

7. Molekulide kogumite modelleerimine

Käsitlus põhineb raamatu “C.J.Cramer, Essentials of Computational Chemistry” 3. peatükil.

Energiamiinimumid potentsiaalse energia pinnal (PES) on arvutuskeemias mõnes mõttes ületähtsustatud. Isegi 0 K juures esineb molekulide nullvõnkumine, ning kõrgemal temperatuuril võivad lisanduda konformatsioonilised muutused (pöörlemine sidemete ümber) ja muud moonutused kuni keemiliste reaktsioonideni välja.

Suurte, paindlike molekulide korral ei ole isegi korrektne rääkida molekuli kujust või struktuurist, sest eri molekulid ongi eri konformatsioonides, kuigi keemikuna käsitleme neid ühe “ainena”. Mõnede biomolekulide korral on välistest mõjuritest tingitud struktuurimuutused nende funktsiooni aluseks.

Termodünaamikast tuleneb, et piisavalt suure molekulide hulga korral moodustub eri konformeeride *statistiline jaotus*, mis sõltub temperatuurist ja süsteemi iseloomustavast termodünaamilisest suurusest (näit. Helmholtzi vabaenergia). Viimane omakorda sõltub sellest, milliseid katsetingimusi hoitakse konstantsena (nt. temperatuur ja rõhk). Need tingimused defineerivad konkreetse “*ansambli*”.

Näiteks kui süsteemi uuritakse konstantse rõhu, ruumala ja temperatuuri juures, räägitakse *PVT* ansamblast.

Väikeste molekulaarsete mudelite korral on P , V ja T üheaegne kindlal väärtusel hoidmine raskendatud, sest vastavalt ideaalgaasi olekuvõrrandile $PV = (N/N_A)RT$ tuleks süsteemi energia muutumise korral muuta osakeste arvu N , mis ei ole mugav. Seetõttu kasutatakse modelleerimisel eelistatult ansambleid, kus N on konstantne, näit. NVT või NPT ansablit.

Enamus tavakeemia katseid mõõdavad molekulaarseid omadusi *keskmistatuna*, tüüpiliselt nii ajas kui üle ansambli(te). Selliste katsete arvutuslikuks kirjeldamiseks kasutatakse peamiselt Monte Carlo (MC) ja molekulaardünaamika (MD) meetodeid.

Faasiruum

Klassikalises mehaanikas on süsteem täielikult kirjeldatud, kui on teada iga osakese asukoht ja impulss. N osakesest koosneva süsteemi kirjeldamiseks on tarvis $6N$ koordinaati. $6N$ -mõõtmelist ruumi, mis selliselt süsteemi olekuid kirjeldab, nimetatakse *faasiruumiks*.

Igal ajahetkel asub süsteem kindlas faasiruumi punktis:

$$\mathbf{X}' = (x_1, y_1, z_1, p_{x,1}, p_{y,1}, p_{z,1}, x_2, y_2, z_2, p_{x,2}, p_{y,2}, p_{z,2}, \dots) \quad (7.1)$$

Koondame asukoha- ja impulsskoordinaadid vektoriteks

$$\mathbf{q} = (x_1, y_1, z_1, x_2, y_2, z_2, \dots) \quad (7.2)$$

$$\mathbf{p} = (p_{x,1}, p_{y,1}, p_{z,1}, p_{x,2}, p_{y,2}, p_{z,2}, \dots) \quad (7.3)$$

Siis saame punkti faasiruumis üles märkida kujul

$$\mathbf{X} = (\mathbf{q}, \mathbf{p}) \quad (7.4)$$

Keskväärtused faasiruumis

Faasiruum hõlmab kõik süsteemi olekud, nii võimalikud kui võimatud. Suuruse A keskvärtus (“oodatav väärtus” , meenutame kvantkeemia kursust) süsteemis, mis on konstantsel temperatuuril, ruumalal ja kindla osakeste arvuga, väljendub integraalina üle faasiruumi:

$$\langle A \rangle = \int A(\mathbf{q}, \mathbf{p}) P(\mathbf{q}, \mathbf{p}) d\mathbf{q} d\mathbf{p} \quad (7.5)$$

kus P on tõenäosus, et süsteem asub antud faasiruumi punktis.

Statistilises mehaanikas näidatakse, et see tõenäosus on seotud süsteemi energiaga faasiruumi antud punktis:

$$P(\mathbf{q}, \mathbf{p}) = Q^{-1} e^{-E(\mathbf{q}, \mathbf{p})/k_B T} \quad (7.6)$$

kus E on süsteemi kineetilise ja potentsiaalse energia summa, need sõltuvad vastavalt $\mathbf{q}$ -st ja $\mathbf{p}$ -st. k_B on Boltzmanni konstant $k_B = R/N_A$, T on temperatuur ja Q on *statistiline summa* (partition function):

$$Q = \int e^{-E(\mathbf{q}, \mathbf{p})/k_B T} d\mathbf{q} d\mathbf{p} \quad (7.7)$$

mida antud kontekstis võib käsitleda kui normaliseerimiskordajat.

Integraali (7.5) on praktiliselt võimalik leida ainult numbriliselt. Võimalik lähenduskäik hõlmaks integreeritava suuruse arvutamist reas faasiruumi punktides, selle sobitamist mõne sujuva integreeruva lähendusfunktsiooniga ja viimase integreerimist. Punktide arvu võiks suurendada kuni saadav vastus koondub soovitava täpsuseni.

