

Gradu amaierako lana
Fisika eta Ingeniaritza Elektronikoko Gradu bikoitza

Solbatazio energia askeen kalkulua transformazio alkimikoen bitartez

Egilea:
Beñat Berasategui Miguéliz
Zuzendaria:
Aitor Bergara Jauregui
Zuzendarikidea:
Aritz Leonardo Liceranzu

Beñat Berasategui Miguéliz, [CC BY 4.0](#) © ⓘ

Leioa, 2024ko ekainaren 21a

Eskerrak Markel García Ibarluzeari lan hau egiteko prozesuan emandako laguntza eta aholku baliagarri guztiengatik. Baita DIPC-ri ere, simulazioak burutzeko *Hyperion* errendimendu handiko ordenagailua erabiltzea ahalbidetzeagatik.

Beñat Berasategui Miguéliz: *Solbatazio energia askeen kalkulua transformazio alkimikoen bitartez*.
Gradu amaierako lana, © 2024ko ekaina.

AURKIBIDEA

1	Sarrera	1
2	Energia askearen kalkulua	3
2.1	Energia askearen kalkulurako metodoak	3
2.1.1	Free Energy Perturbation (FEP)	3
2.1.2	Benett Acceptance Ratio (BAR)	4
2.2	Solbatazio energia askea	5
2.2.1	Ziklo termodinamikoak	5
2.3	Transformazio alkimikoak	6
3	Dinamika molekularra	7
3.1	Force field enpirikoak	7
3.1.1	Gai kobalenteak	7
3.1.2	Gai ez-kobalenteak	10
3.1.3	Modelo osoa	15
3.2	Interakzioen kalkulua praktikan	15
3.2.1	Mugalde baldintza periodikoak	16
3.2.2	Interakzioen mozketak	16
3.2.3	Irismen luzeko interakzio elektrostatisak	17
3.2.4	Salbuespenak	20
3.3	Higidura ekuazioen integrazioa	21
3.3.1	Tenperatura konstanteko simulazioak	21
3.3.2	Presio konstantepeko simulazioak	22
4	Simulazioen garapena	23
4.1	Simulazioen deskribapena	23
4.1.1	NPT multzoaren simulazioa	24
4.2	Transformazio alkimikoen inplementazioa	25
4.2.1	Potentzial alkimikoak	25
4.2.2	Interpolazio eskema	26
4.3	Produktzio simulazioa	26
5	Emaitzak eta analisisa	27
5.1	Metriken azalpena	27
5.1.1	Batez besteko errore absolutua	27
5.1.2	Batez besteko errore koadratikoa	27
5.1.3	Pearson-en korrelazio koefizientea	28
5.1.4	Kendall-en τ koefizientea	28
5.2	Simulazioen emaitzak	28
5.3	Simulazioen konbergentzia	32
6	Ondorioak	33
	Erreferentziak	35
A	Eranskina	37
A.1	Erabilitako molekulen taula	37
A.2	Erabilitako molekulen irudiak	37

AKRONIMOAK

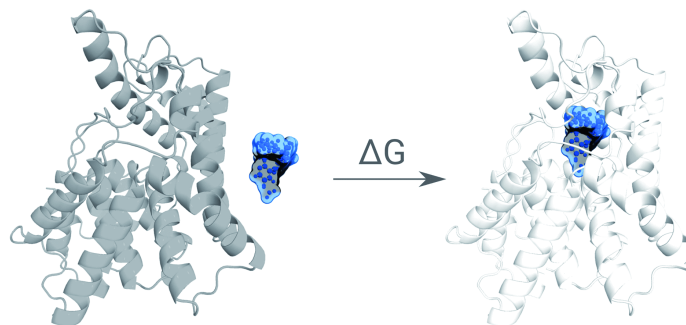
Proiektu honetan jorratzen den gaiari buruz ez da euskarazko bibliografiarik aurkitu. Gauzak horrela, hitz edo kontzeptu gehienentzako itzulpena ahal izan den kontu handienez egin da, baina kasu batzuetan ingelesezko izenak mantentzea erabaki da, irakurleak horiei buruzko informazio gehiago bilatu behar izatekotan errazagoa izan dadin. Bereziki, arlo honetan zabalduko akronimoak dituzten izenen kasuan mantendu da ingelesezko nomenklatura, azpian zerrendatzen direnetako asko tarteko.

FEP	Free Energy Perturbation	3
BAR	Bennett Acceptance Ratio	4
AMBER	Assisted Model Building with Energy Refinement	9
GAFF	General AMBER Force Field	9
MBP	Mugalde Baldintza Periodikoak	16
PME	Particle Mesh Ewald	17
MAE	Mean Absolute Error	27
MSE	Mean Squared Error	27
PCC	Pearson Correlation Coefficient	28

SARRERA

Energia askearen kalkuluak egoera termodinamiko ezberdinen arteko energia diferentziak zehazteko balio duten teknikak biltzen ditu. Energia diferentzia hauek funtsezkoak dira prozesu fisiko eta biokimiko ugari ulertzeko, bereziki farmakoen diseinuan. Kalkulu hauek interakzio molekularrei buruzko iragarpenak egiteko erabil daitezke, hala nola molekula txiki (*ligando*) eta proteina (*hartzaile*) baten lotze-energiaren kasuan [1].

Informazio horrek propietate farmakologiko desiragarriak dituzten molekulak identifikatzeko balio du. Farmakoak diseinatzerakoan, ezinbestekoa da erabiltzen den molekulak hartzaile proteiko jakin batekin era sendo eta selektiboan lotzeko gaitasuna izatea. Identifikazio prozesu hau laborategian egiteko, “saiakera asko, intuizioa eta zortea behar dira” [2]. Horregatik, alternatiba konputazionala onuragarria izan daiteke zenbait kasutan.



1.1. irudia: Proteina-ligando bikotearen lotze-energia funtsezko kantitatea da farmakoen diseinuan. Azken urteetan aurrerapausoak egon dira hori kalkulatzeko metodo konputazionalen arloan [3].

Farmako baten *potentzia*, efektu jakin bat lortzeko behar den farmako kantitatearekin dago erlazionatuta. Zenbat eta dosi txikiagoa beharrezkoa izan, orduan eta potenteagoa dela esaten da. Farmakoaren potentzia lotze-energiaren balioarekin handitzen da. Hau maximizatzeko modu bat molekula jakin baten talde funtzionalak moldatuz hautagai ezberdinak lortzea, aztertzea, eta energia askerik negatiboena duena erabiltzea da. Denbora aldetik zein ekonomikoki hain garestia ez den beste modu bat aukeren arteko aldeak konputazionalki analizatzea da.

Metodo konputazional horien muinean *dinamika molekularra* dago. Aurrerago azalduko denez, denbora tarte batez molekula ezberdinen elkarrekintza simulatzean datza, partikulen ibilbideak bistaratzea eta sistemaren energiaren neurketa ahalbidetuz. Partikulen arteko indarren modelo klasikoa erabiltzen da simulazio hauetan, *force field* bezala ezagutzen dena, eta osagaien higidura ekuazioak zenbakizko metodoen bidez ebatzen dira gero.

Transformazio alkimikoak simulazio horiek egiteko metodo jakin bat dira. Fisikoki esanguratsuak diren bi egoeren arteko energia aske aldea kalkulatzeko tarteko egoera “berezikiak” erabiltzen dituzte, errealitatean ezinezkoak direnak. Esate baterako, molekulen *desagertze-prozesuak* simulatzea ahalbidetzen dute. Energia askea egoera funtzioa delako ematen ditu honek zentzuzko emaitzak, hasiera eta bukaerako egoerak direlako existitu behar diren bakarrak. Teknika tradizionalagoekin alderatuta konbergentzia azkarragoa dutelako dira interesgarriak metodo hauek, koste konputazionala gutxitzen dute eta.

Mota honetako simulazioen emaitzetatik energia askearen balioak lortzeko, analisi estatistikoa egin behar da, estimatzaile ezberdinak erabiliz. Izan ere, dinamika molekularren

bidez sistemaren orekako probabilitate banaketaren laginketa egiten da, eta horien analisi estatistikoa eginez lortzen da energia askearen estimazioa. Beraz, prozesu honetan erabilitako tresnek ere emaitzak baldintzatuko dituzte.

Metodo hauen ospea handituz joan da azkeneko hamarkadan, bai kalkuluak errazagoak egiten dituzten berrikuntzak izan direlako, bai gero eta eskuragarriagoak direlako erabil-tzaile gehiagorentzat, baita konputazio ahalmenaren hazkuntzak sistema mota ezberdinak simulatzea ahalbidetu duelako ere [1].

Dena dela, arretaz inplementatu beharreko teknikak dira, oinarrizko printzipioen ulerme-na eskatzen dutenak. Gainera, zenbait sistema simulatzean kostu konputazionala handiegia da, eta beste batzuetan erabilitako modeloen zehaztasuna ez da behar bezainbestekoa. Zail-tasunak zailtasun, egon diren aurrerapausoak ikusita, baieztatu daiteke etorkizun oparoko arloa dela hau [2].

Aipatutako metodoak aztertzeke, lan honetan ez da 1.1. irudikoa bezalako egoerarik kontsideratu, baizik eta sinpleagoa den *solbatazio energiaren kalkulua* egin da. Hau nahikoa izan da dinamika molekularreko simulazioei buruz ikasteko, eta aztertzen den sistemaren konplexutasuna murriztuz zailtasun gehigarriak saihestea posible izan da.

Helburu hori betetzeko, lehenik eta behin energia askearen kalkuluan funtsezkoak diren bi metodo azalduko dira, eta transformazio alkimikoen bidez solbatazio energia askeak nola kalkulatu. Ondorengo kapituluan, dinamika molekularreko simulazio bat eraikitzeke ezinbesteko osagaiak aztertuko dira ikuspuntu teoriko batetik: *force field*-ak, mugalde baldintzak eta higidura ekuazioen integratzaileak. Horiek denak praktikan nola jarri azalduko da hurrena, simulazioen garapenari buruzko kapituluan. Transformazio alkimikoak inplementatzeko modua ikusiko da bertan. Bukatzeko, aurreko guztia erabiliz egindako simulazioen emaitzak deskribatuko dira, eta haien analisia burutuko da, zenbait metrika erabiliz.

Simulazioetan erabilitako molekulen izenak eta irudiak biltzen dituen A eranskina gehitu da lanaren bukaeran, simulazioen emaitzak aurkezten direnean erabilgarria izan daitekeena. Bestalde, proiektuan zehar erabilitako datuak eta programak eskuragarri jarri dira GitHub-eko errepositorio batean¹: “[Transformazio alkimikoak](https://github.com/BeBerasategi/Transformazio-alkimikoak)”.

¹ Errepositorioaren esteka: <https://github.com/BeBerasategi/Transformazio-alkimikoak>.

ENERGIA ASKEAREN KALKULUA

Magnitude termodinamikorik garrantzitsuena energia askea dela kontsideratzen da askotan. Helmholtz-en funtzioaren (A) edo Gibbs-en funtzioaren (G) bidez adierazi ohi da. Lehenengoa egokia da partikula kopuru, bolumen eta tenperatura (NVT) konstantea denean, eta bigarrena partikula kopuru, presio eta tenperatura (NPT) konstantea denean.

Zoritzarrez, energia askea ezin da era zehatzean kalkulatu dinamika molekular *estandarra* erabiliz, simulazio horiek ez dituztelako behar den moduan lagintzen ekarpen garrantzitsuak egiten dituzten fase espazioko zenbait zonalde. Energia altukoak izaten dira hauek, eta dinamika molekularreko simulazioek energia baxuko eremuak lagintzen dituzte nagusiki. Hortaz, metodo horren bidez lortutako datuak erabiliz kalkulaturako energia askeak zehaztasun txikikoak izango dira.

2.1 ENERGIA ASKEAREN KALKULURAKO METODOAK

2.1.1 Free Energy Perturbation (FEP)

Erlazionaturako problema bat, energia aske *diferentzien* kalkulua da. Ongi definitutako A eta B egoerak hartuko dira. Bien arteko $\Delta G_{A \rightarrow B}$ energia aske aldea honela kalkulatu da:

$$\Delta G_{A \rightarrow B} = -\frac{1}{\beta} \ln \frac{Z_B}{Z_A} = -\frac{1}{\beta} \ln \frac{\int_{\Gamma_B} \exp[-\beta \mathcal{H}_B(\mathbf{r}, \mathbf{p})] d\mathbf{r}d\mathbf{p}}{\int_{\Gamma_A} \exp[-\beta \mathcal{H}_A(\mathbf{r}, \mathbf{p})] d\mathbf{r}d\mathbf{p}}, \quad (2.1)$$

Z_A eta Z_B bi egoeren partizio funtzioak izanik. Hauek lortzeko beharko litzatekeen denbora luzea dela eta, praktikan ez da posible ekuazio hau zuzenean erabiltzea.

Horren ordez, aukera bat Free Energy Perturbation (FEP) teknika erabiltzea da, lan honetan egin diren motako simulazioetan energia aske diferentziak kalkulatzeko metodo hedatuenetakoa [4]. B egoeraren hamiltondarra A egoeraren hamiltondarraren perturbazio moduan adieraztean datza: $\mathcal{H}_B = \mathcal{H}_A + \Delta \mathcal{H}$. Justifika daiteke energia potentziala dela bi egoeren arteko aldeak sortuko dituen, energia zinetikoaren ekarpena bi egoeretan bera dela [4]. Beraz, energia potentziala bakarrik kontsideratzen da:

$$\mathcal{U}_B = \mathcal{U}_A + \Delta \mathcal{U}. \quad (2.2)$$

Hau (2.1) ekuazioan ordezkatuz,

$$\Delta G_{A \rightarrow B} = -\frac{1}{\beta} \ln \frac{\int_{\Gamma_B} \exp[-\beta \mathcal{U}_B(\mathbf{r})] d\mathbf{r}}{\int_{\Gamma_A} \exp[-\beta \mathcal{U}_A(\mathbf{r})] d\mathbf{r}} = -\frac{1}{\beta} \ln \frac{\int_{\Gamma_B} \exp[-\beta \mathcal{U}_A(\mathbf{r})] \exp[-\beta \Delta \mathcal{U}(\mathbf{r})] d\mathbf{r}}{\int_{\Gamma_A} \exp[-\beta \mathcal{U}_A(\mathbf{r})] d\mathbf{r}}, \quad (2.3)$$

non Γ_A eta Γ_B A eta B egoeren konfigurazio guztien fase espazioa diren, momentu linealaren ekarpena kontuan hartu gabe.

Jakina da multzo kanonikoan (NVT) konfigurazio baten probabilitatea ondoko moduan kalkulatu dela:

$$P_A(\mathbf{r}) = \frac{\exp[-\beta \mathcal{U}_A(\mathbf{r})]}{\int_{\Gamma_A} \exp[-\beta \mathcal{U}_A(\mathbf{r})] d\mathbf{r}}. \quad (2.4)$$

Hortaz, fase espazioak bat datozenean, $\Gamma_A = \Gamma_B$ da eta (2.3) ekuazioak itxura hau hartzen du:

$$\Delta G_{A \rightarrow B} = -\frac{1}{\beta} \ln \int_{\Gamma_A} \exp[-\beta \Delta U(\mathbf{r})] P_A(\mathbf{r}) d\mathbf{r} = -\frac{1}{\beta} \ln \langle \exp[-\beta \Delta U(\mathbf{r})] \rangle_A \quad (2.5)$$

Bukaerako A azpiindizeak esan nahi du batezbestekoa A egoerako konfigurazioekin kalkulatu dela. A egoeran simulazio bat egin behar da, bertan $\Delta U(\mathbf{r})$ lagindu, eta horiek erabiliz batezbestekoa kalkulatu.

Analogoa da kalkulua $B \rightarrow A$ noranzkoan eginez prozesuari dagokion adierazpena lortzeko bidea:

$$\Delta G_{A \rightarrow B} = \frac{1}{\beta} \ln \langle \exp[\beta \Delta U(\mathbf{r})] \rangle_B. \quad (2.6)$$

Kasu honetan B -ko konfigurazioekin kalkulatu da batezbestekoa.

Praktikan $\Gamma_A = \Gamma_B$ baldintza ez da betetzen, eta fase espazioen arteko ezberdintasunak handiak direnean estimatzaileen erroreak ere handiak izango dira. Horregatik, zuzenean $A \rightarrow B$ edo $B \rightarrow A$ prozesuei dagozkien energia askearen aldaketak kalkulatu beharrean, bitarteko N egoera erabiltzen dira: $A \rightarrow I_1 \rightarrow I_2 \rightarrow \dots \rightarrow I_N \rightarrow B$. Hauekin,

$$\Delta G_{A \rightarrow B} = \Delta G_{A \rightarrow I_1} + \Delta G_{I_1 \rightarrow I_2} + \Delta G_{I_2 \rightarrow I_3} + \dots + \Delta G_{I_N \rightarrow B}. \quad (2.7)$$

Tarteko N egoerak era egokian aukeratzen badira, $\Gamma_A \sim \Gamma_{I_1} \sim \Gamma_{I_2} \sim \dots \sim \Gamma_{I_N} \sim \Gamma_B$ antzekoak izatea lor daiteke, eta energia askearen kalkulua zehatzagoa izango da.

Orokorrean, N tarteko pausu hartuta, era honetan kalkulatu da bi egoeren arteko energia askeen desberdintasuna:

$$\Delta G_{A \rightarrow B} = \sum_{s \in \mathcal{S}} -\frac{1}{\beta} \ln \langle \exp[-\beta (\mathcal{U}_{s+1}(\mathbf{r}) - \mathcal{U}_s(\mathbf{r}))] \rangle_s, \quad (2.8)$$

$$\Delta G_{A \rightarrow B} = \sum_{s \in \mathcal{S}} \frac{1}{\beta} \ln \langle \exp[\beta (\mathcal{U}_s(\mathbf{r}) - \mathcal{U}_{s-1}(\mathbf{r}))] \rangle_s. \quad (2.9)$$

non $\mathcal{S} = \{A, I_1, I_2, \dots, I_N, B\}$ egoera guztien multzo ordenatua den.

