2. A szerkezetek összehasonlítása

A részgráf-izomorfia egzakt megoldásai exponenciális időigényűek. A megoldás során egy sztochasztikus megközelítést választottam. Véletlenszerű megfeleltetéseket generáltam a két gráf pontjai között, majd a 42. egyenlet segítségével számítottam a legjobb illeszkedést. Ennek a módszernek előnye, hogy időigénye tetszés szerint megválasztható. Az implementáláskor egy $O(n^4)$ -es ciklus mellett döntöttem (ahol n a kisebb gráf illesztendő csúcsainak száma). Ez pontosan adta vissza a hasonlóság értékét $n < 12$ esetén, és kielégítően kicsit nagyobb csúcscsámokra. A gyakorlatban azonban a felhasznált 4-mélységű környezetek esetén az $n < 12$ egy racionális választásnak bizonyult. A vizsgált molekulák esetén általában $n < 10$ teljesült.

Egy adott megfeleltetés esetén a hasonlóság mértékének számításánál egyrészt összegeznünk kell a csúcsok, másrészt az élek egyezéseit, tehát a hasonlósági index definíciója a következőképpen alakul:

$$H = \left(1 - \frac{\sum_{i \in \text{atom}} a_i}{\sum_{i \in \text{atom}} 1}\right) \left(1 - \frac{\sum_{i,j \in \text{atom}} k_{ij}}{\sum_{i,j \in \text{atom}} 1}\right) \quad (42)$$

Az egyenletben $a_i = 1$, ha a két atomtípus különböző, $a_i = 0$, ha azonos és $a_i = 0,7$, ha mindkettő szénatom, csak különböző konfigurációjú. Hasonlóan a kötések esetén $k_{ij} = 0$, ha azonos a kötésrend, $k_{ij} = 0,3$, ha különböző és egyik kötésrend sem 0 (tehát létezik), $k_{ij} = 1$ egyébként. Az a_i és k_{ij} tört értékeit a különböző tautomer formák összehasonlítása indokolja.

3.3.3. A megjelenési intenzitások számítása

Ahogy az előző fejezetben már szerepelt, a megjelenési intenzitások az adott csúcs intenzitásának környezetéhez képest mért relatív magasságát jelentik. A számítások figyelembe kellett venni, hogy az adott csúcs intenzitásába nemcsak a fragmens alapizotóptömegénél mért intenzitás számít bele, hanem a természetes izotópeloszlás alapján számított többi izotópcsúcs is, amelyek nem tekinthetők a csúcs környezetében látható „zajnak”.

Az izotópcsúcsok figyelembe vételének másik problémája, hogy egy adott fragmens spektrumban való megjelenéséhez nemcsak azt kell megvizsgálnunk, hogy az adott tömegénél mekkora csúcsot találunk a spektrumban, hanem azt is, hogy a természetes izotópeloszlás alapján számolt többi izotópcsúcsa szintén „elfér-e” a mért spektrumba. Feltételezve tehát,

hogy az $M, M+1, M-1 \dots$ formának megfelelő alapizotóp tömegnél $I_{\text{exp},0}$ a mért intenzitás, eggyel magasabb tömegnél $I_{\text{exp},1} \dots$, továbbá hogy az adott részszerkezet természetes izotópgyakoriságok alapján számított intenzitás-eloszlása $I_{\text{calc},0}, I_{\text{calc},1} \dots$, a megjelenés maximális lehetséges abszolút intenzitása a következő formula szerint számítható:

$$I_M = \min \left(\frac{I_{\text{exp},i}}{I_{\text{calc},i}/I_{\text{calc},0}}, i = 0, 1, \dots \right). \quad (43)$$

Természetesen a gyakorlatban – a kis csúcsintenzitások mérésének hibái miatt – az izotópeloszlást érdemes valamilyen racionális küszöb alatt elhanyagolni, erre önkényesen 1%-ot választottam. Az így kapott I_M abszolút intenzitás-érték utáná átszámolható a jel/zaj viszony alapú reprezentációba.

A megjelenési értékek kezelésének másik aspektusa az irodalmi részben bemutatott intervallum-valószínűségek mintájára kidolgozott megoldás. Minthogy a tömegspektrumokban a relatív intenzitás értéke számottevő hibával terhelt, mindenképpen szükség van tolerancia-tartományok megadására. Ennek egyik módja, az egyedi előfordulásokból számított átlag/szórás értékpárok megadása lenne. Ugyanakkor az intervallum-valószínűségek analógiájára, a számított gyakoriság paraméterek átlaga helyett is megadhatjuk azok intervallumait. A választott megoldásnak két előnye van. Egyrészt egyszerű az implementációja és a szabályok finomításának módja. Másrészt egyszerű a szabályok utólagos szűrése is. A nem szignifikáns szabályok kiszűrééhez elegendő törölni azokat a szabályokat, amelyeknél az intervallum mérete meghalad egy adott küszöböt. Az intervallum méret-korlátait érdemes logaritmikus skálán kezelni, ez közelebb áll a gyakorlatban használt „szemmértékhez”. Tehát törölnődök azok a szabályok, ahol a megjelenési intenzitás intervallum több mint egy nagyságrendet fog át.

3.3.4. A megvalósított program

A fenti algoritmusrészletek együttes alkalmazásával fejlesztett program képes tetszőleges típusú ionizációs technikával képződött tömegspektrumokból intenzitásértékeket is figyelembe vevő fragmentációs szabályok indukciójára. A program elsősorban spektrumok kötegelte (batch) feldolgozására készült, háttérben, emberi közreműködés nélkül használható. A forráskódot objektumorientált C++ [94; 103] nyelven írtam, a forrás összesen 2900 sornyi kódot tartalmaz. A forráskód megtalálható a <http://alsp.sourceforge.net> web helyen.

