Matematikai módszerek a tömegspektrometriában, a modellezés különböző szintjei

Doktori értekezés tézisei

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

1. Bevezetés, célkitűzések

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) betűszavak analógiájára egy CAMS (Computer Aided Mass Spectrometry – számítógéppel segített tömegspektrometria) rövidítést. Érdemes megjegyezni ugyanakkor, hogy nem-CAMS, tehát számítástechnikát mellőző tömegspektrometria ma már lényegében nem létezik.

A tapasztalt tendencia több kérdést is felvet. Vajon a különböző tömegspektrometriás 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.

Dolgozatomban a tömegspektrometria számítógéppel segített elméleti modellezésének három különböző szintjét mutatom be, a legpontosabbtól 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, termodinamikában használt 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.

Az általánosan elerjedt 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 méréstechnikák 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-kémiai folyamatok 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 méréstechnika, 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. Célom az emberi spektrumértékelési folyamat algoritmikus modellezése volt.

2. A vizsgált modellezési módszerek

SORI-CID folyamatok modellezése

Dolgozatom első részében a Fourier transzformációs tömegspektrometriában (FT-MS) gyakran használt hosszantartó nem rezonáns gerjesztésű ütközéses aktiválást (sustained off-resonance irradiation collision induced dissociation, SORI-CID) modelleztem. A SORI

kísérletek során az ion-ciklotron (ICR) cellában a körpályán keringő ionokat a keringési frekvenciától eltérő rádiófrekvenciás gerjesztéssel periodikusan gyorsítják, illetve lassítják, amelyek így az ütközőgáz (jelen esetben argon) részecskéivel való ütközések során magasabb belső energiára tesznek szert.

A SORI kísérletek modellezése elvileg lehetséges, azonban az egy kísérlethez tartozó nagyszámú gerjesztési ciklus követése a ma elérhető számítási kapacitások esetén is több hónapot igényel. Mivel a SORI-CID folyamat során az ütközések véletlenszerűen következnek be, a számítások egyszerűsítése érdekében a kinetikus energia időfüggése helyett a valószínűségi sűrűségfüggvényét használtam. Ez a közelítés analóg a termodinamikában használatos ergodicitási feltétellel.

Vizsgálataim során egy öttagú peptid – leucin-enkefalin – fragmentációját modelleztem argon ütköző gáz jelenlétében. A Guo és társai által[1] a hőmérséklet, kinetikus energia és idő függvényében felvett kísérleti adatokat (a bomlatlan molekulaionok arányát) modelleztem a MassKinetics[2] program segítségével. A 122 pontból álló kísérleti adatkészletre a szimulált fagyasztás módszere és Powell többdimenziós optimalizáló algoritmusa segítségével illesztettem a szimulációk eredményét, az ütközési energia-átadás és az infravörös hűlés optimális értékeit keresve.

A kinetikus módszer alkalmazhatóságának és hibájának vizsgálata

A dolgozat második részében egy termodinamikából átvett összefüggés nem-termodinamikai rendszerben – a tömegspektrométerben – való alkalmazhatóságát tanulmányoztam. A vizsgált kinetikus módszer a tömegspektrométerben mért intenzitás adatokból a reakciók termokémiai paramétereit becsli.

A kinetikus módszer entalpia- és entrópia-becsléseinek hibáját vizsgáltam a MassKinetics program segítségével generált „virtuális kísérleti adatokból". Az általam kidolgozott szoftver segítségével minden hibavizsgálathoz 1000-3000 virtuális kísérletet generáltam és értékeltem

[1] *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.

[2] **Drahos, L and Vékey, K.** *MassKinetics computer program.* [Online] 2004. [Cited: April 19, 2008.] http://www.chemres.hu/ms/masskinetics

ki. A kiértékeléskor az irodalomban általánosan alkalmazott lineáris regressziós módszer mellett az ortogonális távolság regressziót[3] is alkalmaztam.

A módszerben található közelítésekből származó hibák mellett a kísérleti adatok szórásából és a referencia-anyagok termokémiai paramétereinek pontatlan ismeretéből eredő hibát is tanulmányoztam.

Valószínűségi fragmentációs szabályok gépi tanulása

Az emberi spektrumfejtés módszereinek modellezésével egy valószínűségi fragmentációs szabályokat tanuló algoritmust fejlesztettem ki, amely a molekula szerkezeti gráfjának vizsgálatával a hozzá tartozó spektrumból indukálja a megfelelő fragmentációs szabályokat.