Praktikas see ei ole teostatav, sest faasiruum on suure mõõtmete arvu tõttu hiigelsuur. Näiteks kui võtta igal teljel ainult kaks (!) punkti, on punktide koguarv 2^{6N} , mis 100 osakese korral tähendaks $4,16 \cdot 10^{180}$ punkti!

Teisalt näitab võrrand (7.6), et suurema energiaga punktide kaal on väga madal. Integraali (7.5) arvutamiseks vajame algoritmi, mis leiaks üksnes *olulisi* (suure kaaluga) punkte faasiruumis.

Ergoodilisuse hüpotees

Kujutleme algoritmi, mis alustab mõnest suure tõenäosusega P (madala energiaga) algpunktist faasiruumis ja genereerib *trajektoori*, näiteks Newtoni liikumisseadusi järgides, kirjeldades süsteemi arengut ajas. Siis võime perioodilise intervalli järel ajahetkedel t_i leida suuruse A väärtuse ja arvutada keskväärtuse kui

$$\langle A \rangle = \frac{1}{M} \sum_i^M A(t_i) \quad (7.8)$$

Kui ajaintervall muuta asümptootiliselt väiksemaks, saame summalt üle minna integraalile

$$\langle A \rangle = \lim_{t \rightarrow \infty} \frac{1}{t} \int_{t_0}^{t_0+t} A(\tau) d\tau \quad (7.9)$$

Ergoodilisuse hüpotees väidab, et võrrand (7.9) kehtib mistahes algushetke t_0 korral. Ergoodilisuse hüpoteesi ei ole suudetud tõestada (seega pole tegemist teoreemiga), v.a. mõnedel erijuhtudel. Siiski näitab mahukas kogemus, et see kehtib “mõistlikult” läbi viidud molekulaarsete modelleerimiste korral.

Molekulaardünaamika

Deterministlikus (Newtoni) maailmapildis määrab iga faasiruumi punkt üheselt kogu süsteemi tulevase trajektoori, st osakeste asukohad ja kiiruse tulevastel ajahetkedel. Lisaks, kuna aeg on matemaatiliselt pööratav, saab ühest punktist välja arvutada ka kogu süsteemi mineviku.

Paneme tähele, et trajektoorid faasiruumis ei lõiku kunagi, sest sel juhul eksisteeriks punkt, millel oleks kaks erinevat minevikku ja kaks erinevat tulevikku. Küll aga on võimalikud kinnised trajektoorid, kus süsteem jõuab tagasi punkti, milles ta on varem olnud, ja kordab varasemat käitumist.

Loengus vaatleme harmoonilist ostsillaatorit kui kinnise trajektooriga süsteemi näidet.

Enamuse reaalsete süsteemide jaoks me ei suuda leida analüütilist lahendit trajektoori arvutamiseks. Küll aga saame kasutada Newtoni mehaanikast ($p = mv$; $F = ma$) tulenevat, nn. Euleri lähendust:

$$q(t + \Delta t) = q(t) + \frac{p(t)}{m} \Delta t \quad (7.10)$$

$$p(t + \Delta t) = p(t) + ma(t) \Delta t \quad (7.11)$$

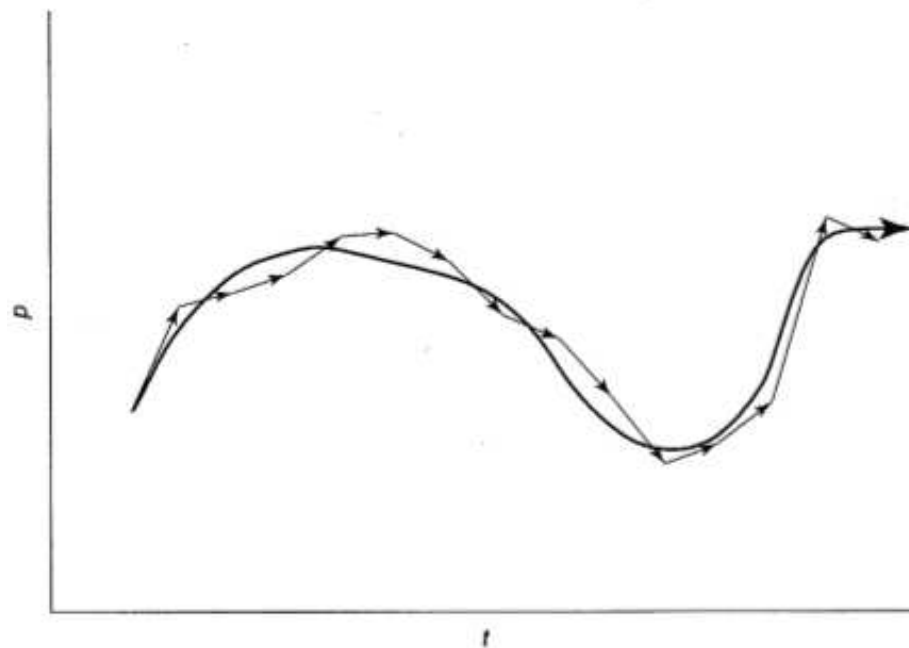
Praktikas kasutatav algoritm lähtub mõnest “mõistlikust” osakeste lähteasendist (impulsid omistatakse sageli juhuslikult) ja kasutab molekulaarmehaanikat, et arvutada osakestele mõjuvad jõud ja nende kaudu kiirendused, ja lähendada süsteemi trajektoori faasiruumis.