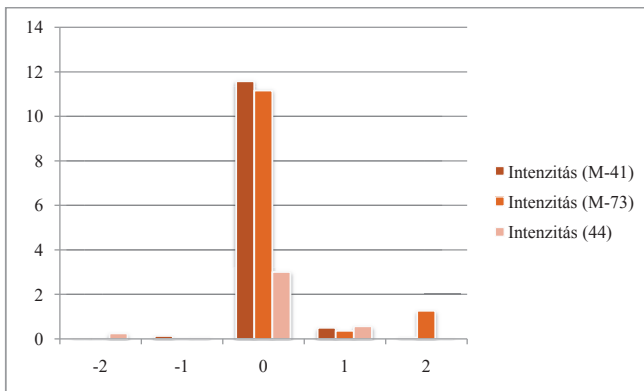
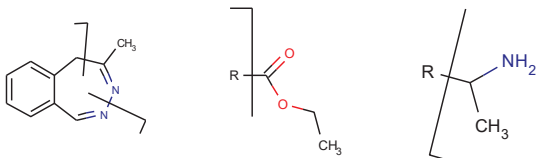
Az algoritmus sebesség-meghatározó lépése a szerkezetek hasonlóságának számítása. A tesztelés során 63 molekula spektrum-szerkezet párijainak feldolgozása mintegy 5 órát vett igénybe egy ma elérhető átlagos asztali személyi számítógépen (Intel P4 3GHz, 1 GB RAM).

3.3.5. A meghatározott szabályok

A vizsgált 63 vegyületből alkotott csoport, vegyületenként közelítőleg 10 jelentősebb fragmensének, tehát összesen mintegy 1000 fragmentáció jellemzésére az algoritmus sikeresen generált 195 szabályt, amelyből szűrés után 57 maradt.

Ez azt mutatja, hogy az emberi kiértékelés mintájára valóban sikerült a vegyületcsaládokon belül kevés szabállyal leírni a megfelelő tömegspektrumokat. A generált szabályok vizsgálata megmutatta, hogy az ismert fragmentációs utaknak megfelelő szabályok felfedezhetők a kapott listákban.

A 27. ábra három kiragadott szabályt mutat a kapott halmazból. A grafikonon a három szabálynak a megjelenési intenzitás oldalát ábrázoltam, tehát azt, hogy milyen intenzitással jelenik meg a vizsgált spektrumokban az ábrázolt fragmentáció-típus. A vízszintes tengelyen a -2, -1, 0, 1, 2 számok rendre a -2 H-atom, -1 H-atom, ... spektrumban való előfordulását jelentik. A grafikon felett a kémiai szerkezetek mutatják a tényleges fragmentációt. Jelen esetben egy benzodiazepin szerkezetből való acetonitril kilépését, egy etilészter csoport lehasadását illetve egy alifás amino-funkció hasadását láthatjuk. Az *y* skála a grafikonon abszolút skálának tekinthető, tehát a vizsgált vegyületek körében háromszor intenzívebbnek mutatkozott a két első fragmentációs reakció.



27. ábra. A tanuló adatkészletből az algoritmus által generált három jellemző fragmentáció, szerkezetek és a hozzájuk tartozó intenzitás-értékek.

Mindhárom előbb bemutatott szabály az adott szerkezetekre irodalomban leírt, jellemző, ismert fragmentáció. Az aminok α -hasadása, az észter-csoport könnyű fragmentálódása és a benzodiazepinek acetonitril-vesztése napi gyakorlatban gyakran előforduló folyamat. A módszer teljesítőképességét mutatja, hogy minden előzetes kémiai-tömegspektrometriai ismeret nélkül, pusztán a szerkezetek és a spektrumok együttesében lévő „minták” felismerésével képes volt a program az ismert fragmentációs szabályok reprodukálására. A meghatározott fragmentációs intenzitások természetesen az adott mintahalmazra jellemzőek. A vizsgált gyógyszermolekulák típusából adódik, hogy a benzodiazepinek 41-vesztése ilyen dominánsan jelentkezik.

Összefoglalásképpen elmondható, hogy a módszer az emberi spektrumfejtés technikáit modellezve sikerült egy olyan valószínűségi szabályokat tanuló algoritmust létrehozni, amellyel adott vegyületsaládokon belül fragmentációs szabályok nyerhetők. A programnak semmilyen kémiai előismeret nem áll rendelkezésre a szerkezetek és a spektrumok felismerésén kívül. A kapott fragmentációs szabályok intenzitásokat is figyelembe vesznek, mégpedig szintén az emberi feldolgozáshoz hasonló módon: kicsi, intenzív, közepes, zajos

jellemzőket tanulva. A meghatározott szabályok összhangban vannak az irodalomban található fragmentációs mechanizmusokkal. [104; 105]

A módszer erőssége a kereskedelmi forgalomban kapható társaival szemben elsősorban az, hogy azok több tízezresre duzzadt fragmentációs adatbázisával szemben bizonyos vegyületsaládokon belül képes felismerni az arra a családra jellemző nagyságrendekkel kisebb számú az intenzitásokat is figyelembe vevő hasadási szabályt. Ugyanakkor bizonyos bonyolult heterociklusos esetektől eltekintve azt is megmutatta, hogy a tömegspektrumok csúcseinak jelentős része jó hatásfokkal azonosítható átrendeződési reakciók figyelembe vétele nélkül, legfeljebb két-három egymást követő hasadási reakció megengedésével. Az asszignált csúcsok aránya a vizsgált gyógyszermolekulák esetén átlagosan 92 %.

