

Doktori értekezés tézisei

Hoffmanné Szalay Zsófia

Dinamikus NMR spektrumok szimulációjának
elmélete és gyakorlata csatolt spinrendszerkben

Témavezető:

Rohonczy János, PhD

Eötvös Loránd Tudományegyetem Kémiai Doktori Iskola

Doktori iskola vezetője: Inzelt György

Elmélet és fizikai kémia, anyagszerkezetkutatás doktori program

Doktori program vezetője: Surján Péter



Eötvös Loránd Tudományegyetem

Természettudományi Kar

Budapest, 2010

1. Bevezetés

A gyakorlati NMR spektroszkópiában gyakran találkozunk dinamikus folyamatokkal. Ilyen például savas protonok cseréje vizes (vagy savas) oldatban vagy a klasszikus vonalkiszéledési és koaleszcencia jelenség. A vonalkiszéledés szimulációja lehetőséget ad a kérdéses cserefolyamat kvantitatív vizsgálatára is. Ezt a módszert az NMR spektroszkópia és a kémia számos területén alkalmazzák, így a komplexek stabilitásának és izomerizációjának vizsgálatára vagy fehérjék dinamikájának tanulmányozására.

A dinamikus NMR spektrumok szimulációja (kémiai eltolódások és csatolási állandók alapján) 10-15 magnál nagyobb rendszerekre már csak speciális esetekben (molekulaszimmetriát figyelembe vevő egyszerűsítésekkel) lehetséges. A spektrumszimuláció széles körben elterjedt módja az egykvantumkoherenciákat tartalmazó átlag sűrűségmátrix időbeli változásának szimulációja. Ezen módszer fő problémája a szimuláció memóriaigénye, ezért a spektrumszimuláló programok fejlesztésének fő iránya a sűrűségmátrix effektív méretének csökkentésére irányul (pl. triviálisan nulla elemek elhagyása, szimmetria kihasználása).

A dinamikus NMR jelenségnek ismert egy egyszerű értelmezése, amely azonban csak egyetlen spint különböző környezetekben tartalmazó rendszerek kvantitatív leírására alkalmas, bonyolultabb rendszerek esetében csak fenomenologikus képet ad. Eszerint a detektálható jelet a vektormodellből ismert eredő mágnesezettségvektor adja. Amennyiben a mag környezetet vált, ennek a vektornak megváltozik a precessziós frekvenciája, így a detektált jelben törés keletkezik. A frekvenciaváltozások gyakorisága hatással van a jel szélességére és alakjára: ritka csere esetén két éles jel van, ami a cseresebesség növelésével szélesedik. Amikor a cseresebesség meghaladja a két jel frekvenciakülönbségét a gyakori frekvenciaváltások miatt már csak a két frekvencia átlagát lehet detektálni: kezdetben egy széles, majd egyre keskenyebb jelet kapunk. Egy szimulációban az eredeti modellben a csereidőpontokat úgy határozzák meg, hogy a sebességi együttható és egy véletlen szám viszonya alapján adott időközönként eldöntik, hogy legyen-e csere vagy nem. Több száz ilyen módon generált jelet összeadva a kapott görbe, megfelelő paraméterválasztás esetén, jól közelíti a kísérletekben mért spektrumot. A véletlen számok szerepe miatt ezt a számolást a dinamikus NMR spektrumok Monte Carlo szimulációjának szokták nevezni. A dolgozat elméleti részében ezt a modellt fejlesztettem tovább.

A Monte Carlo szimulációk egyik fontos hátránya, hogy a pontos eredményhez sok, eltérő paraméterekkel számolt adatpontra van szükség, ami általában időigényes. A szimuláció

sebességének növelésének legelterjedtebb módja a program többszálásítása. Ez a programozástechnikai trükk, az ismétlődő és egymástól független számolások miatt Monte Carlo szimulációkra általában hatékony. A többszálú programokkal hagyományosan egy vagy több gép processzorain párhuzamosan futtathatók a program különböző részletei, így a számolás időigénye ezzel arányosan csökken.

