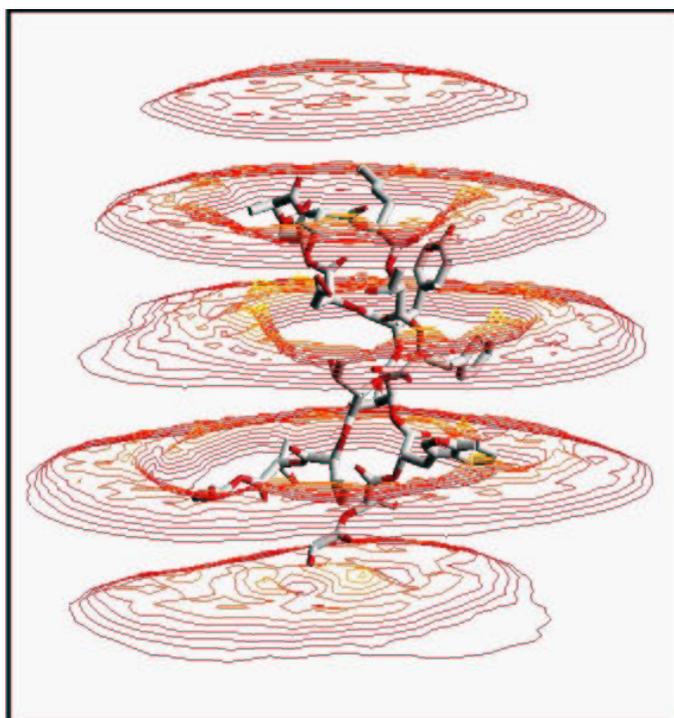


Danilo Roccatano

Dipartimento di Chimica, Ingegneria Chimica e Materiali.
Università de L'Aquila.
email: roccata@caspur.it

INTRODUZIONE ALLA DINAMICA MOLECOLARE



Lex. I.

Corpus omne perseverare in statu suo quiescendi vel movendi
uniformiter in directum, nili quatenus a viribus impressis cogitur
statum illum mutare.

Lex. II.

Mutationem motus proportionalem esse vi motrici impressae, et fieri
fecundum lineam rectam qua vis illa imprimitur.

Lex. III.

Actioni contrariam semper et equalem esse reactionem: sive
corporum duorum actiones in se mutuo semper esse aequales et in
partes contrarias dirigi.

Isaac Newton.

Philosophiae Naturalis Principia Mathematica. London, 1686.

Indice

Prefazione	i
1 La Dinamica Molecolare	1
1.1 Storia della Dinamica Molecolare	2
1.2 Limiti e approssimazioni	3
1.3 Struttura di un programma di DM	5
2 Il campo di Forze	7
2.1 Interazioni di legame	8
2.1.1 Vibrazioni di legame	8
2.1.2 Vibrazione degli angoli di valenza	9
2.1.3 Interazioni a 4 corpi	9
2.2 Interazioni di non legame	10
2.2.1 Interazioni di van der Waals	11
2.2.2 Interazioni Elettrostatiche	13
2.2.3 Il legame a idrogeno	14
2.3 I termini speciali	14
2.3.1 Potenziali di vincolo delle posizioni	15
2.3.2 Potenziali di vincolo delle distanze	15
2.3.3 Potenziali di vincolo sui fattori di struttura cristallografici	16
2.4 Campi di forze e librerie di parametri	16
2.5 Ottimizzazione della geometria molecolare	18
2.6 Appendice	20
2.6.1 Calcolo delle cariche	20
2.6.2 Calcolo di un profilo di energia potenziale per rotazione di un legame.	23
3 Equazioni del moto	27
3.1 Metodi d'integrazione delle equazioni del moto	29
3.2 Passo d'integrazione e vincoli olonomici	31
4 Trattamento delle interazioni a lungo raggio	35
4.1 Il Cutoff	35
4.2 Metodi basati sulle somme di Ewalds	37

5	Simulazione di sistemi in fase condensata	39
5.1	Condizioni al contorno	39
5.2	Calcolo della temperatura	41
5.3	Calcolo della pressione	42
5.4	Metodi per il controllo della temperatura e della pressione	43
5.4.1	Controllo della temperatura	44
5.4.2	Metodo isoterma	44
5.4.3	Metodo di Berendsen	44
5.4.4	Metodi di Nosé-Hoover	45
5.4.5	Controllo della pressione	45
6	Analisi delle traiettorie	47
6.1	Ipotesi Ergotica	47
6.2	Determinazione delle proprietà termodinamiche	48
6.2.1	Entalpia di evaporazione	48
6.2.2	Calcolo di funzioni di risposta termodinamiche	49
6.2.3	Capacità termica	49
6.2.4	Compressibilità isoterma	51
6.2.5	Coefficiente di dilatazione termico	51
6.2.6	Calcolo delle differenze di energia libera	51
6.3	Determinazione delle proprietà strutturali	54
6.3.1	Funzioni di distribuzione atomiche	54
6.4	Determinazione di proprietà dinamiche	55
6.4.1	Calcolo della costante di diffusione	55
6.4.2	Il tempo di residenza e numero d'idratazione	57
7	Calcolo parallelo e la DM	61
7.1	I calcolatori paralleli	61
7.2	Algoritmi per simulazioni di DM su calcolatori paralleli	62
7.2.1	Decomposizione geometrica.	62
7.2.2	Decomposizione dei dati e il <i>systolic loop</i>	63
8	Simulazione di un sistema molecolare	65
8.0.3	Il programma Gromacs	65
8.0.4	Struttura di Gromacs	66
8.0.5	Schema operativo per la simulazione di Dinamica Molecolare di un sistema molecolare	66
8.0.6	Le coordinate atomiche della molecola.	69
8.0.7	Creazione della topologia per molecole semplici	70
8.0.8	Creazione della topologia per biopolimeri	73
8.0.9	Minimizzazione della struttura iniziale nel vuoto	75
8.0.10	Simulazione della fase liquida	77

Prefazione

Nel 1648 Isaac Newton diede alle stampe la sua prima edizione dei Principia Mathematica, uno dei più grandi capolavori scientifici di tutti i tempi. Alla pagina 12 di questa summa scientifica si trovano enunciate le famose tre leggi che portano il suo nome, dalle quali ha avuto inizio la fisica-matematica classica. A 350 anni da quell'evento, le stesse leggi, usate per descrivere il moto delle stelle e dei pianeti, ci tornano ancora utili per semplificare la descrizione del mondo atomico.