4. Összefoglalás

Doktori munkám során a tömegspektrometria modellezési lehetőségeinek különböző szintjeit vizsgáltam, a legpontosabb fizikai-kémiai alapokat használó szimulációktól az emberi spektrumfejtést modellező mesterséges intelligencia módszerekig.

Munkám során módszert dolgoztam ki a SORI-CID kísérletek modellezésére. Valószínűségi eloszlásfüggvények alkalmazásával meghatároztam az átlagos ütközési energiát, így lehetővé tettem a SORI-CID folyamatok gyors szimulációját. Meghatároztam az ütközési energia-átadás és az infravörös hűlés mértékét leucin-enkefalin – argon rendszer esetén. Kimutattam, hogy az ütközési energia-átadás önmagában nem képes magyarázni a fragmentációs görbék idő szerinti lefutását. Vizsgálataim rámutattak az ütközési energia-átadás és az infravörös hűlés független meghatározásának problémájára, a két paraméter korreláltságára.

A kiterjesztett kinetikus módszer vizsgálatokor számítógépes programot fejlesztettem ki, amely alkalmas „virtuális kísérleti adatok” több ezres készletének generálására, illetve az ezekből számított entalpia és entrópia értékek statisztikai vizsgálatára. Munkám során meghatároztam a kiterjesztett kinetikus módszer közelítéseiből, a kísérleti hibák és a referencia anyagok termokémiai adatainak pontatlan ismeretéből adódó hibák nagyságát. Kimutattam, hogy az entalpia és entrópia tagok korrelációja miatt a látszólagos gázfázisú bázicitás az előzőeknél pontosabban határozható meg. Eredményeim bizonyították, hogy a kiterjesztett kinetikus módszer meghatározott feltételek mellett pontos becslést adja az entalpia és entrópia tagoknak.

Az emberi spektrumfejtés folyamatait modellező, valószínűségi fragmentációs szabályokat tanuló algoritmust dolgoztam ki. A módszert sikerrel alkalmaztam gyógyszermolekulák elektronütközéses spektrumai esetén fragmentációs szabályok gépi tanulására. A meghatározott szabályok összhangban vannak az irodalom alapján elvártakkal, segítségükkel a spektrumok csúcsai jó hatásfokkal azonosíthatók.

Eredményeim rámutattak arra, hogy a tömegspektrometriával kapcsolatos különböző feladatok eltérő modellezési módszereket és matematikai eszközöket igényelnek. Minden kérdésfelvetéshez érdemes kiválasztani az adott feladatot kielégítő pontossággal megoldó – nem feltétlenül bonyolult – számítási módszert.

5. Summary

During my Ph.D. work I studied different levels of modeling in mass spectrometry, from the simulations that utilize the most accurate physico-chemical background, to the artificial intelligence methods that mimic the human spectrum processing routine.

I worked out a method to model SORI-CID experiments. Kinetic energy distribution function was derived and used to achieve reasonable MassKinetics run times. I determined the collisional energy transfer and the infrared cooling rate modeling SORI-CID processes in a leucine-enkefaline – argon system. The collisional energy transfer cannot explain the shape of the survival yield curves by itself, as pointed out in the thesis. The study also showed the correlation and the difficulty of the deconvolution of the energy transfer and IR cooling.

I developed a computer program in order to examine the applicability of the extended kinetic method. The program is capable of generating thousands of “virtual experimental datasets” and performing their statistical evaluation afterwards. I determined the effect of the approximations in the kinetic method, experimental errors and poor knowledge of thermochemical properties of reference compounds on the deviation of the results. I pointed out that the apparent gas-phase basicity can be computed quite accurately, due to the cross-correlation between the entropy and enthalpy values. The results suggest that the extended kinetic method – under appropriate conditions – can be used to estimate accurate thermochemical properties.

I designed a probabilistic fragmentation rule induction algorithm based upon the modeling of human interpretation of mass spectra. The learning algorithm does not include any prior chemical knowledge. The developed method was successfully applied to the learning of intensity aware fragmentation rules from electron impact mass spectra of pharmaceutical compounds. The extracted rules are in good agreement with the literature and the processed spectra have a reasonably high assignment ratio.

The results suggest that different applications of mass spectrometry require different mathematical and modeling methods. In every case one should choose the appropriate – not the most complex – theoretical method to achieve the optimal solution of the problem.

Függelék. A kinetikus energia eloszlásának meghatározása

Az eloszlás meghatározásához tekintsünk egy olyan $E(t)$ energia-idő függvényt, amelyre a következő feltételezéseket tesszük:

1. $E(t)$ periodikus,
2. $E(t)$ folytonos,
3. $E'(t) = \frac{dE}{dt}$ bármely véges intervallumon legfeljebb véges sok izolált pontban vagy intervallumon 0 (tehát szemléletesen fogalmazva véges sok alkalommal vált irányt).

Ezek a feltételezések mind érvényesek a SORI kinetikus energia-idő függvényünkre.

Az 1. feltételből következik, hogy elegendő egy $[0, T]$ intervallumon vizsgálnunk a függvényt, ahol T a periódusidő.