A hagyományosnak paralellizálás mellett a legmodernebb párhuzamosítási technikát GPGPU videokártyák (*General Purpose computing on Graphics Processing Unit*) jelentik. A technológia lényege, hogy a párhuzamos programszálak (akár több ezer) különböző adatokon ugyanazokat a műveleteket nagyon hatékonyan tudják egyszerre elvégezni. Ezen tulajdonsága miatt ez a technológia kimondottan alkalmas mátrix- és vektorműveletek végzésére, valamint olyan szimulációs feladatok megoldására, amelyekben egy műveletsort egymástól függetlenül többféle paraméterlistával kell kiszámolni. Az irodalomban fellelhető példák azt mutatják, hogy az ilyen számítások a GPGPU technológiával akár 50-200-szor gyorsabbak lehetnek, mint hagyományos CPU-n.

2. Alkalmazott módszerek

A szimuláció matematikai formalizmusának kidolgozása során a Liouville-von Neumann egyenletet oldottuk meg különböző peremfeltételekkel és az irodalomból ismert összefüggések alapján vezettem le a szükséges képleteket. A szimulációban a kinetikus Monte Carlo módszert egy speciális esetre alkalmaztam. Ezen felül az elmélet értelmezéséhez az NMR irodalomban elterjedt vektormodell kereteit feszegettem.

A ProMoCS programot alapvetően Java nyelven írtam. A program párhuzamosítását a JDK Thread interface felhasználásával végeztem. A távoli gépeken való futtatáshoz a Java RMI (*Remote Method Invocation*) technológiáját vettem igénybe. Az NVidia® videokártyán futó változatot a C programnyelv a CUDA kiterjesztésének segítségével írtam. Az szimuláció ellenőrzéséhez az eredményeket MEXICO programmal számolt görbékhez hasonlítottam.

A számolásokat Microsoft® Windows™ XP ill. RedHat® Fedora™ 10 operációs rendszer alatt végeztük (Intel® Core2 Duo 2,20 GHz processzor, 1 Gb RAM). A CUDA™ szimulációkat NVidia® GeForce 8600 GT (4 multiprocesszor, 32 mag, 256 Mb RAM) és NVidia® GeForce GTS 250 (32 processzor, 128 mag, 1 Gb RAM) videokártyákon futtattuk.

3. Eredmények

3.1. Elméleti eredmények

1. A dinamikus jelenséget és a spinek kölcsönhatásait különválasztva értelmeztem a cserefolyamatokat. A két kölcsönhatás éles szétválasztásával a szimulációs tér dimenziója általában lényegesen csökken. Az értelmezéshez új definíciókra volt szükség (spinhalmaz, konformer értelmezése). A spinhalmaz olyan atommagok rendezett halmaza, amelyek egymáson kívül más magokkal nem csatolnak. Dinamikus jelenség esetén a spinhalmaz többféle kémiai környezetben fordul elő, ezeket nevezzük konformereknek. Az egy szimulációban szereplő spinhalmazok között a cserefolyamatok időpontjaiban és a pillanatnyi konformációkban (azaz a spinhalmazok életútjában) van a különbség. Ugyanakkor a spinhalmazok mérete, konformerei és a csereidőpontjaik átlagos eloszlása azonos.
2. A spinhalmazok viselkedésének leírására bevezettem az egyedi sűrűségmátrix fogalmát. Ez a mátrix azon állapotok átlagát tartalmazza, amelyek életútja azonos. Ezek az állapotok a kvantummechanika törvényei szerint megkülönböztethetetlenek, tehát az átlag makroszkopikusan egyedinek tekinthető, erre utal az elnevezés. Ezzel szemben a dinamikus NMR szimulációkban általában használt átlag sűrűségmátrix elméletileg megkülönböztethető állapotokat átlagol, mivel a cserefolyamatok átlagos eredményét is tartalmazza, ami elvileg klasszikus módszerekkel leírható.
3. A pseudoelsőrendű reakciók elmélete és a termikus egyensúly feltétele alapján levezettem azon összefüggéseket, amelyek alapján a spinhalmazok életútja meghatározható. Ezen összefüggések szerint tisztán a sebességi együtthatók alapján meghatározható az egyes konformerek előfordulási valószínűsége, a csereidőpontok közötti időintervallumok (időszektelek) hosszának eloszlása valamint az egyes reakciók lehetséges kimeneteleinek valószínűsége.
4. A dinamikus cserefolyamat által befolyásolt jel (*fid*) szimulációjának legfontosabb lépése annak meghatározása, hogy a szabad precesszió és a cserefolyamatok során hogyan változik az egyedi sűrűségmátrix. A dolgozatban megadtam ezen két folyamat valamint a detektált jel számolásának matematikai leírását. A képletek kulcsa, hogy minden esetben egy időszekteleken belüli változásokat írnak le, és a cserefolyamatok nem a sűrűségmátrixot, hanem a propagáló operátorokat érintik. Az eredményül kapott

képletek felhasználásával egy egyszerű példán bemutatom, hogy a magas hőmérsékletű spektrumban a jelek kiátlagolódása az új elméletnek is egyenes következménye.

