VILNIAUS UNIVERSITETAS
FIZIKOS FAKULTETAS
TEORINĖS FIZIKOS KATEDRA

Matas Lileikis

ALICINO MOLEKULIŲ DNR SEKAI GALIMO POVEIKIO TYRIMAS CAR-PARRINELLO MOLEKULIŲ DINAMIKOS METODU

Pagrindinių studijų baigiamasis darbas

(studijų programa – MODERNIŲJŲ TECHNOLOGIJŲ FIZIKA IR VADYBA)

Studentas	Matas Lileikis
Darbo vadovas	doc. dr. Mindaugas Mačernis
Recenzentas	dr. Stepas Toliautas
Katedros vedėjas	prof. habil. dr. Leonas Valkūnas

Vilnius 2017

Turinys

Įvadas 2
1 Modeliavimo teorija 4
1.1 Ab initio metodo teorija 4
1.2 Tankio funkcionalo teorija (DFT) 5
1.3 Molekulių dinamika 6
1.4 Car-Parrinello metodo teorija 7
1.5 Temperatūros modeliavimas 8
2 Skaičiavimo priemonės ir metodika 9
2.1 Superkompiuteris „HPC Saulėtekis" 9
2.2 Modeliavimas su NWChem programų paketu 10
2.3 Grafinis atvaizdavimas su VMD ir ChemCraft programomis 12
3 DNR sekos ir alicino molekulių modeliavimo rezultatai ir jų aptarimas 13
3.1 Kelių alicino molekulių modeliavimas 14
3.2 DNR sekos modeliavimas 16
3.3 DNR sekos kartu su alicino molekule modeliavimas 20
Išvados 21
Literatūros sąrašas 23
Santrauka 24
Summary 25
Priedas Nr. 1. 26

Įvadas

Česnakas ir svogūnas yra vieni seniausiai kultivuojamų augalų. Tūkstančius metų jie naudojami liaudies medicinoje, dar XVI a. pr. m. e. egiptiečių papirusuose aptinkami gydomieji receptai, kuriuose česnakas nurodomas kaip pagrindinis ingredientas [1]. Tokios savybės kaip aštrus skonis, ašarojimo skatinimas, antibakterinis bei priešgrybelinis efektas – tai mokslinį susidomėjimą skatinantys objektai.

1944 metais amerikiečių mokslininkas Chester John Cavallito atrado, kad pažeidus česnakinio augalo audinį susiformuoja mažos molekulinės masės medžiaga, kurios cheminė formulė yra $C_6H_{10}OS_2$ [1]. Chester J. Cavallito molekulę pavadino alicinu. Ši molekulė yra chemiškai lengvai reaguojanti, turi gamtoje retai aptinkamų sieros jungčių ir spontaniškai sudaro įvairius sieros junginius. Per pastaruosius metus atlikti tyrimai apie alicino molekulės teigiamą poveikį organizmui [2] teigia, kad alicinas tiesiogiai veikia DNR molekulę [3]. Tai skatina detaliai nagrinėti alicino ir DNR molekulių sąveiką ir išsiaiškinti ar susidaro būdingi kompleksai.

Šio darbo pagrindinis tikslas yra ištirti alicino molekulės poveikį DNR grandinės struktūrai. Pagrindinių studijų kursinio darbo „Alicino molekulės kompleksų modeliavimas Car ir Parrinello molekulių dinamikos metodu“ [4] metu gauti alicino modeliavimo rezultatai padėjo įgyvendinti šį tikslą. Toliau išvardijami pagrindiniai uždaviniai:

1. modeliuoti dviejų ir trijų alicino molekulių sistemas Car-Parrinello metodu,
2. padalinti DNR grandinę į segmentus ir modeliuojant nustatyti, kuriuos iš jų būtų galima panaudoti DNR sekos ir alicino molekulių simuliacijoje,
3. sudaryti keletą sistemų iš DNR segmento ir alicino molekulių ir nagrinėti jų struktūrinius pokyčius modeliuojant Car-Parrinello molekulių dinamikos metodu.

Ruošiant modeliavimo uždavinių įvesties kodą, buvo remtasi Lenkijos Vroclavo universiteto mokslininkų straipsniu „Teorinis tyrimas apie alicino molekulinę struktūrą, topologiją ir virpesius“ [5]. Jame pateikta informacija apie alicino simuliacijoje naudojamus parametrus. Žinoma, modeliuojant sistemą su DNR segmentu, kai kurie nustatymai turėjo būti pataisomi, pavyzdžiui, modeliuojamos gardelės tūris. Uždaviniai superkompiuteriu „HPC Saulėtekis” buvo skaičiuojami po kelis kartus, koreguojant skaičiavimo tikslumą, kuris yra tiesiogiai susijęs su modeliavimo sparta – alicino uždaviniams modeliuoti puse pikosekundės užtekdavo poros dienų, o DNR segmento kartu su alicinu uždaviniams prireikdavo daugiau nei savaitės.

1 Modeliavimo teorija

1.1 Ab initio metodo teorija

Terminas ab initio yra kilęs iš lotynų kalbos ir reiškia „nuo pradžių“. Kadangi taip pavadintas kvantinės chemijos metodas, gali pasirodyti, jog skaičiavimai yra atliekami be aproksimavimo. Tačiau tokia prielaida būtų neteisinga, nes Schrödinger lygtį dviejų kūnų sistemai yra neįmanoma išspręsti be aproksimavimo. Šiame kontekste „ab initio“ turi kitą prasmę – Schrödinger lygties integralai yra sprendžiami nenaudojant empirinių parametrų [6].

Pirmiausia yra taikoma Born–Oppenheimer aproksimacija. Ji nusako atomų branduolių ir elektronų judėjimą kaip atskirus veiksmus – taip Schrödinger lygtis molekulinei sistemai supaprastinama iki elektrono judėjimo konkrečiai branduolių konfigūracijai. Dėl to sumažėja skaičiavimo ciklų skaičius ieškant pusiausvyros struktūros ir molekulės energijos:

$$\hat{H}^{elektrono}\psi^{elektrono} = (E^{elektrono} + E^{branduolio})\psi^{elektrono}, \tag{1}$$

čia $E^{branduolio}$ yra narys atsiradęs dėl branduolio stūmos.

Antroji yra taikoma Hartree–Fock aproksimacija. Ja elektronų judėjimas yra atskiriamas tarpusavyje, o tai lemia tų elektronų pasiskirstymą orbitalėse. Šiuo būdu elektronams bendra banginė funkcija perrašoma vieno elektrono erdvės banginėmis funckijomis f_i ir sukinio funkcijomis α bei β:

$$\psi = f_i(1)\alpha f_i(2)\beta \cdots f_{N/2}(N)\beta. \tag{2}$$

Dėl Hartree–Fock aproksimacijos atskirtas elektronų judėjimas kelia elektronų koreliacijos problemą. Ši problema sprendžiama įvairiais būdais, pavyzdžiui, taikant Coupled-Cluster (CC) ir Møller-Plesset (MPn) modelius [6].

Elektrono tam tikroje orbitalėje energija ε_i yra suskaičiuojama iš tos orbitalės koeficientų c_{ik} funkcijos. Šios lygtys dar vadinamos Roothaan-Hall lygtimis:

$$\sum_k c_{ik}\hat{H}^{HF}\varphi_k = \varepsilon_i \sum_k c_{ik}\varphi_k, \tag{3}$$

čia $\hat{H}^{HF}$ yra Hartree–Fock hamiltonianas.

Norint suskaičiuoti $\hat{H}^{HF}$, reikia „spėti“ kitų orbitalių koeficientų reikšmes, su kuriomis bus lyginami koeficientai, gauti pritaikius variacinius kvantinės mechanikos optimizavimo metodus. Jei tarp šių koeficientų yra gaunamas skirtumas, skaičiavimai vykdomi iš naujo, pakeitus orbitalių koeficientus prieš tai optimizuotais. Procesas tęsiamas, kol pasiekiamas norimas skirtumas tarp nustatytų ir gautų koeficientų – šis iteracinis metodas naudojamas elektronų koreliacijos problemoms spręsti ir yra vadinamas SCF arba suderintinio lauko procedūra.

1.2 Tankio funkcionalo teorija (DFT)

Sprendžiant modeliavimo užduotis kompiuteriu su sudėtingomis molekulėmis, ab initio metodai pradeda naudoti vis daugiau kompiuterio resursų, o tai yra pagrindinis ab initio trūkumas. Tankio funkcionalo metodai (DFT) yra alternatyvus būdas atlikti tuos pačius skaičiavimus, kurių tikslumas prilygsta ab initio su CC ir MPn modeliais, bet skaičiavimo greitis didesnis [6]. Tiriamos sistemos pilnutinė energija nagrinėjama kaip elektronų tankio funkcija. Kadangi elektronų tankio pasiskirstymo funkcija priklauso tik nuo trijų koordinačių, DFT skaičiavimo greitis yra artimas Hartree–Fock metodo skaičiavimo greičiui.

Taikant tankio funkcionalo metodus gaunama tokia energijos išraiška:

$$E^{TF} = E^{EB} + E^{branduolio} + E^{Kulono} + E_{XC}[\rho], \tag{4}$$

čia E^{EB}, $E^{branduolio}$, E^{Kulono}, $E_{XC}[\rho]$ atitinkamai yra elektrono kartu su branduoliu, stūmos tarp branduolių, stūmos tarp elektronų ir koreliacijos energijos.

Lokalaus tankio funkcionalo teorija koreliacijos energiją apibrėžia kaip elektronų tankio funkcijos integralą:

$$E_{XC} = \int \rho(\vec{r}) \varepsilon_{XC}[\rho(\vec{r})] d\vec{r}, \tag{5}$$

čia $\varepsilon_{XC}[\rho(\vec{r})]$ yra elektrono koreliacijos energija homogeniškai pasiskirsčiusių nekintamo tankio elektronų debesyje. Elektronų tankio matrica $\rho(\vec{r})$ N elektronų sistemoje yra nustatoma pagal Kohn–Sham orbitales ψ_i:

$$\rho(\vec{r}) = \sum_{i=1}^{N} |\psi_i|^2. \tag{6}$$

Taigi, kad būtų galima suskaičiuoti koreliacijos energiją E_{XC}, reikia žinoti elektronų tankį. Jam sužinoti yra „spėjamos“ Kohn–Sham orbitalių vertės. Kaip ir ab initio metode, gauti sprendiniai įvertinami atsižvelgiant į realią sistemą. Norint sumažinti paklaidas ir sužinoti tikslesnį elektronų tankį ir koreliacijos energiją, Kohn–Sham orbitalės yra perskaičiuojamos Kohn–Sham lygtimis:

$$\left\{ -\frac{1}{2}\nabla_1^2 - \sum_A \frac{Z_A}{r_{A,1}} + \int \frac{\rho(r_2)}{r_{12}} dr_2 + V_{XC}(r_1) \right\} \psi_1(r_1) = \varepsilon_i \psi_i(r_1), \tag{7}$$

čia ε_i yra Kohn–Sham orbitalės energija, o V_{XC} – koreliacijos potencialas, kuris lygus koreliacijos energijos išvestinei pagal elektronų tankį.