2.1.2 Bennett Acceptance Ratio (BAR)

Energia askearen estimazioa egiteko metodo eraginkorragoa da Bennett Acceptance Ratio (BAR) ekuazioa erabiltzea. Neurketa berdinak erabiliz, estimazio zehatzagoak lor daitezke. Gehiegi sakondu gabe, aurreranzko eta atzeranzko zentzuetan lortutako balioak konbinatzen dira, errore estatistikoa minimizatzeke eran [4].

Funtsezko ekuazioa honakoa da:

$$\exp(-\beta \Delta G) = \frac{\langle \{1 + \exp[\beta(\Delta \mathcal{U} - C)]\}^{-1} \rangle_A}{\langle \{1 + \exp[-\beta(\Delta \mathcal{U}' - C)]\}^{-1} \rangle_B} \exp(-\beta C). \quad (2.10)$$

$A \rightarrow B$ norabidean, $k-1$, k eta $k+1$ egoeretan kalkulatu bada energia, $\Delta \mathcal{U} = \mathcal{U}_{k+1} - \mathcal{U}_k$ eta $\Delta \mathcal{U}' = \mathcal{U}_k - \mathcal{U}_{k-1}$ dira. Bennett-ek aztertu zuenez,

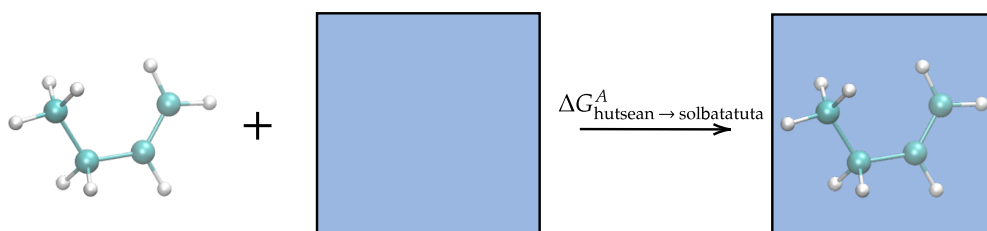
$$C = \Delta G \quad (2.11)$$

denean metodo honek errore estatistiko minimoa ematen du. Beraz hori da C -ren aukerarik eraginkorra.

Simulazioa egin heinean pausu bakoitzari dagozkion saiakera guztiekin lortutako datuak gordetzen dira. Gero, (2.10) eta (2.11) ekuazioak iteratiboki ebatziz askatzen da ΔG .

2.2 SOLBATAZIO ENERGIA ASKEA

Lan honetan kalkulatu diren energia aske diferentziak *solbatazio energia askeak* dira. Beraz, era honetako prozesuak aztertu dira:



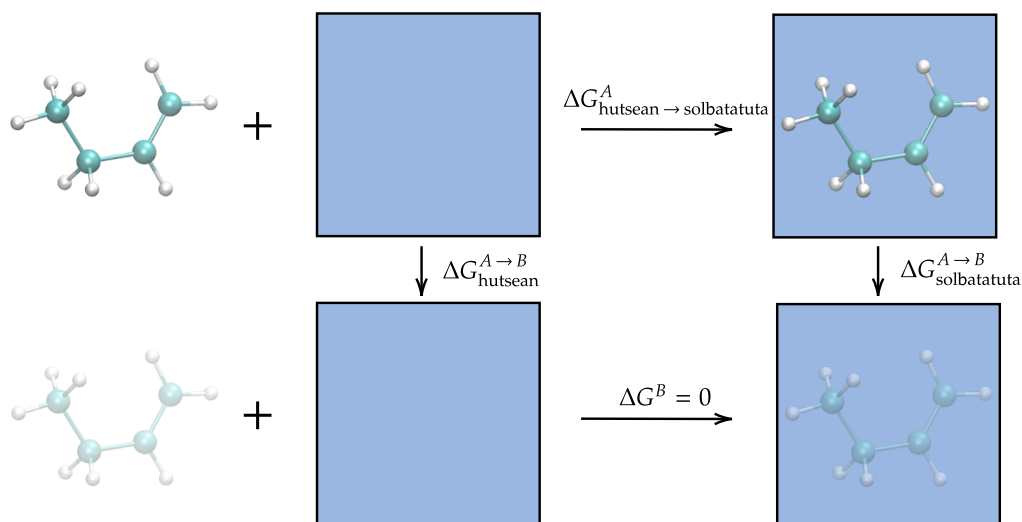
2.1. irudia: Solbatatze prozesuari dagokion energia aske aldaketa adierazten duen eskema.

Dena dela, prozesu honen azterketa konplexua da, hasiera eta bukaerako egoerak lotzen dituen ibilbidea behar bezainbeste lagintzeak koste konputazional handia baitu.

2.2.1 Ziklo termodinamikoak

Energia askea egoera funtzio bat izanik, bi egoeren arteko energia aske aldea kalkulatzeko bestelako bideak erabiltzea posible da, lortutako balioa hasiera eta bukaerako egoeren menpekota izango baita soilik. Hess-en legea izan zen ideia honen aitzindari.

Beste era batean esanda, energia askea egoera funtzioa denez, bere balioak nulua izan behar du ziklo termodinamiko oso bat igarotzean.



2.2. irudia: Solbatazio energia askea kalkulatzeko bide alternatibo bat adierazten duen ziklo termodinamikoak. Energia askea egoera funtzio bat denez, hasiera eta amaierako egoeren menpekotasuna baino ez du, ez haiek lotzen dituen bidearena.

Informazio hau 2.2. irudiko ziklo termodinamikoan baliatzen da. Gezi bertikalek molekula baten *desagertze-prozesua* adierazten dute, hutsean zein uretan. Fisikoki gertaezina

den prozesua bada ere, hasiera eta amaierako egoerak fisikoki esanguratsuak direnez, zentzuzkoa da haien arteko energia aske alde kalkulatzeko.

Prozesuaren behealdeko bi egoerak berdinak dira molekula guztiz desagertu dela kontsideratzen bada, biak urez betetako kutxa soilak baitira. Sistemak inolako aldaketarik jasaten ez duenez, bien arteko $\Delta G = 0$ da.

Aurreko guztia kontuan harturik, solbatazio prozesuari dagokion energia aske aldaketa ondoko eran adieraz daiteke:

$$\Delta G_{\text{hutsen} \rightarrow \text{solbatatuta}}^A = \Delta G_{\text{hutsen}}^{A \rightarrow B} - \Delta G_{\text{solbatatuta}}^{A \rightarrow B}. \quad (2.12)$$

2.3 TRANSFORMAZIO ALKIMIKOAK

2.2. irudiko desagertze prozesuak era zehatzagoan definituko dira hurrena. Dinamika molekularrean sistemaren energia potentzialak adierazpen orokor hau du, hurrengo kapituluan ikusiko denez:

$$\mathcal{U}(\mathbf{r}) = \mathcal{U}_{\text{kobalente}}(\mathbf{r}) + \mathcal{U}_{\text{ez-kobalente}}(\mathbf{r}). \quad (2.13)$$

Bi gai horiek interakzio bikoteak baino ez dituzte kontuan hartzen. Horrek, ondoko deskonposaketa ahalbidetzen du, ur molekulen edo aztertu nahi den molekularekin duten erlazioaren arabera:

$$\begin{aligned} \mathcal{U}(\mathbf{r}) = & \mathcal{U}_{\text{kobalente}}^{\text{ura}}(\mathbf{r}) + \mathcal{U}_{\text{kobalente}}^{\text{molekula}}(\mathbf{r}) + \mathcal{U}_{\text{ez-kobalente}}^{\text{ura-ura}}(\mathbf{r}) \\ & + \mathcal{U}_{\text{ez-kobalente}}^{\text{ura-molekula}}(\mathbf{r}) + \mathcal{U}_{\text{ez-kobalente}}^{\text{molekula-molekula}}(\mathbf{r}). \end{aligned} \quad (2.14)$$

Molekula *desagertzea* horrela interpreta daiteke: sistemako gainontzeko molekulek (urak) aztertu nahi den molekularekin inolako elkarrekintzarik ez dutenean, sistemaren ikuspuntutik molekula hori ez da existitzen.

Desagertze-prozesu hau λ interpolazio-parametro bat erabiliz adieraz daiteke, ondoko eran:

$$\begin{aligned} \mathcal{U}(\mathbf{r}) = & \mathcal{U}_{\text{kobalente}}^{\text{ura}}(\mathbf{r}) + \mathcal{U}_{\text{kobalente}}^{\text{molekula}}(\mathbf{r}) + \mathcal{U}_{\text{ez-kobalente}}^{\text{ura-ura}}(\mathbf{r}) \\ & + (1 - \lambda)\mathcal{U}_{\text{ez-kobalente}}^{\text{ura-molekula}}(\mathbf{r}) + (1 - \lambda)\mathcal{U}_{\text{ez-kobalente}}^{\text{molekula-molekula}}(\mathbf{r}). \end{aligned} \quad (2.15)$$

$\lambda = 0$ denean potentziala (2.14) ekuazioan adierazitakoa izango da, eta $\lambda = 1$ denean berriz, ez da molekularen eta sistemaren arteko elkarrekintzarik egongo eta molekula desagertu dela kontsideratuko da.

$\mathcal{U}_{\text{kobalente}}^{\text{molekula}}(\mathbf{r})$ gaia bere horretan uzten da, gai honen ekarpena berdina izango delako hasiera zein bukaerako egoeretan, eta energia askearen aldaketa kalkulatzeko deuseztatuko da. Gainera, gai hau deuseztatzeotan, molekularen egitura guztiz galduko litzateke, eta ezingo lirarteke konfigurazio errealistak lagindu. Honek konbergentzian eragin ezkorra izango luke.

Azaldu berri dena *transformazio alkimiko* baten adibidea da. Ingurune batetik besterako molekula transferentziari dagokion energia aske aldaketa kalkulatzeko tresnak dira hauek. Bereizten dituen ezaugarria tarteko egoera *alkimikoen* erabilera da, fisikoki existitu ezin daitezkeenak, baina egoera fisikoen arteko *zubi* moduan diharduten energia potentzial funtzioak eraikitzea ahalbidetzen dutenak. Metodo honekin egindako simulazioen konbergentzia azkarragoa delako dira erabilgarriak transformazio alkimikoak.

DINAMIKA MOLEKULARRA

Dinamika molekularra ordenagailu bidezko simulazio teknika bat da. Denbora tarte batez atomo eta molekulen elkarrekintza simulatzean datza, partikulen ibilbideak bistaratzea ahalbidetuz. Horretarako, partikulen arteko indarrak eta energia potentzialak *force field* deitutakoen bitartez modelatzen dira lehenik, eta osagaien higidura ekuazioak zenbakizko metodoen bidez ebazten dira ondoren.

Mekanika molekularra erabiltzen da molekula sistemak mekanika klasikoa erabiliz eredu-tatzeko [5]. Elektroien higidura kontsideratu beharrean, molekulan zehar karga puntual finko batzuk kokatuz hurbiltzen da banaketa elektronikoa¹. Horrela, mekanika kuantikoa erabiltzen duten metodoak baino azkarragoak dira hauek, partikula gutxiago kontsideratzen baitituzte.

Era honetako modeloek atomoen arteko loturak luzatzea, haien arteko angeluak aldatzea edo loturen inguruko biratzea aztertzen dituzte, eta potentzial harmonikoa bezalako funtzio sinpleak erabiliz emaitza onargarriak ekoizteko gai dira.

3.1 FORCE FIELD ENPIRIKOAK

Sistema klasikoki aztertzeaz gain, simulazio hauetan atomoak partikula puntualtzat hartzen dira. Atomo batzuen artean lotura kobalenteak egongo dira, eta horren arabera energia potentzialaren ereduak bi gai nagusi izango ditu:

$$U(\mathbf{r}) = U_{\text{kobalente}}(\mathbf{r}) + U_{\text{ez-kobalente}}(\mathbf{r}) \quad (3.1)$$

Lehenengo gaiak kobalenteki lotutako atomoen arteko elkarrekintzak modelatzen ditu, eta bigarrenak, kobalenteki lotuta ez daudenen artekoak.

3.1.1 Gai kobalenteak

Force fieldaren sofistikazioaren arabera gai gehigarriak ager badaitezke ere, orokorrean hirutan banatzen dira kobalenteki lotutako atomoen interakzioak (ikusi 3.1. irudia):

$$U_{\text{kobalente}}(\mathbf{r}) = U_{\text{loturak}}(\mathbf{r}) + U_{\text{angeluak}}(\mathbf{r}) + U_{\text{tortsioa}}(\mathbf{r}) \quad (3.2)$$

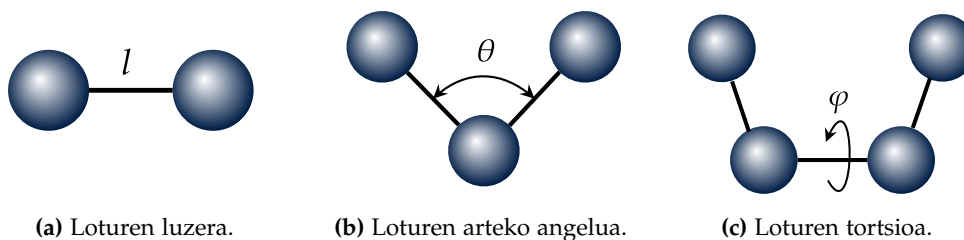
Lehenengo gaiak loturen luzatze edo uzkurtzea kontuan adierazten du, bigarrenak loturen arteko angeluen bibrazioak eta azkenak lotura ezberdinen arteko tortsio-mugimenduak.

3.1.1.1 Loturen luzaketa

Lotura baten energia potentziala modelatzeko eredu ezagunetako bat Morse-ren potentziala da, 3.2a. irudian adierazitakoa:

$$U_{\text{lotura}}(l) = D_e \left[1 - e^{-a(l-l_0)} \right]^2. \quad (3.3)$$

¹ Honen ondorioz, metodo hauek ez dira egokiak molekula baten banaketa elektronikoen araberakoak diren propietateak aztertzeke.



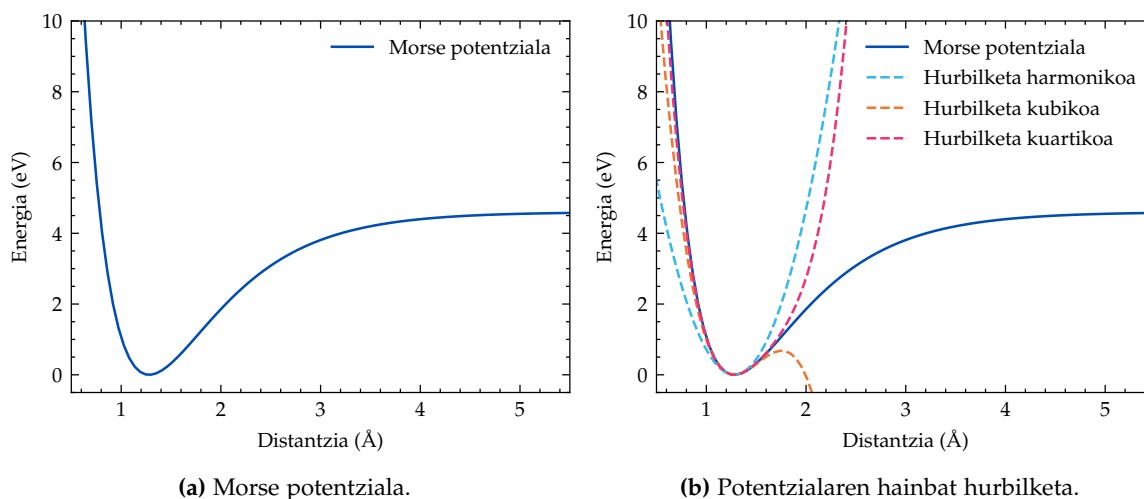
3.1. irudia: Force fieldeko gai kobalenteen hiru ekarpen moten eskema.

D_e energia potentzialaren minimoaren sakonera da, l_0 loturaren erreferentziazko luzera eta $a = \sqrt{k/2D_e}$, non k loturaren indar konstantea den.

Dena dela, eredu hau ez da mekanika molekularreko force fieldetan erabiltzen. Alde batetik, lotura bakoitzeko hiru parametro zehaztea eskatzen duelako, eta bestetik, konputazionalki kalkulu bereziki eraginkorra ez delako. Ondorioz, adierazpen sinpleagoak erabili ohi dira. Lehenengo hurbilketa moduan lotura malguki batekin irudikatzen bada, energia potentzialaren gai hau potentzial harmoniko baten bitartez adieraziko da:

$$\mathcal{U}_{\text{loturak}}(l) = \sum_{\text{loturak}} \frac{1}{2} K_l (l - l_0)^2. \quad (3.4)$$

Energia era koadratikoan aldatzen da l_0 loturaren erreferentziazko luzerarekiko desplazamenduarekin. Loturaren *erreferentziazko luzera* esaten da eta ez *orekako* luzera, kontzeptu ezberdinak direlako. Force fieldeko gainontzeko gai guztiak deuseztatzen direnean loturak duen luzera da erreferentziazko luzera. Orekako luzerak, aldiz, energia minimoko egoeran luzeraren balioa adierazten du, gainontzeko gaien ekarpena kontuan hartuta.



3.2. irudia: H-Cl loturaren energia potentziala modelatzeko hainbat modu. Irudiak egiteko erabilitako balioak hauek izan dira: $D_e = 4.6 \text{ eV}$, $a = 1.4 \text{ \AA}^{-1}$ eta $l_0 = 1.28 \text{ \AA}$.