A fragmentációs szabályok egy szerkezet-párból és egy megjelenési intenzitás adat-párból állnak. A molekula adott kötésének hasítása után a kapott két fragmensnek tekintjük az elhasított kötés(ek) körüli környezetét. Az általam kidolgozott algoritmus ebből határozza meg a két részszerkezetet, a kapcsolódó tömegspektrumból pedig a megjelenési intenzitást.

A szabályok és az aktuális fragmensek összehasonlítására egy heurisztikus, valószínűségi jellegű módszert használtam. A szerkezeti gráfok egymásra való illesztése számításigényes feladat, azonban ezzel a módszerrel kézben tartható a futási idő. Kis szerkezetek esetén az alkalmazott paraméterekkel visszakapjuk a pontos illeszkedést, nagyobb szerkezeti gráfoknál pedig a méret növekedésével egyre pontatlanabbul közelíti azt.

3. Új tudományos eredmények, következtetések

1. Módszert dolgoztam ki SORI-CID kísérletek modellezésére. A periodikusan változó kinetikus energia – idő függvényből kiindulva meghatároztam a kinetikus energia eloszlásának sűrűségfüggvényét, majd ebből a laboratóriumi koordinátarendszerben megadott ütközési energia eloszlását. Az adott ICR cellageometria esetén ez a következő alakban volt megadható:

[3] **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.

$$P(E) = \frac{\sqrt{E}}{N}\frac{1}{\pi}\frac{E_{\max}^{\mathrm{kin}}}{\sqrt{(E_{\max}^{\mathrm{kin}} - E)E}},$$

ahol $E_{\max}^{\mathrm{kin}}$ a periodikusan változó kinetikus maximális értékét jelöli, N pedig egy alkalmasan választott normáló faktor. A levezetések eredményeként az ütközési energia átlagos értéke a maximális kinetikus energia 2/3-ának adódott.

Ez azt jelenti, hogy – meghatározott feltételek teljesülése esetén – a SORI kísérlet eredménye megfeleltethető annak a kísérletnek, ahol minden ion ütközési energiája egységesen a maximális kinetikus energia 2/3-a. A fenti átlagos ütközési energia használatával a gyakorlatban is lehetőség nyílt a SORI-CID gerjesztés szimulációjára.

2. Hőmérséklet, kinetikus energia és idő függvényében felvett kísérleti adatkészletből az infravörös hűlés és az energia-átadás hatékonyságának mértékét határoztam meg. Az 1. pontban bemutatott eredmények felhasználásával a MassKinetics program algoritmusával a leucin-enkefalin – argon rendszerben kiszámítottam a bomlatlan molekulaionok arányát, majd a kapott eredményeket illesztettem a kísérleti adatkészletre. Az illesztést két paraméter, az ütközési energia-átadás hatékonysága ($\eta_{\mathrm{single,com}}$) és az infravörös hűlés sebessége (k_{cool}) függvényében végeztem.

A meghatározott optimális paraméterek a következők: $\eta_{\mathrm{single,com}} = 12{,}8\ \% \pm 2{,}1\ \%$ és $k_{\mathrm{cool}} = 7{,}5 \pm 0{,}5\ \mathrm{s}^{-1}$. Tanulmányoztam az ütközési energia-átadás függvényalakjának hatását a szimulált adatokra, illetve az infravörös hűlés figyelembevételének szükségességét is. A számítások azt igazolták, hogy az energia-átadás átlagos értéke és függvényalakja önmagában nem elégséges a kísérleti adatok megfelelő reprodukálásához, az infravörös hűlés nem hanyagolható el.

3. A két vizsgált paraméter – az energia-átadás hatékonysága és az infravörös hűlés sebessége – vizsgálatakor megállapítottam, hogy a két folyamat egymástól nem választható el, a paraméterek korrelálnak. A számítások azt mutatják, hogy $\ln(k_{cool})$ és η függvényében meghatározva az illesztés jóságát, az ábrázolt felületen egy hosszú „völgy" látható. A paraméterek számos kombinációja elfogadható illeszkedést ad a számított és a mért adatok között, vagyis k_{cool} alacsonyabb értékeit η alacsonyabb értékei kompenzálják. Erre utal az is, hogy az IR hűlés figyelembevétele nélküli modellezés esetén a becsült átlagos energia-átadás mindössze 6,6 %-nak adódott.

4. Kidolgoztam egy automatikus kiértékelő algoritmust a kinetikus módszer hibáinak vizsgálatára. A kifejlesztett program alkalmas „virtuális kísérleti adatok" több ezres nagyságrendben való generálására, majd a kinetikus módszer szerinti kiértékelésére és a kapott eredmények vizsgálatára.