Nei primi decenni dello scorso secolo, la nascita della meccanica quantistica, ha segnato l'inizio dell'esplorazione teorica della realtà a livello atomico. L'equazione di Schrödinger, al pari delle equazioni di Newton, ha permesso di riassumere in forma matematicamente elegante e concisa, le brillanti intuizioni teoriche e i dati sperimentali accumulati nei precedenti decenni. In linea di principio, questa equazione, pur potendo essere usata per descrivere il comportamento chimico-fisico di qualunque sistema molecolare, risulta impossibile da risolvere analiticamente, quando il numero degli elettroni in gioco è superiore a due. La nascita dei calcolatori elettronici nel secondo dopo guerra, ha permesso di risolvere numericamente questa equazione per sistemi poliatomici. Tuttavia, anche con il continuo e rapido sviluppo della potenza degli elaboratori, l'equazione di Schrödinger non può essere ancora usata per descrivere la dinamica di sistemi molecolari formati da centinaia o migliaia di atomi, quali le macromolecole biologiche. Per questo motivo, si è cercato di ridurre, attraverso opportune approssimazioni, la descrizione del comportamento dinamico degli atomi a un modello classico, in cui gli elettroni (gli elementi quantistici) non sono esplicitamente considerati ma si considera il loro effetto medio. Nasce così, circa 40 anni fa, la Dinamica Molecolare (DM). Grazie al continuo sviluppo di calcolatori sempre più veloci, questa tecnica si è progressivamente sviluppata estendendo il suo campo di applicazione dalla simulazione di fluidi semplici a complessi sistemi biologici, quali proteine e membrane cellulari. In tal modo, la DM sta diventando un potente e flessibile strumento d'indagine negli ambiti più disparati della biologia strutturale e della scienza dei materiali.

In questo libro verranno fornite le basi teoriche e applicative di questa tecnica.

Capitolo 1

La Dinamica Molecolare

La Dinamica Molecolare (DM) è una tecnica della chimica computazionale che consente di simulare il moto dei singoli atomi in sistemi atomici o molecolari. La descrizione completa del moto di un sistema di particelle può essere fatta usando l'equazione di Schrödinger dipendente dal tempo

$$\hat{H}\Psi(\mathbf{r}, t) = -i\frac{\hbar}{2\pi}\frac{\partial}{\partial t}\Psi(\mathbf{r}, t) \quad (1.1)$$

dove l'operatore di Hamilton \hat{H} è composto dal termine dell'energia cinetica $-(\hbar^2/8\pi^2m_i)\partial^2/\partial\mathbf{r}_i^2$ per ognuna delle particelle i (con massa m_i) del sistema e da un termine $V(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) \equiv V(\mathbf{r})$ che rappresenta l'interazione tra le particelle. Il comportamento del sistema è dato dalla funzione d'onda $\Psi(\mathbf{r}, t)$ che soddisfa l'equazione 1.1. L'uso di una descrizione con questo livello di accuratezza, per sistemi molto grandi, non è ancora possibile, per cui si deve ricorrere a delle approssimazioni. Il comportamento dinamico di un sistema molecolare può essere decomposto nel moto dei nuclei e in quello degli elettroni. Usando l'approssimazione di Born-Oppenheimer i due moti possono essere separati. In questo modo il sistema molecolare può essere descritto come un sistema di masse puntiformi che si muovono in virtù di un campo di potenziale *effettivo* prodotto dagli elettroni. Se si escludono gli atomi di H ed He, il moto dei nuclei può essere descritto con ragionevole accuratezza dalla meccanica classica. Con queste approssimazioni è possibile ridurre la complessità dei calcoli (altrimenti richiesta per una descrizione quantistica delle interazioni tra le particelle) e quindi simulare sistemi molecolari di grandi dimensioni per tempi dell'ordine dei nanosecondi. La possibilità di descrivere sistemi molecolari complessi spiega la grande diffusione della DM nell'ambito della biochimica strutturale, per lo studio delle proprietà dinamiche e strutturali di proteine, e della scienza dei materiali. Infatti, è possibile utilizzare i risultati ottenuti dalle simulazioni per comprendere il comportamento dinamico e strutturale in condizioni non accessibili a tecniche sperimentali. Nel caso di proteine, è per esempio possibile studiare come la loro struttura vari in conseguenza dell'idratazione da parte del solvente, evidenziare il meccanismo di cambiamento conformazionale dovuto al

legame con un substrato, esplorare il modo in cui un farmaco possa interagire con il sito attivo di una proteina, e così via.

Lo sviluppo della DM è fortemente influenzato dalla costruzione di calcolatori sempre più veloci e dalla introduzione di nuove metodologie simulate che consentano di migliorare l'accuratezza dei risultati. In tal senso, la costruzione di calcolatori basati su architetture parallele, ha aperto nuovi orizzonti alla DM, consentendo lo studio di sistemi molecolari sempre più complessi e per tempi di simulazione sempre più lunghi. Nel paragrafo che segue daremo un racconto della storia di questa tecnica per meglio comprenderne il suo sviluppo.

1.1 Storia della Dinamica Molecolare

Le prime simulazioni di dinamica molecolare furono effettuate negli anni '50, agli albori dell'era informatica, da Alder e Wrainwright. [2] Questi svilupparono e usarono il metodo per studiare fluidi semplici mediante modelli che rappresentavano gli atomi come dischi e sfere rigide.* Il successivo uso di funzioni di potenziale più realistiche ha permesso di ottenere dati simulati direttamente confrontabili con dati sperimentali, mostrando come la DM possa essere usata come un valido strumento di indagine delle proprietà microscopiche di sistemi reali. Le prime simulazioni di questo tipo furono effettuate da Rahman e Verlet [70]: in queste simulazioni fu utilizzato un potenziale di tipo Lennard-Jones per descrivere le interazioni atomiche dell'argon allo stato liquido. Lo sviluppo di calcolatori più veloci e versatili ha consentito la simulazione di sistemi fisici più complessi. Si iniziarono, quindi, a studiare sistemi come l'acqua [71], alcani flessibili [74] e piccole proteine, come, per esempio, la BPTI (Bovine Pancreatic Tripsin Inhibitor) [59]. Inoltre, nello stesso periodo, furono sviluppati nuovi metodi per calcolare in modo più accurato le interazioni elettrostatiche nello studio di sistemi ionici [98]. La simulazione di biomolecole è iniziata alla fine degli anni '70, ed i successi ottenuti nel riprodurre dati sperimentali strutturali di proteine e macromolecole in generale, hanno portato, negli anni successivi, a una grande diffusione della DM nell'ambito degli studi di biochimica strutturale. Il continuo aumento della potenza dei calcolatori, ha consentito, in questo specifico settore, di simulare biomolecole sempre più grandi e, con l'inclusione di modelli di solvente, in condizioni sempre più realistiche [87]. Attualmente la tecnica è in continuo sviluppo e la complessità dei sistemi studiati, così come la lunghezza delle simulazioni, è legata alla potenza dei calcolatori disponibili: sistemi contenenti fino a 10^5 atomi per tempi fino a decine di nanosecondi sono alla portata delle attuali capacità di calcolo. La DM è usata con successo

*Per dischi o sfere rigide si intendono particelle che danno luogo a collisioni perfettamente elastiche in due o tre dimensioni spaziali. Particelle di questo tipo interagiscono con un potenziale *hard core*, ovvero descritto dalle seguenti equazioni:

$$\begin{aligned} v(r) &= \infty & r \leq r_s \\ &= 0 & r > r_s \end{aligned} \tag{1.2}$$

con r_s il raggio della sfera, che produce collisioni elastiche.

Tabella 1.1: Livelli di approssimazione di modelli utilizzati in DM.