A 2. és 3. feltétel alapján a $[0, T]$ intervallum felosztható véges sok intervallumra úgy, hogy ezek fölött az $E(t)$ függvény konstans, vagy szigorúan monoton. Jelöljük az osztópontokat a következőképpen: $0 = a_0, a_1, \dots, a_n = T$. Egyenletesen mintavételezett $E(t)$ esetén az E eloszlásfüggvényének meghatározásához nyilvánvalóan elegendő ezeken az $[a_i, a_j]$ intervallumokon meghatározni az eloszlásfüggvényt, majd összesíteni.

Amennyiben egy adott $[a_i, a_{j=i+1}]$ intervallumon $E(t) = E_{ij}$ konstans, akkor az $F(x)$ eloszlásfüggvényre

$$F(x) = 0, \tag{44}$$

ha $x < E_0$, és

$$F(x) = 1, \tag{45}$$

ha $x \geq E_0$. A sűrűségfüggvény pedig ebből adódóan

$$f(x) = \delta(x - E_0), \tag{46}$$

ahol δ a Dirac-delta függvény.

Amennyiben $[a_i, a_j]$ intervallumon $E(t)$ szigorúan monoton nő, tehát $E(a_i) < E(a_j)$, akkor

$$F(x) = \frac{t - a_i}{a_j - a_i}, \quad (47)$$

ahol $t = E^{-1}(x)$. (Megjegyzés: a szigorú monotonitás miatt az inverz függvény létezik.)

A sűrűségfüggvény így ebben az esetben

$$f(x) = F'(x) = \frac{1}{a_j - a_i} (E^{-1}(x))'. \quad (48)$$

Szigorúan monoton csökkenő esetben hasonlóan adódik

$$F(x) = \frac{a_j - t}{a_j - a_i}, \quad (49)$$

$$f(x) = \frac{-1}{a_j - a_i} (E^{-1}(x))'.$$

Amennyiben az esetleges konstans szakaszoktól eltekintünk, a teljes $[0, T]$ periódusidőre összesítve

$$f(x) = \frac{1}{T} \sum_{i=0}^{n-1} (-1)^{s(i)} \frac{1}{E_i'(E_i^{-1}(x))}, \quad (50)$$

ahol $s(i)$ értéke 1, ha szigorúan monoton nő a függvény az adott intervallumon és -1 , ha csökken. Amennyiben vannak konstans szakaszok is, azokra a megfelelő Dirac-delta függvényeket még hozzá kell venni, majd – minthogy egy valószínűségi sűrűségfüggvényről van szó – 1-re normálni.

Irodalomjegyzék

- [1]. **Middleditch, BS, [ed.]**. *Practical mass spectrometry. A contemporary introduction*. New York : Plenum Press, 1979.
- [2]. *Principles of collisional activation in analytical mass spectrometry*. **McLuckey, SA**. 1992, Journal of American Society for Mass Spectrometry, Vol. 3, p. 599.
- [3]. *Electrospray: principles and practice*. **Gaskell, SJ**. 1997, Journal of Mass Spectrometry, Vol. 32, p. 677.
- [4]. *Fourier transform ion cyclotron resonance mass spectrometry: A primer*. **Marshall, AG, Hendrickson, CL and Jackson, GS**. 1998, Mass Spectrometry Reviews, Vol. 17, p. 1.
- [5]. *Examples of Fourier transform ion cyclotron resonance mass spectrometry developments: From ion physics to remote access biochemical mass spectrometry*. **Rompp, A, et al**. 2005, European Journal of Mass Spectrometry, Vol. 11, p. 443.
- [6]. *Ion cyclotron resonance spectroscopy*. **Beauchamp, JL**. 1971, Annual Review of Physical Chemistry, Vol. 22, p. 527.
- [7]. **Korn, GA és Korn, TM**. *Matematikai kézikönyv műszakiaknak*. Budapest : Műszaki könyvkiadó, 1975.
- [8]. **Press, WH, et al**. *Numerical recipes in C++: The art of scientific computing*. UK : Cambridge University Press, 2002. p. 398.
- [9]. *Mass spectrometry in proteomics*. **Aebersold, R and Goodlett, DR**. 2001, Chemical Reviews, Vol. 101, p. 269.
- [10]. **James, Peter, [ed.]**. *Proteome research: mass spectrometry*. Berlin : Springer, 2001.
- [11]. *Proteomics by FTICR mass spectrometry*. **Bogdanov, D and Smith, RD**. 2005, Mass Spectrometry Reviews, Vol. 24, p. 168.
- [12]. *Fragmentation energetics of small peptides from multiple-collision activation and surface-induced dissociation in FT-ICR MS*. **Laskin, J, Denisov, E and Futrell, JH**. 2002, International Journal of Mass Spectrometry, Vol. 219, p. 189.