A virtuális kísérlet előnye, hogy olyan kísérleti körülmények is tanulmányozhatók, amelynek tényleges megvalósítása nehéz vagy lehetetlen lenne.

5. Vizsgáltam, hogy a kinetikus módszer segítségével számított entalpia- és entrópia-paraméterek hibája milyen mértékben függ a kezdeti paraméterek bizonytalanságától, a módszer eredendő hibájától és a kiértékelés módjától. A számítások eredményeképpen az alábbi táblázatban látható szórások adódtak.

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 ($Jmol^{-1}K^{-1}$, szórás)	$\Delta G_{app,298}$ 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**

6. Meghatároztam a módszer alkalmazhatóságának korlátait. A kiértékelések során bebizonyítottam, hogy a kinetikus módszer – meghatározott feltételek mellett – alkalmas termokémiai paraméterek meghatározására. A módszer segítségével az entalpia- és entrópia-értékek a referencia-anyagok megfelelően kicsi entrópia-különbségei esetén (35 J mol^{-1} K^{-1} alatt) a fenti táblázatban megadott hibákkal számíthatók. Az ennél nagyobb entrópia-különbségeknél a módszer szisztematikusan alulbecsli az entalpia- és az entrópia-változás értékét. Amennyiben szükséges, az alulbecsült tartományban használható egy empirikus skálázó faktor, de feltétlenül szem előtt kell tartani azt, hogy az ilyen korrekciós megoldások egy hamis biztonság látszatát keltik. Az alulbecsült tartományon a kinetikus módszer alapvetően hibás eredményt ad, mivel a módszer alkalmazásának az alapfeltételei nem teljesülnek.

Kimutattam, hogy az entalpia és entrópia tagok korrelációja miatt a látszólagos gázfázisú bázicitás az előzőeknél pontosabban meghatározható, akár extrém nagy entrópia-különbségek esetén is. A kiértékelés módja, a lineáris regresszió és az *x* adatok hibáját is figyelembe vevő ortogonális távolság regresszió nem vezetett eltérő számszerű eredményre.

7. Kifejlesztettem egy valószínűségi fragmentációs szabályokat tanuló algoritmust, az emberi spektrumfejtési folyamatok modellezésével. Az algoritmus fizikai-kémiai előismeretet nem tartalmaz a szabályok felállításához, csak gráf-algoritmusok és tömegspektrum-feldolgozási módszerek alkalmazásával történik a szabályok indukciója.

Fontos hangsúlyozni, hogy az általam kidolgozott algoritmus a spektrumok intenzitás adatait is figyelembe veszi egy jel/zaj viszony jellegű mértéket használva. Ez az abszolút illetve relatív intenzitásértékkel ellentétben jól modellezi az ember spektrum-felismerési gondolkodásmódját: felismerhető egy fragmens, ha nincs környezetében intenzív háttér, ugyanakkor nem jellemző, ha egy zsúfolt csúcsköteg közepén szerepel, bármilyen intenzív legyen is.

Az elkészült algoritmust sikerrel alkalmaztam gyógyszermolekulák elektronütközéses spektrumából fragmentációs szabályok tanulására. A meghatározott szabályok jellemzőek a vizsgált vegyületcsaládra, megfelelnek a tömegspektrometriás irodalom alapján várható fragmentációs utaknak. Az elkészített programmal az Egis Gyógyszergyár Nyrt. Szerkezetkutatási osztályán mért gyógyszermolekulák elektronütközéses spektrumait dolgoztam fel. Az algoritmus tesztelése során 63 spektrum – szerkezet párból generáltam fragmentációs szabályokat.

A kidolgozott szabálytanuló algoritmus legfontosabb felhasználási területe a korábban nem vizsgált – azaz spektrumkönyvtárakban nem szereplő – vegyületcsaládok tömegspektrumának értékelése. Példaként említhetjük a gyógyszergyári kutató-laboratóriumokat, ahol új gyógyszermolekulák, származékaik és metabolitjaik szerkezetvizsgálata során nyújthat segítséget.

Eredményeim összefoglalásaképpen elmondható, hogy a tömegspektrometriával kapcsolatos különböző feladatok eltérő modellezési módszereket és matematikai eszközöket igényelnek. Minden kérdésfelvetéshez az adott feladatot kielégítő pontossággal megoldó – nem feltétlenül bonyolult – számítási módszert kell választani.

4. Publikációs jegyzék

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 formaldehide.* J. Lengyel, H. Kalász, T. Szarvas, **Cs. Peltz**, A. Szarkáné-Bolehovszky. Journal of Chromatographic Science, 2003, 41, 177-181

Konferencia előadások és 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 (poszter)
- *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)
- *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)