Tipo di Modello	Gradi di libertà		Esempio di proprietà predicibile
	considerati	rimossi	
<i>Quantistico</i>	nuclei ed elettroni	nucleoni	reazioni chimiche
<i>Con atomi polarizzabili</i>	atomi, dipoli	elettroni	legami di substrati carichi
<i>Con atomi non polarizzabili</i>	atomi di soluti e solvente	dipoli	idratazione
<i>Con solvente implicito</i>	atomi del soluto	atomi del solvente	struttura terziaria di macromolecole
<i>Con gruppi di atomi trattati come sfere rigide</i>	tutti i gruppi atomici	i singoli atomi	L' avvolgimento delle proteine

per studi strutturali e dinamici di gas [79], liquidi [69], fibre polimeriche [13], polisaccaridi [51], membrane [21], acidi nucleici (DNA e RNA) [77, 32] e proteine [35]. Le simulazioni permettono di studiare processi chimici e biologici, di predire le proprietà di proteine modificate e le interazioni tra piccole molecole e macromolecole [72, 90]. Nel campo dell'ingegneria delle proteine e nello studio dei farmaci, le tecniche di simulazione stanno avendo una grande importanza, infatti, le simulazioni di DM permettono di studiare processi chimici e biologici, di predire le proprietà di proteine modificate e le interazioni tra piccole molecole e macromolecole [72, 90]. Lo sviluppo esponenziale della potenza dei calcolatori (un ordine di grandezza nella velocità di calcolo ogni 5 anni), dovuto tra l'altro allo sviluppo di calcolatori paralleli, consentirà nei prossimi anni di migliorare i risultati ottenuti introducendo modelli sempre più accurati. Questo renderà possibile lo studio di processi molto complessi che coinvolgono membrane, complessi enzimatici e reazioni enzimatiche.

1.2 Limiti e approssimazioni

Come tutte le tecniche di calcolo con cui si studiano modelli di sistemi fisici reali anche la DM è soggetta ad assunzioni e approssimazioni che ne limitano l'accuratezza. Pertanto è necessario conoscere e valutare molto bene questi limiti per evitare di sovrastimare le capacità del metodo e quindi i risultati che questo può fornire. I problemi fondamentali che si devono considerare nelle simulazioni di DM si possono riassumere in due aspetti fondamentali:

1. L'accuratezza del modello molecolare che viene usato per simulare il sistema.

I modelli fisico-matematici che descrivono la materia possono essere classificati in base al livello di dettaglio con cui descrivono il sistema che si sta

studiando. Nella Tabella 1.2 sono riportati i vari livelli di approssimazione in ordine di complessità decrescente. Aumentando la complessità del fenomeno da studiare, diminuisce il livello di dettaglio con cui il sistema viene descritto attraverso l'omissione della esplicita rappresentazione di opportuni gradi di libertà. Come è stato già detto nel precedente paragrafo, quando si escludono le reazioni chimiche, le basse temperature o i dettagli del moto degli atomi di idrogeno, si può ragionevolmente assumere che il moto degli atomi del sistema possa essere descritto dalle leggi della meccanica classica (seconda riga). Nella DM classica, le interazioni tra gli atomi sono descritte da un potenziale di *interazione effettiva*. In questi potenziali, l'effetto medio del grado di libertà elettronico omesso è stato incorporato nel grado di libertà atomico esplicitamente presente nel modello.

2. Le dimensioni dello spazio delle fasi accessibile al sistema simulato.

La simulazione di un sistema molecolare, a temperature maggiori dello zero assoluto, genera un insieme di configurazioni rappresentative di un particolare insieme statistico (vedi Capitolo VIII). Le proprietà termodinamiche di tale sistema sono definite dal loro valor medio, valutato sull'insieme delle configurazioni generate. Per l'ipotesi ergodica, la media su tali traiettorie su un tempo infinito:

$$\langle P \rangle = \lim_{\tau \rightarrow +\infty} \frac{1}{\tau} \int_{t_0}^{t_0+\tau} P(p, q) d\tau \quad (1.3)$$

dove il tempo totale durante il quale la quantità F è misurata varia da t_0 a $t_0 + \tau$, equivale alla media effettuata sull'insieme statistico, [3, 29]. L'applicazione di questa ipotesi in modo rigoroso è praticamente impossibile per via della traiettoria di lunghezza infinita. Tuttavia, se lo spazio configurazionale del sistema non è molto esteso (sistema con gradi di libertà ridotti), il tempo necessario a esplorare buona parte dello spazio configurazionale del sistema potrebbe diventare accessibile alla simulazione. Pertanto, per rendere possibile un'efficace esplorazione di questo spazio nei tempi di simulazione attualmente realizzabili, si introducono delle approssimazioni che riducono il numero di gradi di libertà del sistema (vedi Tabella 1.2), semplificando in tal modo la sua descrizione. È ovvio che tali approssimazioni sono tanto più valide quanto più è realistico assumere che la proprietà di cui si vuole conoscere il valore medio non sia legata ai gradi di libertà esclusi.

1.3 Struttura di un programma di DM

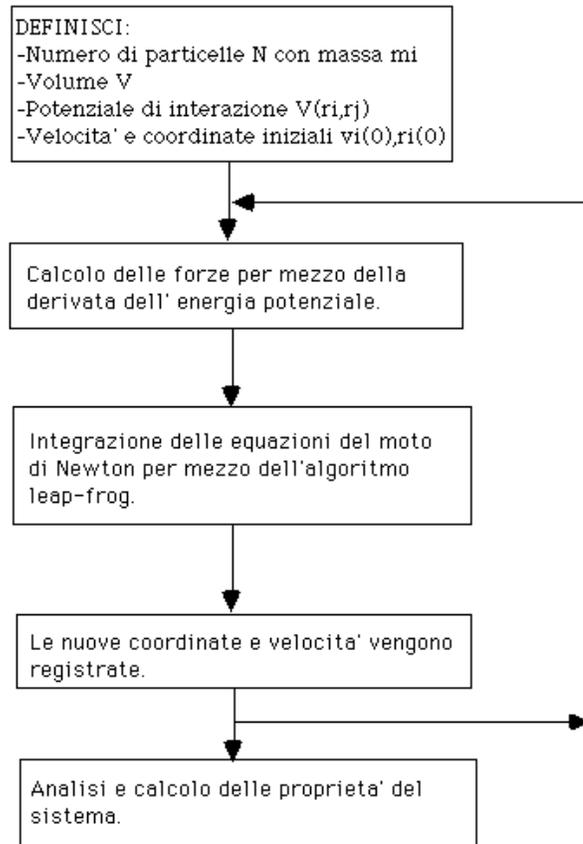


Figura 1.1: Schema a blocchi di un programma di MD.

Nella figura 1.1 è riportato in modo schematico la struttura di un programma di DM. Dopo la lettura delle coordinate iniziali e la definizione dei parametri per la simulazione, si procede dapprima alla generazione delle velocità iniziali (ottenute da una distribuzione di Maxwell-Boltzmann alla temperatura desiderata) e quindi si entra in un ciclo di calcolo la cui lunghezza dipende dal tempo di simulazione che si vuole ottenere. In questo ciclo vengono prima calcolati i potenziali e quindi, dalle derivate di quest'ultimi, le forze che agiscono su ciascun atomo. Le forze vengono quindi integrate per ottenere le posizioni e le velocità

al tempo successivo. Coordinate, velocità, forze e potenziali vengono memorizzate su disco con una certa frequenza che dipende dalla accuratezza con cui si vogliono ottenere informazioni sulla traiettoria. Nei capitoli successivi verranno approfonditi gli aspetti teorici e pratici dei vari blocchi di questo schema.