1.3 Molekulių dinamika

Molekulių dinamika yra vadinami skaitiniai modeliavimo kompiuteriu metodai, kurie leidžia sužinoti daugelio molekulių sistemos parametrus, tokius kaip atskiros molekulės energija ir struktūra [7].

Molekulių dinamika modeliuoja sudėtingos sistemos priklausomybę nuo laiko, pavyzdys – Brauno judėjimas. Žemiau pateikti bendri molekulių dinamikos vykdymo žingsniai sistemai, kuri yra pusiausvyros būsenoje:

1. nustatomos pradinės atomų koordinatės,
2. parenkamas kiekvieno atomo greitis taip, kad galiotų Bolcmano pasiskirstymas,
3. suskaičiuojamas kiekvieno atomo judesio kiekis pagal atomo greitį ir masę,
4. kadangi visi atomai yra nustatytame molekulių dinamikos jėgų lauke, suskaičiuojamos kiekvieną atomą veikiančios jėgos,
5. po trumpo laiko tarpo, vadinamu „laiko žingsniu“, taikant Niutono dėsnius, suskaičiuojamos naujos atomų koordinatės,
6. suskaičiuojamas pasikeitęs kiekvieno atomo greitis ir pagreitis,
7. kartojami 3-6 žingsniai,
8. iteruojama, kol pasiekiamas naujas sistemos pusiausvyros taškas, kuris yra parenkamas ne pagal mažiausią energiją, o pagal tai, kiek realistiška tokios energijos sistemai yra egzistuoti,
9. iteravimo procese yra renkama informacija – išsaugoma kas 5-25 „laiko žingsnius“
10. iteravimas tęsiamas, kol suskaičiuojamas užsibrėžtas uždavinys
11. vykdoma sistemos analizė – skaičiuojami difuzijos koeficientai, gradientai, spinduliavimo funkcijos ar bet kokie kiti dydžiai, kuriuos galima nustatyti iš surinktos informacijos.

Molekulių dinamikoje labai svarbu nustatyti reikiamą, bendrą visai sistemai jėgų lauką, nes nuo jo priklauso viso modeliavimo eiga. Būtent šio lauko jėgų linijos aprašo tarpatomines jėgas, dėl kurių veikimo sistema gali vystytis. Modeliavimo užduotims yra sugalvoti ir plačiai naudojami GROMOS ir OPLS jėgų laukai.

Yra keletas algoritmų, kurie naudojami skaitiniuose modeliavimo metoduose. Ypač plačiai naudojamas yra Verleto integravimo metodas, nes jo efektyvumas yra didžiausias kompiuterio resursų naudojimo atžvilgiu – reikalauja mažiausiai atminties ir procesoriaus darbo laiko.

1.4 Car-Parrinello metodo teorija

Lietuviškai – Karo ir Parinelo procedūra, tai – elektronų struktūrų metodų, kaip ab initio ir DFT, pritaikymas molekulių dinamikos modeliavime. Bendru atveju, ab initio ir DFT metodai yra labiau tinkami nuosekliam ir tiksliam mažų sistemų modeliavimui, o molekulių dinamika pasižymi skaitiniais metodais, kurie skaičiuoja ne taip tiksliai, tačiau geba modeliuoti dideles molekulių sistemas ir nustatyti įvairius dinaminius sistemos dydžius [8].

Elektronų struktūrų ir molekulių dinamikos metodus praktiškai įmanoma sujungti suskaičiavus jėgas, veikiančias atomus pradinėje sistemos konfigūracijoje. Tai galima padaryti pritaikius Hellmann–Feynman teoremą:

$$\vec{F}_I = -\left\langle \psi \left| \frac{dH}{d\vec{R}_I} \right| \psi \right\rangle. \tag{8}$$

Šias jėgas galime įsistatyti į Niutono lygtis ir taip sistemos konfigūracija pasikeis vienu „laiko žingsniu“. Kartojant šį procesą yra įmanomas sistemos modeliavimas, tačiau tokiam būdui reikia labai galingų kompiuterių, nes kiekvienu laiko momentu yra perskaičiuojama visa sistemos elektroninė struktūra.

Kad to daryti nereikėtų, 1985 metais buvo pasiūlytas alternatyvus Car-Parrinello metodas. Juo skaičiavimuose yra įvedama elektronų banginė funkcija, kuria yra įgalinamas lygiagretus atomų judėjimo skaičiavimams elektronų struktūrų perskaičiavimas. Taikant šį procesą, Niutono dėsniai yra pakeičiami Lagranžo klasikine mechanika. Ji pasižymi apibendrintomis koordinatėmis q_i ir judesio kiekiu $p_i = m dq_i/dt$, o pagrindinis dydis, kuris aprašo elektronų dinamiką, yra papildytas lagranžianas:

$$L^{papildytas} = K_N + K_e - E_{pot}, \tag{9}$$

čia K_N yra branduolių kinetinė energija, K_e yra elektronų kinetinė energija, o E_{pot} – potencinė energija, kuri priklauso nuo branduolių ir elektronų pozicijos.

Car-Parrinello procedūra remiasi tankio funkcionalo metodu (spartesnis elektronų koreliacijos problemos sprendimas Kohn-Sham orbitalių metodu) ir yra naudojama efektyviau išnaudoti kompiuterio resursus molekulių dinamikos modeliavimo uždaviniuose.

1.5 Temperatūros modeliavimas

Sistemos, sudarytos iš N atomų, temperatūra yra nusakoma jų judesio kiekiais:

$$T(t) = \frac{1}{(3N-n)k_B} \sum_{i=1}^{N} \frac{|\vec{p}_i(t)|^2}{m_i}, \tag{10}$$

Kadangi temperatūra priklauso nuo atomo greičio, teoriškai galima sistemos temperatūrą nustatyti atitinkamai reguliuojant greičio parametrus [9]. Tačiau toks sprendimas būtų klaidingas, nes norint išlaikyti nekintančią sistemos temperatūrą, reikėtų daryti daug greičio pakeitimų. Tai aprašytų trajektorijas, neatitinkančias klasikinės mechanikos dėsnių.

Momentiškai pradėti sudėtingos sistemos modeliavimą, kurios būsena atitiktų aukštos tikimybės tašką fazinėje erdvėje, yra praktiškai neįmanoma. Prieš pradedant modeliuoti yra įprastas pusiausvyros būsenos (angl. equilibration) paruošimo etapas, per kurį sistema relaksuojama iki realistiškos konfigūracijos. Nustatant pusiausvyros būseną, sistemos temperatūra yra keliama iki pageidaujamos, pradedant nuo 0 K. Per šį laiką dalelių judėjimas ribojamas, kad būtų pasiekiama reikiama temperatūra. Tai suteikia pranašumo tiksliau modeliuojant sistemą, nes pradėti simuliaciją be pusiausvyros būsenos nustatymo reikštų padaryti tikėtinai didelę pradinės konfigūracijos spėjimo paklaidą.

Alternatyvus temperatūros nustatymo procesas tai plačiai paplitusi Nose-Hoover procedūra. Jos esmė ta, kad temperatūrai kontroliuoti yra įvedami nepriklausomi kintamieji. Modeliuojant sistemą, šie kintamieji yra perskaičiuojami greta judėjimo kintamųjų, o tai leidžia atskirti judėjimo bei temperatūros nustatymo logiką. Žinoma, ir toliau reikalingas pusiausvyros būsenos nustatymo periodas, tačiau naudojant Nose-Hoover procedūrą, sumažėja simuliacijos laikas.

NWChem programų paketas, dirbant Car-Parrinello metodu, suteikia galimybę taikyti Nose-Hoover procedūrą [10]. Tai įgyvendinama taikant pagalbinius pilnutinės energijos jonams ir elektronams funkcionalus:

$$\begin{aligned} E_j(x_j) &= \frac{1}{2} Q_j \dot{x_j} + E_{j0} x_j, \\ E_e(x_e) &= \frac{1}{2} Q_e \dot{x_e} + E_{e0} x_e. \end{aligned} \tag{11}$$

Lygtyse esantis narys $E_{j0} = \frac{1}{2} f k_B T$ yra vidutinė kinetinė energija. Šiame naryje ir slypi sistemos modeliuojama temperatūra, kurios norimą vertę reguliuoja kintamieji x_j bei x_e.

2 Skaičiavimo priemonės ir metodika

2.1 Superkompiuteris „HPC Saulėtekis”

Šio kursinio darbo pagrindinė priemonė buvo superkompiuteris „HPC Saulėtekis”. Jis yra Nacionalinio fizinių ir technologijos mokslų centro infrastruktūros dalis. Superkompiuterio mazgai yra padalinti vartotojų prisijungimui (telkinio grupė LOGIN1) ir uždavinių skaičiavimui, bet šiame darbe buvo dirbama su SUPER telkinio grupės mazgais. Juos sudaro 96 GB operatyviosios atminties, 12 branduolių procesoriai, galintys atlikti lygiagrečius skaičiavimus kitų procesorių atžvilgiu. Superkompiuteriu didžiausia pasiekiama sparta yra 25,7 teraflopai su 15 TB operatyviosios atminties [11].

Jungiantis per WebMO sistemą, uždavinius galima valdyti per patogią grafinę vartotojo prieigą. Siekiant išbandyti šią funkciją, buvo paleisti vandens ir etileno molekulių struktūros optimizavimo uždaviniai, kurie sėkmingai buvo baigti per mažiau nei minutę.

Šiuo metu WebMO leidžia skaičiuoti uždavinius su Gaussian, QChem, NWChem ir kitais programų paketais, tačiau norint pilnai kontroliuoti uždavinio procesą, tai reikia daryti komandine eilute. Ja turime daugiau pasirinkimų ir komandinių nustatymų, pavyzdžiui, galime tiksliai nurodyti, kiek ir kurių mazgų procesorius užsakysime darbui, tačiau svarbiausias pranašumas, kai dirbame prisijungę nuotoliniu būdu su komandine eilute, yra tas, kad galime naudoti paketus, kurie nepalaiko grafinės prieigos. Turint leidimą, galima diegti naujus programų paketus, todėl pilnaverčiam darbui superkompiuteriu būtina tai daryti komandine eilute.

Prisijungti nuotoliniu būdu reikia dviejų dalykų: SSH programos prisijungimui (gali būti naudojama putty programa) ir privataus rakto bylos, kurią gali suteikti sistemos administratoriai. Kai privataus rakto byla yra nurodoma autentifikavimo nustatymuose, žinant vartotojo prisijungimo slaptažodžius, galima prisijungti prie superkompiuterio vartotojų mazgo super1.ff.vu.lt.

2.2 Modeliavimas su NWChem programų paketu

NWChem programų paketas yra skirtas didelei uždavinių įvairovei vykdyti [12]. Jis buvo projektuotas darbui su didelio našumo superkompiuteriais, todėl skaičiavimo greitis turėtų nemažėti uždavinį dalijant daugeliui branduolių lygiagrečiam darbui. Praktikoje šis teiginys yra sunkiai patvirtinamas, bent jau iš to, kiek su NWChem teko dirbti šio darbo metu. Tačiau platus NWChem funkcijų ir metodų pasirinkimas bei dokumentacijos kokybė lenkia alternatyvius programų paketus, galinčius modeliuoti Car-Parrinello metodu, kaip, pavyzdžiui, CPMD [13].