Morseren potentziala zehatzago modelatzeko, gai kubikoa edo ordena altuagokoak har daitezke. 3.2b. irudiak erakusten duenez, hurbilketa kubikoa maximo lokal bat du erreferentziazko baliotik urruntzerakoan, eta horrek loturen gehiegizko luzatzea bezalako ondorio baztergarriak ekar ditzake. Hurbilketa hauekin ez da lotura kobalenteen apurketa

kontuan hartzen, baina izatez lan honetan aztertutako egoeretan ez dago fenomeno hori simulatzeko beharrik.

Egoera horietan loturen luzerak ez dira orekako baliotik gehiegi urruntzen, eta beraz, Assisted Model Building with Energy Refinement (AMBER) eta General AMBER Force Field (GAFF) force fieldek hurbilketa harmonikoa baino ez dute erabiltzen.

3.1.1.2 Angeluen tolestura

Hurrengo gaia angeluen uzkuertze edo luzatzeari dagokio. Kasu honetan ere egokia da askatasun gradu hau osziladore harmoniko baten bidez adieraztea:

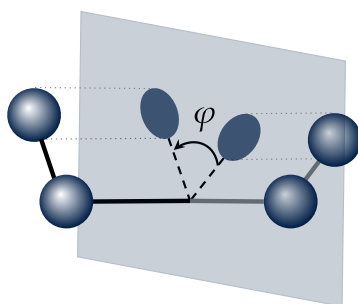
$$U_{\text{angeluak}}(\theta) = \sum_{\text{angeluak}} \frac{1}{2} K_{\theta} (\theta - \theta_0)^2. \quad (3.5)$$

K_{θ} indar konstantea eta θ_0 erreferentziazko balioa, angelu bakoitzaren bereizgarri dira. Oraingoan ere ordena altuagoko gaiak hartuz force fieldaren zehaztasuna hobetzeko aukera dago, baina lan honetan ez da beharrezkoa izan.

Aipagarria da askoz energia gutxiago behar dela loturen arteko angelua aldatzeko, lotura luzatzeko baino. Bi karbono atomoen arteko loturari dagokion indar konstantea $K_l = 317 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ da eta erreferentziazko luzera $l_0 = 1.523 \text{ \AA}$. Lotura 0.2 \AA luzatzeko, $6.34 \text{ kcal mol}^{-1}$ behar dira. Aitzitik, karbonoen arteko lotura bakunen angeluari $K_{\theta} = 0.0099 \text{ kcal mol}^{-1} \text{ deg}^{-2}$ eta $\theta_0 = 109.47^{\circ}$ dagozkio, eta $6.34 \text{ kcal mol}^{-1}$ erabilita 36° -ko aldaketa sortzen da [5].

3.1.1.3 Angeluen tortsioa

Hirugarren gai kobalenteak *angelu diedroen* arabera energia aldaketak kontuan hartzeko balio du. Loturen luzaketa eta angeluen tolesturari dagozkien askatasun graduak "gogorrak" direla esan ohi da, energia handia behar delako erreferentziazko balioetatik mugitzeko. Angelu diedroen biraketa energiak, ordea, izaera periodikoa du, eta ondorioz posible da simulazioetan zehar φ -k balio sorta handiagoa hartzea.

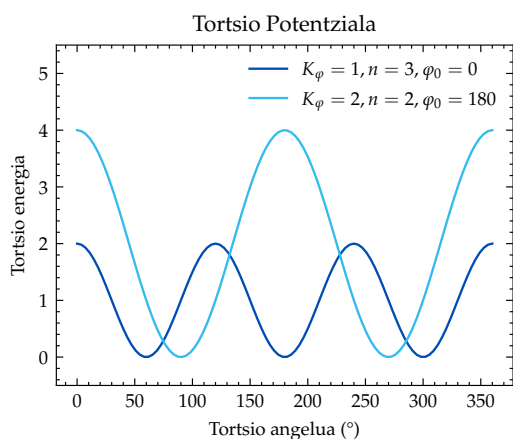


3.3. irudia: Angelu diedroa irudikatzen duen eskema.

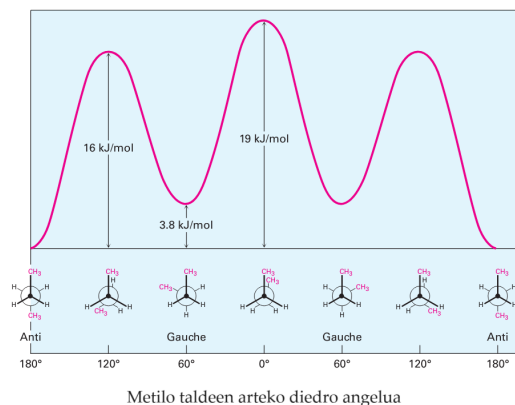
Angeluen tortsioa kontuan hartzen duten force fieldetan, kobalentekei lotutako A-B-C-D erako atomo laukote bakoitzak bere ekarpena egiten du. Etanoaren kasuan esaterako (C_2H_6), horrelako sei ekarpen izango lirarteke. Fourier-en serieak erabiltzen dira potentzial hauek adierazteko:

$$U_{\text{tortsioa}}(\varphi) = \sum_{\text{diedroak}} K_{\varphi} [1 + \cos(n\varphi - \varphi_0)], \quad (3.6)$$

non n anizkoiztasuna eta φ_0 fase faktoreak diren. Kasu honetan n parametroak lotura 360° biratzean funtzioak izan beharreko minimo kopurua baldintzatzen du. Fase faktoreak zehazten du minimoa non dagoen. K_φ potentzial-langaren garaierarekin dago erlazionatuta.



(a) Tortsio potentzialaren bilakaera K_φ , n eta φ_0 -ren balio ezberdinetarako.



(b) Butanoaren isomero konformazionalen energia potentziala, C_2C_3 loturaren tortsioaren arabera [6].

3.4. irudia: Tortsio potentziala adierazteko askotan nahikoa izaten da kosinu bidezko garapeneko gai bakarria erabiltzea. Dena den, zenbait loturen kasuan energia langa ez simetrikoak erabili behar izaten dira, eta horretarako, kosinu gehiagoren ekarpena kontuan hartzen da. Butanoaren (C_4H_{10}) kasuan adibidez, energia maximoa da metilo taldeak lerrokatuta daudenean, metilo talde bat hidrogeno batekin lerrokatuta dagoenean baino handiagoa, aldarapen-indarrak ezberdinak direlako.

3.1.2 Gai ez-kobalenteak

Molekula eta atomo independenteen arteko interakzioa ez-kobalentea da, eta honek ere garrantzia du molekula bakoitzaren egitura zehazterakoan. *Urruneko interakzioak* direla onartuta, distantziaren alderantzizkoaren berretura jakin batekin deskribatu ohi dira. Bi taldetan sailkatzen dira gai hauek: interakzio elektrostatikoak eta van der Waals-en interakzioak.

3.1.2.1 Interakzio elektrostatikoak

Elektronegativotasunak atomo batek lotura bateko elektroiak erakartzeko duen ahalmena neurtzen du [7]. Elementu elektronegatiboek gehiago erakartzen dituzte elektroiak hain elektronegatiboak ez diren beste elementuek baino, eta honen ondorioz karga-banaketa ez uniformeak sortzen dira molekuletan. Lan honetan sistemaren eredu klasikoa erabiliko da, eta beraz, ezezagunak dira molekulen dentsitate elektronikoak, ez baitira interakzio kuantikoak aztertuko.

Karga dentsitatea adierazteko *karga partzialak* erabiliko dira, molekulan zehar kokatutako karga puntualak. Ez dira egoera ionikoaren adierazle, baizik eta molekularen propietate elektrostatikoak erreproduzitzeko modu bat. Era egokian aukeratuta, interakzio elektrostatikoak zehaztasun handiz kalkulatzeko ahalbidetzen dute.

Gauzak horrela, Coulomb-en legea erabil daiteke bi molekulen (edo molekula bereko bi zatiren) arteko elkarrekintzak kalkulatzeko, karga partzial bikoteen interakzioen batura gisa:

$$\mathcal{U}_{\text{Coulomb}} = \sum_i^{N_A} \sum_j^{N_B} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}. \quad (3.7)$$

Metodo honen abantaila nagusietakoa da konputazionalki merkeagoa dela dentsitate elektronikoen elkarrekintza kalkulatzeko baina.

Aipagarria den elkarrekintza elektrostatikoen kalkulurako beste metodo bat *garapen multipolar zentrala* da [8, 9], momentu elektriko edo multipoloak darabilzkiena: karga (q), dipoloa (μ), kuadrupoloa (Θ), oktopoloa (Φ) eta abar. Multipolo bakoitza karga puntual egoki batekin deskriba daiteke, eta askotan nahikoa izaten da momentu elektriko ez-nulu txikienaren ekarpena kontuan hartzea.

Garapen honen eta sistema jakin baten karga banaketaren arteko lotura erakusteko, q_1 eta q_2 kargak baino ez dituen molekula kontsideratuko da (3.5a. irudia). Jatorritik r distantziara dagoen puntuko potentzial elektrostatikoa honakoa da:

$$\phi(r) = \frac{1}{4\pi\epsilon_0} \left(\frac{q_1}{r_1} + \frac{q_2}{r_2} \right). \quad (3.8)$$

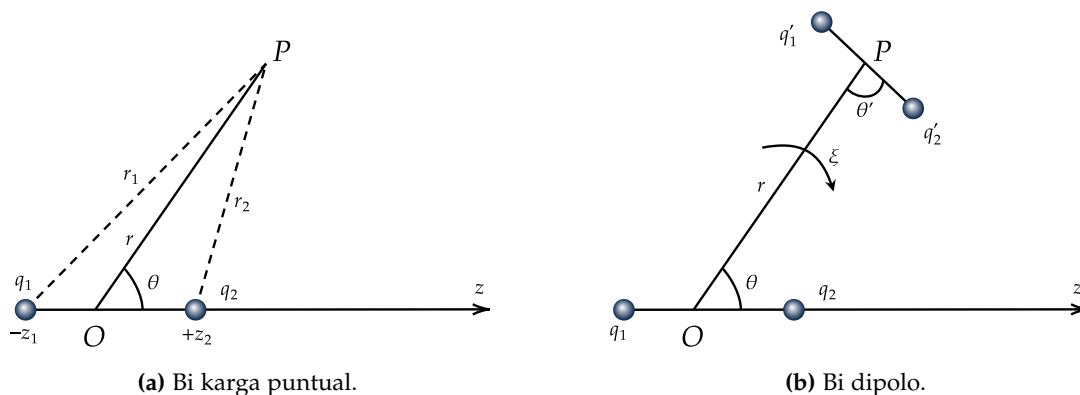
Kosinuaren teorema erabiliz,

$$\begin{aligned} r_1 &= \sqrt{r^2 + z_1^2 + 2rz_1 \cos \theta}, \\ r_2 &= \sqrt{r^2 + z_2^2 - 2rz_2 \cos \theta}. \end{aligned} \quad (3.9)$$

$r \gg z_1$ eta $r \gg z_2$ denean, ondoko moduan garatu daiteke (3.7) ekuazioa:

$$\phi(r) = \frac{1}{4\pi\epsilon_0} \left(\frac{q}{r} + \frac{\mu \cos \theta}{r^2} + \frac{\Theta (3 \cos^2 \theta - 1)}{2r^3} + \dots \right), \quad (3.10)$$

non $q = q_1 + q_2$ karga, $\mu = q_2 z_2 - q_1 z_1$ dipoloa eta $\Theta = q_1 z_1^2 + q_2 z_2^2$ kuadrupoloa definitu diren.



3.5. irudia: Garapen multipolarra aztertzeko bi adibide.

Jarraitzeko, mota bereko bigarren molekula bat kokatuko da, haren masa zentroa P puntuarekin bat datorrelarik (3.5b. irudia). Bi molekulen posizio erlatiboa lau parametrok zehazten dute: r , θ , θ' eta ζ . Bien arteko elkarrekintzaren energia kalkulatzeko, karga

bakoitza puntu horretako potentzialarekin biderkatzen da, eta gero emaitzak batu. Modu honetan lortzen den adierazpena azpikoa da [5]:

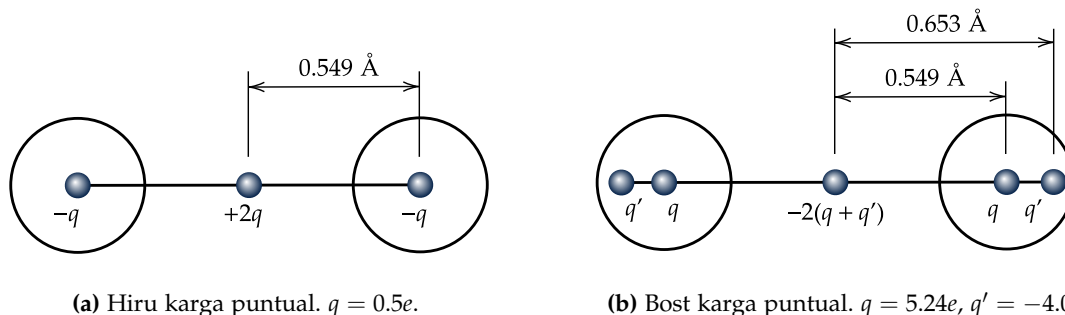
$$\begin{aligned} \mathcal{U}_{Elec} = \frac{1}{4\pi\epsilon_0} \left\{ \frac{qq'}{r} + \frac{1}{r^2} (q\mu' \cos \theta + q'\mu \cos \theta') \right. \\ \left. + \frac{\mu\mu'}{r^3} (2 \cos \theta \cos \theta' + \sin \theta \sin \theta' \cos \xi) \right. \\ \left. + \frac{1}{2r^3} [q\Theta' (3 \cos^2 \theta' - 1) + q'\Theta (3 \cos^2 \theta - 1)] + \dots \right\}. \end{aligned} \quad (3.11)$$

Ikus daitekeenez, bi karga banaketan arteko elkarrekintza energia serie infinitu batek ematen du, zeinetan karga-karga, karga-dipolo, dipolo-dipolo, karga-kuadrupolo... gaiak identifika daitezkeen. Gai bakoitzak r distantziaren alderantzizkoarekiko menpekotasun ezberdina du. Orokorrean, n eta m ordenako bi multipoloren arteko interakzio energia $r^{-(n+m+1)}$ bezala txikitzen da. Nabarmentzeko modukoa da adierazpen hauek baliozgarriak izateko ezinbestekoa dela molekulen arteko r distantzia molekula bakoitzaren tamaina baino askoz handiagoa izatea.

Hortaz, garapen multipolarrak bi molekulen arteko elkarrekintza elektrostatikoa kalkulatzeko modu bat ematen du, eta momentu multipolarrak mekanika kuantikoa erabiliz edo esperimenduetatik lor daitezke. Metodo honen abantaila bat efizientzia da. Esate baterako, bentzenoak (C_6H_6) ez du ez karga ez momentu dipolarrik, baina bai momentu kuadrupolarra. Bi bentzeno molekulen arteko karga-karga elkarrekintza energia kalkulatzeko 144 gai indibidual kontuan hartu beharko liriateke, kuadrupolo-kuadrupolo gai bakarra kalkulatzeko baino askoz ere garestiagoa dena [5].

Zoritxarrez, esan bezala, molekulak oso hurbil daudenean metodo hau ez da erabilgarria. Gainera, konbergentzia mantsokoa izan daiteke. Bestalde, momentu multipolarrak molekula osoaren propietateak direnez, ezin dira interakzio intramolekularrak aztertzeko erabili. Hau guztia dela eta, molekula txikiak eta konformazio finkokoak agertzen diren kalkuluetan bakarrik erabili ohi da metodo hau.

Aurreko arrazoiengatik, karga partzialen metodoa izango da baliatuko dena. Gakoa da karga puntualen banaketa egoki bat eraikita posible dela momentu multipolarrak era zehatzean ordezkatzeko, eta horrela (3.11) ekuazioak emandako elkarrekintza energia karga puntual bikoteen interakzioen batura baten bidez hurbiltzea, (3.7) ekuazioa erabiliz.



3.6. irudia: N_2 molekularen bi modelo, karga partzialak erabiliz. Loturaren luzera 1.098 \AA da, eta esperimentalki neurtutako momentu kuadrupolarra $\Theta = -4.65 \cdot 10^{-40} \text{ C m}^2$ [10]. Sortzen dituzten potentzial elektrostatikoa 3.7a eta 3.7b. irudietan daude irudikatuta.

Ohikoa da karga partzial hauek atomoen nukleoaren posizioetan kokatzea, baina zenbait kasutan beharrezkoa izaten da konfigurazio konplexuagoak erabiltzea, momentu multipo-

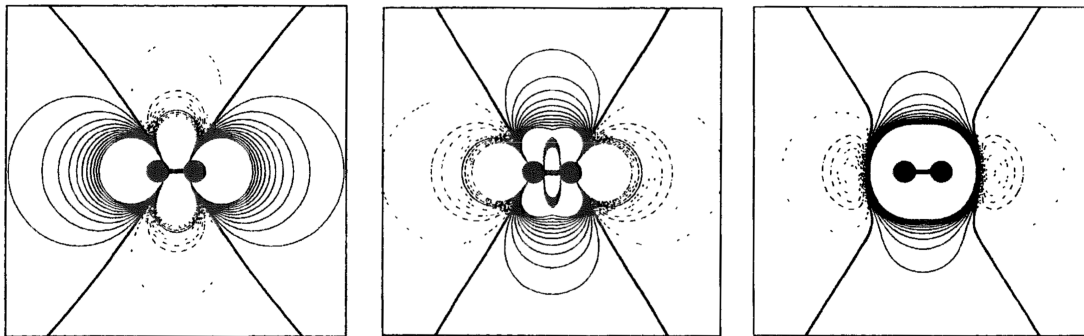
larrak hobeto adierazteko. N_2 nitrogeno molekularra honen adibide ona da (3.6. irudia). Ez du karga neto edo momentu dipolarrik, baina bai bere propietateak baldintzatzen dituen momentu kuadrupolarra. Hau modelatzeko modurik errazena molekularen loturaren norabidean hiru karga lerrokatuz kuadrupolo lineal bat osatzea da, $-q$ karga nukleo bakoitzean eta $+2q$ masa zentroan [9]. Loturaren luzera l baldin bada, masa zentrotik r distantziara eta θ norabideko potentzial elektrostatiokoaren adierazpena

$$\phi(r) = \frac{2q(l/2)^2}{4\pi\epsilon_0} \frac{(3\cos^2\theta - 1)}{2r^3} \quad (r \gg l/2), \quad (3.12)$$

da. (3.10) ekuazioarekin alderatuta, Θ momentu kuadrupolarraren eta q kargaren arteko erlazioa honakoa dela ikus daiteke:

$$\Theta = 2q(l/2)^2. \quad (3.13)$$

Esperimentalki lortutako momentu kuadrupolarrarekin kontsistentea izateko, $q \simeq 0.5e$ izan behar da, e elektroien karga izanik. Izatez, N_2 molekularen inguruko potentzial elektrostatiokoaren deskribapen egokiago bat lortzen da 3.6b. irudian erakusten den 5 kargako banaketa erabiliz.