Capitolo 2

Il campo di Forze

La funzione di potenziale usata nella DM per descrivere le interazioni tra gli atomi del sistema è chiamata *campo di forze*. Si tratta di funzioni analitiche che dipendono dalle coordinate atomiche \mathbf{r}_i del sistema e da un numero M di parametri $s \equiv (s_1, s_2, \dots, s_M)$. Sono costituite da una somma di termini ognuno dei quali descrive un tipo di interazione classificata in base al numero di atomi coinvolti.

$$V(\mathbf{r}, s) = \underbrace{\sum_{i,j} V^{(2)}(\mathbf{r}_i, \mathbf{r}_j; s)}_{\text{Inter. a 2 atomi}} + \dots + \underbrace{\sum_{\text{tutti}} V^{(N)}(\mathbf{r}; s)}_{\text{Inter. a N atomi}} \quad (2.1)$$

Un esempio di *campo di forza*, utilizzato in molti programmi per simulazioni di DM, è il seguente:

$$V(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N; s) = \left. \begin{aligned} & \underbrace{\sum_{\text{legami}} \frac{1}{2} K_b (b - b_0)^2}_{V_s} \\ & + \underbrace{\sum_{\text{angoli}} \frac{1}{2} K_\theta (\theta - \theta_0)^2}_{V_b} \\ & + \underbrace{\sum_{\text{diedri impropri}} \frac{1}{2} K_\xi (\xi - \xi_0)^2}_{V_{imp}} \\ & + \underbrace{\sum_{\text{diedri}} K_\phi [1 + \cos(n\phi - \delta)]}_{V_{pr}} \end{aligned} \right\} \text{Interazioni di legame} \quad (2.2)$$

$$\left. \begin{aligned}
 &+ \underbrace{\sum_{coppie(i,j)} \left(\frac{C_{ij}^{12}}{r_{ij}^{12}} - \frac{C_{ij}^6}{r_{ij}^6} \right)}_{V_{vdw}} \\
 &+ \underbrace{\sum_{coppie(i,j)} \left(\frac{q_i q_j}{4\pi\epsilon_0\epsilon_r r_{ij}} \right)}_{V_{ei}} \\
 &+ \text{Termini Speciali}
 \end{aligned} \right\} \text{Interazioni di non legame} \quad (2.3)$$

2.1 Interazioni di legame

Sono descritte da termini a 2, 3 e 4 corpi, generalmente armonici, che contengono costanti ottenute da dati sperimentali cristallografici e spettroscopici.

2.1.1 Vibrazioni di legame

Il primo termine (V_s) rappresenta l'energia di vibrazione dei legami. Si usa un potenziale armonico in cui b_0 , che rappresenta la lunghezza del legame corrispondente al minimo di energia, e la costante di forza K , dipendono dal tipo di

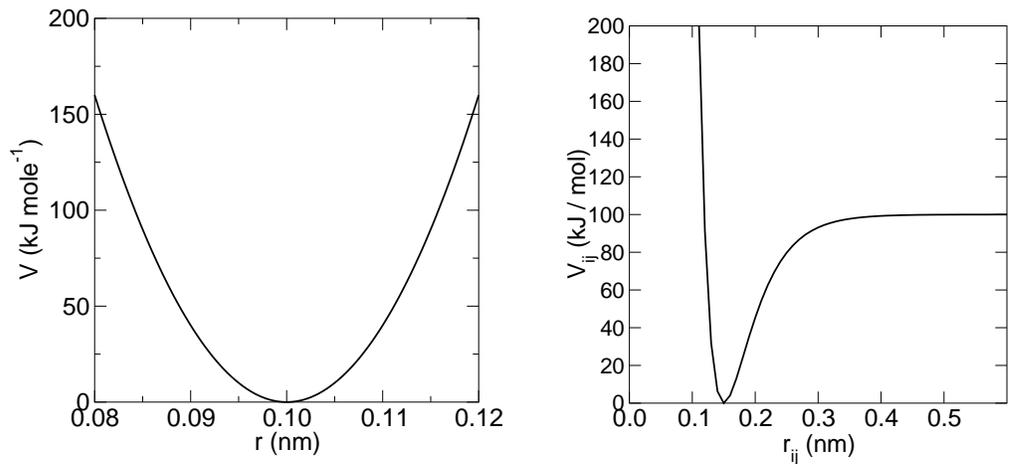


Figura 2.1: Curve di potenziale per le vibrazione di legame: armonico (sinistra); Morse (destra).

In alcuni campi di forze si utilizza la funzione di Morse:

$$V_M(b) = \sum_{legami} D_e \left\{ e^{[-k_b(b-b_0)]} - 1 \right\}^2 \quad (2.4)$$

dove il parametro D_e rappresenta l'energia di dissociazione del legame e la costante $k_b = \omega\sqrt{\mu/2D_e}$, dove ω è la frequenza di vibrazione e μ la massa ridotta. Questo tipo di potenziale descrive meglio vibrazioni su valori di b più ampi e fornisce risultati più accurati quando si vogliono riprodurre gli spettri vibrazionali di sistemi molecolari.

2.1.2 Vibrazione degli angoli di valenza

Il secondo termine nella 2.2 si riferisce all'energia dovuta alla deformazione dell'angolo di valenza (interazione a tre corpi). Anche questo ha la forma di potenziale armonico dove θ_0 rappresenta l'angolo di valenza di riferimento e K_θ la costante di forza. Esiste un gran numero di varianti dell'espressione 2.2. Alcuni

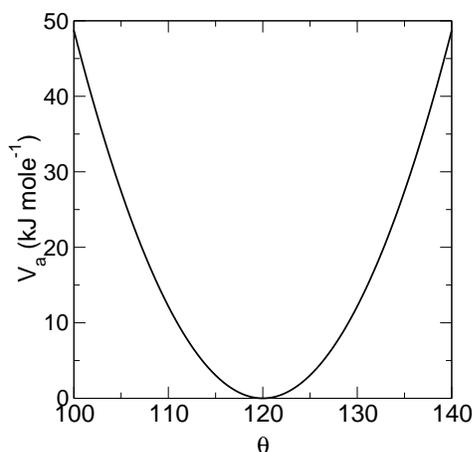


Figura 2.2: Potenziale armonico di vibrazione dell'angolo di legame.

campi di forza prevedono termini misti tipo $K_{b\theta}[b - b_0][\theta - \theta_0]$ che accoppiano direttamente variazioni della lunghezza dei legami e degli angoli di legame. [19] Altri usano termini più complessi per descrivere gli angoli diedri. [61, 52]

2.1.3 Interazioni a 4 corpi

Due termini sono utilizzati per descrivere le interazioni a quattro corpi.

Vibrazione di angoli diedri impropri.