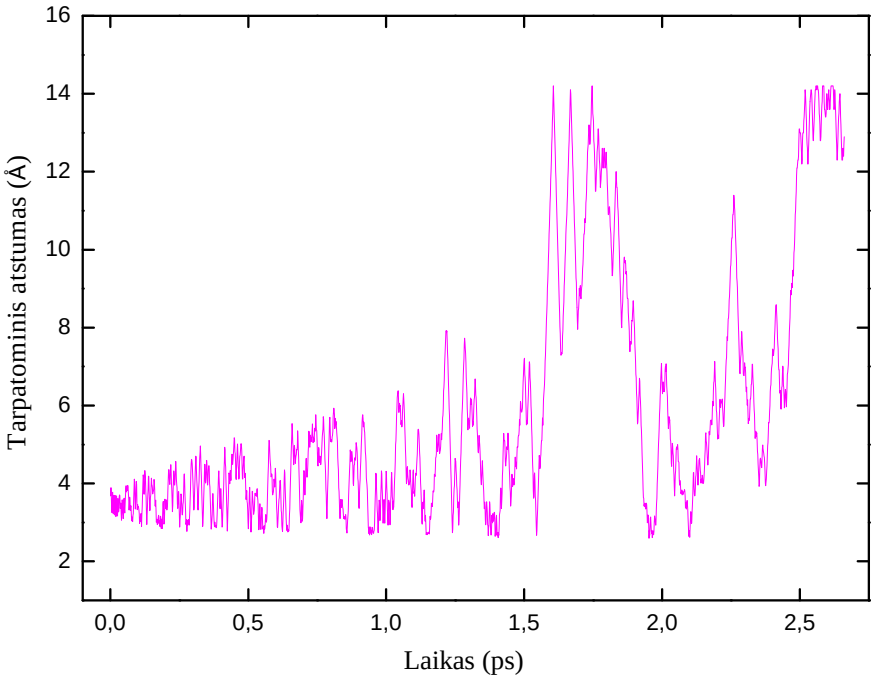
Modeliavimo uždavinys yra vykdomas skaitant instrukcijų įvesties failą. NWChem programos įvesties kodas yra skaidomas į atskirus segmentus, kaip parodyta pavyzdyje (1 pav.), kuriame kodas yra paruoštas alicino struktūros optimizavimo užduočiai spręsti.

```
title "Alicino geometrijos optimizavimo uždavinys"
geometry
 C                    -0.68637536     1.98477810     0.37730864
 C                    -0.08290980     3.34664142     0.19111556
 C                    -0.73003113     4.37750929    -0.32454386
 H                    -0.25219530     5.34348079    -0.42538366
 H                    -1.75892966     4.30238765    -0.65791294
 H                     0.94629869     3.44983553     0.51374636
 S                     0.26755537     0.76389054    -0.58684315
 S                    -0.85962004    -0.92006012     0.07089685
 C                     0.32545195    -2.21542271    -0.48537083
 C                     0.09027607    -3.47052341     0.30291017
 C                    -0.34854954    -4.60266937    -0.21988926
 H                    -0.58187941    -4.69330807    -1.27466576
 H                    -0.48611482    -5.48583075     0.39068729
 H                     0.30697648    -3.41094012     1.36401673
 H                     0.19844327    -2.38748641    -1.55205557
 H                     1.32117926    -1.81855700    -0.29225738
 O                     1.60963739     0.70444891     0.06825293
 H                    -1.72331428     1.94379934     0.04416132
 H                    -0.61255010     1.65372905     1.41310665
end
basis
 C library cc-pvdz
 H library cc-pvdz
 S library cc-pvdz
 O library cc-pvdz
end
task scf optimize
```

1 pav. NWChem įvesties duomenų failo pavyzdys

Čia pirmojoje eilutėje pateiktas uždavinio pavadinimas, kuris neturi įtakos skaičiavimams. Antrame, raudonai pažymėtame, segmente yra aprašoma molekulės pradinė struktūra, kurią šiuo uždaviniu optimizuosime. Kiekvienam molekulės atomui yra nurodytos x, y ir z koordinatės. Trečiame, mėlynai pažymėtame, segmente yra nurodytos kiekvieno atomo bazinės funkcijos, kurios padeda konstruoti molekulės orbitales. Paskutiniame segmente yra nurodytas uždavinio vykdymo kodas, pagal kurį NWChem programų paketas pradės atitinkamus skaičiavimus, šiuo atveju – struktūros optimizavimą.

Sudėtingesniems uždaviniams spręsti buvo mokomasi iš NWChem dokumentacijoje pateiktų pavyzdžių [14]. Vienas jų – disociacijos reakcijos $CH_3Cl \rightarrow CH_3^+ + Cl^-$ modeliavimas. Šiuo uždaviniu buvo išbandytas NWChem programų paketo Karo ir Parinelo metodo taikymas, įvairių parametrų naudojimas ir duomenų išvesties failų spausdinimas, kuriuos naudojant buvo atvaizduoti sprendimo rezultatai (2 pav.).



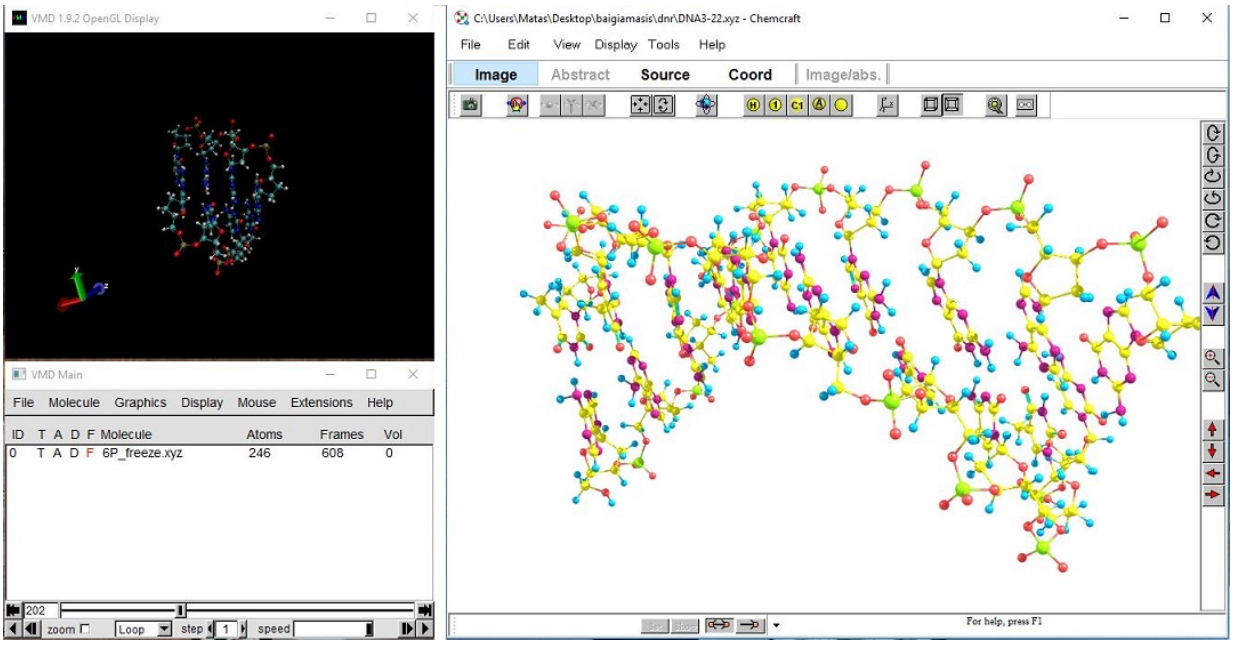
2 pav. C–Cl ryšio kitimas vykstant disociacijai

2.3 Grafinis atvaizdavimas su VMD ir ChemCraft programomis

NWChem programų paketu galima nurodyti, kad simuliacijos metu skaičiuojamos sistemos struktūros koordinatės būtų spausdinamos į atskirą duomenų failą. Tokiu būdu galima išsaugoti kiekvieno modeliuoto kadro geometriją ir šią informaciją panaudoti grafiniam modeliuojamos sistemos atvaizdavimui.

Sprendžiant šiame darbe aprašytus Car-Parrinello dinamikos uždavinius, NWChem programos įvesties kode (žr. 1 priedą) nurodyta geometrinę informaciją spausdinti į meta.xyz bylą. Šį duomenų formatą palaiko VMD (angl. Visual Molecular Dynamics) [15] programa, kuria galima animuoti simuliacijos kadrus, nustatyti matuojamą atstumą tarp atomų 3D erdvėje ir kitaip apdoroti vaizdinę informaciją. VMD programos veikiamasis langas pavaizduotas 3 pav. kairėje pusėje.

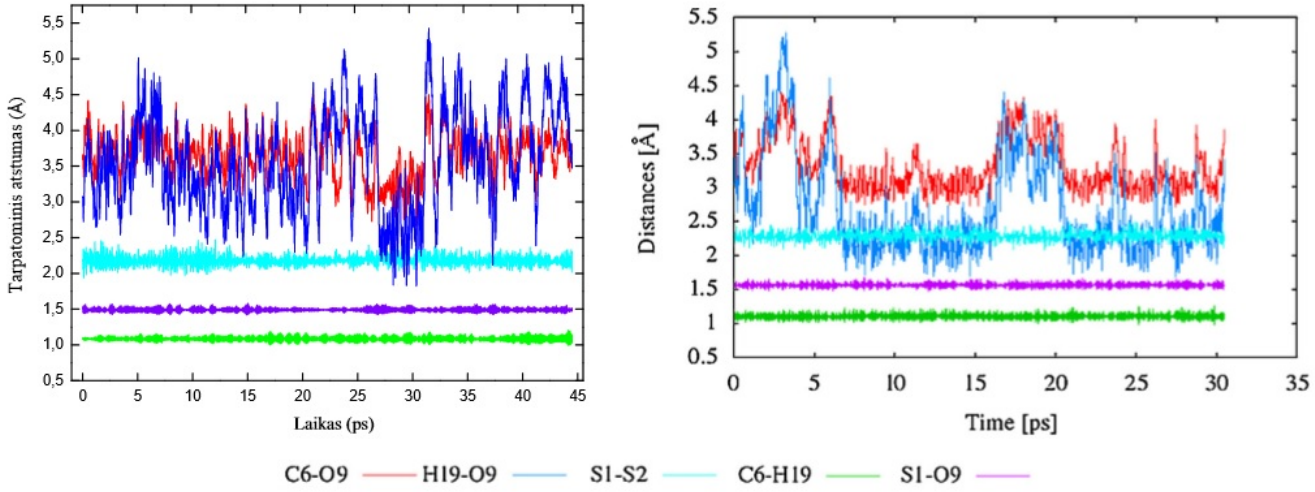
ChemCraft [16] taip pat yra skaitinių uždavinių vizualizacijai skirta programa ir palaiko .xyz plėtinio rinkmenas. Šia programa patogu ruošti modeliuojamos sistemos koordinates, nes vartotojo sąsaja suteikia galimybę manipuliuoti atomais ir molekulėmis 3D erdvėje. Ruošiant dviejų ir trijų alicino molekulių sistemas bei formuojant DNR segmento kartu su alicino molekule struktūras, ChemCraft programa naudota sudaryti koordinates. Jos buvo eksportuojamos .xyz plėtinio byla, o ši naudojama NWChem programų paketu modeliuojamų uždavinių įvesties kode. ChemCraft programos veikiamasis langas pavaizduotas 3 pav. dešinėje pusėje.