Impulsside väärtustest on võimalik arvutada temperatuur:

$$T(t) = \frac{1}{(3N - n)k_B} \sum_{i=1}^N \frac{|\mathbf{p}_i(t)|^2}{m_i} \quad (7.12)$$

kus n on liikumispiiranguga vabadusastmete arv (nt. keemilised sidemed).

Praktikas põhjustab probleeme lõpliku ajasammu Δt kasutamise vajadus. Joonis illustreerib tegeliku (sujuv kõver) ja lähendatud (sirglõigud) trajektoori erinevust.



Joonis 3.2 raamatust "Essentials of Comp. Chem."

Reaalse molekulaarse süsteemi korral tähendab liiga suur kõrvalekaldumine, et aatomid satuvad üksteisele liiga lähedale, põhjustades suuri jõude, mis omakorda paiskavad nad teistele aatomitele liiga lähedale, jne. Süsteem jääb liikuma kõrge energia ja järelkult madala tõenäosusega olekutesse.

Praktikas on kiireim oluline protsess vesiniku aatomite võnkumine sidemete ot-sas, mis paneb paika realistlikult pikima ajasammu 0,1 fs. Suurte süsteemide (biopolümeerid koos ümbritseva lahustiga) mudeleid jooksutatakse maksimaalselt 10^7 sammu, mis teeb 10 ns. See jääb suurusjärke alla mitmete oluliste protsesside (näit. valkude pakkimine) toimumisajale (mitu millisekundit).

Süsteemi arengu modelleerimise praktilised küljed

Eksisteerib Euleri lähendusest tunduvalt paremaid arengu modelleerimise (ingl. propagation) algoritme.

Nii näiteks *Verlet' algoritm* (1967) kasutab jooksvaid ja eelmise punkti ruumikoordinaate ning kiirendusi, arvutamaks uusi asukohti. Kiirusi (ja seega impulsse) ei arvutata. Seda algoritmi kasutatakse, kui pole vaja leida impulssidest sõltuvaid suurusid, näiteks temperatuuri.

Nn *konnahüppe meetod* (leapfrog method) arvutab impulsid Verlet' algoritmi punktide vahe-hetkedel, $0,5\Delta t$ kaugusel asukohtade omadest. See on efektiivsem kui nende arvutamine samade ajahetkede jaoks. Puuduseks on, et kõik koordinaadid faasiruumis ei ole ühegi hetke jaoks korraga teada, samuti on kasutatavad valemid jämedamad lähendused tegelikule trajektoorige kui Verlet' omad.

On olemas ka palju täpsemaid, näit. Gear'i prediktor-korrektor meetod (1971), kuid need on ka arvutusmahukamad ja selle tõttu harva kasutusel.

Olulist kokkuhoidu arvutusajas annab, kui osad vabadusastmed külmutada. Nii näiteks fikseeritakse sageli solvendimolekulide (vee) O–H sidemepikkused ja -nurgad, samuti vesinike sidemed ja nurgad uuritavas biomolekulis. See võimaldab vähendada faasiruumi koordinaatide arvu ning kasutada pikemaid ajasamme, sest nüüd pole vaja modelleerida kõige kiiremaid, vesiniku aatomite, võnkumisi.

Osade vabadusastmete fikseerimise algoritmid töötati välja 1970-ndate aastate teises pooles. Üks levinumaid selliseid kannab nime SHAKE (detailid on leitavad õpikutest ja internetist).

Arvutusaega vähendab ka Newtoni kolmanda seaduse ärakasutamine: kuna iga jõu komponent on põhjustatud mõne teise osakese poolt, siis saame ühte arvutatud komponenti kasutada mõlemale osakesele mõjuva kogujõu arvutamisel, mis vähendab töö mahtu ligi kahekordselt.

Stohhastiline dünaamika

Kui arvutuse peamiseks eesmärgiks on süsteemi käitumise kirjeldamine ajas ning termodünaamilised parameetrid on vähe olulised, võib solvendi molekulide asemel kasutada *kontiinumi mudelit*, kus solventi vaadeldakse ajas keskmistatud homogeense ollusena.

Sellist lahustit iseloomustatakse tema elektrostaatiliste ja hõõrdumisomaduste kaudu. Kaks levinumat matemaatilist lähenemist kannavad vastavalt Langevini ja Browni dünaamikate nimetusi.

Monte Carlo meetod

Monte Carlo meetodi puhul ei kasutata süsteemi osakeste vahelisi jõude nende liikumise arvutamiseks. Põhimõtteliselt võiks osakeste asendeid genereerida täiesti juhuslikult, kuid sel juhul oleks valdav enamus konfiguratsioone väga kõrge energiaga ja vastavalt sellele ülimadala kaaluga.

Seetõttu on eelistatud algoritmid, kus mingist mõistlikust lähtegeomeetriast lähtudes genereeritakse uusi geomeetriaid, mis püsivad mõistlikult madala energia juures. Metropolis *et al* sõnastasid lähenemise aastal 1953 selliselt: “Selle asemel, et genereerida juhuslikke konfiguratsioone ja võtta neid arvesse kaaluga $\exp(-E/k_B T)$, valime me konfiguratsioone tõenäosusega $\exp(-E/k_B T)$ ja võtame need arvesse võrdse kaaluga.”

Kui suurus A sõltub ainult osakeste koordinaatidest, siis saame selle keskväärtuse arvutada kui

$$\langle A \rangle = \frac{1}{X} \sum_i^X A(\mathbf{q}_i) \quad (7.13)$$

kus andmed võetakse X koordinaatide kogumi $\mathbf{q}$ jaoks, mis on leitud Metropolis'i algoritmi alusel. Võime täheldada sarnasust võrrandiga (7.8), kuid tuleb rõhutada, et Monte Carlo meetodi korral ei eksisteeri ajakoordinaati — punktid lihtsalt järgnevad üksteisele ilma ajalist suhet omamata.