- [13]. *A comparative study of collision-induced and surface-induced dissociation. I. Fragmentation of protonated dialanine.* **Laskin, J, Denisov, E and Futrell, JH.** 2000, Journal of American Chemical Society, Vol. 122, p. 9703.
- [14]. *Internal energy distributions resulting from sustained off-resonance excitation in FTMS. I. Fragmentation of the bromobenzene radical cation.* **Laskin, J, Byrd, M and Futrell, JH.** 2000, International Journal of Mass Spectrometry, Vol. 195/196, p. 285.
- [15]. *Activation of large ions in FT-ICR mass spectrometry.* **Laskin, J and Futrell, JH.** 2005, Mass Spectrometry Reviews, Vol. 24, p. 135.
- [16]. *BIRD (blackbody infrared radiative dissociation): Evolution, principles, and applications.* **Dunbar, RC.** 2004, Mass Spectrometry Reviews, Vol. 23, p. 127.
- [17]. *Sustained off-resonance irradiation for collision-activated dissociation involving Fourier transform mass spectrometry. Collision-activated dissociation technique that emulates infrared multiphoton dissociation.* **Gauthier, JW, Trautman, TR and Jacobson, DB.** 1991, Analytica Chimica Acta, Vol. 246, p. 211.
- [18]. *Pressure limited sustained off-resonance irradiation for collision-activated dissociation in Fourier transform mass spectrometry.* **Gorshkov, MV, Pasa-Tolic, L and Smith, RD.** 1999, Journal of American Society for Mass Spectrometry, Vol. 10, p. 15.
- [19]. *Internal energy distributions resulting from sustained off-resonance excitation in Fourier transform ion cyclotron resonance mass spectrometry. II. Fragmentation of the 1-bromonaphthalene radical cation.* **Laskin, J and Futrell, JH.** 2000, Journal of Physical Chemistry A, Vol. 104, p. 5484.
- [20]. *Theory of collisional activation of macromolecules. Impulsive collisions of organic ions.* **Uggerud, E and Derrick, PJ.** 1991, Journal of Physical Chemistry, Vol. 95, p. 1430.
- [21]. *The geometrical factor of infinitely long cylindrical ICR cells for collision energy-resolved mass spectrometry: appearance energies of $[E12]^+$ ($E = P, As, Sb, \text{ and } Bi$) from collision-induced dissociation of $[E13]^+$ and $[E12.ligand]^+$ complexes.* **Sievers, HL, Grützmacher, HF and Caravatti, P.** 1996, International Journal of Mass Spectrometry and Ion Processes, Vol. 157/158, p. 233.

- [22]. *The effective temperature of peptide ions dissociated by sustained off-resonance irradiation collisional activation in Fourier transform mass spectrometry.* **Schnier, PD, Jurchen, JC and Williams, ER.** 1999, Journal of Physical Chemistry B, Vol. 103, p. 737.
- [23]. *Simulation for internal energy deposition in sustained off-resonance irradiation collisional activation using a Monte Carlo method.* **Fujiwara, M and Naito, Y.** 1999, Rapid Communications in Mass Spectrometry, Vol. 13, p. 1633.
- [24]. *Infrared radiative cooling of gas-phase ions.* **Dunbar, RC.** 1992, Mass Spectrometry Reviews, Vol. 11, p. 309.
- [25]. *Activation of unimolecular reactions by ambient blackbody radiation.* **Dunbar, RC and McMahan, TB.** 1998, Science, Vol. 279, p. 194.
- [26]. *Thermometric study of CO₂ laser heating and radiative cooling of n-butylbenzene ions.* **Uechi, GT and Dunbar, RC.** 1993, Journal of Chemical Physics, Vol. 98, p. 7888.
- [27]. *Energetics from slow infrared multiphoton dissociation of biomolecules.* **Jockusch, RA, Paech, K and Williams, ER.** 2000, Journal of Physical Chemistry A, Vol. 104, p. 3188.
- [28]. *The kinetic method of making thermochemical determinations.* **Cooks, RG, Koskinen, JT and Thomas, PD.** 1999, Journal of Mass Spectrometry, Vol. 34, p. 85.
- [29]. *Experimental techniques in gas-phase ion thermochemistry.* **Ervin, KM.** 2001, Chemical Reviews, Vol. 101, p. 391.
- [30]. *The theoretical basis of the kinetic method from the point of view of finite bath theory.* **Laskin, J and Futrell, JH.** 2000, Journal of Physical Chemistry A, Vol. 104, p. 8829.
- [31]. *Microcanonical analysis of the kinetic method. The meaning of the "effective temperature".* **Ervin, KM.** 2000, International Journal of Mass Spectrometry, Vol. 195/196, p. 271.
- [32]. *How closely related are the effective and the real temperature.* **Drahos, L and Vékely, K.** 1999, Journal of Mass Spectrometry, Vol. 34, p. 79.
- [33]. *Entropy considerations in kinetic method experiments.* **Wesdemiotis, C.** 2004, Journal of Mass Spectrometry, Vol. 39, p. 998.

- [34]. *Kinetic energy release of protonated methanol clusters using the low-temperature fast-atom bombardment: Experiment and theory combined.* **Gömöry, Á, et al.** 2004, European Journal of Mass Spectrometry, Vol. 10, p. 213.
- [35]. *Ionization energy determination by the kinetic method.* **Wong, PSH, Ma, S and Cooks, RG.** 1996, Analytical Chemistry, Vol. 68, p. 4254.
- [36]. *Proton affinity of deuterated acetonitrile estimated by the kinetic method with full entropy analysis.* **Williams, TI, Denault, JW and Cooks, RG.** 2001, International Journal of Mass Spectrometry, Vol. 210/211, p. 133.
- [37]. *The gas-phase basicities and proton affinities of amino acids and peptides.* **Harrison, AG.** 1997, Mass Spectrometry Reviews, Vol. 16, p. 201.
- [38]. *Thermochemical aspects of proton transfer in the gas phase.* **Gal, JF, Maria, PC and Raczynska, ED.** 2001, Journal of Mass Spectrometry, Vol. 36, p. 699.
- [39]. *Entropy evaluation using the kinetic method: is it feasible?* **Drahos, L and Vékey, K.** 2003, Journal of Mass Spectrometry, Vol. 38, p. 1025.
- [40]. *Systematic and random errors in ion affinities and activation entropies from the extended kinetic method.* **Ervin, KM and Armentrout, PB.** 2004, Journal of Mass Spectrometry, Vol. 39, p. 1004.
- [41]. *Accuracy of enthalpy and entropy determination using the kinetic method: are we approaching a consensus?* **Drahos, L, Peltz, C and Vékey, K.** 2004, Journal of Mass Spectrometry, Vol. 39, p. 1016.
- [42]. **Dahl, DA.** SIMION 3D. 7th edition Idaho : Idaho National Engineering and Environmental Laboratory, 1995.
- [43]. *Simulation of ion internal energy evolution during collisional processes in quadrupole ion traps.* **Plass, WR, et al.** 1996, Journal of Chemical Physics, Vol. 104, p. 2214.
- [44]. *Windows version of the ion trap simulation program ITSIM: a powerful heuristic and predictive tool in ion trap mass spectrometry.* **Bui, HA and Cooks, RG.** 1998, Journal of Mass Spectrometry, Vol. 33, p. 297.