3 pav. VMD ir ChemCraft programų veikiamieji langai

3 DNR sekos ir alicino molekulių modeliavimo rezultatai ir jų aptarimas

Pagrindinių studijų kursiniame darbe „Alicino molekulės kompleksų modeliavimas Car ir Parrinello molekulių dinamikos metodu“ [4] buvo modeliuojama viena alicino molekulė ir analizuojama jos struktūra. Pavyzdžiui, taikant straipsnyje apie alicino tyrimą [5] nurodytus parametrus, buvo renkami duomenys apie kintančią alicino struktūrą: 4 pav. kairėje pusėje yra kursinio darbo rezultatai, o dešinėje – straipsnio.



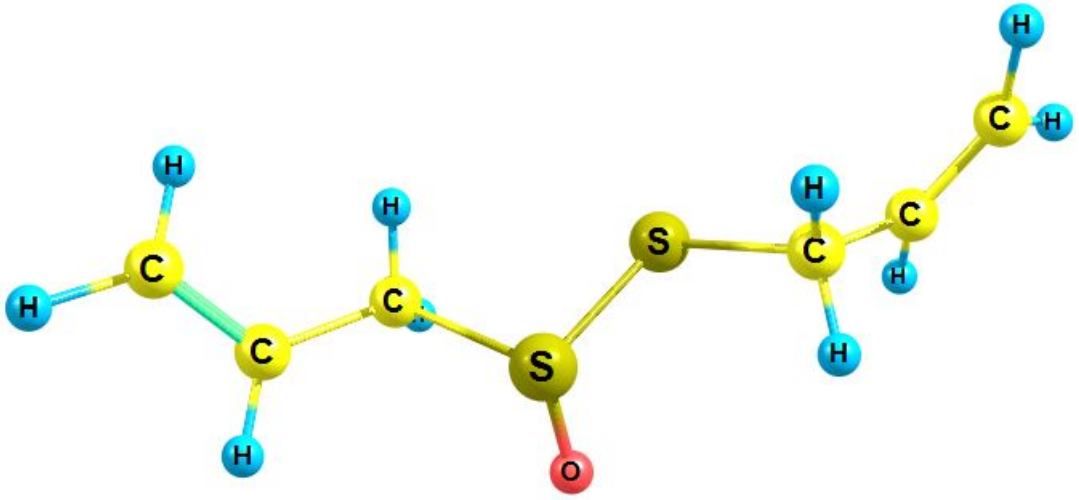
4 pav. Alicino molekulės modeliavimo rezultatų palyginimas

Šiame darbe alicino nagrinėjimas tęsiamas keliais etapais. Pirmiausia Car-Parrinello metodu modeliuojamos alicino molekulių grupės. Tada atskirai tiriamos DNR sekos, padalintos modeliavimui parankiais segmentais. Trečiame etape modeliuojami DNR sekos segmentai kartu su alicino molekulėmis.

Svarbiausi gauti rezultatai ir jų įžvalgos toliau pateikiami atitinkamuose skyreliuose.

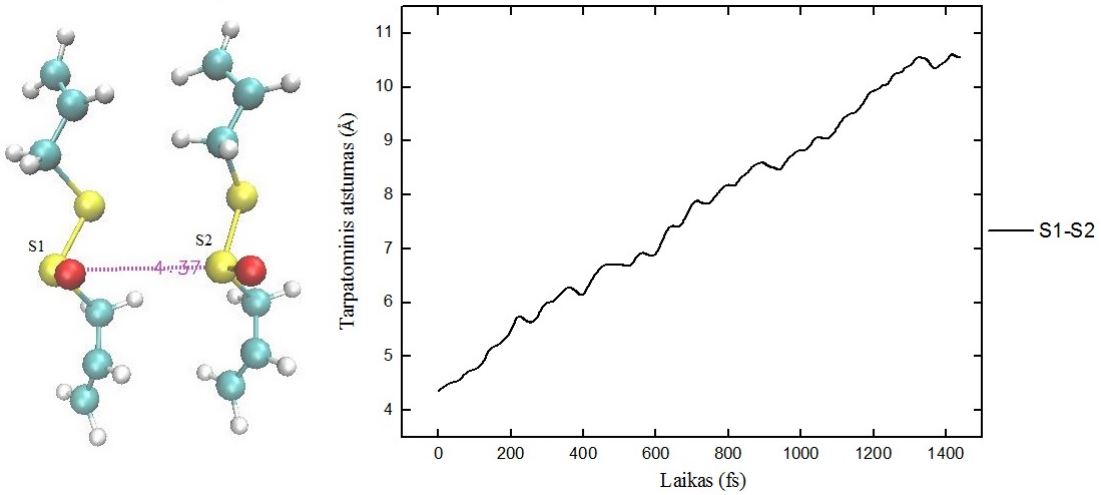
3.1 Kelių alicino molekulių modeliavimas

Dviejų alicino molekulių modeliavimas buvo vykdomas Car-Parrinello metodu, naudojant NWChem programų paketą. Molekulės modeliuotos gardelėje, kurios viena kraštinė yra 35 angstremų ilgio. Nustatyta pastovi 300 K sistemos temperatūra ir modeliuojama neperiodinėmis sąlygomis, kurios neleidžia sistemai plėstis iš už aprašytos gardelės ribų. Pradinė alicino molekulės konfigūracija pasirinkta pagal pagrindinių studijų kursinio darbo [4] rezultatuose pateiktą mažiausios energijos struktūrą (5 pav.).



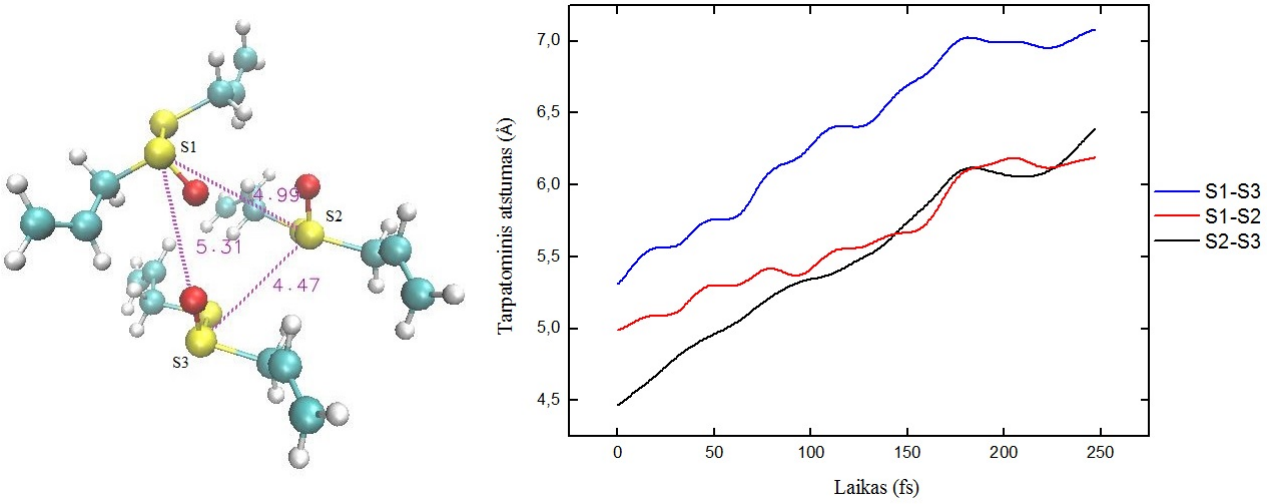
5 pav. Pradinė alicino molekulės struktūra

6 pav. pavaizduota, kaip modeliuojant 1,4 pikosekundes keitėsi nuotolis tarp dviejų alicino molekulių. Atstumas S1-S2 per 1,4 pikosekundes padidėjo nuo 4,3 iki 10,5 angstremų. Dėl šio tiesinio atstumo didėjimo galima teigti, kad kompleksai tarp dviejų alicino molekulių nesusidaro, kai jos yra žemiausios energijos būsenose uždaroje sistemoje.



6 pav. Dviejų alicino molekulių sąveika

Trijų alicino molekulių modeliavimas vyko analogiškomis dviejų molekulių uždaviniui sąlygomis. Šiuo atveju taip pat pastebimas visų tarpmolekulinių atstumų (S1-S2, S1-S3, S2-S3) didėjimas (7 pav.). Todėl tarp trijų alicino molekulių kompleksai nesusidaro, kai molekulės yra žemiausios energijos būsenose uždaroje sistemoje.

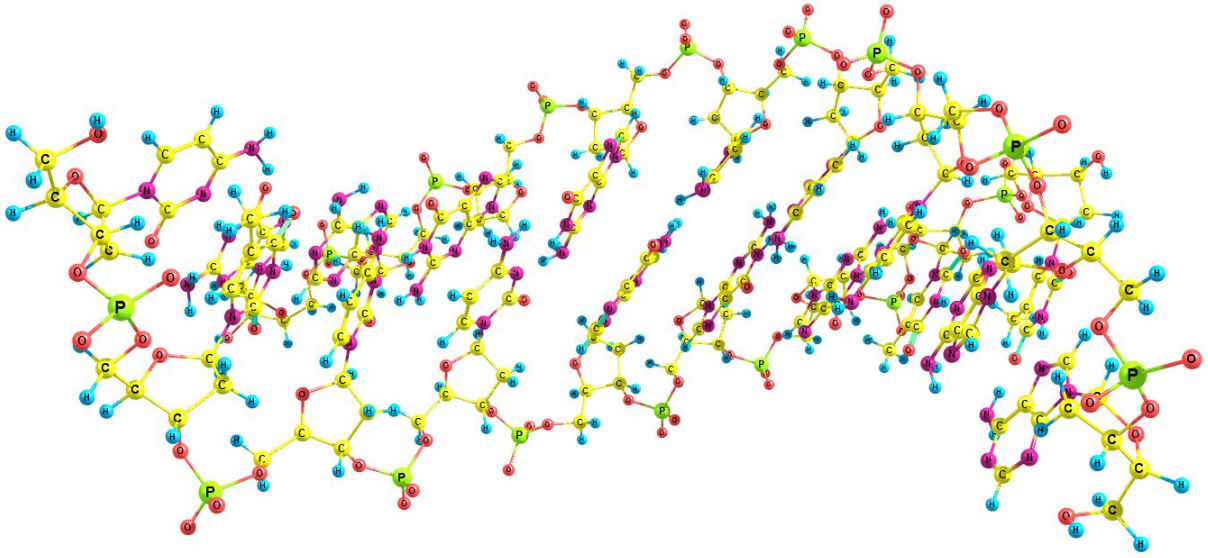


7 pav. Trijų alicino molekulių sąveika

NWChem programų paketas trijų alicino molekulių uždavinį modeliavo 250 femtosekundes. Per šį laiką S1-S2 didėja nuo 5,0 iki 6,2 Å, S1-S3 didėja nuo 5,3 iki 7,1 Å ir S2-S3 didėja nuo 4,5 iki 6,4 Å.