Praktiliselt lähtub algoritm algkonfiguratsioonist q_1 , millest juhusliku häirimise teel genereeritakse uus konfiguratsioon (“punkt”) q_2 , jne. Iga punkti jaoks arvutatakse sellele vastav energia, mida võrreldakse eelnenuga ja punkt “aktsepteeritakse” tõenäosusega, mis sõltub selle energiast. NVT ansambli korral on vastav võrrand

$$p = \min \left[1, \frac{\exp(-E_2/k_B T)}{\exp(-E_1/k_B T)} \right] \quad (7.14)$$

Seega kui uues punktis on süsteemi energia madalam kui eelmises, siis see aktsepteeritakse alati. Kui aga süsteemi energia tõusis, on selle aktsepteerimine seda vähem tõenäoline, mida rohkem energia suurenes.

Kui punkti ei aktsepteeritud, siis liidetakse summale (7.8) uuesti eelmise punkti A väärtus. Sellist ahelat, kus iga uus punkt sõltub üksnes vahetult eelmisest, nimetatakse *Markovi ahelaks*.

Paljudest molekulidest koosnevate süsteemide modelleerimisel seisneb üks samm tavaliselt ühe molekuli asukoha ja konformatsiooni juhuslikus muutmises. Muutmise määr valitakse selline, et aktsepteerimise tõenäosus oleks 20...50%.

Esimese loenguvideo lõpp.

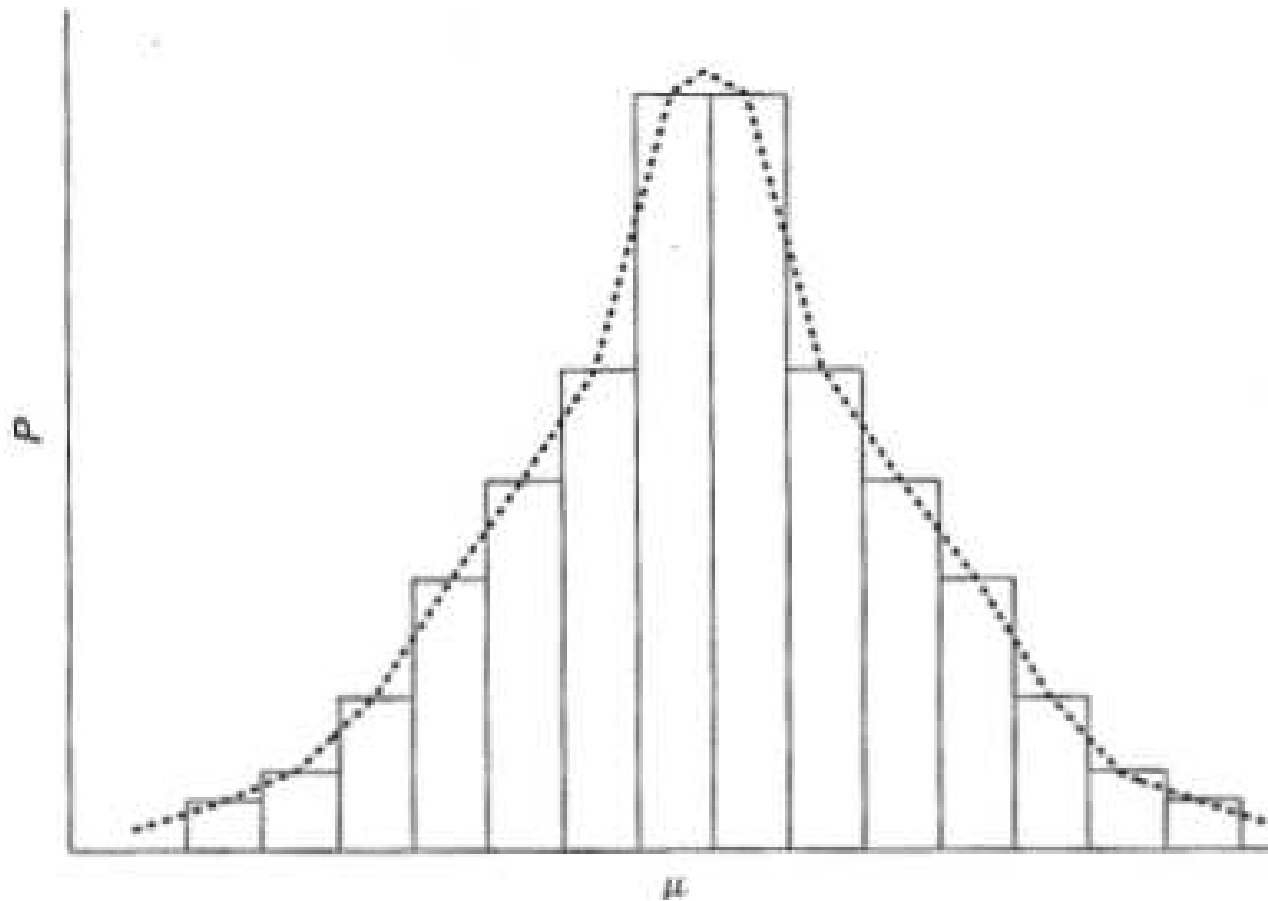
Näiteid makroskoopilistest ja dünaamilistest omadustest