(a) Hiru karga puntual erabilia (3.6a. irudia).

(b) Bost karga puntual erabilia (3.6b. irudia).

(c) *Ab initio* metodoak erabiliz.

3.7. irudia: N_2 molekularen bi modelo, karga partzialak erabiliz. Alderatzeko, *ab initio* metodoak erabiliz kalkulaturako potentzial elektrostatiokoak. Marra lodiz zero potentzialeko gainazalak adierazten dira, eta marra etenez potentzial negatibokoak [5].

Karga partzialak ez dira esperimentalki behagarriak diren kantitateak, eta beraz, ezin dira era unibokoan lortu uhin-funtzio batetik. Hori dela eta, molekula bera erduztatzeko modu bat baino gehiago dago, aukera zuzena egiteko kontuz aztertu beharrekoak.

3.1.2.2 Van der Waalsen interakzioak

Aurreko atalean aipaturakoez gain, beharrezkoa da van der Waals-en elkarrekintzak ere kontuan hartzea, interakzio ez-kobalenteak beren osotasunean deskribatzeko. Hauen jatorria gertu dauden molekulen polarizazioen fluktuazioa da, eta azalpen zehatz bat emateko ezinbestekoa da mekanika kuantikoa erabiltzea. Lan honetan nahikoa izango da laburki deskribatzea.

Oro har anisotropoak direnez, jarraian emango diren adierazpenak espazioko batezbestekoak dira. Hirutan banatzen dira van der Waalsen indarrak [7, 11]:

- *Keesom-en indarrak.* Bi dipolo iraunkorren artekoak. (3.11) ekuazioaren arabera bi dipolo estatikoren arteko interakzioaren energia $1/r^3$ -ren proportzionala da. Dena

dela, dipoloek biratzeko askatasuna dutenean batezbesteko energia potentzialaren adierazpena ezberdina da:

$$\langle \mathcal{U}_{\text{Keesom}} \rangle = -\frac{2\mu_1^2\mu_2^2}{3(4\pi\epsilon_0)^2 k_B T r^6}. \quad (3.14)$$

- *Debye-ren indarrak.* Dipolo iraunkor batek eta honek inguruko molekula batean induzitutako dipoloaren artekoak. Dagokion adierazpena:

$$\langle \mathcal{U}_{\text{Debye}} \rangle = -\frac{\mu_1^2\alpha_2}{(4\pi\epsilon_0)^2 r^6}, \quad (3.15)$$

non α_2 bigarren molekularen polarizagarritasuna den.

- *London-en indarrak.* Molekula bateko elektroien fluktuazioen ondorioz sortutako aldiuneko dipolo baten eta honek ondoko molekula batean induzitutako dipoloaren artekoak. Adierazpen honekin deskribatzen dira:

$$\langle \mathcal{U}_{\text{London}} \rangle = -\frac{3\alpha_1\alpha_2}{2(4\pi\epsilon_0)^2} \frac{I_1 I_2}{I_1 + I_2} \frac{1}{r^6}, \quad (3.16)$$

α_i eta I_i molekula bakoitzaren polarizagarritasuna eta ionizazio-energia izanik.

3.1. taula: Elkarrekintza intermolekularren distantziarekiko menpekotasuna [7].

ELKARREKINTZA MOTA	r -REKIKO MENPEKOTASUNA
Ioi–ioi	$1/r$
Ioi–dipolo	$1/r^2$
Dipolo–dipolo	$1/r^3$
Keesom	$1/r^6$
Debye	$1/r^6$
London	$1/r^6$

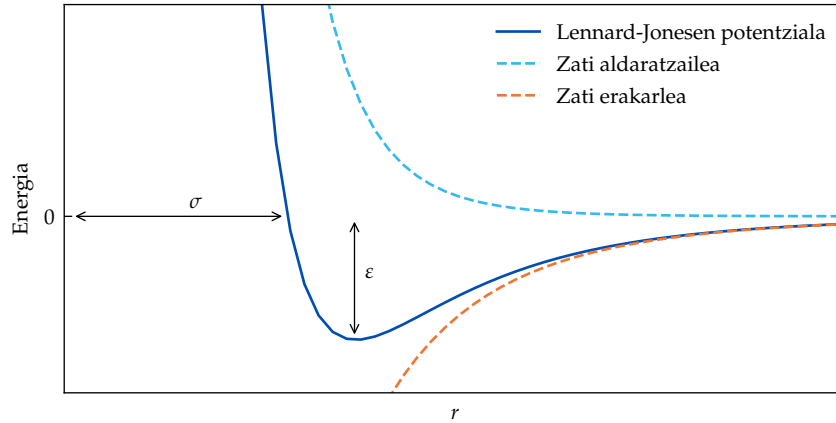
Zerrendatutako hiru ekarpenak erakarleak direla ikus daiteke, baina horiez gain badira aldarapen indarrak ere. Molekulak asko hurbiltzen direnean gertatzen da hau, eta zenbat eta gehiago hurbildu sendoagoa da indar hori. Orbitalak gainezartzen diren eremuetan elektroien arteko aldarapen coulombdar zein Pauliren baztertze printzipioarekin lotu daiteke fenomeno hau.

Praktikan, ekarpen hauek guztiak force fieledean sartzeko, energia potentzialaren irudikapen sinplifikatu bat erabiltzen da. Aipatutako gai bakoitzaren xehetasunak arbuiatzen dira, eta ezaugarri orokorrak parametro gutxi batzuen bidez doitzen dira. Testuinguru honetan erabilienetako bat Lennard-Jones-en potentziala da:

$$\mathcal{U}_{\text{Lennard-Jones}} = 4\epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right], \quad (3.17)$$

non ϵ energia potentzialaren balio minimoa den eta σ -k potentziala nulua izateko r -k hartu beharreko balioa. Potentzial putzuaren minimoa $r = 2^{1/6}\sigma$ balioan dago. 3.8. irudian ikus daitekeenez, potentzial honek r^{-12} -ren proportzionala den zati aldaratzaile bat du eta r^{-6} -ren proportzionala den beste zati erakarle bat.

Zati erakarlearen distantziarekiko menpekotasuna van der Waalsen elkarrekintzak deskribatzerakoan ikusitakoarekin (3.1. taula) bateragarritasuna bermatzeko aukeraten da, baina ez dago zati aldaratzailearen forma justifikatzen duen argudio teorikorik. Izatez, uhin-funtzio atomikoen joera aztertuta, e^{-ar} erako esponentzial bat egokiagoa izango litza-teke. Dena dela, helburuetako bat kalkulu azkarrak ahalbidetzen dituen forma bat izatea da, asko baitira ohiko simulazioetan kalkulatu beharreko van der Waalsen interakzioak. r^{-12} gaia azkar lor daiteke r^{-6} gaiaren karratua eginez, eta ematen duen zehaztasun maila nahikoa denez, (3.17) ekuazioa da gehien erabiltzen dena.



3.8. irudia: Lennard-Jonesen potentziala. Lerro etenez adierazi dira bi ekarpenak nagusiak.

Aipagarriak dira bi molekula ezberdinen arteko interakzioa kalkulatzeko erabiltzen diren Lorentz-Berthelot-en *konbinazio arauak*:

$$\sigma_{ij} = \frac{\sigma_{ii} + \sigma_{jj}}{2}, \quad \varepsilon_{ij} = \sqrt{\varepsilon_{ii} \cdot \varepsilon_{jj}}. \quad (3.18)$$

3.1.3 Modelo osoa

Ekarpen guztiak aztertu ondoren, horrela berridatz daiteke atal honen hasieran aipatutako (3.1) ekuazioa, batugai bakoitza esplizituki adieraziz:

$$\begin{aligned} \mathcal{U}(\mathbf{r}) = & \sum_{\text{loturak}} \frac{1}{2} K_l (l - l_0)^2 + \sum_{\text{angeluak}} \frac{1}{2} K_\theta (\theta - \theta_0)^2 + \sum_{\text{diedroak}} K_\varphi [1 + \cos(n\varphi - \varphi_0)] \\ & + \sum_i^N \sum_j^N \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} + \sum_i^N \sum_j^N 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right]. \end{aligned} \quad (3.19)$$

3.2 INTERAKZIOEN KALKULUA PRAKTIKAN

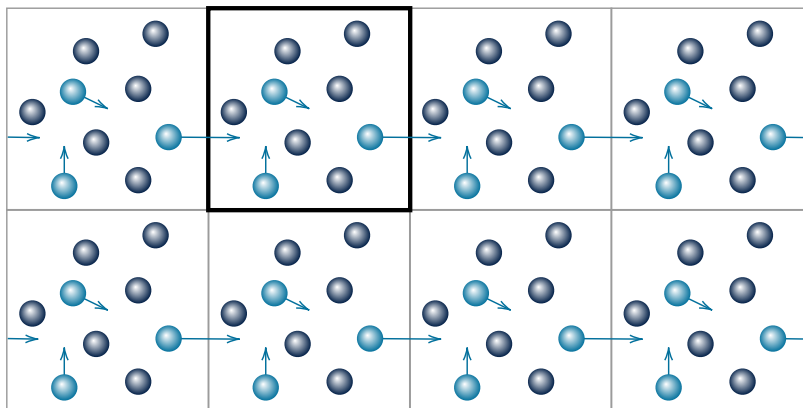
Force field egoki bat erabiltzeaz gain, badira praktikan kontuan hartu beharreko beste zenbait alderdi, hala nola erabilitako mugalde baldintzen egokitasuna eta irismen luzeko interakzioen tratamendua. Ondoko ataletan azalduko dira hauek.

3.2.1 Mugalde baldintza periodikoak

Alde batetik, atomo kopurua ahalik eta txikiena izatea komeni da, simulazioak konputazionalki efizienteak izan daitezten. Bestetik, simulatzen ari den sistema aztertzeke, laborategian haren ezaugarriak neurtzean izango liratekeen baldintzak ahalik eta hobekien imitatu behar dira, baina esperimentalki erabiltzen diren laginen atomo kopurua ordenagailuz simula daitekeena baino askoz handiagoa da.

Atomo gutxiegia erabilia, zenbait efektu galtzen dira. Hauen artean garrantzitsuenetakoa, *irismen luzeko interakzio elektrostatisak* dira. Interakzio elektrostatisaren magnitudea distantziarekin azkar gutxitzen bada ere, atomo askoko sistemetan irismen luzeko interakzio hauen eragina nabarmena izan daiteke. Bestalde, tamaina txikiko simulazioetan ertzetako atomoek jasaten dituzten interakzioak barrualdekoek jasaten dituzten ezberdinak dira.

Araza hauek konpontzeko Mugalde Baldintza Periodikoak (MBP) erabiltzen dira: simulazioko atomoak tamaina finkoko "kutxa" batean sartzen dira, eta ondoren, kristal sare bateko gelaxka primitiboa izango balitz bezala, jatorrizko kutxa horren kopia ugari egiten dira.



3.9. irudia: Mugalde baldintza periodikoak bi dimentsiotan.

Simulazioan zehar kutxaren alde batetik atomo bat ateratzen bada, atomo horren kopia bat sartuko da kontrako aldetik, kutxako atomo kopurua simulazioan zehar konstante mantenduz. Era honetan, jatorrizko kutxaren ertzetan dauden atomoek ere inguruko kutxetako atomoekin izango dute elkarrekintza, eta oro har, simulazioak hobeto hurbilduko du laborategiko egoera.

3.2.2 Interakzioen mozketak

N atomoko sisteman, interakzio ez-kobalenteak kalkulatzen diren aldiro, beharrezkoa da partikula bikote bakoitzeko indarrak kalkulatzeko. Hau da, N^2 indar kalkulatu behar dira. Gainera, MBPak erabilia, jatorrizko kutxako atomoen arteko elkarrekintzez gain, kutxaren beste kopia guztien artekoak ere kalkulatu beharko lirateke, teorian.

Hau praktikan egitea ez da posible, ezta beharrezkoa ere. Izan ere, kontuan hartu diren interakzioak azkar desagertzen dira distantziarekin, eta beraz, simulazioak azkartzeko

ohiko teknika da *interakzioen mozket*a. Distantzia jakin batetik aurrera interakzioak zerotzat hartzen dira:

$$\mathcal{U}_{\text{ez-kobalente}}(r) = \begin{cases} \sum_i^N \sum_j^N \frac{Q_i Q_j}{4\pi\epsilon_0 r_{ij}} + \sum_i^N \sum_j^N 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right], & r \leq r_{\text{moztu}} \\ 0, & r > r_{\text{moztu}} \end{cases} \quad (3.20)$$

Praktikan atomo batek bere ingurukoekin dituen interakzioak kontsideratuko dira bakarrik. r_{moztu} delakoa programa idazterakoan zehazten da. Lan honetan 0.9 nm erabili da.

3.2.3 Irismen luzeko interakzio elektrostatikoak

Lehenago aipatutako moduan, irismen luzeko interakzioen eragina ez da arbuigarria atomo askoko sistemetan; karga-karga interakzioa bereziki arazotsua izan ohi da. Ondorioz, aurreko ataleko interakzioen mozketa nolabait zuzendu behar da hauek kontuan har ditzan.

Helburu horrekin erabilitako metodoak Particle Mesh Ewald (PME) du izena [12]. Honen oinarria *Ewald batura* da, jarraian azalduko dena.

3.2.3.1 Ewald batura

Teknika hau P. P. Ewald-ek erabili zuen aurreneko aldiz 1921an, kristal ionikoen energiak aztertzeko, baina dinamika molekularreko simulazioetan ere baliagarria da. MBPak azaltzerakoan aipatu denez, simulazio kutxa sare kristalino baten gelaxka primitibotzat har daiteke. Bertako N karga partzialek irudi periodikoekin batera \mathbf{r} posizioan sortutako potentziala ondoko adierazpenak ematen du:

$$\phi(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \sum_{\mathbf{n}} \sum_{j=1}^N \frac{Q_j}{|\mathbf{r} - \mathbf{r}_j + \mathbf{n}|}. \quad (3.21)$$

Adierazpen horretan $\mathbf{n} = L \cdot (n_x, n_y, n_z)$ da, L gelaxka primitiboaren aldea izanik. Hau erabiliz energia potentziala adierazten da, sareko interakzio bikote guztien batura gisa:

$$\mathcal{U} = \frac{1}{4\pi\epsilon_0} \frac{1}{2} \sum_{\mathbf{n}=0}^N \sum_{j=1}^N \sum_{i=1}^N \frac{Q_i Q_j}{|\mathbf{r}_{ij} + \mathbf{n}|}, \quad (3.22)$$

non $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$ den. $1/2$ gaiak bitan zenbatzen diren elkarrekintzak konpentsatzen ditu, eta $\mathbf{n} = 0$ denean $i = j$ gaia baturan sartzen ez dela adierazten du batukariko primak (').

(3.22) ekuazioan ikusten denez, energia osoak gelaxka primitiboko interakzioen zein gelaxka primitiboko eta gainontzeko irudien arteko interakzioen ekarpenak kontuan hartzen ditu. Honek duen arazoa da baturaren konbergentzia oso mantsoa dela, eta dibergentea izan daiteke zenbait baldintzatan, gai positiboak edo negatiboak bakarrik hartuta serie dibergenteak lortzen baitira.

Ewald batura aurreko seriea bitan banatzean datza, bi serie berrien konbergentzia azkarragoa izateko moduan. Ondoko identitatea kontuan hartuta,

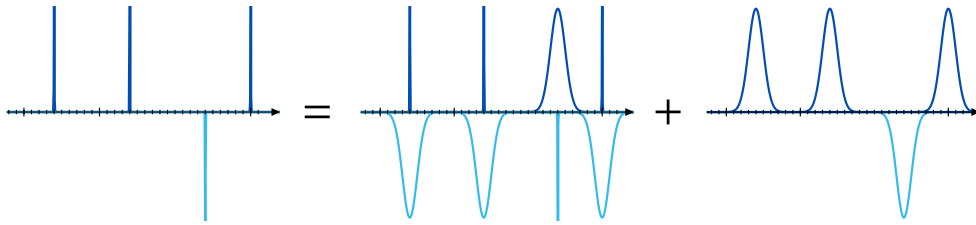
$$\frac{1}{r} = \frac{f(r)}{r} + \frac{1-f(r)}{r}, \quad (3.23)$$

helburua $f(r)$ egoki bat hautatzea da, $1/r$ -ren aldaketak (txikiak r handia denean, handiak r txikia denean) kontuan hartzeko. Energia ere bitan banatzen da, $\mathcal{U} = \mathcal{U}^{\text{labur}} + \mathcal{U}^{\text{luzer}}$, lehenengo gaien irismen laburreko interakzioak sartzen direlarik eta bigarreanean luzekoak. Irismen laburreko gaien batura espazio zuzenean egiten da, eta irismen luzekoen gaiena elkarrekiko espazioan, Fourierren transformatua erabiliz.