- [45]. *Kinetics of collision-induced dissociation in the Paul trap: a first-order model*. **Goeringer, DE and McLuckey, SA**. 1996, Rapid Communications in Mass Spectrometry, Vol. 10, p. 328.
- [46]. *Relaxation of internally excited high-mass ions simulated under typical quadrupole ion trap storage conditions*. **Goeringer, DE and McLuckey, SA**. 1998, International Journal of Mass Spectrometry, Vol. 177, p. 163.
- [47]. *MassKinetics: a theoretical model of mass spectra incorporating physical processes, reaction kinetics and mathematical descriptions*. **Drahos, L and Vékey, K**. 2001, Journal of Mass Spectrometry, Vol. 36, p. 237.
- [48]. **Drahos, L and Vékey, K**. *MassKinetics computer program*. [Online] 2004. [Cited: April 19, 2008.] <http://www.chemres.hu/ms/masskinetics>.
- [49]. *Determination of the thermal energy and its distribution in peptides*. **Drahos, L and Vékey, K**. 1999, Journal of American Society for Mass Spectrometry, Vol. 10, p. 323.
- [50]. *Energetics and dynamics of peptide fragmentation from multiple-collision activation and surface-induced dissociation studies*. **Laskin, J**. 2004, European Journal of Mass Spectrometry, Vol. 10, p. 259.
- [51]. *Landmarks in the theory of mass spectra*. **Lorquet, JC**. 2000, International Journal of Mass Spectrometry, Vol. 200, p. 43.
- [52]. **Tournès, Dominique**. *L'intégration approchée des équations différentielles ordinaires (1617-1914)*. l'Université Paris 7-Denis Diderot. 1996. Doktori értekezés.
- [53]. *Mass spectrometry-based proteomics*. **Aebersold, R and Mann, M**. 2003, Nature, Vol. 422, p. 198.
- [54]. *LC/MS applications in drug development*. **Lee, MS and Kerns, EH**. 1999, Mass Spectrometry Reviews, Vol. 18, p. 187.
- [55]. *Code developments to improve the efficiency of automated MS/MS spectra interpretation*. **Sadygov, RG, et al**. 2002, Journal of Proteome Research, Vol. 1, p. 211.

- [56]. *A hypergeometric probability model for protein identification and validation using tandem mass spectra data and protein sequence databases.* **Sadygov, RG and Yates, JR.** 2003, Analytical Chemistry, Vol. 75, p. 3792.
- [57]. *Peptide sequence motif analysis of tandem MS data with the SALSA algorithm.* **Liebler, DC, et al.** 2002, Analytical Chemistry, Vol. 74, p. 203.
- [58]. *A statistical model for identifying proteins by tandem mass spectrometry.* **Nezvizhskii, AI, et al.** 2003, Analytical Chemistry, Vol. 75, p. 4646.
- [59]. *Large-scale protein identification using mass spectrometry.* **Lin, D, Tabb, DL and Yates, JR.** 2003, Biochimica et Biophysica Acta, Vol. 1646, p. 1.
- [60]. *MS2Assign, automated assignment and nomenclature of tandem mass spectra of chemically crosslinked peptides.* **Schilling, B, et al.** 2003, Journal of American Society for Mass Spectrometry, Vol. 14, p. 834.
- [61]. *De novo peptide sequencing based on a divide-and-conquer algorithm and peptide tandem spectrum simulation.* **Zhang, Z.** 2004, Analytical Chemistry, Vol. 76, p. 6374.
- [62]. *De novo peptide sequencing by two-dimensional fragment correlation mass spectrometry.* **Zhang, Z and McElvain, JS.** 2000, Analytical Chemistry, Vol. 72, p. 2337.
- [63]. *MSNovo: A dynamic programming algorithm for de novo peptide sequencing via tandem mass spectrometry.* **Mo, L, et al.** 2007, Analytical Chemistry, Vol. 79, p. 4870.
- [64]. **Kowalski, BR, [ed.].** *Chemometrics. Mathematics and statistics in chemistry.* Dordrecht : D. Reidel Publishing Co., 1983.
- [65]. **Dinya, Elek.** *Biometria az orvosi gyakorlatban.* Budapest : Medicina, 2001.
- [66]. **Boggs, PT and Rogers, JE.** Orthogonal distance regression. [ed.] PJ Brown and WA Fuller. *Contemporary mathematics, volume 112: Statistical analysis of measurement error models and their applications.* Providence, Rhode Island : American Mathematical Society, 1990, pp. 183-194.
- [67]. **Rogers, JE.** *ODRPACK: Software for orthogonal distance regression.* [Online] 1997. [Cited: April 19, 2008.] <http://www.boulder.nist.gov/mcsd/Staff/JRogers/odrpac.html>.