5. A *fid* képlete alapján többféle szimulációs algoritmusra van lehetőség, ezek közül a leggyorsabb a jel diszkrét pontjainak kiszámolása, majd az ezt követő diszkrét Fourier-transzformáció. Ezen a megoldáson kívül lehetőség van a *fid* analitikus Fourier-transzformációjára, ami egy elméleti szempontból érdekes megközelítést tesz lehetővé. A transzformációban az integrálást időszeletek szerint felbontva meghatározható egy-egy időszelvény spektruma. Az időszeletek hosszának valószínűségi eloszlásának ismeretében pedig meghatározható a teljes spektrum várható értéke, ami elméleti megegyezik a végtelen számú spinhalmaz spektrumának összegével. Ezen képlet segítségével az NMR spektrumok jelalakváltozása is értelmezhető. Ezt kvalitatíven egy egyszerű példán mutatom be.
6. A 4. pontban említett képleteket az egyszerűbb kezelhetőség miatt az egykvantumátmenetek terében vezettem le. A cserefolyamatok és a spinek kölcsönhatásainak elkülönítése miatt a képletek az állapotfüggvények terében, azaz a Hilbert-térben is levezethetők a Liouville-von Neumann-egyenlet egy másik megoldását használva. A kétféle megoldás ekvivalenciáját a mátrixműveletek részletes kifejtésével be is bizonyítottam. További számolástechnikai egyszerűsítést tesz lehetővé, hogy a spinek Hamilton-operátora totálspinkvantszám szerinti blokkdiagonális, így az egyedi sűrűségmátrix is kisebb blokkokban számolható. Így a szimulációs tér dimenziója és a program memóriagénye lényegesen lecsökken. Emellett a detektált jel számolását gyorsítja, ha az egyedi sűrűségmátrixot minden időszelvényben a pillanatnyi konformer sajátfüggvényeinek bázisára transzformáljuk, mivel ezen a bázison a szabad precesszió operátora diagonális.
7. A dolgozatban az NMR spektroszkópia elméletében használt vektormodellt értelmeztem erősen csatolt spinrendszerekre is. Eszerint a modellben nem egy makroszkopikus mágnesezettségvektor van, hanem minden bázispárra jut egy vektor és ezek lineáris kombinációja adja a detektált jelet. Ezek a vektorok nem vonhatók össze, mivel az egyes események (cserefolyamat, precesszió) hatása egyedileg befolyásolja őket. A csatolt spinrendszerekben két ilyen vektorkészlet van, egy a bázis-, egy pedig a sajátfüggvényeket illetve, -koherenciákat reprezentálja. A két vektorkészlet közötti átváltást lehetővé teszi, hogy a sajátfüggvények kifejezhetők a bázisfüggvények lineáris kombinációjaként és fordítva. Dinamikus folyamatok értelmezéséhez mindkét

vektorkészletre szükség van, mivel a cserefolyamatok során a bázisfüggvények, míg a precesszió során a sajátfüggvények viselkedését tudjuk nyomon követni a kvantummechanika szabályai szerint. Ezen modell segítségével az egyes lépésekre kapott képletek jól értelmezhetők.