3.2 DNR sekos modeliavimas

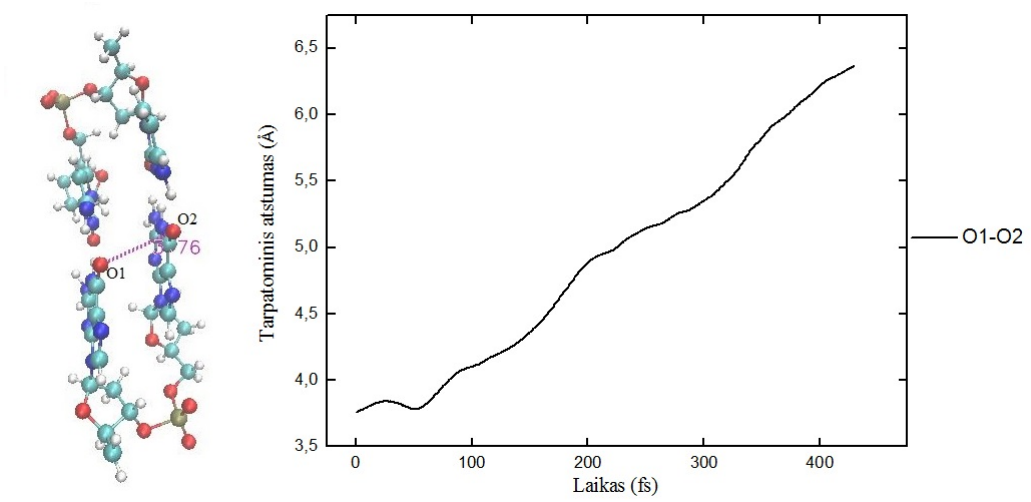
DNR grandinę sudaro nukleotidų bazių poros. Kiekvienas nukleotidas DNR sekoje yra susietas su fosforo rūgščių junginiu (fosfatu), kuris suteikia neigiamą elektros krūvį -1. DNR grandinės modeliavime tai yra labai svarbu, nes nagrinėjant iškirptą DNR sekos segmentą, sistemos elektros krūvis turi būti nurodomas atsižvelgiant į tai, kiek fosfato junginių yra sistemoje.



8 pav. DNR sekos dalis, iš kuriuos buvo pasirenkami segmentai modeliavimo uždaviniuose

Uždaviniuose su DNR gardelės kraštinės ilgis yra 35 angstremai. Šis atstumas optimalus, nes $35 \cdot 35 \cdot 35$ Å^3 tūrio gardelėje užtenka vietos didžiausio tūrio (ilgis 20 Å, aukštis 18 Å, plotis 6 Å) DNR segmentui, suteikiant vietos jam plėstis dėl Car-Parrinello metodu modeliuojamų tarpatominių sąveikų. Kiekvienas uždavinys pradedamas modeliuojant sistemos pusiausvyros būseną, šis procesas užtrunka 72 femtosekundes. Per šį laiką sistemos temperatūra pakeliama nuo 0 iki 300 K. Pradinė DNR sekos konfigūracija (8 pav.), kuri šiame darbe dalinama segmentais, pasirinkta remiantis straipsniu apie nukleazės poveikį DNR sekai [17].

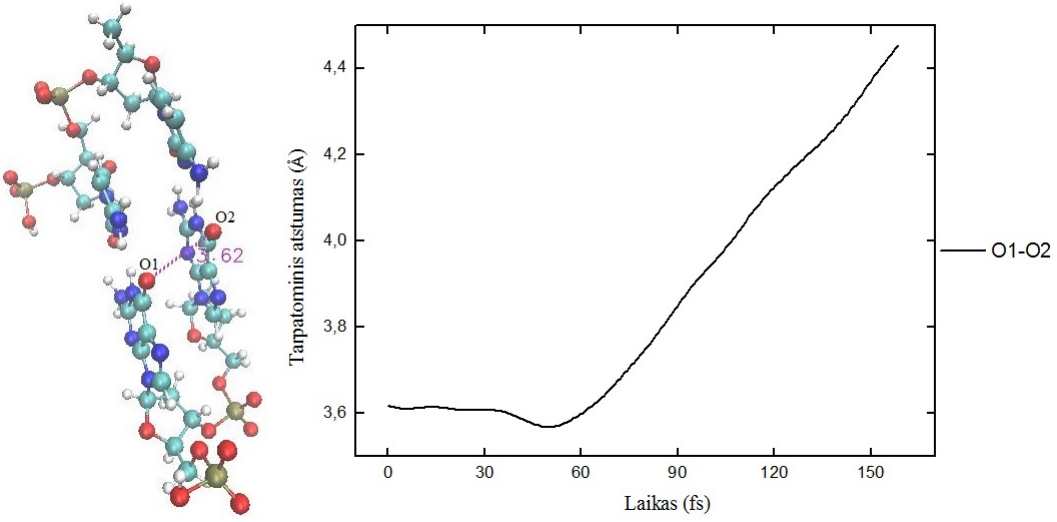
Segmentas, sudarytas iš dviejų bazinių DNR porų, modeliuotas 430 femtosekundžių (9 pav.). Viena iš porų neturi fosforo atomų, todėl šios sistemos elektros krūvis, nurodomas uždavinio įvesties kode, yra -2.



9 pav. Dvi DNR bazinės grupės su dviem fosforo atomais

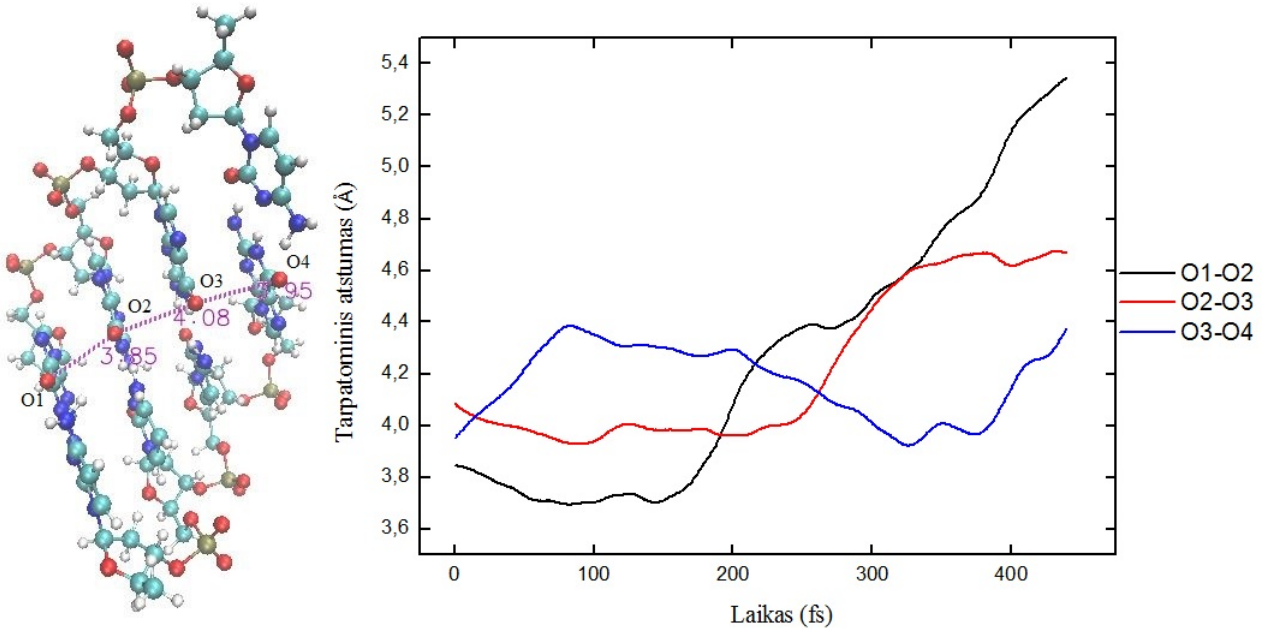
Matuojamas atstumas tarp deguonies atomų O1 ir O2 didėja nuo 3,8 iki 6,4 Å.

Kitas DNR segmentas pasirinktas su pilnom bazinėm porom, dėl to sistema turi du papildomus fosfato junginius, matomus segmento kairėje (10 pav.). Šiuo atveju krūvis, nurodomas programos įvesties kode, yra -4. Šalia esančiame grafike vaizduojamas 160 femtosekundžių modeliuotas atstumas O1-O2. Kitaip nei modeliuojant DNR segmentą su dviem fosfatais, čia per pirmas 48 femtosekundes atstumas nedidėjo, o laikėsi ties 3,6 Å. Likusį laiką bazių poros pradėjo tolti viena nuo kitos iki 4,5 Å.



10 pav. Dvi DNR bazinės grupės su keturiais fosforo atomais

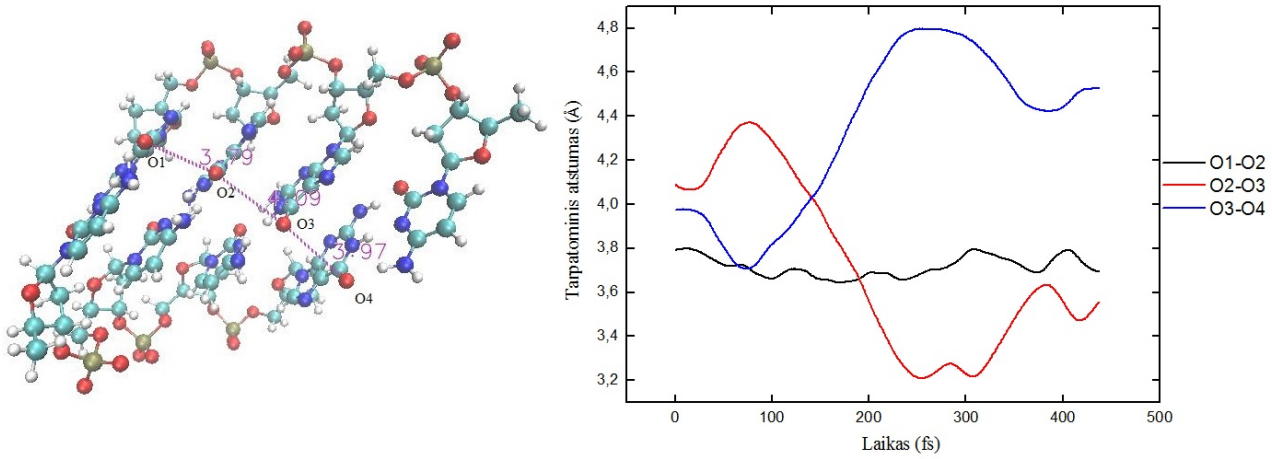
Pasirinktas vienas didesnis DNR segmentas, sudarytas iš keturių bazinių porų ir šešių fosfato junginių (11 pav.). Tai lemia bendrą sistemos elektros krūvį: -6. Uždavinys modeliuotas NWChem programų paketu, taip pat kaip ir alicino bei DNR iš dviejų bazinių porų atvejais. Kadangi čia aprašomas segmentas sudarytas iš 246 atomų, buvo pastebėtas žymiai padidėjęs superkompiuterio skaičiavimo laikas – 440 fs modeliuotos apie savaitę. Prieš tai nagrinėtoms sistemoms užtekdavo perpus mažiau laiko.