- [68]. **Obádovics, Gyula J és Szarka, Zoltán.** *Felsőbb matematika.* Második kiadás. Budapest : Scholar, 1999.
- [69]. **Stoyan, Gisbert és Takó, Galina.** *Numerikus módszerek.* Budapest : ELTE - TypoTEX, 1993. 1. kötet.
- [70]. **Papadimitriou, CH.** *Számítási bonyolultság.* Győr : Novadat Bt., 1999.
- [71]. **Bar-Yam, Yaneer.** *Dynamics of complex systems.* Reading, Massachusetts : Addison-Wesley, 1997.
- [72]. **Jain, LC and Martin, NM, [ed.].** *Fusion of neural networks, fuzzy sets, and genetic algorithms.* Boca Raton : CRC Press, 1999.
- [73]. *Artificial neural network analysis for evaluation of peptide MS/MS spectra in proteomics.* **Baczek, T, et al.** 2004, Analytical Chemistry, Vol. 76, p. 1726.
- [74]. **Cox, Earl.** *Fuzzy modeling and genetic algorithms for data mining and exploration.* San Francisco : Morgan Kaufmann Publishers, 2005.
- [75]. **Frasconi, Paolo.** *Neural networks and kernels for learning discrete data structures.* Department of Systems and Computer Science, University of Florence. 2005. <http://www.dsi.unifi.it/neural/>.
- [76]. *Artificial neural network modeling of Kováts retention indices for noncyclic and monocyclic terpenes.* **Jalali-Heravi, M and Fatemi, MH.** 2001, Journal of Chromatography A, Vol. 915, p. 177.
- [77]. *A machine learning perspective on the development of clinical decision support systems utilizing mass spectra of blood samples.* **Shin, H and Markey, MK.** 2006, Journal of Biomedical Informatics, Vol. 39, p. 227.
- [78]. **Buchanan, BG and Shortliffe, EH, [ed.].** *Rule-based expert systems. The MYCIN experiments of the Stanford Heuristic Programming Project.* Reading, Massachusetts : Addison-Wesley Publishing Co., 1984.
- [79]. **Buchanan, B, Sutherland, G and Feigenbaum, EA.** HEURISTIC DENDRAL: a program for generating explanatory hypotheses in organic chemistry. [ed.] B Meltzer and D Michie. *Machine Intelligence 4.* Edinburgh : Edinburgh University Press, 1969, p. 209.

- [80]. *DENDRAL: a case study of the first expert system for scientific hypotheses formation*. **Lindsay, RK, et al.** 1993, Artificial Intelligence, Vol. 61, p. 209.
- [81]. **Pál, Lénárd.** *A valószínűségszámítás és a statisztika alapjai*. Budapest : Akadémiai Kiadó, 1995. 1-2. kötet.
- [82]. *The refinement of probabilistic rule sets: sociopathic interactions*. **Wilkins, DC and Ma, Y.** 1994, Artificial Intelligence, Vol. 70, p. 1.
- [83]. **Kyburg Jr, HE.** Interval-valued probabilities. *Imprecise probabilities project*. [Online] 1998. [Cited: April 19, 2008.] <http://citeseer.ist.psu.edu/kyburg98intervalvalued.html>.
- [84]. **Ferson, S, et al.** Dependence in probabilistic modeling, Dempster-Shafer theory and probability bounds analysis. [Online] 2004. [Cited: April 19, 2008.] <http://www.ramas.com/depend.pdf>.
- [85]. *Experimental calibration of the SORI-CID internal energy scale: Energy uptake and loss*. **Guo, XH, et al.** 2003, International Journal of Mass Spectrometry, Vol. 225, p. 71.
- [86]. *Design and performance of a new FT-ICR cell operating at a temperature range of 77 to 438 K*. **Guo, XH, et al.** 2004, International Journal of Mass Spectrometry, Vol. 231, p. 37.
- [87]. *SORI excitation: collisional and radiative processes*. **Peltz, Cs, Drahos, L and Vékey, K.** 2007, Journal of American Society for Mass Spectrometry, Vol. 18, p. 2119.
- [88]. *Density-functional exchange energy approximation with correct asymptotic behavior*. **Becke, AD.** 1988, Physical Reviews A, Vol. 38, p. 3098.
- [89]. *Density-functional thermochemistry. 2. The effect of the Perdew-Wang generalized gradient correlation correction*. **Becke, AD.** 1992, Journal of Chemical Physics, Vol. 97, p. 9173.
- [90]. *Development of the Colle-Salvetti correlation energy formula into a functional of electron density*. **Lee, C, Yang, W and Parr, RG.** 1988, Physical Reviews B, Vol. 37, p. 785.
- [91]. *Novel method to determine collisional energy transfer efficiency by Fourier transform ion cyclotron resonance mass spectrometry*. **Heeren, RMA and Vékey, K.** 1998, Rapid Communications in Mass Spectrometry, Vol. 12, p. 1175.