Il primo viene usato per descrivere i cosiddetti *angoli diedri impropri*, che sono introdotti per correggere due artefatti: la deformazione delle geometrie tetraedriche e planari. La deformazione delle geometrie tetraedriche è dovuta al fatto che nella maggior parte dei programmi di DM, gli idrogeni apolari (cioè

quelli legati agli atomi di carbonio) non vengono trattati esplicitamente, ma viene incluso il loro effetto modificando le funzioni di potenziale degli atomi a cui sono legati. In tal modo un carbonio chirale, legato a un idrogeno, viene simulato con solo tre legami; questo comporta che durante la simulazione si possono avere delle inversioni di chiralità rendendo, quindi, necessario l'uso di questo ulteriore termine di potenziale per vincolare la geometria del sito. L'angolo diedro improprio, A-X-Y-B, è definito come l'angolo tra il piano che passa per gli atomi A, X e Y e il piano passante per gli atomi X, Y e B.

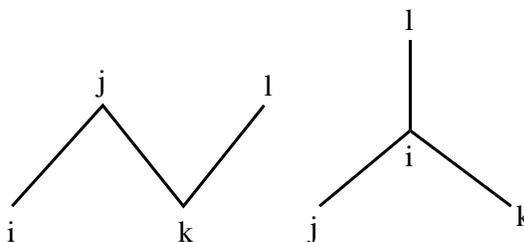


Figura 2.3: Definizione dei diedri Impropri.

Diedri propri.

L'altro termine per le interazioni a quattro atomi è quello relativo ai normali angoli diedri ed è una funzione sinusoidale, K_ϕ è la costante di forza, n la molteplicità dei minimi e δ la fase.

Parametri per le interazioni di legame.

Per determinare i parametri per i campi di forza si possono seguire due strade. Il metodo più elegante è tarare questi parametri con i risultati di calcoli quantistici *ab-initio* su piccoli aggregati molecolari. I campi di forza ottenuti in tal modo, molto spesso, non danno risultati soddisfacenti date le molte approssimazioni che devono essere fatte in questo tipo di procedura. In alternativa è possibile tarare i parametri del campo di forza con dati sperimentali quali: strutture cristallografiche, energie e dinamiche di reticolo, dati spettroscopici, proprietà di liquidi tipo densità e entalpia di vaporizzazione, energia libera di solvatazione, dati di RMN, ecc. [46, 53, 27, 20] In questo modo viene garantita una maggiore riproducibilità delle proprietà macroscopiche di massa di liquidi, in quanto i parametri sperimentali tengono conto implicitamente dell'effetto del campo medio delle molecole del solvente.

2.2 Interazioni di non legame

L'ultimo termine nella equazione 2.2 descrive le interazioni tra coppie di atomi non legati. Queste sono le interazioni di van der Waals e Coulombiane tra gli atomi i e j con carica q_i e q_j alla distanza r_{ij} .

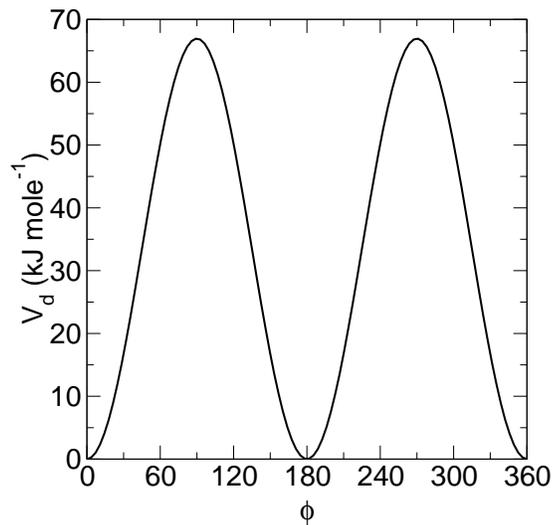


Figura 2.4: Andamento della funzione di potenziale di rotazione intorno a un diedro proprio.

2.2.1 Interazioni di van der Waals

Una delle piú utilizzate espressioni per la descrizione del potenziale d'interazione di non legame é il potenziale di Lennard-Jones. Il potenziale di Lennard-Jones é formato da due termini che descrivono una interazione di tipo repulsivo e una di tipo attrattivo:

$$V_{LJ} = \underbrace{\frac{C^{12}}{r^{12}}}_{\text{Termine repulsivo}} - \underbrace{\frac{C^6}{r^6}}_{\text{Termine attrattivo}} \quad (2.5)$$

C^{12} e C^6 sono delle costanti che possono essere ricavate da dati sperimentali o da calcoli quantomeccanici. In queste funzioni il termine di repulsione dovuto alle forze repulsive a corto raggio viene rappresentato con un termine C_{ij}^{12}/r^{12} , mentre nelle funzioni di Buckingham [16] si utilizza un termine esponenziale del tipo $b \exp[-ar]$, con a e b due costanti. Il termine C_{ij}^6/r^6 fornisce il contributo dovuto alle interazioni di tipo dispersivo o forze di London [57, 44]. I valori delle costanti C_{ij}^6 per la funzione di Lennard-Jones sono ottenuti per i singoli atomi

dalla equazione di Slater-Kirkwood, [58]

$$C_{ij}^6 = \frac{3}{2} \frac{\alpha_i \alpha_j}{\sqrt{\frac{\alpha_j}{N_j}} + \sqrt{\frac{\alpha_i}{N_i}}} = \frac{\alpha_i \alpha_j}{\alpha_j^2 C_{jj}^6 + \alpha_i^2 C_{ii}^6} \quad (2.6)$$

in cui N rappresenta il numero effettivo di elettroni e α le polarizzabilità atomiche: il termine nella seconda eguaglianza deriva dalla relazione: $N_i = 16(C_{jj}^6)^2 / 9\alpha_i^3$

I valori di C_{ij}^{12} si possono ottenere da misure sperimentali del secondo coefficiente del viriale (vedi Box) o da dati cristallografici e termodinamici.

L'equazione di stato di un gas reale può essere espressa nella forma del viriale

$$\frac{PV}{nRT} = 1 + \frac{B}{V} + \frac{C}{V^2} + \dots \quad (2.7)$$

dove B e C sono il secondo e il terzo coefficiente del viriale. In genere è sufficiente considerare solo il coefficiente B per avere una buona descrizione del comportamento ideale del gas.

Applicando la teoria cinetica dei gas, Rayleigh, ha trovato una espressione analitica che lega B al potenziale di interazione atomica:

$$B = 2\pi N \int_0^{\infty} \left(1 - \exp\left(-\frac{V(r)}{RT}\right)\right) r^2 dr \quad (2.8)$$

dove r è la distanza da una molecola scelta come origine. Sostituendo in $V(r)$ l'espressione della equazione di Lennard-Jones, si ottengono espressioni che permettono di determinare i parametri C_{jj}^6 e C_{jj}^{12} , dato che B può essere misurato a differenti temperature. In realtà, le soluzioni che si ottengono per i valori dei parametri non sono univoche. Per ridurre tale indeterminazione, solo il parametro C_{jj}^{12} viene determinato, mentre il parametro C_{jj}^6 si ottiene dalla equazione di Slater-Kirkwood.

Una forma molto usata del potenziale di Lennard-Jones, sostituisce a C^{12} e C^6 , i valori di σ ed ϵ che rappresentano la distanza che annulla il potenziale e il valore dell'energia nel punto di minimo della funzione (vedi Figura).

$$V_{LJ} = 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^6 \right] \quad (2.9)$$

I valori di σ ed ϵ relativi a due tipi di atomo, possono essere combinati per ottenere i corrispondenti parametri per l'interazione tra la coppia.