11 pav. Keturios DNR bazinės grupės su šešiais fosforo atomais

11 pav. grafike pavaizduota, kaip keitėsi atstumai tarp DNR bazių porų. Segmento kairėje esantis atstumas O1-O2 (juoda kreivė) didėjo sparčiausiai, per 440 fs pakito nuo 3,8 iki 5,3 Å. Centrinė segmento dalis O2-O3 (raudona kreivė) per pirmas 260 fs išliko stabili – nuo 4,0 iki 4,1 Å, bet vėliau pradėjo plėstis ir pasiekė 4,7 Å. Mėlyna kreivė vaizduoja O3-O4 nuotolį, kuris keitėsi trimis etapais: per pirmas 100 fs didėjo nuo 4,0 iki 4,4 Å, tada traukėsi iki 3,9 Å, galiausia vėl plėtėsi iki 4,4 Å.

Iki šiol nagrinėti DNR segmentai NWChem programų paketo interpretuoti kaip molekulės su įvesties kode apibrėžta geometrine konfigūracija. Kad modeliuojamas segmentas iš tikrųjų reprezentuotų DNR sekos dalį, vykdytas dar vienas uždavinys, kuriame aprašyta DNR struktūra tokia pat kaip ir 11 pav., bet segmento kraštuose esančios bazių poros yra neaktyvios modeliavime. Dėl šio nustatymo galima interpretuoti, kad tiriamas segmentas yra toliau besitęsiančios DNR sekos dalis.



12 pav. DNR segmentas su kraštuose neaktyviomis bazių poromis

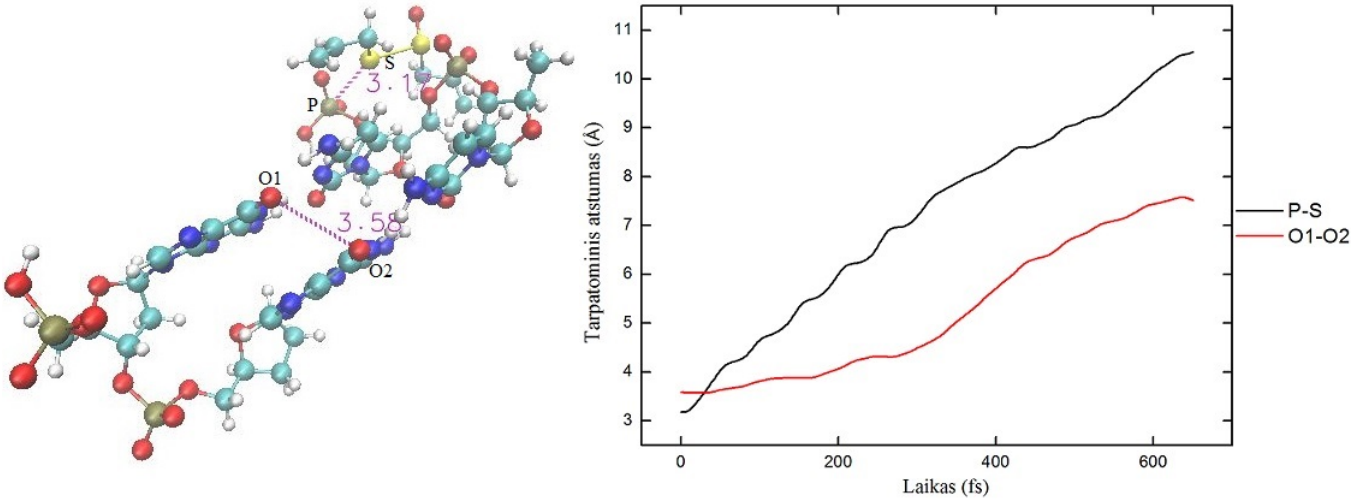
12 pav. grafike galima matyti kaip keitėsi DNR segmento su kraštuose neaktyviomis bazių poromis struktūra. Modeliuota apie 440 fs. Atstumas tarp pirmų dviejų bazių porų O1-O2 (juoda kreivė) išliko pastovus, apie 3,7 Å. Dvi vidurinės poros O2-O3 priartėjo viena prie kitos iki 3,6 Å. Paskutinės dvi poros O3-O4, priešingu žingsniu nei O2-O3, tolo viena nuo kitos iki 4,5 Å. Kadangi kraštuose esančios poros nustatytos kaip neaktyvios, o atstumas O1-O2 išliko pastovus, šiame uždavinyje esanti trečia bazių pora su deguonies atomu O3 pasislinko link bazių poros su atomu O2. Šis poslinkis lemia 12 pav. grafike pavaizduotą kreivių formą.

3.3 DNR sekos kartu su alicino molekule modeliavimas

Alicino molekulė kartu su DNR segmentu sudaro sudėtingą sistemą, kurios modeliavimas NWChem programų paketu, skiriant keturis superkompiuterio „HPC Saulėtekis" studqc eilės procesorius, gali užtrukti daugiau nei porą savaičių. Alicino molekulės struktūra, kuri naudojama šio skyrelio eksperimentuose, atitinka pagrindinių studijų kursinio darbo [4] rezultatuose pateiktą mažiausios energijos struktūrą (5 pav.). DNR sekos struktūra atitinka „Journal of Molecular Biology" aprašytą DNR struktūrą [17].

DNR segmentas, sudarytas iš dviejų bazinių nukleotidų porų, modeliuotas 650 femtosekundžių kartu su alicino molekule, kurios pradinė padėtis – 3,2 Å nuo artimiausio fosforo atomo, kaip pavaizduota 13 pav. Iš šiame paveiksle esančio grafiko matyti, jog visos simuliacijos metu alicino molekulė tolo nuo DNR segmento. Tai vaizduoja juoda kreivė, žyminti atstumą tarp alicino sieros ir DNR fosforo atomų – atstumas didėjo iki 10,6 Å. DNR segmento bazių poros taip pat tolo viena nuo kitos. Atstumas O1-O2 (raudona kreivė) keitėsi nuo 3,6 iki 7,5 Å.

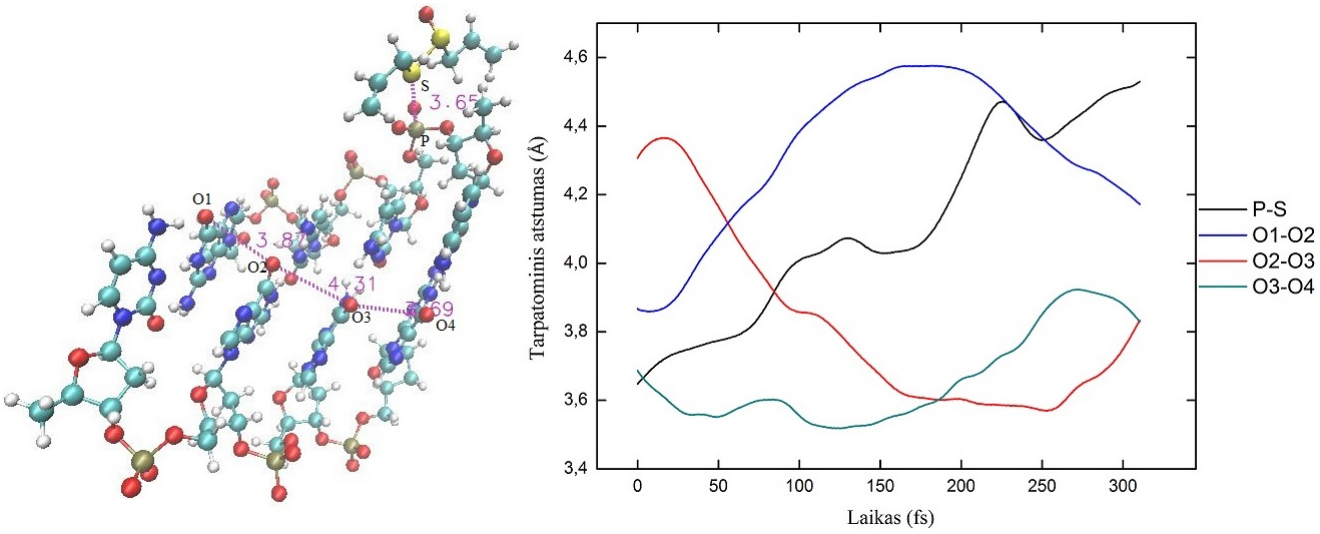
Aprašyti rezultatai atitinka skyreliuose apie alicino bei DNR segmentų modeliavimą aprašytus rezultatų dėsningumus. Visais atvejais nagrinėjami dariniai tolsta vieni nuo kitų ir kompleksai nesusidaro.



13 pav. Dviejų bazinių porų DNR segmentas su alicino molekule

Uždaviniai su alicino molekule ir DNR seka iš keturių bazinių porų vykdyti tokiomis pat sąlygomis (keičiasi tik sistemos elektros krūvis), kaip ir kituose šio darbo rezultatų skyrelyje aprašytuose sistemų modeliavimo uždaviniuose. Pagrindiniai nustatymai: 35 Å gardelė, temperatūra palaikoma apie 300 K, simuliacijoje aktyvių atomų sąrašas. Kadangi DNR seka iš tiesų yra ilgesnė nei modeliuojamas jos segmentas, kraštuose esančios segmento bazinės nukleutidų poros yra nurodomos kaip neaktyvios simuliacijoje. Tai imituoja likusios DNR grandinės poveikį modeliuojamai sistemai.

14 pav. pavaizduota alicino molekulė, esanti virš DNR segmento. Šalia pateiktas grafikas, kuriame matyti, kaip per 310 femtosekundžių keitėsi sistemos tarpatominiai atstumai. Svarbiausias jų – grafike juoda kreive pažymėtas atstumas P-S. Jis rodo, kad alicino molekulė nuo DNR segmento tolo iki 4,5 Å. Pradinėje konfigūracijoje šis atstumas buvo 3,6 Å. Tuo pačiu metu, DNR segmento struktūra keitėsi, kaip vaizduoja mėlyna, raudona ir žalia kreivės. Atstumai kito atitinkamai: O1-O2 nuo 3,9 iki 4,2 Å, O2-O3 nuo 4,3 iki 3,8 Å, O3-O4 nuo 3,7 iki 3,8 Å. Šie rezultatai aprašo DNR segmento dviejų centrinėje dalyje esančių bazių porų (O2-O3) artėjimą.