Gai hauek kalkulatzeko, karga puntualak Dirac-en orrazi baten bidez adierazten dira:

$$\rho(r) = \sum_{i=1}^N Q_i \delta(\mathbf{r} - \mathbf{r}_i) \quad (3.24)$$

Karga puntual bakoitzaren inguruan zentratutako funtzio gaussdarren banaketa bat gehitzen da hurrena, magnitude bereko baino aurkako zeinuko gaiak dituen, 3.10. irudian bezala. Bukatzeko, funtzio gaussdarren beste banaketa bat batzen da, aurrekoaren efektua neutralizatzen duena, osotara jatorrizko karga puntualen banaketaren eragina bakarrik izateko.



3.10. irudia: Ewald baturan jatorrizko karga banaketa funtzio gaussdarren banaketa batekin inguratzen da (espazio zuzenean kalkulatu). Hori neutralizatzen duen beste karga banaketa bat gehitzen da (elkarrekiko espazioan kalkulatu dena).

Era honetan, \mathbf{r}_i puntuko karga banaketa bi zatitan dago banatuta, $\rho_i(r) = \rho_i^{\text{labur}}(\mathbf{r}) + \rho_i^{\text{luzer}}(\mathbf{r})$. Ohiko funtzio gaussdarra $G_\sigma(\mathbf{r})$ ikurrak adierazten bada,

$$G_\sigma(\mathbf{r}) = \frac{1}{(2\pi\sigma^2)^{3/2}} \exp\left[-\frac{|\mathbf{r}|^2}{2\sigma^2}\right], \quad (3.25)$$

itxura hau hartzen dute bi batugaiek:

$$\begin{aligned} \rho_i^{\text{labur}}(r) &= Q_i \delta(\mathbf{r} - \mathbf{r}_i) - Q_i G_\sigma(\mathbf{r} - \mathbf{r}_i), \\ \rho_i^{\text{luzer}}(r) &= Q_i G_\sigma(\mathbf{r} - \mathbf{r}_i). \end{aligned} \quad (3.26)$$

Aipatutako moduan, lehenengo batugaia irismen laburreko gaien potentziala kalkulatzeko erabiliko da, eta bigarrena irismen luzeko gaien potentzialerako.

$$\begin{aligned} \phi_i^{\text{labur}}(r) &= \frac{Q_i}{4\pi\epsilon_0} \int \frac{\delta(\mathbf{r} - \mathbf{r}') - G_\sigma(\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d^3\mathbf{r}', \\ \phi_i^{\text{luzer}}(r) &= \frac{Q_i}{4\pi\epsilon_0} \int \frac{G_\sigma(\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d^3\mathbf{r}'. \end{aligned} \quad (3.27)$$

Dirac-en delta bat tartean dagoenean, elkarrekiko espazioko kalkuluak ez dira erabilgarriak, infinitu osagai baititu. Lortu berri den $\phi^{\text{luzer}}(r)$ -n, aldiz, funtzio gaussdar bat baino ez da agertzen. Behar bezain zabala bada, elkarrekiko espazioan osagai gutxi izango ditu eta baturaren konbergentzia azkarra izango da.

Potentzialen adierazpen esplizitua honakoa da:

$$\begin{aligned}\phi_i^{\text{labur}}(r) &= \frac{1}{4\pi\epsilon_0} \frac{Q_i}{|\mathbf{r} - \mathbf{r}_i|} \operatorname{erfc}\left(\frac{|\mathbf{r} - \mathbf{r}_i|}{\sqrt{2}\sigma}\right), \\ \phi_i^{\text{luzer}}(r) &= \frac{1}{4\pi\epsilon_0} \frac{Q_i}{|\mathbf{r} - \mathbf{r}_i|} \operatorname{erf}\left(\frac{|\mathbf{r} - \mathbf{r}_i|}{\sqrt{2}\sigma}\right).\end{aligned}\quad (3.28)$$

non $\operatorname{erfc}(x) = 1 - \operatorname{erf}(x)$ den, eta $\operatorname{erf}(x)$ errore funtzioa den². $x \rightarrow \infty$ limitean $\operatorname{erf}(x) \rightarrow 1$ eta $\operatorname{erfc}(x) \rightarrow 0$ direnez, irismen laburreko gaia distantzia handitan nulua da, eta irismen luzekoa ez. (3.23) ekuazioko $f(r) = \operatorname{erf}(r)$ da.

Hau dela eta, irismen luzeko gaia ezin da espazio zuzenean kalkulatu, baina karga banaketa gaussdar baten funtzio periodikoa denez, elkarrekiko espazioan egin daiteke kalkulua. Karga dentsitatearen Fourierren transformatua:

$$\hat{\rho}^{\text{luzer}}(\mathbf{k}) = \sum_{j=1}^N Q_j e^{-i\mathbf{k}\mathbf{r}_j} e^{-\sigma^2 k^2/2}. \quad (3.29)$$

Poissonen ekuazioa $\nabla^2 \phi(r) = -\rho(r)/\epsilon_0$ da. Haren Fourierren transformatuak $k^2 \hat{\phi}(\mathbf{k}) = \frac{\hat{\rho}(\mathbf{k})}{\epsilon_0}$ adierazpena ematen du. Hortaz, elkarrekiko espazioan irismen luzeko potentziala kalkulatzeko:

$$\hat{\phi}^{\text{luzer}}(\mathbf{k}) = \frac{1}{\epsilon_0} \sum_{j=1}^N Q_j e^{-i\mathbf{k}\mathbf{r}_j} \frac{e^{-\sigma^2 k^2/2}}{k^2}. \quad (3.30)$$

Alderantzizko Fourierren transformatua eginez berreskuratzen da $\phi^{\text{luzer}}(\mathbf{r})$:

$$\phi^{\text{luzer}}(\mathbf{r}) = \frac{1}{V} \sum_{\mathbf{k} \neq 0} \hat{\phi}^{\text{luzer}}(\mathbf{k}) e^{i\mathbf{k}\mathbf{r}} = \frac{1}{V\epsilon_0} \sum_{\mathbf{k} \neq 0} \sum_{j=1}^N \frac{Q_j}{k^2} e^{i\mathbf{k}(\mathbf{r}-\mathbf{r}_j)} e^{-\sigma^2 k^2/2}, \quad (3.31)$$

non V gelaxka primitiboaren bolumena den.

Bukatzeko, zuzenketa bat egin behar da, irismen luzeko gaian puntu batek bere buruari egindako ekarpena sartuta dagoelako, funtzioa periodikoa izan dadin. (3.22) ekuazioan, prima (') erabiliz adierazi da gai hori baztertzen zela. (3.34) ekuazioan esplizituki idatziko da kenketa. $\mathcal{U}^{\text{auto}}$ gai hori honela kalkulatu da:

$$\phi_i^{\text{luzer}}(\mathbf{r}_i) = \lim_{\mathbf{r} \rightarrow \mathbf{r}_i} \frac{1}{4\pi\epsilon_0} \frac{Q_i}{|\mathbf{r} - \mathbf{r}_i|} \operatorname{erf}\left(\frac{|\mathbf{r} - \mathbf{r}_i|}{\sqrt{2}\sigma}\right) = \frac{Q_i}{4\pi\epsilon_0} \sqrt{\frac{2}{\pi}} \frac{1}{\sigma}. \quad (3.32)$$

Limitea ebazteko $\operatorname{erf}(z)$ -ren serie bidezko garapena erabili da. Horrekin,

$$\mathcal{U}^{\text{auto}} = \frac{1}{4\pi\epsilon_0} \frac{1}{\sqrt{2\pi}\sigma} \sum_{i=1}^N Q_i^2. \quad (3.33)$$

Aurreko guztia bilduz,

$$\begin{aligned}\mathcal{U} &= \frac{1}{4\pi\epsilon_0} \frac{1}{2} \sum_{\mathbf{n}} \sum_{j=1}^N \sum_{i=1}^N \frac{Q_i Q_j}{|\mathbf{r}_{ij} + \mathbf{n}|} \operatorname{erfc}\left(\frac{|\mathbf{r}_{ij} + \mathbf{n}|}{\sqrt{2}\sigma}\right) + \frac{1}{2V\epsilon_0} \sum_{\mathbf{k} \neq 0} \sum_{i=1}^N \sum_{j=1}^N \frac{Q_i Q_j}{k^2} e^{-\sigma^2 k^2/2} \cos(\mathbf{k} \cdot \mathbf{r}_{ij}) \\ &\quad - \frac{1}{4\pi\epsilon_0} \frac{1}{\sqrt{2\pi}\sigma} \sum_{i=1}^N Q_i^2.\end{aligned}\quad (3.34)$$

² $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2) dt = \operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \left(z - \frac{z^3}{3} + \frac{z^5}{10} - \frac{z^7}{42} + \frac{z^9}{216} - \dots \right)$.

Lehenengo batugaia irismen laburrekoa da espazio zuzenean, $\operatorname{erfc}(x)$ funtzioak mozten du. Bigarren batugaia irismen laburrekoa da elkarrekiko espazioan, eta $e^{-\sigma^2 k^2/2}$ esponentzialak mozten du.

Ewald baturaren erabilerak duen abantaila nagusia Fourierren transformatuen parte hartzea da. Izan ere, hauen kalkuluan *Fast Fourier Transform* (FFT) algoritmoa erabil daiteke. Honi esker, PME-ren jatorrizko $\mathcal{O}(N^2)$ konplexutasun konputazionalaren ordez, $\mathcal{O}(N \ln N)$ konplexutasuna izango du irismen luzeko interakzioen kalkuluak [12, 13].

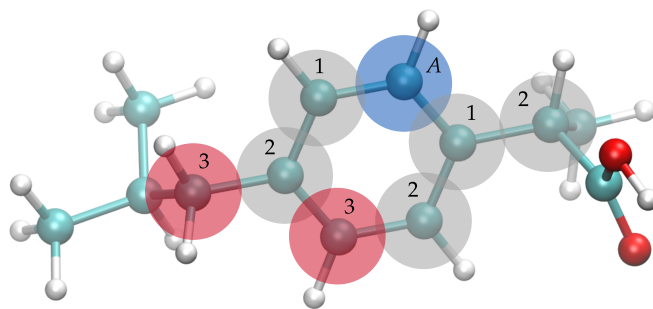
3.2.4 *Salbuespenak*

Interakzio ez-kobalenteen kalkuluan, kontu handiz egin behar da molekula bereko atomoen azterketa. Molekula bereko bi atomo kobalentekei lotuta ez badaude ere, dagoeneko haien elkarrekintzaren araberako gaiak daude force fieledean, hala nola loturen arteko angelua edo angelu diedroen menpekoak direnak. Gai hauek molekularen geometriaren araberakoak badira ere, atomoen arteko interakzio elektrostatikoak modelatzen dituzte.

Horrelakoetan, ezinbestekoa da simulatu nahi den molekularen topologia kontuan hartzea, eta horretarako *salbuespenak* erabiltzen dira. Force field bakoitzaren araberakoa da haien funtzionamendua, baina oro har ondoko arauak jarraitzen dituzte:

- Kobalentekei lotuta dauden atomoen (lehenengo auzokoen) arteko eta bigarren auzokoen arteko elkarrekintza ez-kobalenteak baztertu.
- Hirugarren auzokoak direnen elkarrekintza ez-kobalenteak konstante batez eskalatu.

Era honetan, lehenengo eta bigarren auzokoen arteko interakzioak loturak eta angeluak deskribatzen dituzten gaiak jasotzen dituztela suposatzen da. Hirugarren auzokoenak, aldiz, neurri batean hartzen dira kontuan.



3.11. irudia: Ibuprofeno molekula, salbuespenen adibide moduan. A karbono atomoaren lehenengo, bigarren eta hirugarren auzokoak diren karbono atomo batzuk irudikatu dira. VMD erabiliz egindako irudia [14].

3.11. irudian adibide bat ageri da. A karbono atomoa erreferentziazat hartuta, grisez bere lehenengo eta bigarren auzokoak diren beste karbono atomoak irudikatu dira. Atomo hauekin angelu ezberdinak eratzen ditu A atomoak; angelu horien araberakoak diren energia gaiak kalkulatzean dagoeneko kontuan hartzen ari dira atomo horien arteko interakzio elektrostatikoak. Hortaz, ez litzateke egokia izango atomo hauen arteko elkarrekintza ez-kobalenteak energia osoari gehitzea.

Aipatutako arauen arabera, A atomoaren eta grisez inguratutako atomoen arteko elkarrekintzak baztertzen dira, lehenengo eta bigarren auzokoak baitira. Hirugarren auzoko batzuk gorritz adierazi dira, eta A atomoarekin dituzten elkarrekintza ez-kobalenteak kontuan hartuko dira, baina konstante batez eskalatuta.

3.3 HIGIDURA EKUAZIOEN INTEGRAZIOA

Interakzioen araberako energiaren deskribapen zuzen bat izatea bezain garrantzitsua da gero hau erabiliz higidura ekuazioak ebazteko gai izatea. Horretarako, diferentzia finituen metodoak erabiltzen dira. Hauen funtsa integrazioa etapa txiki anitzetan moztea da, ondoko ondoko bi etapen arteko denbora tarteak Δt izanik. Kontsidera bedi Newtonen 2. legea:

$$\mathbf{F} = -\vec{\nabla}\mathcal{U} = m\frac{d^2\mathbf{r}}{dt^2}, \quad (3.35)$$

non \mathcal{U} aztergai den sistemaren energia potentziala deskribatzen duen force fielda den.

Gorputz askoko dinamika aztertzean ezinezkoa da soluzio analitikoak lortzea. Integratzaile numeriko guztiak ere ez dira egokiak, Hamilton-en higidura ekuazioen zenbait ezaugarri mantentzen dituztenak dira behar direnak: *integratzaile sinplektikoak*. Ezaugarri horiek dira:

- *Itzulgarritasuna*: jatorrizko higidura ekuazioen antzera, diskretizatutako ekuazioak ere denborarekiko itzulgarriak izan behar dira.
- *Fase espazioaren bolumenaren kontserbazioa*: Liouvilleren teoremarekin lortuta, energiaren kontserbazioa behartzen du. Higidura kaotikoa saihesten da era honetan.

Integratzaile sinplektikoena gai zabala da, lan honetan sakonduko ez dena. Kontuan hartu beharrekoa da ezaugarri hauek gabeko integrazio metodoek erroreak pilatuko dituztela denbora luzean.

Metodo zabalduenetakoa *Verlet-en algoritmoa* da, L. Verlet-ek 1967an ezagutarazi zuena [15]. $\mathbf{r}(t + \Delta t)$ eta $\mathbf{r}(t - \Delta t)$ -en Taylorren garapenak batuz lortzen da ondoko adierazpena:

$$\mathbf{r}(t + \Delta t) = 2\mathbf{r}(t) - \mathbf{r}(t - \Delta t) + \ddot{\mathbf{r}}(t)\Delta t^2 + \mathcal{O}(\Delta t^4). \quad (3.36)$$

Posizioetatik abiadura lortzeko:

$$\dot{\mathbf{r}}(t + \Delta t) = \frac{\mathbf{r}(t + \Delta t) - \mathbf{r}(t - \Delta t)}{2\Delta t}. \quad (3.37)$$

Hasierako baldintzak zehazteko, biomolekula konplexuen kasuan ohikoa da posizioak espermentalki neurtutako konfigurazioetatik hartzea. Lan honetan erabilitako molekula sinplegoen kasuan, posible da bi dimentsiotako konfigurazioa ezagutuz hiru dimentsiotakoa ondorioztatzea, energia minimizatuz. Abiadurak ezartzeko Maxwell-Boltzmann-en banaketa bat lagintzen da. Gero, sistemaren masa zentroaren momentu lineala nulua izateko eta tenperatura ezarritakoa izateko balio horiek doitu egiten dira.

Verleten integratzailearen bitartez multzo mikrokanokikoa (NVE) simulatzen da, integrazio prozesuan zehar kutxako atomo kopurua, kutxaren bolumena eta energia konstante mantentzen direlako. Simulazioen bidez lortutako emaitzak esperimentuekin alderatzeko, esperimuntuetako baldintzak mantentzea izaten da beharrezkoa. Hauetan tenperatura edota presioa izan ohi dira konstanteak, ez energia. Hortaz, praktikan, erabilgarriagoa izaten da multzo kanonikoa (NVT) edo isotermo-isobarikoa (NPT) simulatzea.

3.3.1 Tenperatura konstanteko simulazioak

NVT edo NPT multzoak simulatzeko, tenperatura konstante mantenduko duen teknika bat erabili behar da. Hau egiteko modu sinpleenetakoa Newtonen higidura ekuazioa (3.35) simulatu beharrean Langevin-en higidura ekuazioa simulatzea da. Honakoa da:

$$m\ddot{\mathbf{r}} = -\vec{\nabla}\mathcal{U} - \gamma m\dot{\mathbf{r}} + \sigma W(t). \quad (3.38)$$

Disoluzio batean disolbatzaileak solutuarekin dituen bat-bateko talkak eta sortutako marruskadura kontuan hartzen ditu ekuazio honek. (3.38) ekuazioko bigarren gaiak, $-\gamma m \dot{r}$, ingurunearen biskositateak higitzen ari den partikulari eragindako marruskadura indarra adierazten du. Hirugarren gaiak berriz, $\sigma W(t)$, partikulak inguruneko partikula txikiekin jasandako ausazko talkak irudikatzen ditu, eta horren ondoriozko fluktuazioak. $W(t)$ zorizko aldagai bat da, banaketa normal estandarra duena, $\mathcal{N}(0,1)$. σ parametroarekin eskalatuz aldagaiaren desbideraketa estandarra σ izatea lortzen da.

Ekuazio honekin bero iturri baten eragina simulatzea lortzen da. Sistemak bere tenperatura konstante mantentzen ahalko du, bero iturriarekin energia trukaturik.