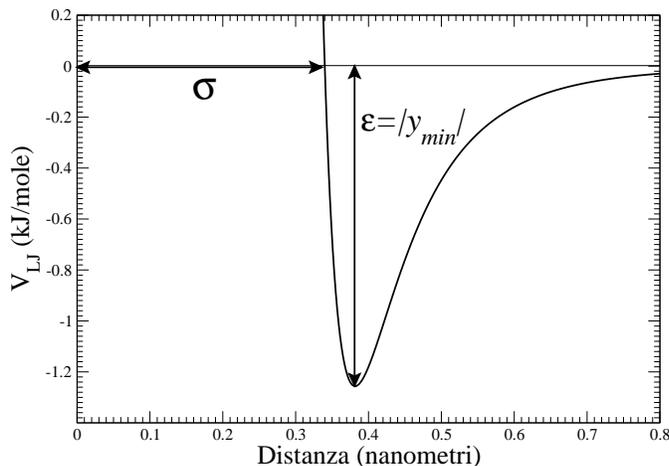


Figura 2.5: Curva di Lennard-Jones con i parametri.

Esistono diversi tipi di regole di combinazione, in una delle più usate i nuovi parametri si ottengono in questo modo:

$$\sigma_{ij} = \frac{\sigma_{ii} + \sigma_{jj}}{2} \quad (2.10)$$

$$\epsilon_{ij} = \sqrt{\epsilon_{ii}\epsilon_{jj}} \quad (2.11)$$

Le interazioni di non legame tra atomi legati covalentemente o separati da due legami non vengono generalmente calcolate. Nel caso, invece, di interazioni tra atomi separati da tre legami i parametri delle funzioni di LJ sono opportunamente ridotti per evitare le forti repulsioni che si avrebbero a distanze corte. I parametri C_{ij}^{12} di atomi uniti (CH-, CH₂-, CH₃-, carboni aromatici) sono valutati in modo da riprodurre le corrette densità cristalline di piccole molecole quando sono simulate a pressione costante.

2.2.2 Interazioni Elettrostatiche

L'interazione elettrostatica viene generalmente descritta attraverso un termine coulombiano. Ad ogni atomo del sistema viene quindi attribuita una carica atomica parziale che vengono calcolate con metodi quantomeccanici (vedi appendice 2.6.1). La scelta della costante dielettrica relativa ϵ_r e' ancora materia di discussione. Sono stati utilizzati valori che vanno da $\epsilon_r = 1$ [48, 84] a $\epsilon_r = 8$ [41], mentre altri prendono ϵ_r proporzionale alla distanza r_{ij} [59, 11, 95]. Queste approssimazioni sono fisicamente poco realistiche: la soluzione migliore a questo problema sarebbe l'inclusione nel potenziale della polarizzabilità atomica, che permetterebbe di ottenere il corretto campo elettrico in ogni punto del sistema

simulato. Per far questo occorre introdurre l'equazione di campo elettrico che tiene conto della presenza di dipoli indotti sui singoli atomi.

$$\Delta \mathbf{p}_i = \alpha_i \mathbf{E}_i = \alpha_i \sum_i^N \sum_{j \neq i}^N \mathbf{T}_{ij} (\mathbf{p}_j + \Delta \mathbf{p}_j) \quad (2.12)$$

dove N è il numero di atomi del sistema, \mathbf{p}_i sono il momento di dipolo e $\Delta \mathbf{p}_i$ il momento di dipolo indotto dell'atomo i -esimo, e

$$\mathbf{T}_{ij} = \frac{1}{4\pi\epsilon_0 r^3} \left(\frac{3\mathbf{r}_i \mathbf{r}_j}{r_{ij}^2} - 1 \right) \quad (2.13)$$

le precedenti equazioni danno luogo a un sistema di $3N$ equazioni in $3N$ incognite (ciascuna delle componenti x, y, z di ogni atomo) che può essere risolto algebricamente o iterativamente. Un altro modo elegante di introdurre la polarizzabilità nel sistema è quello di usare delle cariche puntiformi, localizzate su ciascun atomo e legate da tra loro da oscillatori di opportuna costante di forza (oscillatori di Drude), la loro deformazione in seguito alla interazioni con altri atomi produce una separazione delle cariche e quindi un dipolo indotto

2.2.3 Il legame a idrogeno

Un'altra importante interazione molecolare è il legame idrogeno. Il termine *legame idrogeno* viene usato per indicare quasi tutte quelle situazioni in cui le distanze internucleari, R_{H-X} e R_{H-Y} tra un atomo d'idrogeno H e due altri atomi X e Y , sono significativamente più piccole della somma dei loro raggi di van der Waals ($R_{H-X} < r_H^{vw} + r_X^{vw}$, e $R_{H-Y} < r_H^{vw} + r_Y^{vw}$). Il legame idrogeno dà origine a delle interazioni specifiche la cui energia è minore di quella dovuta alle sole forze di dispersione e mostra delle caratteristiche dipendenze angolari [56]. In alcuni campi di forza tale tipo di interazione viene riprodotta utilizzando opportuni valori per le cariche e per i parametri della funzione di LJ degli atomi coinvolti in questo tipo di legame [54, 48, 28, 36]. In altri campi di forza sono utilizzati, per garantire il legame idrogeno, speciali termini di potenziale [52, 11, 95]. Un tipico termine di potenziale è:

$$V(r) = \frac{A}{r^{12}} - \frac{C}{r^{10}} \quad (2.14)$$

che viene usato per descrivere l'interazione tra il donatore dell'atomo di idrogeno e l'atomo accettore. In altri potenziali (vedi paragrafo 2.4) la precedente funzione viene moltiplicata per un termine che tiene conto dell'orientamento tra gli atomi coinvolti.

2.3 I termini speciali

All'espressione 2.2 possono essere aggiunti, alcuni termini speciali che, vincolando alcuni gradi di libertà del sistema limitano, lo spazio conformazionale

esplorato sulla base dei dati strutturali ottenuti sperimentalmente. Questi termini costituiscono la base per l' affinamento delle strutture molecolari ottenute da misure di risonanza magnetica nucleare (RMN) o di diffrazione dei raggi X.

2.3.1 Potenziali di vincolo delle posizioni

Se nel corso di una simulazione si vogliono tenere bloccati le posizioni nello spazio di alcuni atomi del sistema, si può ricorrere all'uso di potenziali armonici del tipo:

$$V(\mathbf{r}) = \frac{1}{2} \sum_N (r_i - r_i^{ref})^2 \quad (2.15)$$

dove N indica il numero di atomi a cui applicare il potenziale di vincolo, r_i sono le coordinate attuali e r_i^{ref} quelle di riferimento dell'atomo i-esimo. Questi potenziali sono generalmente usati nella fase di equilibratura di una proteina, per consentire alle molecole del box di simulazione di equilibrarsi intorno al soluto senza perturbarne la struttura iniziale.

2.3.2 Potenziali di vincolo delle distanze

La spettroscopia RMN permette di ottenere informazioni sulle distanze interprotoniche (tramite l'effetto *Overhauser* interprotonico) e su angoli diedri (dalle costanti di accoppiamento J) e sugli spostamenti chimici.