14 pav. Keturių bazinių porų DNR segmentas su alicino molekule

Išvados

Vykdant modeliavimo uždavinius superkompiuteriu „HPC Saulėtekis” ir naudojant programų paketą NWChem nustatyta, kad kompleksai iš dviejų ir trijų alicino molekulių nesusidaro.

Skaičiuojant DNR grandinės modeliavimo uždavinius Car-Parrinello metodu nustatyta, kad DNR segmentas, sudarytas iš dviejų arba keturių bazinių nukleotidų porų, yra nestabilus – atstumai tarp porų didėja ir segmentas deformuojasi.

Paruošus DNR uždavinio įvesties kodą taip, kad DNR segmento kraštuose esančios bazinės nukleotidų poros būtų fiksuotos, pastebėta, kad DNR segmentas išlieka stabilus, todėl tokią konfigūraciją galima naudoti sprendžiant DNR ir alicino molekulės modeliavimo uždavinius.

Modeliuojant uždavinius su viena alicino molekule ir DNR segmentu iš dviejų bei keturių bazinių nukleotidų porų nustatyta, kad kompleksai nagrinėtomis sąlygomis nesiformuoja.

Literatūros sąrašas

[1] E. Block, The Chemistry of Garlic and Onions, Scientific American **252**, 114–119 (1985).

[2] J. Borlinghaus, F. Albrecht, M. C. H. Gruhlke, I. D. Nwachukwu, A. J. Slusarenko, Allicin: Chemistry and Biological Properties, Chemical Physics Letters **19**, 12591–12618 (2014).

[3] E. D. Tekin, Investigation of Biologically Important Small Molecules: Quantum Chemical and Molecular Dynamics Calculations. In Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Physics. (2010).

[4] M. Lileikis, Alicino molekulės kompleksų modeliavimas Car ir Parrinello molekulių dinamikos metodu (2017).

[5] P. Durlak, S. Berski, Z. Latajka, Theoretical Studies on the Molecular Structure, Conformational Preferences, Topological and Vibrational Analysis of Allicin, Molecules **644**, 5–13 (2016).

[6] M. Mueller, *Fundamentals of Quantum Chemistry* (Kluwer Academic Publishers, 2001).

[7] D. C. Young, *Computational Chemistry: A Practical Guide for Applying Techniques to Real-World Problems* (John Wiley & Sons, 2004).

[8] P. Carloni, F. Alber, *Quantum Medicinal Chemistry* (Wiley-VCH, 2003).

[9] C. J. Cramer, *Essentials of Computational Chemistry* (John Wiley & Sons, 2001).

[10] http://www.nwchem-sw.org/index.php/release66:plane-wave_density_functional_theory#car-parrinello_scheme_for_ab_initio_molecular_dynamics.

[11] http://www.supercomputing.vu.lt/hpc-sauletekis/.

[12] http://www.nwchem-sw.org/index.php/release66:capabilities.

[13] http://cpmd.org/cpmd-tutorial.

[14] http://www.nwchem-sw.org/images/nwchem-new-pmf.pdf.

[15] http://www.ks.uiuc.edu/research/vmd/.

[16] http://www.chemcraftprog.com.

[17] M. Sokolowska, M. Kaus-Drobek, H. Czapinska, G. Tamulaitis, R. Szczepanowski, C. Urbanke, V. Siksnys, M. Bochtler, Monomeric Restriction Endonuclease BcnI in the Apo Form and in an Asymmetric Complex with Target DNA, Journal of Molecular Biology **369**, 722–734 (2007).

Matas Lileikis

ALICINO MOLEKULIŲ DNR SEKAI GALIMO POVEIKIO TYRIMAS CAR-PARRINELLO MOLEKULIŲ DINAMIKOS METODU

Santrauka

Nuo seno žinoma, kad česnakas turi gydomųjų savybių. Jį pjaunant susidaro alicino molekulė, kurios struktūrinės savybės yra aktyviai nagrinėjamos ([4], [5]). Alicino molekulė yra chemiškai lengvai reaguojanti, todėl įdomu tirti, kokiomis sąlygomis ir kokius kompleksus ši molekulė gali sudarinėti.

Šio darbo pagrindinis tikslas yra nagrinėti alicino molekulės poveikį DNR grandinės struktūrai. Remiantis pagrindinių studijų kursinio darbo [4] rezultatais apie alicino modeliavimą, išsikelti uždaviniai yra: modeliuoti dvi ir tris alicino molekules Car-Parrinello molekulių dinamikos metodu, modeliuoti skirtingų struktūrų DNR segmentus, modeliuoti DNR segmento ir alicino molekulės sistemą.

Modeliavimo uždaviniai buvo vykdomi Lietuvos Nacionalinio fizinių technologijų mokslų centro superkompiuteriu „HPC Saulėtekis", NWCHem programų paketu. Ruošiant uždavinių įvesties kodą, buvo remtasi straipsniu „Teorinis tyrimas apie alicino molekulinę struktūrą, topologiją ir virpesius" [5] ir NWChem Car-Parrinello metodo dokumentacija [12].

Šio darbo rezultatai:

1. kompleksai iš dviejų ir trijų alicino molekulių nesusidaro, kai modeliuojama Car-Parrinello metodu, 300 K temperatūros ir 30 Å kraštinės ilgio gardelėje,

2. DNR segmentai iš dviejų ir keturių bazinių nukleotidų porų yra nestabilūs, kai modeliuojama Car-Parrinello metodu, 300 K temperatūros, 35 ir 40 Å kraštinės ilgio paprastose gardelėse,

3. užfiksavus DNR segmento kraštuose esančias bazių poras, pasiekta stabili segmento konfigūracija, kurią galima naudoti modeliuojant DNR ir alicino molekulės sistemą,

4. DNR segmento ir alicino molekulės kompleksai nesusidaro, kai modeliuojama Car-Parrinello metodu, 300 K temperatūros ir 40 Å kraštinės ilgio gardelėje.

Matas Lileikis

MODELING ALLICIN AND DNA INTERACTIONS USING CAR-PARRINELLO MOLECULAR DYNAMICS

Summary

It is known that garlic has healing properties. After being chopped garlic produces allicin molecule, whose structural characteristics are of most interest ([4], [5]). Because allicin is highly reactive: it changes spontaneously into other organic sulfur compounds [1], it is compelling to study complexes which could be influenced by this molecule.

The main objective of this study was to investigate structural changes in systems composed of DNA and allicin molecules. To achieve this, research was done in three stages while referring to the results of previous study "Allicin Complexes Modeling by Car-Parrinello Molecular Dynamics" [4]. The goal in each stage was: to simulate systems of two and three allicin molecules, to simulate different parts of DNA molecule, that consist of different number of nucleobase pairs, to simulate systems composed of allicin and selected parts of DNA molecules.

Computations were performed on resources at the High Performance Computing Center "HPC Sauletekis" in Vilnius University Faculty of Physics. Computational software NWChem was used to perform Car-Parrinello tasks [12]. To prepare input parameters for each problem, article "Theoretical studies on the molecular structure, conformational preferences, topological and vibrational analysis of allicin" [5] was analysed.

The results described in this paper are as follows:

1. complexes of two or three allicin molecules do not form, when simulation is performed at 300 K temperature, in the simple cubic box of dimension 30 Å,

2. systems of two and four DNA base pairs are unstable, when simulation is performed at 300 K temperature, in the simple cubic box of dimension 35 or 40 Å,

3. by restricting movement of DNA base pairs, which are on the side of the segment, stable DNA systems were simulated,

4. molecular binding between DNA and allicin molecules does not occur, when simulation is performed at 300 K temperature, in the simple cubic box of dimension 40 Å.

Priedas Nr. 1.
Įvesties kodo šablonas

```
title "DNR segmento kartu su alicino molekule modeliavimo uždavinys"
echo
scratch_dir ./betaPPdir
permanent_dir ./betaPPdir
memory 90000 mb
charge -6
geometry noautoz
C 65.601000000 19.262000000 38.239000000
C 64.259000000 19.223000000 37.543000000
O 64.047000000 17.937000000 36.921000000
C 63.067000000 19.383000000 38.474000000
O 62.861000000 20.777000000 38.729000000
C 61.942000000 18.783000000 37.652000000
C 62.647000000 17.641000000 36.914000000
N 62.455000000 16.290000000 37.484000000
C 61.182000000 15.713000000 37.429000000
O 60.246000000 16.372000000 36.958000000
N 61.003000000 14.452000000 37.892000000
N 61.820000000 12.528000000 38.820000000
C 62.035000000 13.776000000 38.403000000
C 63.337000000 14.349000000 38.501000000
C 63.500000000 15.599000000 38.037000000
P 62.066000000 21.242000000 40.046000000
O 62.082000000 22.726000000 40.059000000
O 62.596000000 20.498000000 41.217000000
O 60.571000000 20.756000000 39.786000000
C 59.768000000 21.349000000 38.771000000
C 58.417000000 20.678000000 38.735000000
O 58.541000000 19.288000000 38.366000000
C 57.719000000 20.630000000 40.083000000
O 57.117000000 21.891000000 40.363000000
C 56.700000000 19.531000000 39.865000000
C 57.488000000 18.549000000 38.994000000
N 58.090000000 17.415000000 39.719000000
C 57.249000000 16.397000000 40.168000000
O 56.025000000 16.532000000 40.026000000
N 57.785000000 15.301000000 40.755000000
N 59.597000000 14.100000000 41.472000000
C 59.107000000 15.216000000 40.922000000
C 59.987000000 16.269000000 40.524000000
C 59.440000000 17.339000000 39.929000000
```

P 56.848000000 22.321000000 41.882000000
O 56.198000000 23.659000000 41.821000000
O 58.095000000 22.148000000 42.662000000
O 55.799000000 21.244000000 42.405000000
C 54.456000000 21.254000000 41.933000000
C 53.664000000 20.150000000 42.594000000
O 54.085000000 18.839000000 42.152000000
C 53.795000000 20.071000000 44.105000000
O 53.066000000 21.133000000 44.732000000
C 53.218000000 18.694000000 44.368000000
C 53.813000000 17.893000000 43.197000000
N 55.076000000 17.214000000 43.530000000
C 55.025000000 15.950000000 44.122000000
O 53.921000000 15.428000000 44.326000000
N 56.178000000 15.328000000 44.454000000
N 58.468000000 15.272000000 44.558000000
C 57.350000000 15.920000000 44.210000000
C 57.431000000 17.203000000 43.599000000
C 56.281000000 17.808000000 43.277000000
P 53.348000000 21.484000000 46.275000000
O 52.494000000 22.651000000 46.634000000
O 54.816000000 21.554000000 46.498000000
O 52.779000000 20.196000000 47.008000000
C 53.332000000 19.709000000 48.219000000
C 53.053000000 18.230000000 48.320000000
O 53.968000000 17.473000000 47.502000000
C 53.183000000 17.609000000 49.698000000
C 53.233000000 16.137000000 49.334000000
C 54.077000000 16.155000000 48.047000000
N 55.486000000 15.878000000 48.289000000
C 56.551000000 16.743000000 48.195000000
N 57.684000000 16.198000000 48.541000000
C 57.348000000 14.894000000 48.874000000
C 58.153000000 13.826000000 49.351000000
O 59.369000000 13.823000000 49.602000000
N 57.399000000 12.669000000 49.549000000
C 56.046000000 12.558000000 49.330000000
N 55.495000000 11.358000000 49.573000000
N 55.287000000 13.550000000 48.904000000
C 56.001000000 14.679000000 48.700000000
C 58.354000000 3.819000000 53.433000000
C 57.578000000 4.916000000 52.744000000
O 58.219000000 6.194000000 52.900000000
C 57.410000000 4.740000000 51.234000000
O 56.122000000 4.185000000 50.953000000
C 57.540000000 6.154000000 50.674000000