Tenperatura handitzeak atomoen abiadura handitzea dakar, eta ondorioz, jasaten duten marruskadura indarra. Era honetan, bero iturriak energia disipatuko du sistemaren energia (eta ondorioz tenperatura) txikituz. Energia galera hori ausazko talkek gehitutakoaren bidez konpentsatzen da. Bi ekarpen hauei esker osotara sistemaren tenperatura konstante mantenduko da, nahiz eta fluktuazioak egon.

Azalpen intuitibo horretatik haratago joanez, froga daiteke $\sigma = \sqrt{2m\gamma k_B T}$ balioa erabilita zehazki multzo kanonikoa erreproduzitzeko dela [16]. Honen froga ez da lan honetan aztertu, baina simulazioan zehar behar den multzo termodinamikoa lagintzen dela ziurtatzen du.

(3.38) ekuazioa diskretizatzen duen integratzaile sinplektikoa nola lortu azaltzea ere lan honen irismenetik kanpo dago. OpenMM-ek inplementatuta duen LangevinMiddleIntegrator metodoa erabili da, $T = 300$ K, $\gamma = 1$ ps⁻¹ eta $\Delta t = 4$ fs balioekin [17].

3.3.2 Presio konstantepeko simulazioak

Multzo isoterma isobarikoko (NPT) simulazioak egiteko, azkeneko osagai bat gehitu behar da: presioa konstante mantenduko duen teknika bat. Lan honetan *Monte Carlo barostatoa* erabili da. Metodo honen bidez simulazio kuxtaren bolumena aldatzen da denboran zehar, presioa konstante mantentzeko. Hau egiten denean atomoen posizioak ere moldatzen dira, kutxatik atera ez daitezela. Monte Carlo barostatoak honako pausuak jarraituko ditu:

1. Zorizko eran, simulazio kuxtaren bolumenaren $\Delta V = V - V_0$ aldaketa bat aukeratu.
2. Aldaketa onartzen den erabakitzeke, ΔH kalkulatu da:

$$\Delta H = \Delta U + P\Delta V - Nk_B T \ln \left(\frac{V}{V_0} \right). \quad (3.39)$$

3. ΔH negatiboa bada, konfigurazio berriaren energia txikiagoa da, eta aldaketa zuzenean onartzen da. Bestela, 0 eta 1 arteko ausazko aldagai bat aukeratzen da, eta azpikoa betetzen bada bakarrik onartzen da aldaketa:

$$\text{random}(0,1) \leq e^{-\frac{\Delta H}{k_B T}}. \quad (3.40)$$

Algoritmo honekin (3.39) ekuazioan ezarritako P -ren balio bera lortzen da simulazioan zehar, fluktuazio batzuk gorabehera. Prozesu honek duen kostu konputazionala dela eta, bolumen aldaketak ez dira integrazio pauso bakoitzean egiten, maiztasun gutxiagorekin baizik. Lan honetan $P = 1$ atm presioa erabili da, eta proposamen berriak 25 pausuan behin egin dira.

SIMULAZIOEN GARAPENA

Aurreko kapituluaz aztertutakoa praktikan jartzeko erabilitako softwareak OpenMM du izena, eta simulazio molekularrerako errendimendu handiko tresna da [18]. Python programazio lengoia idatzi dira programak, eta haien exekuzioarako DIPC-ren *Hyperion* zein BCB taldearen *Calmodulin* ordenagailuak baliatu dira [19, 20]. Programa guztiak GitHub-eko errepositorio batean¹ eskuragarri jarri badira ere, atal honetan ezaugarririk garrantzitsuenak azpimarratuko dira. Azalpena kode zatiekin hornitzea erabaki da, argigarria izango delakoan. Lan honen gainontzeko zatietan hori egitea saihestu den arren, kasu honetan komenigarria dela pentsatu da.

4.1 SIMULAZIOEN DESKRIBAPENA

Beharrezko liburutegiak kargatuz hasten dira simulazio guztiak.

```
1 import openmm as mm
2 from openmm import app, unit
3 from openff.toolkit.topology import Molecule
```

Aztertuko diren molekulei buruzko informazioa .sdf fitxategietan gordetzen da, eta atomoen hasierako posizioak .pdb fitxategietan.

```
4 ibuprofen = Molecule('Ibuprofen.sdf')
5 ibuprofen_pdb = app.PDBFile('Ibuprofen.pdb')
```

Hurrena, force fielda definitzen da, eta hautatutako molekularen parametroak gehitzen zaizkio. TIP3P modeloak ura ereduizatzen du, eta izeneko 3 zenbakiak molekula bakoitzeko karga partzial kopurua adierazten du. Kasu honetan, karga bat atomo bakoitzeko.

```
6 forcefield = app.ForceField('amber14/tip3pfb.xml')
7 gaff = GAFFTemplateGenerator(ibuprofen, forcefield='gaff-2.11')
8 forcefield.registerTemplateGenerator(gaff.generator)
```

Sistemaren loturak eta hasierako posizioak ezartzen dira jarraian. Uretan solbatatutako kasuan adibidez, molekula kokatu eta gero ura gehitzen da.

```
9 solvated_model = app.Modeller(ibuprofen_pdb.topology, ibuprofen_pdb.positions)
10 solvated_model.addSolvent(forcefield, model='tip3p', padding=1.5 * unit.nanometer)
```

Aurreko objektu guztiak elkartuz osatzen da simulatu nahi den sistema.

```
11 system = forcefield.createSystem(topology=solvated_model.topology,
12     nonbondedMethod=app.PME, nonbondedCutoff=0.9 * unit.nanometer,
13     removeCMMotion=True, rigidWater=True, constraints=app.HBonds)
```

Pausu honetan rigidWater parametroa ezartzen da, erabilitako ur modeloan loturen luzerak eta hidrogenoen arteko angelua konstante mantenduko direla zehazteko, SETTLE algoritmoaren bidez [21]. constraints parametroak simulazioan zehar hidrogeno eta bestelako atomoen arteko loturak finko mantenduko direla ezartzen da. Hau LINCS algoritmoa erabiliz egiten da [22]. Lotura hauen bibrazio-maiztasuna hain handia denez, askatasun gradu hauek izoztu ezean integrazio pausua txikitzea beharrezkoa da.

¹ Esteka: <https://github.com/BeBerasategi/Transformazio-alkimikoak>.

Era analogoan sortzen dira hutseango sistemari dagokion modeloa (`vacuum_model`) eta sistema (`vacuum_system`). Erabilitako gelaxka primitiboa kubikoa da, 3 nm-ko aldekoa.

Integratzailea definitzerakoan simulazioaren tenperatura, (3.38) ekuazioko γ marruskadura-koefizientea eta integrazio-pausua ezartzen dira: 300 K, 1 ps^{-1} eta 4 fs.

```

14 temperature = 300 * unit.kelvin
15 friction_coefficient = 1 / unit.picosecond
16 time_step = 4 * unit.femtosecond
17 integrator = mm.LangevinMiddleIntegrator(temperature, friction_coefficient, time_step)

```

Honekin, simulazioa egiteko beharrezko guztia prest dago.

```

18 simulation = app.Simulation(solvated_model.topology, system, integrator, platform=mm.
    Platform.getPlatformByName('CUDA'))

```

Hasierako posizioen balioetan egon daitezkeen akatsak konpontzeko, energia minimizatzea da simulazioaren aurreneko pausua, *gradiente jaitziera* motako algoritmo baten bidez.

```

19 simulation.minimizeEnergy()

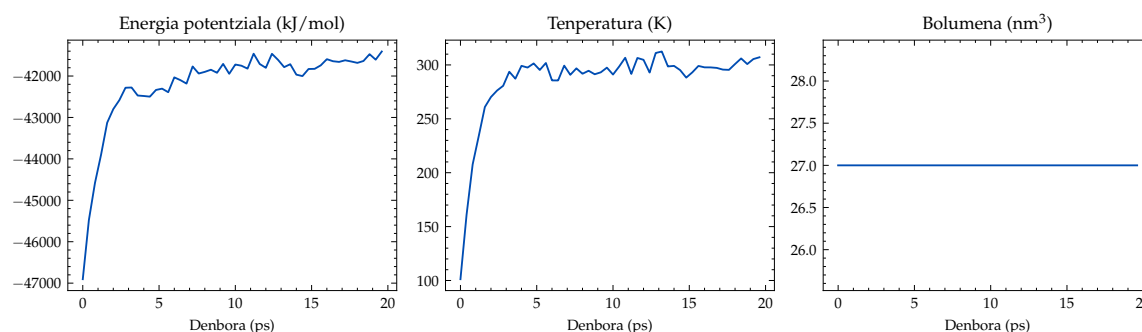
```

Ondoren, orekatzeko denbora ematen zaio simulazioari, efektu iragankorrak baztertzeko. Simulazio pausua 4 fs-koa denez, 5000 pausu emanaz orekatzeko 20 ps simulatzen dira.

```

20 simulation.step(5000)

```



4.1. irudia: Ibuprofeno molekula, uretan. Orekatzeko 4 fs-ko 5000 pausu eman dira, eta 100 pausuro energia potentziala, tenperatura eta bolumena neurtu dira. Horrelako irudiek simulazioa ongi hasten dela ziurtatzeko eta arazoak detektatzeko balio dute.

4.1.1 NPT multzoaren simulazioa

Teorian azaldutako Monte Carlo barostatoa implementatzen da ondoren, presioa konstante mantentzeko. 1 atm presio konstantea erabiliko da, eta gelaxkaren bolumena aldatzeko proposamenak 25 simulazio pausuro egingo ditu barostatoak.

```

21 pressure = 1.0 * unit.atmosphere
22 mc_frequency = 25
23 simulation.system.addForce(mm.MonteCarloBarostat(pressure, temperature, mc_frequency))

```

Aldaketa hau egin ondoren ere, orekatzeko astia eman behar zaio simulazioari. Komenigarria da lehenengo NVT-n orekatzea, gero barostatoa gehitu eta ondoren NPT-n berriro orekatzea. Izan ere, hasieratik barostatoa gehituta orekatzen saiatuz gero, arazoak izan daitezke, NPT-ko simulazioak NVT-koak baino ezegonkorragoak izan daitezkeelako.

```

24 simulation.step(5000)

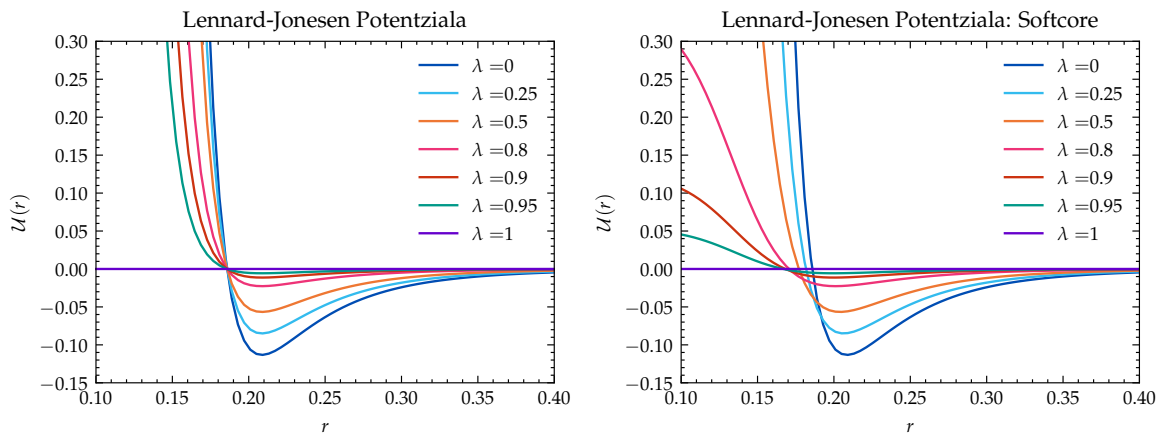
```

4.2 TRANSFORMAZIO ALKIMIKOEN INPLEMENTAZIOA

Aurreko pausuk orokorrak dira dinamika molekularreko simulazioetan. Ondoren azalduko direnak bereziak dira, transformazio alkimikoaren inplementaziorako beharrezkoak.

4.2.1 Potentzial alkimikoak

“Energia askearen kalkulua” kapituluko (2.15) ekuazioan adierazienez, molekula baten *desagertzea* simulatu nahi da, λ interpolazio parametro baten arabera. $\lambda = 0$ baliorako, molekula uretan disolbatuta dago. λ handitu ahala molekula eta uraren arteko interakzioen indarra gutxitzen da, eta $\lambda = 1$ denean molekula desagertu dela kontsideratzen da. Ideia sinplea bada ere, praktikan hau zuzenean egiteak arazoak sor ditzake.



(a) Lennard-Jonesen potentzial estandarra.

(b) Lennard-Jonesen potentzial leuna.

4.2. irudia: Transformazio alkimikoak egiteko egoera ezberdinen arteko interpolaketa leuna beharrezkoa da, bat-bateko aldaketarik gabea, simulazioetan ezegonkortasunak ekiditeko.

Kontsideratu 4.2a. irudia. $\lambda = 1$ denean egon daitezke arazoak, egoera horretan potentziala nulua delako, baina λ -ren balioa apur bat txikitzeak ($\lambda = 0.95$ -ra, adibidez) potentzialaren bat-bateko hazkundera eragiten du. Potentzialik gabeko simulazioan bi atomo asko hurbiltzen badira, $\lambda = 1$ -tik $\lambda = 0.95$ -ra aldatzean aldarapen indar bortitzak agertuko dira bikote horren artean, simulazioa ezegonkortuz.

$\lambda = 0$ eta $\lambda = 1$ -en artean interpolazio leuna behar da, 4.2a. irudiko bat-bateko aldaketarik gabea. Hau dela eta, transformazio alkimikoak egiteko *softcore* potentzialak erabiltzen dira Lennard-Jones-en potentzial estandarren ordez. 4.2b. irudian ikus daitekeenez, potentzial hauen balioa jatorrizkoaren berdina da muturretako egoeretan, baina tartekoetan interakzioak nahita ahultzen dira. Energia askea egoera funtzioa denez, dagoeneko aipatu da hasiera eta bukaerako egoerak direla fisikoki esanguratsuak izan behar direnak.

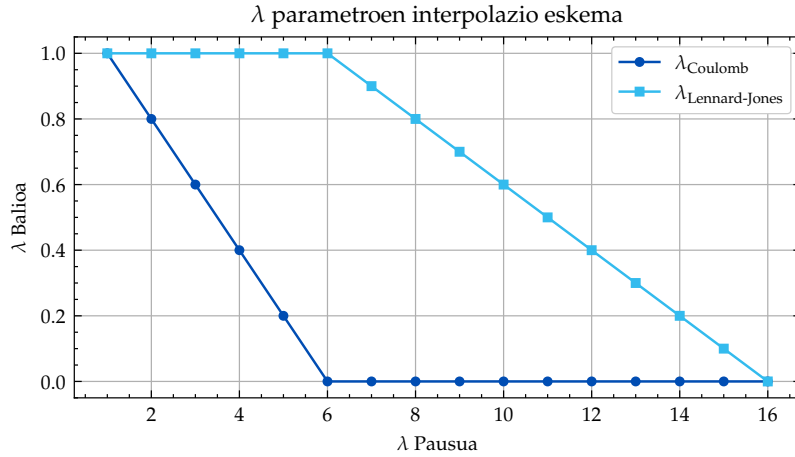
$r_{ij}^{\text{softcore}} = (0.5\lambda_{LJ}\sigma^6 + r_{ij}^6)^{1/6}$ distantzia definituta, ondorengo forma dauka Lennard-Jonesen potentzial leunak:

$$\mathcal{U}_{\text{Lennard-Jones}}^{\text{softcore}} = (1 - \lambda_{LJ}) 4\epsilon \left[\left(\frac{\sigma}{r_{ij}^{\text{softcore}}} \right)^{12} - \left(\frac{\sigma}{r_{ij}^{\text{softcore}}} \right)^6 \right]. \quad (4.1)$$

Coulomben potentzialaren kasuan ere konponbide bera erabiltzen da. Potentzialaren adierazpenean λ_{Coulomb} -en arabera $r_{ij}^{\text{softcore}} = (a \cdot \lambda_C + r_{ij})$ distantzia ordezkatzuz interpolazio leunagoa erdiesten da.

4.2.2 Interpolazio eskema

Mota honetako simulazioetan agertzen den beste arazo bat Coulomben interakzioek sortzen dute. Orokorrean, simulazioan zehar distantzia interatomikoa oso txikia izan daitekeenean, zaila da hauek era egonkorrean simulatzea. λ -ren balioa handitzen den heinean aldarapen indarrak txikitu, eta atomoak elkarrengana asko hurbildu daitezke. Elkarrekintza coulombdarra neurri berean ahultzen bada ere, bi molekula gehiegi hurbiltzen badira interakzio hauek asko handitu daitezke bat-batean, simulazioa ezegonkortuz.



4.3. irudia: Elkarrekintzen desagertzea erduztatzeko erabiliko 16 pausuko interpolazio eskema. OpenMM-n hasierako egoeran $\lambda = 1$ da, eta molekula desagertu denean $\lambda = 0$.