Le intensità NOE (*Nuclear Overhauser Effect*) possono essere convertite in un insieme di limiti superiori r_{ij}^{ub} per le distanze tra gli atomi d'idrogeno i e j . Queste informazioni possono essere incluse nel campo di forza usando l'espressione: [49]

$$V(\mathbf{r}) = \frac{1}{2} \sum_{NOE} [MAX(0, \langle r_{ij}^{-1/3} \rangle - r_{ij}^{ub})]^2 \quad (2.16)$$

dove la funzione MAX fornisce il valore del maggiore tra i due argomenti.

I valori delle costanti di accoppiamento J possono essere incluse nel campo di forza usando la relazione [81]:

$$V_J(\mathbf{r}) = \frac{1}{2} K_J \sum_{\phi_i} [\langle J(\phi_i(\mathbf{r})) \rangle - J_i^{spe}]^2 \quad (2.17)$$

Le costanti di accoppiamento J dipendono dall'angolo di torsione ϕ_i formato dagli atomi per i quali viene misurato il valore di J .

Infine, gli spostamenti chimici σ possono essere vincolati ad assumere i valori misurati σ^{spe} usando la funzione [33]:

$$V_\sigma(\mathbf{r}) = \frac{1}{2} K_\sigma \sum_i [\langle \sigma_i(\mathbf{r}) \rangle - \sigma_i^{spe}]^2 \quad (2.18)$$

dove la sommatoria è effettuata su tutte le risonanze osservate. La media $\langle \dots \rangle$ presente nelle precedenti equazioni viene valutata sulle traiettorie.

2.3.3 Potenziali di vincolo sui fattori di struttura cristallografici

Nel caso in cui si disponga di dati strutturali ottenuti da misure di diffrazione dei raggi X (ampiezza dei fattori di struttura $F(hkl)$ e, se disponibili, le fasi $\alpha(hkl)$) il termine da aggiungere è:

$$V_X = \frac{1}{2} K_X \sum_{hkl} [F_{calc}(hkl) - F_{obs}(hkl)]^2 \quad (2.19)$$

dove $F_{calc}(hkl)$ sono le ampiezze dei fattori di struttura calcolati dalla DM. [15, 14] I valori delle costanti di forza presenti in tali espressioni possono essere cambiati durante l' affinamento della struttura. L'aumento del valore delle costanti porterà a soddisfare maggiormente le osservazioni sperimentali, ma se i valori diventano molto elevati, la flessibilità della molecola viene fortemente ridotta [85] rendendo l' affinamento poco efficace.

L'uso dei valori medi nelle precedenti funzioni di vincolo le rende non conservative (poichè diventano dipendenti dal tempo), questo rende necessario dissipare il calore generato accoppiando il sistema a un bagno termico esterno [82].

2.4 Campi di forze e librerie di parametri

Esistono vari tipi di campi di forza che sono stati sviluppati da diversi gruppi di ricerca in tutto il mondo. Questi campi di forza sono costituiti da librerie di parametri ottimizzati per varie classi di molecole. Le librerie di parametri più comuni sono quelle ottimizzate per lo studio di proteine o acidi nucleici. Le librerie più usate sono le seguenti:

- **GROMOS:** il campo di forza GRONingen MOlecular Simulations è stato realizzato all'Università di Groningen da H.J.C. Berendsen e W.F. van Gunsteren. Nella sua attuale versione (GROMOS96) [89], viene utilizzato nel pacchetto di programmi per DM GROMOS96 e *Gromacs*.
- **OPLS:** L' *Optimized Potential for Liquid Simulations* è stato sviluppato da W.L. Jorgensen e J. Tirado-Rives [48] e inizialmente utilizzato per simulazioni Monte Carlo, è diventata una delle librerie più estese e dettagliate. Esistono centinaia di diversi tipi atomici ottimizzati per differenti classi di molecole organiche. Le caratteristiche di questa libreria sono:
 - Le cariche iniziali per la procedura di ottimizzazione dei parametri sono ottenute da calcoli quantomeccanici (set di base 6-31G* o 6-31+G* per anioni).
 - I parametri iniziali della funzione di Lennard-Jones sono gli stessi quelli usati per gli stessi tipi di atomi in molecole simili dello stesso tipo.

- Un primo tentativo per migliorare le cariche parziali e i parametri di LJ, viene fatto modellando queste funzioni sui risultati energetici e strutturali ottenuti da calcoli QM su complessi acqua-molecola usando come set di base il 6-31G* o il 6-31+G*.
 - Per molecole neutre i parametri sono ulteriormente migliorati attraverso simulazioni del liquido puro. La densità, l'entalpia di vaporizzazione, la capacità termica a pressione costante e la costante di diffusione calcolate sono confrontate con i valori sperimentali. I valori delle cariche e dei parametri di Lennard-Jones vengono quindi modificati fino a riprodurre i dati in fase gassosa e nel liquido.
 - I parametri per molecole cariche sono verificati calcolando le energie libere di solvatazione e di legame.
 - Da notare che le funzioni per la descrizione della vibrazione di angoli e legami non sono proprie del OPLS poichè questo campo di forze venne inizialmente utilizzato per simulazioni Monte Carlo che non consentono il cambiamento di questi parametri. Per usare questo campo di forze in programmi di DM si utilizzano come funzioni di legame quelle di AMBER.
- **CHARMM:** Il campo di forza CHARMM (Chemistry at HARvard, Macromolecular mechanics) e l'omonimo programma, sono stati sviluppati da M. Karplus e B. R. Brooks [11, 1] alla Univerità di Harvard. Le caratteristiche di questo campo di forze possono essere riassunte nei seguenti punti:
 - Le cariche parziali vengono determinate in modo che il calore di sublimazione, i momenti di dipolo del composto nonché le energie d'interazioni, le geometrie in fase gassosa di complessi molecola-acqua, vengano riprodotte. Le energie di interazione ottenute dal calcolo QM al livello 6-31G*, vengono scalate di un fattore 1.16, cioè dello stesso fattore che esiste tra le energie di interazione delle molecole di acqua nel modello TIP3P e quelle ottenute dal calcolo QM allo stesso livello.
 - Le costanti di forza per le vibrazioni di legami, angoli e diedri sono tali da riprodurre le proprietà geometriche e vibrazionali del composto, ottenute da strutture cristallografiche, spettri IR e Raman e da calcoli QM nel vuoto.
 - **AMBER:** L' *Assisted Model Building using Energy Refinement* [95, 18] è il campo di forza presente nell'omonimo pacchetto di programmi per DM sviluppato da P. Kollman, all' UCSF. Le caratteristiche di questo campo di forza possono essere riassunte nei seguenti punti:
 - Le cariche originarie furono ottenute modellando le energie di interazione ottenute dal calcolo QM a livello STO-3G su composti modello.
-

- In versioni più recenti del campo di forza, le cariche sono state migliorate attraverso il calcolo della densità elettronica a livello 6-31G* e il fitting di questa con il metodo RESP (vedi appendice).
 - Come parametri di Lennard-Jones si usano quelli del campo di forza OPLS.
 - Le costanti di forza per le vibrazioni di legami e angoli vengono scelti in modo da riprodurre le frequenze sperimentali e teoriche dei modi normali di composti modello.
 - I *lone pairs* presenti sugli atomi di zolfo sono esplicitamente trattati.
- **CVFF:** Il *Consistent-Valence Force Field* è stato proposto da Dauber-Osguthorpe [19] e viene usato nel programma *Discover*. Nella figura 2.6 viene riportato la funzione di potenziale che descrive questo campo di forza. Una grande attenzione è stata data a creare un campo di forze che fosse in grado di riprodurre le caratteristiche vibrazionali di molecole di interesse biochimico. Questo è evidente dalla presenza nel potenziale, di numerosi termini di vibrazione accoppiati (legame-legame, legame-angolo, angolo-angolo, angolo-diedro, diedro-diedro).