3.2. Gyakorlati eredmények

8. Az elméleti eredmények és a matematikai képletek alapján Java nyelvű szimulációs programot írtam dinamikus NMR spektrumok számolására (ProMoCS: *Propagators & Monte Carlo for DNMR Spectrum Simulation*). Ez a program párhuzamosan több szálon, egy vagy több gép processzorain számolja ki az eredményt. A program algoritmusát a dolgozat részletesen ismerteti.
9. A ProMoCS program fő algoritmusát C nyelven is megírtam (ProMoCS-C), ez a Java nyelvű programból meghívható. Ennek elsődleges szerepe, hogy hidat biztosít a CUDA nyelvű algoritmusok felé. A ProMoCS programozható videokártyán futtatható változatának (ProMoCS-CUDA) megírásánál a speciális architektúra miatt külön változat készült a 16×16-nál nagyobb és kisebb sűrűségmátrixok számolására. Előbbi esetben a CPU-ra írt algoritmust a speciális környezetben való alkalmazás mellett csak néhány ponton kellett megváltoztatni, míg a kis mátrixok hatékony számolásához jelentős változtatásokat kellett végezni.
10. A ProMoCS programcsalád teljesítőképességét és az egyes paraméterek futási időre gyakorolt hatását hipotetikus csatolt spinrendszeren teszteltem (1–8 mag 2–4 konformerrel). Az eredményeket egyrészt önmagukkal, másrészt az átlagsűrűségmátrix szimulációján alapuló, széles körben elterjedt a MEXICO-val vettem össze. A szimulált spektrumok helyesnek bizonyultak. A futási idő a várakozásoknak megfelelően függ a magok számától és az átlagos cseresebességtől, nem függ viszont a konformerek számától. A CPU-hoz képest a GPU-n futtatott tesztek 5–10 illetve 10–100-szoros gyorsulást mutattak a két videokártyán.
11. A ProMoCS programmal néhány valós rendszer hőmérsékletfüggő NMR spektrumainak szimulációját is elvégeztem (N,N-diizopropil-karbaminsav-trimetilszililészter, a trimetilszililciklopenta-[l]-fenantrén, az N,N-dimetil-para-nitrozo-anilin és egy aranykomplex). A példák mindegyikét leírták az irodalomban, a szimuláció célja az eredmények reprodukálása volt. Összesen öt spektrumsorozatot szimuláltam (négy ^1H és egy ^{31}P NMR), a spektrumok minden esetben jól egyeztek a referenciával.

A dolgozat alapját képező publikációk

Cikkek

- [1] Zs. Szalay, J. Rohonczy: Monte Carlo Simulation of DNMR spectra of coupled spin systems. *J. Magn. Reson.* **191** (2008) 56–65.
- [2] Zs. Szalay, J. Rohonczy: Simulation of DNMR spectra using propagator formalism and Monte Carlo method. *J. Magn. Reson.* **197** (2009) 48–55.
- [3] Zs. Szalay, J. Rohonczy: Monte Carlo simulation of NMR lineshapes in chemically exchanging spin systems. *Progr. Nucl. Magn. Reson. Spect.* **56** (2010) 198–216.
- [4] Zs. Szalay, J. Rohonczy: Fast Calculation of DNMR Spectra on CUDA Enabled Video Card. *J. Comp. Chem.* (beküldve).

Poszterek

- [5] Zs. Szalay, J. Rohonczy: Multithreaded Simulation Of Dnmr Spectra On Cuda Enabled GPGPU Video Cards. Euromar 2009 Conference (Göteborg, Svédország, 2009. július)
- [6] Zs. Szalay, J. Rohonczy: Monte Carlo simulation of DNMR spectra of mutually exchanging coupled spin systems (poszter). Euromar 2008 Conference (Szentpétervár, Oroszország, 2008. július)
- [7] Zs. Szalay, J. Rohonczy: Monte Carlo simulation of DNMR spectra of coupled spin systems (poszter). Euromar 2007 Conference (Tarragona, Spanyolország, 2007. július)

Előadások

- [8] J. Rohonczy, Zs. Szalay: New Aspects and Tools in the DNMR Spectrum Analysis, MTA NMR Munkabizottság ülése (MTA KKK, Budapest, 2007. január)

Diákköri dolgozatok

- [9] Szalay Zsófia: Folyamatszimulációs módszer kidolgozása összetett spinrendszerek dinamikus NMR spektrumának számolására, XXVIII .OTDK Kémiai és Vegyipari Szekció (Szeged, 2007. április) – Különdíj
- [10] Szalay Zsófia: Folyamatszimulációs módszer kidolgozása összetett spinrendszerek dinamikus NMR spektrumának számolására, ELTE Kémiai Intézet Házi TDK Konferencia (Budapest, 2006 november) – I. díj