Hau konpontzeko estrategia sinplea da. Lennard-Jonesen eta Coulomben interakzioak aldi berean desagerrarazi beharrean, lehenengo Coulomb-en interakzioak desagerrarazten dira, eta ondoren, Lennard-Jonesenak. Hau da, bi interpolazio parametro independente erabiltzen dira, λ_{Coulomb} eta $\lambda_{\text{Lennard-Jones}}$. Bakoitza eskema ezberdina jarraituz aldatzen da, 4.3. irudiak erakusten duenez. Potentziala horrela deskonposatuta geratzen da beraz:

$$\begin{aligned}
 \mathcal{U}(\mathbf{r}) = & \mathcal{U}_{\text{bonded}}^{\text{water}}(\mathbf{r}) + \mathcal{U}_{\text{bonded}}^{\text{molecule}}(\mathbf{r}) + \mathcal{U}_{\text{non-bonded}}^{\text{water-water}}(\mathbf{r}) \\
 & + (1 - \lambda_{\text{Lennard-Jones}}) \cdot \mathcal{U}_{\text{Lennard-Jones}}^{\text{water-molecule}} + (1 - \lambda_{\text{Coulomb}}) \cdot \mathcal{U}_{\text{Coulomb}}^{\text{water-molecule}}(\mathbf{r}) \\
 & + (1 - \lambda_{\text{Lennard-Jones}}) \cdot \mathcal{U}_{\text{Lennard-Jones}}^{\text{molecule-molecule}}(\mathbf{r}) + (1 - \lambda_{\text{Coulomb}}) \cdot \mathcal{U}_{\text{Coulomb}}^{\text{molecule-molecule}}(\mathbf{r}).
 \end{aligned}
 \tag{4.2}$$

4.3 PRODUKZIO SIMULAZIOA

Puntu honetan, iraupen luzeagoko simulazioak egin dira, gerora intereseko magnitudeen estatistikak kalkulatzeko beharrezko datuak emango dituztenak. 4.3. irudiko λ bikote bakoitza 5 ns-z simulatu da eta 16 bikote simulatu dira: 80ns guztira.

Bikote bakoitzerako, balio egokiak ezartzen zaizkio potentzialari, energia minimizatzen da, eta sistemari orekatzeko denbora ematen zaio (20 ps, 4.1. irudian bezala). Gero, 5 ns pasa arte, iteratiboki errepikatzen da ondoko prozesua: 1 ps simulatu, eta lortutako konfigurazioaren energia neurtu eta gorde. λ_k bikotea simulatu bada, txandaka λ_{k+1} eta λ_{k-1} bikoteen balioak ezarri, eta konfigurazio berdinak balio horietarako duen energia kalkulatu. Bukatzeko, λ_k balioa berrezarri eta simulazioak aurrera jarraitzen du.

Era honetan lortzen dira (2.8, 2.9, 2.10) ekuazioen bidez energia aske aldeak kalkulatzeko beharrezko datuak. Bertan agertzen diren batezbestekoak kalkulatzeko hartzen dira horrenbeste neurketa, lan honetan 5000 lagin λ -ren balio bakoitzeko.

EMAITZAK ETA ANALISIA

Molekula baten solbatazio energia askea kalkulatzeko bi simulazio egitea da beharrezkoa, bakoitza 2.2. irudiko gezi bertikal bati dagokiolarik. Urez betetako kutxa batean molekula-
ren desagertzea simulatuz, $\Delta G_{\text{solbatatuta}}^{A \rightarrow B}$ lortzen da. Hutsik dagoen kutxa batean gauza bera
eginez, $\Delta G_{\text{hutsean}}^{A \rightarrow B}$ eskuratzen da. Bi horien diferentziarekin $\Delta G_{\text{hutsean} \rightarrow \text{solbatatuta}}^A$ solbatazio
energia askea kalkulaten da, (2.12) ekuazioak adierazten duen moduan.

Bi txandatan egin dira simulazioak. Lehenengo saiakerak 9 molekulekin egin dira, praktika hartu eta baldintza ezberdinen aldaketen eragina arretaz aztertzeko. Horrela, emaitza egokienak lortzeko ezarpenak aurkitu dira, eta horiek baliatuz beste 20 molekulen simulazioak burutu dira.

Kapitulu honetan lortutako emaitzen eta balio esperimentalen arteko konparaketa egiten da. Horretarako zenbait metrika erabiltzen dira, eta horiek azaltzen dira lehenik. Ondoren, lehenengo 9 molekulen simulazioen emaitzak aztertuko dira, eta bukatzeko, simulatutako molekulen multzo osoarenak.

5.1 METRIKEN AZALPENA

Solbatazio energia askeen balio esperimentalak FreeSolv datu basetik hartu dira [23]. Simulazioen emaitzekin konparaketa sistematikoki egiteko, balio numerikoak zuzenean alderatzeaz gain lau metrika erabili dira: MAE, MSE, PCC eta τ .

5.1.1 Batez besteko errore absolutua

Mean Absolute Error (MAE) da ingelesez. Molekula bakoitzarentzat simulatutako balioaren eta balio esperimentalaren aldea kalkulaten da lehenik, eta gero horien guztien batezbestekoa. Metrikarik intuitiboena da, batez beste nolako erroreak espero daitezkeen adierazten du eta. Bere unitateak kcal/mol dira.

$$MAE = \frac{1}{N} \sum_i^N \left| \Delta G_i^{\text{simulatua}} - \Delta G_i^{\text{esperimentalala}} \right|. \quad (5.1)$$

5.1.2 Batez besteko errore koadratikoa

Ingelesez, Mean Squared Error (MSE). Molekula bakoitzaren balioen errore absolutua kalkulaten da, eta ondoren horien karratuen batezbestekoa. MAE-ren antzekoa da, baina gehiago zigortzen ditu errore altuko kasuak. kcal²/mol² unitateak ditu.

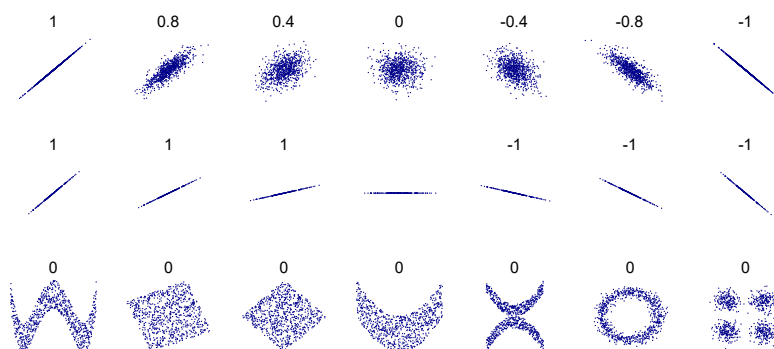
$$MSE = \frac{1}{N} \sum_i^N \left(\Delta G_i^{\text{simulatua}} - \Delta G_i^{\text{esperimentalala}} \right)^2. \quad (5.2)$$

5.1.3 Pearson-en korrelazio koefizientea

Pearson Correlation Coefficient (PCC) da ingelesez. Bi datu multzoren arteko korrelazio lineala neurtzen du. Erroreak egonda ere, simulazioek balio esperimentalei buruz zenbat informazio eman dezaketen neurtzen du. Desbideraketa estandarrekin normalizatutako kobariantza dela uler daiteke. Beti dago -1 eta 1 balioen artean. 5.1. irudian ikus daitezkeen, erlazio lineala bakarrik kontuan hartzen du. $\{(x_1, y_1), \dots, (x_n, y_n)\}$ datu bikoteentzat:

$$PCC = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}, \quad (5.3)$$

non \bar{x} eta \bar{y} batezbestekoak diren.



5.1. irudia: Pearsonen korrelazio koefizientearen balio ezberdinak dituzten puntu banaketen adibideak. (Denis Boigelot, <https://commons.wikimedia.org/w/index.php?curid=15165296>).

5.1.4 Kendall-en τ koefizientea

Simulazioetan lortutako balioak erabiliz egindako zerrenda baten kalitate edo fidagarritasuna neurtzen du metrika honek. Hau garrantzitsua da, askotan molekulen artean hautagairik egokienak aukeratzea baita helburua, ez bakoitzaren balioak zehazki jakitea.

(x_i, y_i) eta (x_j, y_j) datu bikotea hartuta, $i < j$ izanik, bat datorren bikotea (BDB) dela esaten da $x_i > x_j$ eta $y_i > y_j$ edo $x_i < x_j$ eta $y_i < y_j$ badira. Bestela, bat ez datorren bikotea (BEDB) da. Definizio hauekin,

$$\tau = \frac{1}{\binom{n}{2}} (\text{BDB kopurua} - \text{BEDB kopurua}), \quad (5.4)$$

non $\binom{n}{2}$ koefiziente binomialak n elementutik 2 aukeratzeko modu kopurua ematen duen. x_i elementuen zerrendaren ordena guztiz bat badator y_i elementuenarekin, $\tau = 1$ da.

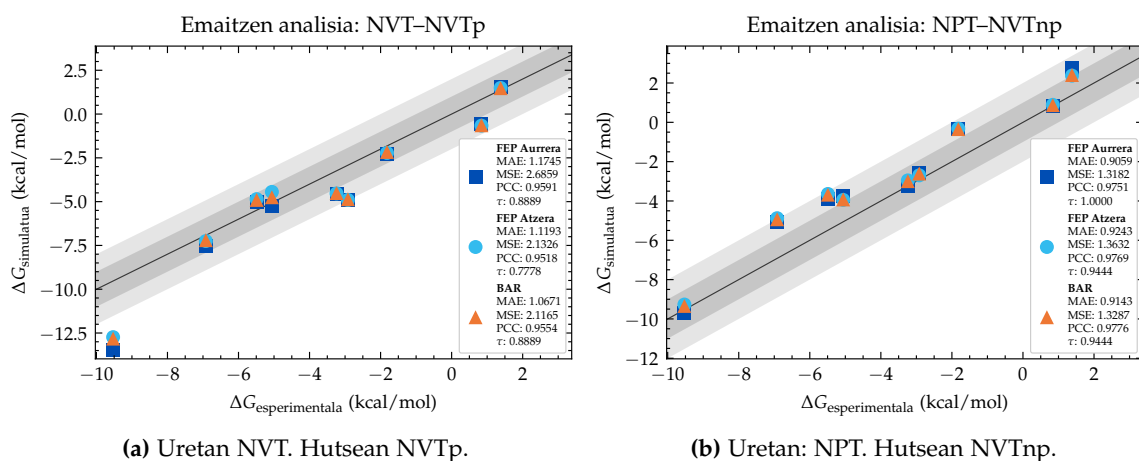
5.2 SIMULAZIOEN EMAITZAK

Eranskineko A.1. taulan erabilitako molekula guztien zerrenda dago. Bertako *Errazak I* multzoa osatzen duten 9 molekulekin egin dira lehenengo simulazioak. Simulazioetan konfigurazio ezberdinen energia neurtzen da, baina horietatik abiatuta energia askearen balioak lortzeko “Energia askearen kalkulua” kapituluaren deskribatutako FEP edo BAR metodoak ezinbestekoak dira.

(2.8) ekuazioak FEP metodoa adierazten du *aurreranzko* zentzuan. (2.9) ekuazioak aldiz metodoa *atzeranzko* zentzuan nola aplikatu erakusten du. Hau da, A eta B egoeren arteko $k - 1$, k eta $k + 1$ etapak izanik, lehenengoak $k \rightarrow k + 1$ norabidean kalkulatu du energia, eta bigarrenak berriz, $k \rightarrow k - 1$ zentzuan. Emaitza bera lortzea espero litekeen arren, praktikan balio ezberdinak eskuratzen dira. BAR metodoak $k - 1$, k eta $k + 1$ etapak kontuan hartzen dituen pausuro, errore estatistiko txikiagoko emaitzak espero dira.

5.2. irudian bi simulazioetan lortutako emaitzak adierazi dira, balio espermentalekin konparatuta. Lerro beltz diagonalak simulazioen emaitza eta balio espermentalak bat etorriko balira lortutako puntuak lotzen ditu. Simulazioen emaitzak aipatu berri diren hiru metodoak erabiliz kalkulatu dira, eta grafikoki ikus daitekeenez, hiru metodoek ia balio bera ematen dute puntu bakoitzean.

Esandako moduan, molekula bakoitzaren solbatazio energia askea lortzeko bi simulazio dira beharrezkoak, bat uretan eta bestea hutsean. 5.2a. irudian uretako sistema NVT multzo termodinamikoa erabiliz simulatu da, eta 5.2b. irudian, NPT multzoa erabiliz. Hutseango sistema NVT multzoan simulatzen da beti, ez baitago konstante mantentzeko presiorik. Horrez gain, mugalde baldintzen eragina ere aztertu da. 5.2a. irudian hutseango sistema mugalde baldintza periodikoak (p) erabiliz simulatu da, eta 5.2b. irudian aldiz, horiek gabe (np).



5.2. irudia: Bi simulazioetan lortutako emaitzak, datu espermentalekin alderatuta. Hiru metodo-rekin kalkulatu dira energiak: FEP aurrera, FEP atzera eta BAR. Barra grisek 1 eta 2 kcal/mol-ko marjina adierazten dute.

Uretako sistema simulatzeko multzoaren aukerak eta hutseango sistemaren mugalde baldintza motak konbinatuz, molekula bakoitzarentzat solbatazio energia askea estimatzeko 4 modu daude. Gainera, simulazioen emaitzetatik abiatuta energiaren balioa hiru metodoren bidez kalkulatu daitekeenez, FEP aurrera, FEP atzera eta BAR erabiliz, guztira 12 modu daude.

5.1. taula: FEP aurrera erabiliz kalkulatuak diren energiaren dagozkien metrikak.

	NVT-NVTp	NVT-NVTnp	NPT-NVTp	NPT-NVTnp
MAE	1.1745	1.1628	0.9102	0.9059
MSE	2.6859	2.6284	1.3238	1.3182
PCC	0.9591	0.9595	0.9744	0.9751
τ	0.8889	0.8889	1.0000	1.0000

Konbinazioen artean egokiena aurkitzea da helburua, gainontzeko molekulekin hori bakarrik erabiltzeko. Konparazioak egiteko aurreko ataleko metrikak kalkulatu dira: MAE eta MSE zenbat eta txikiago izan, orduan eta hobeto. PCC eta τ -ri dagokionez, 1-tik ahalik eta hurbilen egotea da desiragarriena. Emaitzak 5.1, 5.1 eta 5.3 tauletan bildu dira.

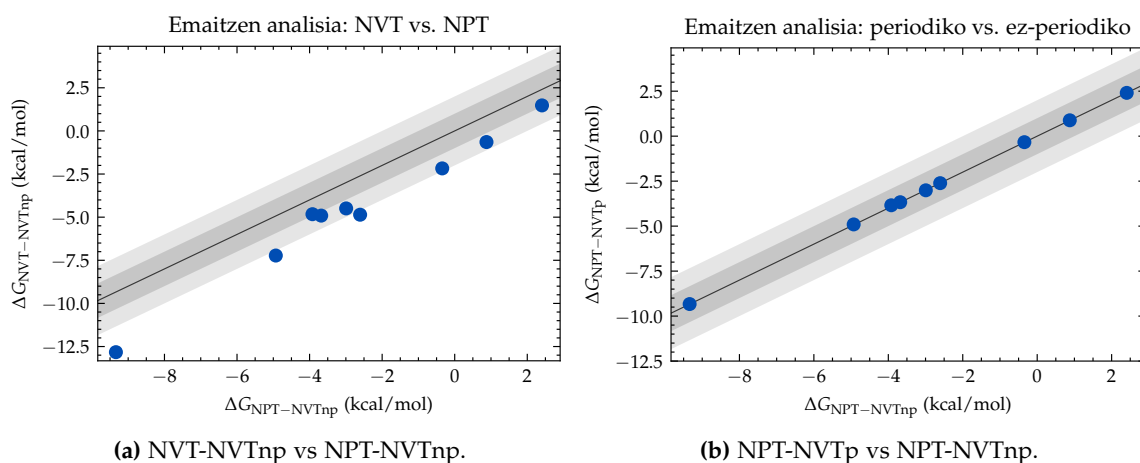
5.2. **taula:** FEP atzera erabiliz kalkulaturako energiari dagozkien metrikak.

	NVT-NVT _p	NVT-NVT _{np}	NPT-NVT _p	NPT-NVT _{np}
MAE	1.1193	1.0945	0.9456	0.9243
MSE	2.1326	2.1144	1.4135	1.3632
PCC	0.9518	0.9539	0.9764	0.9769
τ	0.7778	0.8333	0.9444	0.9444

5.3. **taula:** BAR erabiliz kalkulaturako energiari dagozkien metrikak.

	NVT-NVT _p	NVT-NVT _{np}	NPT-NVT _p	NPT-NVT _{np}
MAE	1.0671	1.0602	0.9271	0.9143
MSE	2.1165	2.113	1.3671	1.3287
PCC	0.9554	0.9565	0.9769	0.9776
τ	0.8889	0.8889	0.9444	0.9444

Aipatu behar da erabilitako molekulak nahiko sinpleak izan direnez, metodoen arteko aldeak ez direla itzelezkoak. Uretako sistema NVT multzoan beharrean NPT-n simulatuta emaitzak apur bat hobekiak direla ikusten da kasu guztietan. Hau lortzea espero zen, disoluzio baten laborategiko baldintzak hobeto hurbiltzen baititu NPT multzoak. Hutseango sisteman mugalde baldintza ez periodikoak erabiltzea egokiagoa da kontzeptualki, baina hobekuntza ia arbuigarria da. Konparaketa hauek grafikoki egiten dira 5.3. irudian.

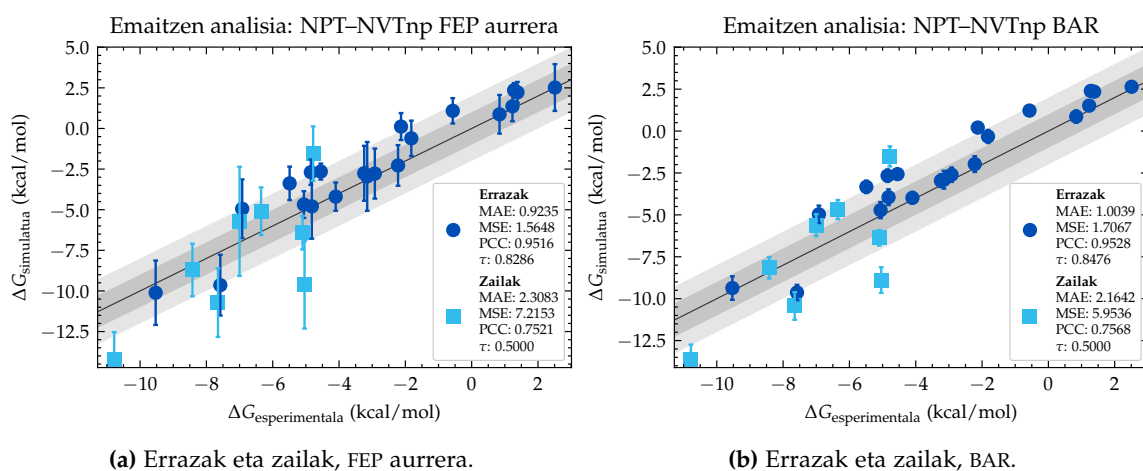


5.3. **irudia:** Konparaketa, konbinazio ezberdinen artean. 5.3 taulako datuak hartzen dira, BAR metodoarekin kalkulaturakoak. Barra grisek 1 eta 2 kcal/mol-eko marjina adierazten dute. 5.3a. irudiak solbataturako sistema NVT edo NPT multzoan simulatzeak duen eragina nabarmentzen du. 5.3b. irudiak aldiz, agerian uzten du hutseango sistemaren mugalde baldintzak aldatzeak duen eragin ia arbuigarria.