Skalaarne omadus, näit. dipoolmoment

Kui täpne olla, siis skalaarne on dipoolmomendi, mis ise on vektor, magnituud. Vee molekuli dipoolmoment gaasifaasis on 1,85 D. Vedelikus on dipoolmomendi vektorid eri suundadega ja otseselt mõõta vedela vee molekulide dipoolmomente ei saa. Mudelist võiks neid aga arvutada:

$$\langle |\mu| \rangle = \frac{1}{N} \sum_{n=1}^N \left| \sum_{i=1}^3 q_{i,n} \mathbf{r}_{i,n} \right| \quad (7.15)$$

Siin N on molekulide arv süsteemis, q on aatomi (vesinik või hapnik) laeng kasutatavas mudelis ja $\mathbf{r}$ vastava aatomi asukohavektor (koordinaadid). Saadud keskmise võib leida igas trajektoori punktis või valitud punktides ja omakorda keskmistada üle nende.



Joonis 3.3 raamatust "Essentials of Comp. Chem."

Keskväärtus koondab väga suure andmehulga ühteainsasse arvu. Detailsema pildi saame, kui jaotame leitud dipoolmomentide väärtused kindla vahemikuga (näit. 0,01 D) klassidesse ja saame *jaotuskõvera*, millele võib vastavusse seada sujuva *jaotusfunktsiooni*.

Lisaks tavalistele skalaarsetele suurustele, nagu dipoolmoment, võime me uurida suurusi, mis sõltuvad mingist parameetrist. Üheks selliseks on *radiaalne jaotusfunktsioon* (r.d.f.), mis on defineeritud kahe aatomi A ja B jaoks sfäärilises ruumielemendis kui

$$\frac{1}{V}g_{AB}(r) = \frac{1}{N_A N_B} \left\langle \sum_{i=1}^{N_A} \sum_{j=1}^{N_B} \delta[r - r_{A_i, B_j}] \right\rangle \quad (7.16)$$

Sisuliselt see loendab aatomipaare, mille vahekaugus r_{AB} on r .

Radiaalne jaotusfunktsioon on seotud tõenäosusega leida kaks osakest teineteisest kaugusel $r \dots r + \Delta r$:

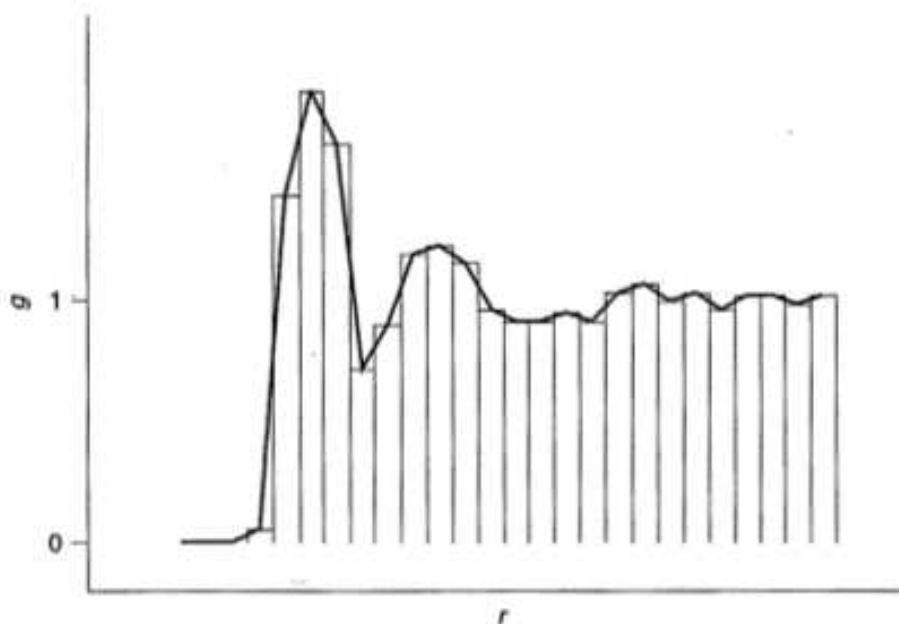
$$P\{A, B, r, \Delta r\} = \frac{4\pi r^2}{V} g_{AB}(r) \Delta r \quad (7.17)$$

Väikesel kaugusel peab $g_{AB}(r)$ kiiresti nulliks muutuma. Suurtel kaugustel peaks see aga olema konstantne, sest kaugete aatomite asukoht ei ole antud aatomi omaga korreleeritud. Kui $g_{AB}(r)$ ei ole suurtel kaugustel konstantne, on süsteemis kaugkorrapära — selle struktuur hakkab lähenema tahkise omale.

Korrapärase tahkise radiaalne jaotusfunktsioon koosneb üksikutest piikidest — aatomid esinevad üksteisest vaid kindlatel vahekaugustel.

Radiaalne jaotusfunktsioon.

Kui $g > 1$, on vastavad vahekaugused eelistatud. Kui $g < 1$, on tõenäosus seal aatomit kohata keskmisest madalam. Toodud näidisfunktsioon ei stabiliseeru suurel kaugusel — süsteemis esineb kaugkorrapära.



Joonis 3.4 raamatust “Essentials of Comp. Chem.”

Sageli huvitab meid eri liiki aatomite vahekaugus, näiteks lahusti molekulide kaugus lahustunud ionist. Jaotades jaotusfunktsiooni miinimumkohtade kohalt lõikudesse, saame pindalasid integreerides teada esimesse, teise jne solvatatsioonisfääri kuuluvate lahustimolekulide arvu.