- [92]. *Guided ion beam study of collision-induced dissociation dynamics: Integral and differential cross sections.* **Muntean, F and Armentrout, PB.** 2001, Journal of Chemical Physics, Vol. 115, p. 1213.
- [93]. *Energy transfer pathways in the collisional activation of peptides.* **Meroueh, O and Hase, WL.** 2000, International Journal of Mass Spectrometry, Vol. 201, p. 233.
- [94]. **Eckel, Bruce.** *Thinking in C++.* 2nd edition. New Jersey : Prentice Hall, 2000.
- [95]. **Hajnal, Péter.** *Gráfelmélet.* Szeged : Polygon, 2003.
- [96]. **Cormen, TH, Leieron, CE és Rivest, RL.** *Algoritmusok.* Harmadik kiadás. Budapest : Műszaki könyvkiadó, 2001.
- [97]. **Karakoc, E, Cherkasov, A and Sahinalp, SC.** Distance based algorithms for small molecule classification and structural similarity search. [Online] [Cited: April 19, 2008.] <http://citeseer.ist.psu.edu/746678.html>.
- [98]. *Automatic identification of molecular similarity using reduced-graph representation of chemical structure.* **Takahashi, Y, Sukekawa, M and Sasaki, SI.** 1992, Journal of Chemical Information and Computer Sciences, Vol. 32, p. 639.
- [99]. **Spriggs, RV.** Identification of β -sheet motifs in three-dimensional protein structures, using a subgraph isomorphism algorithm. *Electronic dissertations library.* [Online] 1992. [Cited: April 19, 2008.] <http://dis.shef.ac.uk/ruth/litreva.html>.
- [100]. **Klinger, S and Austin, J.** Chemical similarity searching using a neural graph matcher. *European Symposium on Artificial Neural Networks proceedings.* Bruges : ESANN, 2005, p. 479.
- [101]. **Bender, Andreas.** *Studies on molecular similarity.* University of Cambridge. 2005. Doktori értekezés.
- [102]. **Sokolsky, O, Kannan, S and Lee, I.** *Simulation-based graph similarity.* Department of Computer and Information Science, University of Pennsylvania. 2006.
- [103]. **Eckel, Bruce.** *Thinking in Java.* 3rd edition. New Jersey : Prentice Hall, 2003.
- [104]. **Budzikiewicz, H, Djerassi, C and Williams, DH.** *Mass spectrometry of organic compounds.* San Francisco : Holden-Day, Inc., 1967.

[105]. **McLafferty, FW and Tureček, F.** *Interpretation of mass spectra.* 4th edition.
Sausalito, CA : University Science Books, 1993.

Közlemények

Az értekezés alapját képező közlemények

- *SORI excitation: collisional and radiative processes.* Cs. **Peltz**, L. Drahos, K. Vékey. Journal of American Society of Mass Spectrometry, 2007, 18, 2119-2126
- *Accuracy of enthalpy and entropy determination using the kinetic method: are we approaching a consensus?* L. Drahos, Cs. **Peltz**, K. Vékey. Journal of Mass Spectrometry, 2004, 39(9): 1016-24

Az értekezés témaköréhez nem kapcsolódó közlemények

- *Microstructure of water at the level of three-particle correlation functions as predicted by classical intermolecular models.* Cs. **Peltz**, A. Baranyai, A.A. Chialvo, P.T. Cummings. Molecular Simulations, 2003, 29, 13-21
- *HPLC analysis of metabolically produced formaldehyde.* J. Lengyel, H. Kalász, T. Szarvas, Cs. **Peltz**, A. Szarkáné-Bolehovszky. Journal of Chromatographic Science, 2003, 41, 177-181

Az értekezés témaköréhez kapcsolódó előadások, poszterek

- *Internal energy effects in the course of SORI-CID experiment.* Cs. **Peltz**, L. Drahos, X. Guo, M.C. Duursma, A. Al-Khalili, R.M.A. Heeren, K. Vékey. 16th International Mass Spectrometry Conference, Edingburgh, UK, 2003 (poszter)
- *Kinetic Method: Enthalpy and Entropy and Accuracy.* L. Drahos, Cs. **Peltz**, K. Vékey. 22nd Informal Meeting in Mass Spectrometry, Tokaj, Hungary, 2004
- *Asymmetric cleavage of macromolecules.* Cs. **Peltz**, K. Vékey, L. Drahos. 23rd Informal Meeting in Mass Spectrometry, Fiera di Primiero, Italy, 2005 (előadás)

- *Automated interpretation of mass spectra by incremental learning or probabilistic fragmentation rules.* **Cs. Peltz**, I. Kövesdi, K. Újszászy, L. Drahos, J. Dombi. 17th International Mass Spectrometry Conference, Prague, Czech Republic, 2006 (poszter)
- *A physico-chemical modeling tool and its application for fragmentation rule induction.* **Cs. Peltz**. 25th Informal Meeting in Mass Spectrometry, Nyíregyháza, Hungary, 2007 (poszter)
- *Investigation of the effects contributing to response factors in gas chromatography–electron impact–mass spectrometry.* K. Rádi, **Cs. Peltz**, K. Újszászy, D. Frigyes. 25th Informal Meeting in Mass Spectrometry, Nyíregyháza, Hungary, 2007 (poszter)

Az értekezés témaköréhez nem kapcsolódó előadások, poszterek

- *Combination of deuteration techniques in the synthesis of an LCMS internal standard.* G. Németh, **Cs. Peltz**, R. Kapiller-Dezsőfi. 14th European Symposium on Organic Chemistry, Helsinki, Finland, 2005 (poszter)
- *Tömegspektrometria a szintetikus kémia támogatásában.* **Cs. Peltz**. Waters Tömegspektrometria Szeminárium, Budapest, 2006 (előadás)