Lortutako emaitzen arabera, aukerarik egokiena uretako sistema NPT multzoan simulatzea da, eta hutseango sistema NVT multzoan mugalde baldintza periodikorik gabe. Uretako sistema simulatzeko beti erabili behar dira mugalde baldintza periodikoak.

Ezarpen hau erabilia egin dira gainontzeko simulazioak, beste 20 molekulekin. A.1. taulako bigarren eta hirugarren zutabeetakoak dira hauek. Molekula txikiak eta lotura birakor gutxikoak simulatzeko errazagoak dira. Hiru lotura birakor edo gehiago dituztenak zailtasun gehiago sortzen dituzte, tortsio askatasun gradu horien ondorioz. Beraz, hasiera batean molekula *errazak* eta *zailak* multzokatuz aztertu dira emaitzak.

Simulazioetan lortutako balioez gain, dagozkien errore estatistikoak ere adierazi dira 5.4. irudian, *bootstrapping* laginketa metodoa erabiliz kalkulatu direnak. Aipagarria da magnitude txikiko erroreak lortzen direla, eta irudietan ikusi ahal izateko 10 aldiz handitzea ezinbestekoa izan dela. 5.4a. irudiko balioak FEP aurrera metodoa erabiliz kalkulatu dira, eta 5.4b. irudikoak BAR metodoaren bitartez. Azpimarratu behar da simulazioekin lortutako balio berdinak erabiltzen dituztela biek, baina hala ere energia askearen estimazio ezberdinak ematen dituztela.



(a) Errazak eta zailak, FEP aurrera.

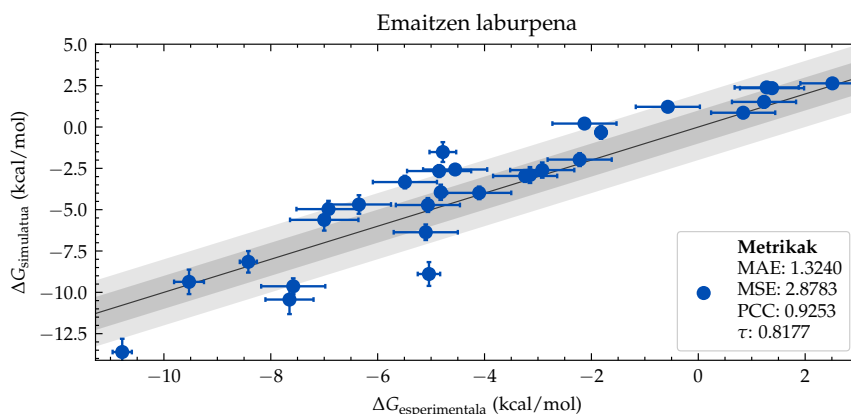
(b) Errazak eta zailak, BAR.

5.4. irudia: Simulatutako 29 molekulen emaitzak, denak batera. “Zailak” eta “errazak” bereizi dira, molekularen malgutasunaren arabera. Zenbat eta lotura birakor gehiago izan, zailagoa izaten da molekula simulatzea. Simulazioen errore estatistikoa adierazi da, irudietan ikusi ahal izateko 10 aldiz handituta. Azpimarragarria da BAR metodoa erabilia asko txikitzen dela.

Agerikoa da BAR metodoaren bidez lortutako emaitzek errore txikiagoa dutela. Metrikak aztertuta, molekula zailen kasuan nabaritzen da gehiago hobekuntza. Honetaz gain, ikus daiteke *zail* bezala sailkatutako metrikei balio zatarragoak dagozkiela.

Egindako analisi zatikatu honekin ondorioztatu da zeintzuk diren simulazioak egiteko baldintzarik onenak, eta lortutako balioekin energia kalkulatzeko BAR metodoa dela egokiena. Hau kontuan hartuta simulatutako 29 molekulen emaitzak 5.5. irudian adierazi dira, zailen eta errazen arteko banaketarik egin gabe.

Irudi horretan, simulazioetan lortutako balioen errore estatistikoek gain datu esperimentalen erroreak ere adierazi dira, handipenik gabe. Metrikak multzo osoarenak dira.



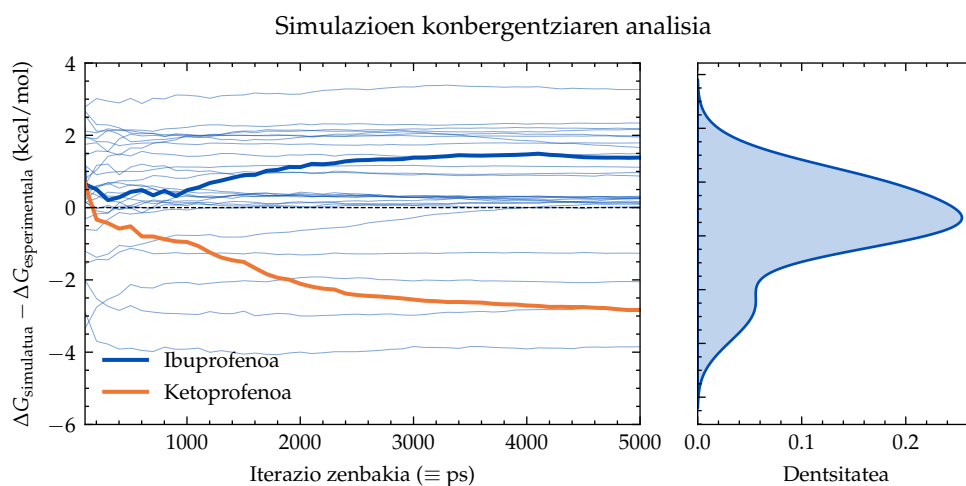
5.5. irudia: Simulatutako 29 molekulekin lortutako emaitzen eta balio esperimentalen arteko konparaketa. Errore estatistikoak 10 aldiz handitu dira, baina hala ere ia ez dira ikusten, kalkulua BAR erabiliz egin delako. Errore esperimentalek ez dute handipenik. Multzo osoaren metrikak adierazten dira.

5.3 SIMULAZIOEN KONBERGENTZIA

Bukatzeko, simulazioen konbergentziaren analisi labur bat egin da. 5.6. irudian molekula bakoitzari dagokion simulazioarekin lortutako energia askearen balioak adierazi dira, edo zehatzago, horien eta balio esperimentalen arteko aldeak. Simulazioaren iraupenaren arabera balio horien bilakaera adierazten du irudi horrek.

Era honetan bi kasu berezi identifikatu dira, ibuprofeno eta ketoprofeno molekulak. Hasieran balio esperimentaletik gertu daude, baina gero hortik urrunago dauden balioetara konbergitzen dute.

Salbuespenak salbuespen, 5.6. irudiaren eskuinaldean dagoen banaketak argi uzten du simulazioen gehiengoan konbergentzia zuzena izan dela, lortutako balioaren eta balio esperimentalaren arteko aldea 2 kcal/mol baino txikiagoa izanik.



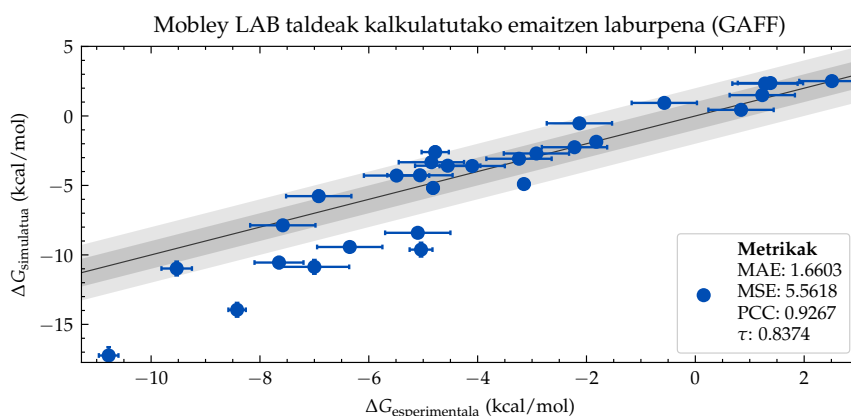
5.6. irudia: Konbergentziaren analisia eta bukaerako balioen banaketa. Simulazioak emandako energia askearen balioaren eta balio esperimentalaren arteko aldea irudikatu da, iterazio kopuruaren menpe. Iterazio bakoitza simulatutako 1 ps-ri dagokio. Ibuprofenoaren eta ketoprofenoaren kasuan, hasieran balio esperimentaletik gertu egon arren, horretatik urruntzen dira denborak aurrera egin ahala.

ONDORIOAK

Lan honen helburu nagusia dinamika molekularra erabiliz transformazio alkimikoak burutzen dituzten simulazioak inplementatzea izan da. Hasierako kapitulu teorikoagoetan azaldutako ideiak baliatuz egin da hau, eta horiekin kontsistenteak diren emaitzak lortu dira.

Horrez gain, energia askeak kalkulatzeko metodo honen egokitasuna aztertu nahi izan da. Argi geratu da funtzionamendua ona dela. 5.5. eta 5.6. irudiek adierazten dutenez, simulazio gehienek era egokian konbergitu dute balio esperimentalera. Salbuespenen portaera ulertzeak bestelako azterketa bat eskatuko luke.

Erreferentzia moduan Mobley LAB taldeak egindako simulazioen emaitzak adierazi dira 6.1. irudian. OpenMM beharrezan GROMACS erabili dute 642 molekula simulatzeko, baina 6.1. irudia egiteko lan honetako 29 molekulak baino ez dira hartu. Lan honetan batez besteko errore absolutu eta koadratiko txikiagoak lortu direla ikusten da. Hori bai, errore barra bertikalei erreparatuz igarri daiteke haien errore estatistikoak txikiagoak direla, ziurrenik energiaren estimatzaile aurreratuago bat erabili dutelako.



6.1. irudia: Lan honetakoak bezalako simulazioak eginez Mobley LAB ikerketa taldeak lortutako emaitzak [23]. FreeSolv datu-basean argitaratu dituzten balio guztietatik, lan honetako 29 molekulei dagozkienak bakarrik adierazi dira irudian. Errore estatistikoak 10 aldiz handitu dira, baina hala ere ia ez dira ikusten. Errore esperimentalek ez dute handipenik.

Lan honen hasieran Free Energy Perturbation ekuazioa erabili da energia askeak estimatzeko, baina gerora Bennett Acceptance Ratio metodo apur bat aurreratuagoaren bidez emaitzen errore estatistikoa txikitzea posible izan da. Bestelako erroreen jatorria ulertzea lan gehigarria izango litzateke, egindakoaren jarraipen posible bat. Hala ere, erroreen zenbait iturri identifika daitezke.

Batetik, simulazioak motzegiak izatea. Gerta daiteke egindako simulazioak intereseko egoerak lagintzeko behar bezain luzeak ez izatea. 5.6. irudian ikusten denez, kasu gehienetan 5 ns nahikoa izan da konbergentzia lortzeko, baina beharbada ibuprofeno edo ketoprofenoaren kasuan iraupen handiagoko simulazioek emaitza hobekak lortzea ahalbidetzen dute.

Bestetik, transformazio alkimikoak egiterakoan hasierako A eta B egoeren arteko I_N etapen aukerak eragina du. Lan honetan egindako simulazioetan 4.3. irudian agertzen

diren bitarteko 16 pausu erabili dira. Gutxiago edo gehiago kontsideratuta emaitzak nola aldatzen diren aztertzea interesgarria izan liteke, fase espazioaren zati garrantzitsuen gainezarmenaren arabera.

Azkenik, erabilitako energia potentzialaren modeloak berak dituen erroreak kontsideratu beharko lirateke. Force fielda bera izan daiteke errorean iturri nagusia, bai baliatzen dituen balio enprikogatik baita egiten diren hurbilketengatik ere. Esan bezala, errorean azterketa zorrotz batek lan osagarria eskatzen du, proiektu honen irismenetik kanpo geratu dena.

Aurreko guztia laburbilduz, transformazio alkimikoen erabilerak konputazionalki garestiegiak ez diren simulazioak gauzatzea ahalbidetzen du. Substantzia baten solbatatze energia askea esperimentalki neurtzeko, substantzia bera sintetizatu behar da lehenik, eta ondoren neurtu. Molekula askorekin egin behar bada hau, alternatiba konputazionalak denbora zein baliabide ekonomikoak aurrez ditzake.

Abantaila hau oraindik nabariagoa da lanaren sarreran aipatzen diren bezalako egoera konplexuagoetan. Proiektu honetan erraztasunagatik erabaki da solbatazio energia askeak kalkulatzeko, baina printzipioz erabilitako tresna berdinekin eginkizun zailagoei aurre egin dakieke. Horrelakoetan, bide konputazionala baliagarria izan daiteke funtzio jakin bat betetzen duten molekulen artean hautagairik interesgarrienak identifikatzeko. Laborategiko lana murriztuko luke honek, aukera guztien azterketa esperimentala egin beharrean nahikoa izango baitzen hautagai horiena egitea.

ERREFERENTZIAK

- [1] M. R. Shirts, "Best practices in free energy calculations for drug design", *Computational drug discovery and design*, 425-467 (2012).
- [2] M. R. Shirts, D. L. Mobley eta S. P. Brown, "Free-energy calculations in structure-based drug design", non *Drug Design: Structure- and Ligand-Based Approaches* (Cambridge University Press, 2010), orrk. 61-86.
- [3] Y. Khalak et. al., "Alchemical absolute protein-ligand binding free energies for drug design", *Chem. Sci.* **12**, 13958-13971 (2021).
- [4] C. Chipot eta A. Pohorille, *Free Energy Calculations. Theory and Applications in Chemistry and Biology*, 1. edizioa, Springer Series in Chemical Physics (Springer Berlin, 2007).
- [5] A. R. Leach, *Molecular Modelling: Principles and Applications*, 2. edizioa (Prentice Hall, 2001).
- [6] J. McMurry, *Organic Chemistry*, 8. edizioa (Brooks/Cole, 2012), orrk. 96-99.
- [7] P. Atkins eta J. De Paula, *Atkins' physical chemistry*, 8. edizioa (Oxford university press, 2006).
- [8] D. J. Griffiths, *Introduction to electrodynamics*, 4. edizioa (Pearson, 2013), orrk. 151-158.
- [9] P. Lorrain eta D. R. Corson, *Electromagnetic fields and waves*, 3. edizioa (W. H. Freeman eta Company, 1988), orrk. 84-98.
- [10] NIST Computational Chemistry Comparison and Benchmark Database, *Experimental data for N₂*, (2024/05/03an kontsultatua), <http://cccbdb.nist.gov/>.
- [11] F. London, "The general theory of molecular forces", *Transactions of the Faraday Society* **33**, 8-26 (1937).
- [12] T. Darden, D. York eta L. Pedersen, "Particle mesh Ewald: An $N \log(N)$ method for Ewald sums in large systems", *The Journal of chemical physics* **98**, 10089-10092 (1993).
- [13] P. Heckbert, "Fourier transforms and the fast Fourier transform (FFT) algorithm", *Computer Graphics* **2**, 15-463 (1995).
- [14] University of Illinois at Urbana-Champaign. Theoretical and Computational Biophysics Group, *Tutorial: Using VMD*, (2024/05/20an kontsultatua), (2019) <http://www.k.s.uiuc.edu/Training/Tutorials/vmd/vmd-tutorial.pdf>.
- [15] L. Verlet, "Computer experiments on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules", *Physical review* **159**, 98-103 (1967).
- [16] D.-Z. Li eta X.-B. Yang, "On numerical stationary distribution of overdamped Langevin equation in harmonic system", *Chinese Physics B* **32**, 080501 (2023).
- [17] Z. Zhang et. al., "Unified efficient thermostat scheme for the canonical ensemble with holonomic or isokinetic constraints via molecular dynamics", *The Journal of Physical Chemistry A* **123**, 6056-6079 (2019).
- [18] P. Eastman et. al., "OpenMM 7: Rapid development of high performance algorithms for molecular dynamics", *PLoS computational biology* **13** (2017).

- [19] DIPC Technical Documentation, *Hyperion system*. (2024/06/06an kontsultatua), (2023) <https://scc.dipc.org/docs/systems/hyperion/>.
- [20] BCB, *Bilbao Computational Biophysics*, (2024/06/06an kontsultatua), (2023) <https://bcb.eus/>.
- [21] S. Miyamoto eta P. A. Kollman, "Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models", *Journal of computational chemistry* **13**, 952-962 (1992).
- [22] B. Hess et. al., "LINCS: A linear constraint solver for molecular simulations", *Journal of computational chemistry* **18**, 1463-1472 (1997).
- [23] D. L. Mobley eta J. P. Guthrie, "FreeSolv: a database of experimental and calculated hydration free energies, with input files", *Journal of computer-aided molecular design* **28**, 711-720 (2014).
- [24] A. S. Mey et. al., "Best practices for alchemical free energy calculations", *Living journal of computational molecular science* **2** (2020).