C 57.628000000 7.022000000 51.922000000
N 58.437000000 8.244000000 51.793000000
C 57.899000000 9.317000000 51.086000000
O 56.769000000 9.200000000 50.597000000
N 58.617000000 10.453000000 50.953000000
N 60.503000000 11.686000000 51.348000000
C 59.831000000 10.541000000 51.500000000
C 60.408000000 9.460000000 52.230000000
C 59.683000000 8.339000000 52.350000000
P 55.883000000 3.389000000 49.580000000
O 54.414000000 3.397000000 49.336000000
O 56.605000000 2.090000000 49.667000000
O 56.610000000 4.318000000 48.504000000
C 56.053000000 4.548000000 47.212000000
C 55.046000000 5.676000000 47.260000000
O 55.529000000 6.826000000 47.980000000
C 54.658000000 6.224000000 45.898000000
O 53.715000000 5.344000000 45.286000000
C 54.050000000 7.567000000 46.257000000
C 54.893000000 8.002000000 47.463000000
N 55.943000000 8.971000000 47.168000000
C 57.286000000 8.810000000 47.410000000
N 57.998000000 9.851000000 47.077000000
C 57.073000000 10.756000000 46.576000000
C 57.258000000 12.063000000 46.066000000
O 58.312000000 12.706000000 45.963000000
N 56.052000000 12.628000000 45.658000000
C 54.825000000 12.015000000 45.742000000
N 53.773000000 12.732000000 45.300000000
N 54.640000000 10.793000000 46.223000000
C 55.799000000 10.227000000 46.618000000
P 53.694000000 5.192000000 43.692000000
O 52.867000000 3.996000000 43.376000000
O 55.095000000 5.276000000 43.201000000
O 52.931000000 6.498000000 43.192000000
C 51.555000000 6.719000000 43.502000000
C 51.134000000 8.087000000 43.020000000
O 51.879000000 9.105000000 43.723000000
C 51.423000000 8.354000000 41.549000000
O 50.388000000 7.792000000 40.728000000
C 51.450000000 9.868000000 41.515000000
C 52.169000000 10.180000000 42.828000000
N 53.619000000 10.285000000 42.705000000
C 54.562000000 9.331000000 43.013000000
N 55.787000000 9.734000000 42.801000000
C 55.642000000 11.031000000 42.323000000

C 56.618000000 11.983000000 41.916000000
O 57.848000000 11.868000000 41.897000000
N 56.029000000 13.171000000 41.496000000
C 54.678000000 13.414000000 41.466000000
N 54.298000000 14.621000000 41.017000000
N 53.761000000 12.539000000 41.845000000
C 54.309000000 11.380000000 42.256000000
P 50.586000000 7.711000000 39.137000000
O 49.341000000 7.106000000 38.590000000
O 51.902000000 7.087000000 38.837000000
O 50.634000000 9.244000000 38.704000000
C 51.223000000 9.664000000 37.475000000
C 51.495000000 11.150000000 37.522000000
O 52.383000000 11.449000000 38.622000000
C 52.167000000 11.733000000 36.285000000
C 52.902000000 12.937000000 36.847000000
C 53.298000000 12.477000000 38.252000000
N 54.647000000 11.931000000 38.381000000
C 54.983000000 10.662000000 38.796000000
N 56.270000000 10.456000000 38.827000000
C 56.821000000 11.661000000 38.410000000
C 58.180000000 12.045000000 38.255000000
O 59.202000000 11.377000000 38.469000000
N 58.291000000 13.359000000 37.809000000
C 57.233000000 14.197000000 37.552000000
N 57.548000000 15.435000000 37.140000000
N 55.962000000 13.851000000 37.692000000
C 55.831000000 12.580000000 38.124000000
H 65.600656660 20.136650970 38.855348570
H 65.745526380 18.297455400 38.679074820
H 64.305349690 20.039573380 36.853101730
H 63.166925200 18.918576000 39.432762380
H 61.226432950 18.356792850 38.323722610
H 61.631442680 19.508111410 36.929042470
H 62.212235800 17.597416160 35.937280730
H 62.557462520 12.250600980 39.435791220
H 64.152985350 13.804121920 38.927820560
H 64.462314470 16.062480500 38.100534960
H 59.600090560 22.364084060 39.064786960
H 60.243716350 21.134073250 37.836976210
H 57.860072290 21.265009400 38.034891800
H 58.356660440 20.438606760 40.920650160
H 56.500591380 19.063165350 40.806417510
H 55.894213630 19.927531040 39.283302950
H 56.791276770 18.118733590 38.305271060
H 59.001692190 13.328531800 41.247399430

H 61.041519160 16.206181650 40.694127030
H 60.068593060 18.145451890 39.613712640
H 54.024058620 22.177523850 42.257700400
H 54.496213830 21.019756200 40.889729810
H 52.662693650 20.403125220 42.314333730
H 54.784810500 20.187663620 44.494313220
H 53.633047870 18.320421740 45.280729180
H 52.156390070 18.740937090 44.242775140
H 53.112992030 17.131862570 42.922120650
H 58.282209040 14.687036430 45.347493090
H 58.375550020 17.666510460 43.404363640
H 56.305400410 18.772332630 42.814001950
H 52.798913460 20.181804350 49.017232290
H 54.393445630 19.821388210 48.144153050
H 52.030147760 18.177846120 48.010236770
H 54.015054760 17.921018720 50.294011060
H 53.785130250 15.621255120 50.091667080
H 52.242880200 15.818286660 49.083032680
H 53.713835160 15.393239110 47.389176590
H 56.459363540 17.758722930 47.871256760
H 56.040423950 10.873038690 50.256611910
H 58.592526950 4.179657220 54.411739630
H 57.784014340 2.920782990 53.318010680
H 56.621087340 4.862999530 53.219824660
H 58.128295690 4.074693610 50.802353500
H 56.629249980 6.394549210 50.166485980
H 58.474389970 6.221147600 50.156982990
H 56.644450740 7.374178920 52.153302530
H 61.441203610 11.763679100 51.685253520
H 61.382183590 9.540810610 52.665127540
H 60.084486260 7.506510840 52.889138740
H 55.508230980 3.664119710 46.953383760
H 56.856089150 4.882888180 46.589265920
H 54.210164120 5.212084840 47.740688160
H 55.460365050 6.315764740 45.196081510
H 53.047174350 7.401815850 46.591596640
H 54.240561360 8.246515280 45.452733040
H 54.223922960 8.481178910 48.146830010
H 57.705772380 7.919724010 47.829642490
H 52.856697040 12.331572460 45.306832160
H 51.478542480 6.732070020 44.569184810
H 50.993677940 6.006747940 42.934096350
H 50.078222210 8.114373890 43.191708830
H 52.324659930 7.916964300 41.173621300
H 50.444576400 10.227456540 41.584385720
H 52.074407030 10.178608680 40.703496390

H 51.823392330 11.133473510 43.169091790
H 54.317276670 8.359052560 43.387604670
H 53.453430090 14.896670890 41.476028570
H 52.174250210 9.178731030 37.407642250
H 50.488316210 9.509309570 36.712628300
H 50.522593560 11.587147240 37.612708310
H 52.814466730 11.059652450 35.763185940
H 52.200560240 13.739057390 36.944909110
H 53.797291610 13.075746650 36.277700170
H 53.273350150 13.346135030 38.875615820
H 54.260928330 9.919893140 39.065796060
H 58.507100780 15.699588070 37.039407500
H 60.067675320 12.465160750 50.897000050
H 54.563972360 11.482046680 49.916219990
H 57.880647650 11.855971150 49.876107980
H 53.908684080 13.664465310 44.965189190
H 56.081336010 13.551001720 45.274324070
H 59.177934460 15.935406970 44.794398500
H 55.015966650 15.289723480 41.210216950
H 56.636555100 13.904957180 41.192382370
H 60.515013700 13.927662300 41.114857790
H 56.823581230 16.092749480 36.933643330
H 59.211931600 13.720948440 37.664506710
H 60.946594210 12.484425780 39.305039810
H 51.448863372 11.975663198 35.515718286
H 66.479409344 19.390145626 37.623877473
H 52.378745350 17.907762058 50.353984520
H 59.317833688 3.556802377 53.022296910

C 54.522523391 23.314692217 44.179193545
C 53.736569039 24.501130239 43.751924280
C 53.767443034 24.852503680 42.474650025
H 53.191061254 25.704121152 42.165212317
H 54.289162176 24.271494927 41.726614647
H 53.157863679 25.063066304 44.469901352
S 55.709780790 23.568528670 45.758750082
S 56.857093064 21.767287643 45.616915262
C 58.358077952 22.312799603 46.612514462
C 59.130887731 21.108927108 47.149862869
C 60.380849592 20.691268919 46.898575430
H 60.989392235 21.161857067 46.139373601
H 60.791903138 19.869519105 47.456380580
H 58.514145863 20.568415002 47.856371939
H 58.870935730 22.998783562 45.974407764
H 57.990176851 22.733065605 47.527824310
O 54.901008508 23.351744813 46.953210803

```
H 55.131544783 23.026637856 43.339231261
H 53.905452266 22.453470118 44.416093784
end
set geometry:actlist 16:53 91:137 154:246 247:265
nwpw
simulation_cell units angstroms
boundary_conditions periodic
SC 40
end
cutoff 40
mult 1
lmbfgs
mapping 2
xc revPBE
lcao_skip
end
nwpw
rotation off
end
#Optimize the wave function
task pspw energy
#CPMD parameters
nwpw
pseudopotentials
C TETER c.teter
H TETER h.teter
O TETER o.teter
S TETER s.teter
end
car-parrinello
Nose-Hoover 1200.0 300.0 1200.0 300.0
time_step 3.0
fake_mass 450.0
scaling 1.0 1.0
end
end
#Briefly equilibrate
nwpw
car-parrinello
loop 10 100
fei equi.fei
xyz_filename equi.xyz
ion_motion_filename equi.ion_motion
emotion_filename equi.emotion
end
end
```

```
task pspw car-parrinello
#Metadynamics begins here
nwpw
#This is not necessary - defines separate metadynamics output files
#Otherwise the metadynamics output will be appended to equi.*
car-parrinello
loop 10 2000 #inner-loop outer-loop
fei meta.fei
xyz_filename meta.xyz
ion_motion_filename meta.ion_motion
emotion_filename meta.emotion
end
#Metadynamics block
metadynamics
print_shift 1 #prints the bias potential every 1000 time steps
update 100
bond 16 17
end
end
task pspw car-parrinello
task pspw car-parrinello
task pspw car-parrinello
```