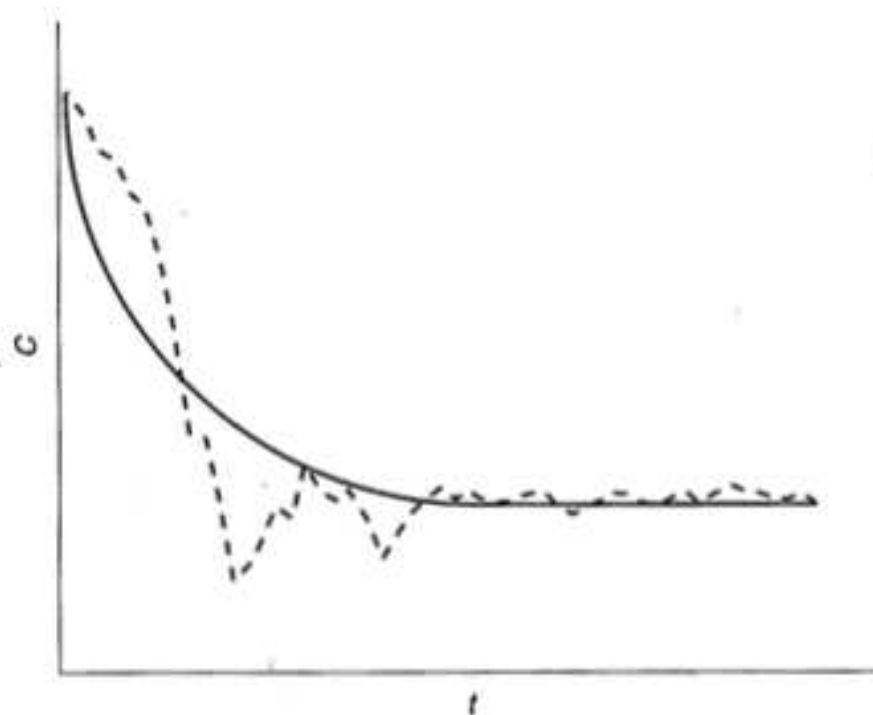
Autokorrelatsiooni funktsioonid

Autokorrelatsiooni funktsioon on defineeritud kui

$$C(t) = \langle a(t_0)a(t_0 + t) \rangle_{t_0} \quad (7.18)$$

ja näitab, mil määral suurus a ajahetkel $t_0 + t$ sõltub selle väärtusest ajahetkel t_0 . Täiesti juhusliku suuruse korral on C alati null. Kui süsteemil esineb “mälu”, siis on alghetkel C maksimaalne ja hakkab tavaliselt seejärel kahanema, näiteks eksponentsiaalselt, või ka perioodiliste võnkumistena.

Pidev joon iseloomustab müra-
vaba langusega autokorrelatsioo-
ni. Kriipsjoonega kujutatud au-
tokorrelatsioonis esineb võnkumi-
si, samuti on alguses väärtuse sta-
biliseerumine aeglane.



Joonis 3.5 raamatust “Essentials of Comp. Chem.”

Eri füüsikalistel suurustel on erinevad iseloomulikud stabiliseerumise ajad. Katsega võrreldavate tulemuste saamiseks peaks mudelit jooksutama vähemalt paar-kolm korda kauem kui on huvipakkuva suuruse stabiliseerumise aeg autokorrelatsiooni funktsiooni põhjal.

Keemiliselt huvipakkuvad autokorrelatsiooni funktsioonid on näiteks:

- Osakeste asukoha ja kiiruse autokorrelatsioonist saab arvutada difusioonikoefitsiente;
- Pinge (*stress*) autokorrelatsiooni funktsioonid on seotud katkemisviskoossusega (*shear viscosity*);
- Dipoolmomendi autokorrelatsioon on infrapunase spektri Fourier' transformatsioon: ühest saab arvutada teist.

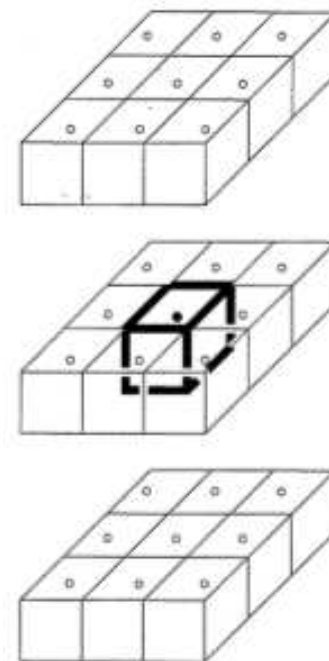
Lisaks on olemas palju kasulikke korrelatsioonifunktsioone kahe *erineva* suuruse vahel.

Praktilisi aspekte

Äralõiked ja rajatingimused

Realistlikes keemilistes süsteemides on nii palju molekule, et süsteemi välispinnal ei ole enamasti olulist mõju (on erandeid!). Modelleerimiseks kohased süsteemid seevastu on väga väikesed ja paljud molekulid on pinna lähedal või pinnal, mis mõjutab oluliselt süsteemi käitumist ja omadusi.

Pinnaefektide elimineerimiseks kasutatakse *perioodilisi rajatingimusi* (periodic boundary conditions, PBC), st süsteemi korratakse kindla vahemaa tagant. Osakesed interakteeruvad kõige tugevamalt endale “lähima” naabriga, mis tegelikult võib asuda mudelsüsteemi teises servas. Kui osake väljub süsteemi ühest küljest, siseneb tema koopia süsteemi teisest küljest.



Joonis 3.6 raamatust “Essentials of Comp. Chem.”

Interaktsioonid lõigatakse tavaliselt ära kuubi küljepikkusest veidi väiksemal kaugusel, et vältida osakese interaktsiooni omaenda koopiaga. Perioodilised rajatingimused võivad süsteemile lisada mitte-füüsikalisi omadusi ja nähtusi, kuid nende mõju on enamasti väiksem kui väikese tilga või klatri korral (otsene kokkupuude vaakumiga). Samuti muutuvad mõjud seda väiksemaks, mida suuremat süsteemi modelleerimiseks kasutada.

Kaugmõjuga vastasmõjud (eelkõige elektrostaatilised) kannatavad märkimisväärselt, kui kasutada äralõiget. Seetõttu eelistatakse nende jaoks kasutada summeerimist üle kogu lõpmatu korduva süsteemi (Ewaldi summad).

Kui süsteem on olemuslikult perioodiline (näiteks kristall), siis on perioodiliste rajatingimuste rakendamine loomulik ja süsteemi olemusega kooskõlas.

Perioodiliste rajatingimuste korral on süsteemi osakeste arv, impulss, ja sellest tulenevalt ka energia, automaatselt konstantsed. Praktikas toob lõplike integreerimissammude kasutamine sisse fluktuatsioonid energias, mida on vaja kunstlikult summutada.

Süsteemi parameetrite vaoshoidmine

Tüüpilises MD mudelis on kõige lihtsam hoida konstantsena osakeste arvu, süsteemi ruumala ja energiat, st kasutada NVE ansamblit. Sellel on aga nõrk seos reaalse katsetingimustega.

Praktilisem oleks energia asemel hoida konstantsena temperatuuri, st kasutada NVT ansamblit. Temperatuuri hoidmiseks on kõige lihtsam igal sammul skaleerida osakeste kiirusi, mis aga rikub Verlet' või konnahüppe meetodite aluseks oleva Newtoni mehaanika seadusi. Saadavad omaduste keskväärtused ei ole enam usaldusväärsed.

Nn *Berendseni meetodil* (1984) on skaleerimine sujuvam, modelleerides süsteemi kontakti suure termostaadiga (bath), mille temperatuur on T_0 . Osakestele rakendatakse täiendav jõud, mis peab nende kiirused sobivaks muutma.

Süsteemi alg-geomeetria ei ole peaagu kunagi faasiruumi suure tõenäosusega piirkonnas. Seetõttu tavaliselt alustatakse modelleerimist madalalt temperatuurilt ja suurte pidurdusjõududega, et saavutada esialgne relaksatsioon enam-vähem mõistliku geomeetria juurde.

Sellele järgneb “tasakaalustamine” (equilibration) periood, mille vältel jälgitakse temperatuuri, erinevaid korrelatsioonifunktsioone jm parameetreid, neid vajadusel väliselt mõjutades, kuni on saavutatud uuritavate suuruste stabiilsus.

Praktikas võib tasakaalustamine olla üsna keerukas ja mitmeastmeline, näiteks alguses relakseeritakse solvent, seejärel järk-järgult lahustunud molekulid ja ioonid, võidakse sobitada need huvipakkuvasse algasendisse, jne.

Alles seejärel algab “tootmine” (production), st arvutuse periood, kus kogutakse arvutuse eesmärgiks vajalikud statistilised andmed.

Paljud keemiakatsed viiakse läbi mitte konstantsel ruumalal, vaid konstantsel rõhul. Rõhku on võimalik arvutada osakesevahelistest jõududest ja jooksvalt korrigeerida, muutes süsteemi ruumala. Sel juhul on tegemist NPT ansambliga. Otsesele ruumala muutmisele eelistatakse tänapäeval *Nosé-Hoover'i skeemi* (1984–1985), mis toob sisse rõhu ja temperatuuriga seotud täiendavad muutujad, mille eesmärk on vastavate suuruste ettenähtud piirides hoidmine.

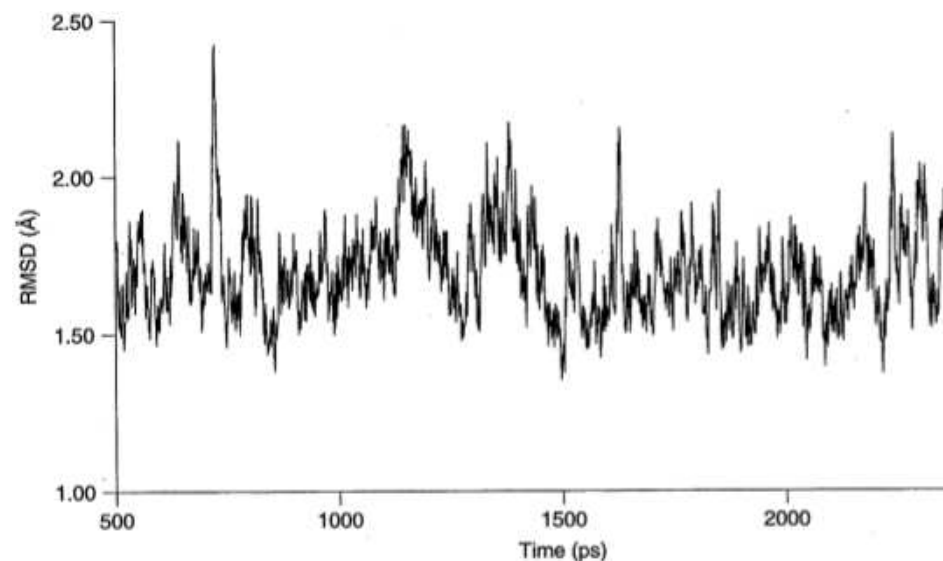
Monte Carlo meetodite puhul on kõige loomulikum töötada NVT ansambliga. Teiste ansamblite kasutamisel peab konfiguratsiooni aktsepteerimise algoritm võtma arvesse (ka) muid muutujaid peale energia. Ka MC modelleerimiste korral tuleb tavaliselt kasutada algset tasakaalustamise ja sellele järgnevat tootmise perioodi.

Koonduvus

Arvutus loetakse koondunuks, kui on kogutud piisav hulk faasiruumi punkte, eeldamaks, et kõik olulised faasiruumi osad on piisavalt läbi käidud (vt ka ergoodilisuse hüpotees, eespool). Üldiselt ei ole võimalik tõestada, et koonduvus on saavutatud, mis osade uurijate jaoks on fundamentaalne murekoht.

“Statistiline koondumine” näidatakse tavaliselt ära ühe-kahe uuritava suuruse graafiku esitamisega, millelt on näha, et see suurus on piisavalt pika aja vältel käitunud ootuspäraselt (näit. olnud enam-vähem konstantne).

Biomolekulide korral on selliseks suuruseks tüüpiliselt osakeste ruutkeskmise kauguse kristallstruktuuriga võrreldes (RMSD).



Joonis 3.7 raamatust “Essentials of Comp. Chem.”

Solvateeritud tRNA mikroheeliksi RMSD pärast 500 ps tasakaalustamisperioodi (Nagan *et al* 1999).

Praktikas ei ületa simulatsiooniajad enamasti $1\ \mu\text{s}$, olles tihti palju lühemad. See loob võimaluse, et süsteem on “kinni” metastabiilses olekus, millest väljumine nii lühikese aja vältel on vähetõenäoline. Seetõttu jooksutatakse tavaliselt mitu eraldi trajektoori erinevast lähtegeomeetriast lähtudes. Kui need annavad sarnased tulemused, lisab see kindlust, et süsteem on tõepoolest koondunud.

Statistilisele koonduvusele lisaks on palju keerulisem “keemilise koonduvuse” saavutamine. Näiteks paljud biomolekulid võivad olla erinevalt protoneeritud. Prootoni lisandumine ja lahkumine ei ole enamikus jõuväljades kasutatava mudeli loomulik osa — kui prooton on mingis asukohas (aminohappe jäägi küljes), siis ta ei kandu üle solvendile ja vastupidi. Prootoniülekanne tuleb soovi korral läbi viia “käsitsi”. Kui protoneerumis-saite on palju, tuleks läbi proovida kõik, pluss võib-olla rida nende kombinatsioone.

Kui arvutuse läbiviija ei suuda seda laadi muutusi süsteemis ette näha, võib kogu arvutus olla keemiliselt mõttetu, sest uurib praktikas mitte ette tulevat faasiruumi osa.

Mitme miinimumi probleem

Süsteem võib omada mitut konformatsiooni või isomeeri, mis on eraldatud suhteliselt kõrge barjääriga. Barjääri ületamise juhtude vaheline aeg võib olla tunduvalt pikem kui lühike simulatsiooniaeg. Sel juhul on süsteem ühes miinimumis “lõksus” ja tuleks taas jooksutada mitu erineva miinimumi piirkonnas paiknevat trajektoori.

Seejuures tuleb saadud tulemusi kaaluda vastavate energiafaktoritega, sest reaalses süsteemis on suurem osa molekule madalama energiaga olekus ja väiksem osa (vastavalt Boltzmanni jaotusele) kõrgema energiaga olekutes.

Vähegi suuremate süsteemide korral ei ole kõigi konformatsioonide käsitsi läbi arvutamine realistlik. Nii näiteks on tsüklodekanaal (10 süsinikku) 18 konformeeeri, kuid tsükloheptadekanaal (17 süsinikku) juba 12513 konformeeeri. Sellest tulenevalt on ka proteiini pakkimise (protein folding) ülesanne — ennustada valgu 3D struktuur üksnes primaarjärjestuse põhjal — siiamani sisuliselt lahenduseta.

Barjääride automaatseks ületamiseks on välja pakutud süsteemi kuumutamist kõrge temperatuurini ja taas maha jahutamist. Kuna jõuväljades sidemed ei katke, ei ohustata sellega süsteemi keemilist iseloomu.

On loodud mudeleid, millesse on sisse toodud täiendav “reaktsioonikoordiaat”, mis võimaldab kunstlikult sundida süsteemi energiabarjääri ületama (Stolovitzky ja Berne 2000) või kasutada logaritmilist temperatuuriskaalat (Nakamura 2002) koos vastava järelkorrektsiooniga.

Samuti on välja töötatud mudeleid, kus eksisteerib mitu “paralleelmaailma”, milles sama molekul on erinevates konformatsioonides (Verkhivker, Elber, Novak, 1992).

Lõplikku kindlust, et keemiline koondumine on saavutatud, ei anna aga ükski meetod ning kõige usaldusväärsemaks peetakse võrdlemist eksperimendiandmetega. Süsteemide modelleerimine, mille sarnaseid ei ole katseliselt kunagi uuritud, või polegi võimalik uurida, on alati seotud suurte riskidega. Samas just need on samadel põhjustel sageli enim huvipakkuvad.