2.5 Ottimizzazione della geometria molecolare

Il campo di forza, descritto nei precedenti paragrafi, viene usato in meccanica molecolare per costruire la superficie di energia conformazionale della molecola e quindi ricercare quelle conformazioni che sono energeticamente più stabili, cioè localizzate in punti di minimo. Questa procedura, come vedremo, è molto importante nella fase di partenza delle simulazioni di DM per evitare che configurazioni molecolari troppo stressate geometricamente, possano creare problemi negli algoritmi di integrazione delle equazione del moto.

Una configurazione molecolare si trova in un minimo locale della sua superficie di energia potenziale (V) se si verifica la seguente condizione:

$$\sum_{i=1}^N \frac{\partial V_i}{\partial \mathbf{r}_i} = 0 \quad (2.20)$$

dove N è il numero totale di atomi del sistema, V_i il potenziale dell'atomo i -esimo e \mathbf{r}_i la posizione dell'atomo i -esimo. La funzione V per N grande è molto complessa e corrugata per cui si comprende bene che la ricerca analitica dei minimi di V è impossibile, occorre pertanto ricorrere a metodi numerici.

Tra i metodi di ricerca dei minimi locali quelli più usati sono:

- Il metodo dello *Steepest descent*.
- Il metodo dei gradienti coniugati (*conjugate gradient*).

Nel primo caso, data una geometria molecolare, definita dalle coordinate \mathbf{r} , si calcola il gradiente $\mathbf{f} = \nabla V_i$ che definisce la direzione lungo cui l'energia

$$\begin{aligned}
E_{pot} = & \sum_b D_b [1 - e^{-\alpha(b-b_0)}] + \\
& \sum_\theta H_\theta (\theta - \theta_0)^2 + \\
& \sum_\phi H_\phi [1 + s \cos(n\phi)] + \\
& \sum_x H_x \chi^2 + \\
& \sum_b \sum_{b'} F_{bb'} (b - b_0)(b' - b'_0) + \\
& \sum_\theta \sum_{\theta'} F_{\theta\theta'} (\theta - \theta_0)(\theta' - \theta'_0) + \\
& \sum_b \sum_\theta F_{b\theta} (b - b_0)(\theta - \theta_0) + \\
& \sum_\phi F_{\phi\theta\theta'} \cos\phi (\theta - \theta_0)(\theta' - \theta'_0) + \\
& \sum_x \sum_{x'} F_{xx'} \chi\chi' + \\
& \sum \epsilon [(r^*/r)^{12} - 2(r^*/r)^6] + \\
& \sum q_i q_j / \epsilon r_{ij}
\end{aligned}$$

Figura 2.6: Funzione di potenziale per il campo di forze CVFF.

potenziale sta cambiando più rapidamente. La ricerca delle nuove coordinate consiste nel muovere le componenti delle coordinate di ciascun atomo, di un passo λ lungo il vettore $\mathbf{r} - \lambda \mathbf{f}$ fino a raggiungere un minimo di V lungo questa direzione. Partendo dalla nuova geometria \mathbf{r}_2 si calcola il nuovo gradiente, \mathbf{f}_2 nel nuovo punto e si muove le coordinate lungo questo gradiente verso un nuovo minimo. Si procede in questo modo fino a quando il valore del gradiente non si riduce ulteriormente (minimo locale). Il metodo è molto veloce ma diventa poco efficiente in prossimità del minimo.

Il metodo dei gradienti coniugati parte come quello dello *steepest descent*, ma dopo aver individuato il minimo \mathbf{r}_2 lungo la direzione $\mathbf{s}_1 = \mathbf{r}_2 - \lambda \mathbf{f}$, procede in una nuova ricerca nella direzione $\mathbf{s}_2 = -\mathbf{f}_2 - \beta \mathbf{s}_1$ dove $\beta = \frac{|\mathbf{f}_2|^2}{|\mathbf{f}_1|^2}$. In questo modo la direzione di ricerca nel punto \mathbf{r}_k non è più il gradiente di \mathbf{f}_k ma la direzione data dalla combinazione lineare del nuovo gradiente e quello ottenuto

dalla precedente direzione di ricerca:

$$\mathbf{s}_k = -\mathbf{f}_k + \beta \mathbf{s}_{k-1} \quad (2.21)$$

con $\beta = \frac{|\mathbf{f}_k|^2}{|\mathbf{f}_{k-1}|^2}$. Pertanto questo metodo, tiene conto del percorso di minimizzazione che consenta, al contrario del metodo dello *steepest descent*, di minimizzare la molecola più efficientemente in prossimità di un minimo locale.

2.6 Appendice

2.6.1 Calcolo delle cariche

Le cariche parziali da usare nel campo di forze si ottengono generalmente da calcoli quantomeccanici. Nel caso di molecole rigide, il calcolo delle cariche é abbastanza semplice. Nel caso di molecole flessibili, occorre valutare quanto le diverse conformazioni influenzano la distribuzione di carica e, quindi, stimare la carica parziale come media pesata tra i vari conformeri. Il calcolo QM fornisce le cosiddette cariche di Mulliken. Questo tipo di cariche possono portare ad una elevata inaccuratezza nel riprodurre proprietà chimico-fisiche di piccole molecole. Per evitare questo inconveniente sono state introdotte varie procedure per ottenere delle cariche parziali che tengano conto della diversa capacità dei singoli atomi di accomodare una diversa distribuzione di carica. Queste procedure vanno sotto il nome di metodi di *Electrostatic potential fitting* tra cui i più usati sono il RESP e il CHELPG. Vediamo come questi metodi funzionano.

Il potenziale elettrostatico $\phi(\mathbf{r})$ nel punto \mathbf{r} può essere determinato dalla densità elettronica attraverso la relazione:

$$\phi(\mathbf{r}) = \phi_N(\mathbf{r}) + \phi_e(\mathbf{r}) = \sum_{i=1}^N \frac{Z_i}{|\mathbf{r} - \mathbf{R}_i|} - \int \frac{d\mathbf{r}' \rho(\mathbf{r}')}{|\mathbf{r}' - \mathbf{r}|} \quad (2.22)$$

dove $\phi_N(\mathbf{r})$ e $\phi_e(\mathbf{r})$ sono i contributi nucleare ed elettronico alla carica totale, N è il numero di atomi, \mathbf{R}_i è la posizione dell'atomo i -esimo. Per ottenere una descrizione a cariche atomiche puntiformi di questo potenziale, è necessario effettuare una procedura di modellazione dei valori di queste cariche in modo da minimizzare la funzione:

$$R = \sum_{i=1}^M w_i (\phi(\mathbf{r}_i)^{Q_M} - \phi(\mathbf{r}_i)^{Coul}) \quad (2.23)$$

ovvero, la somma dei minimi quadrati della differenza tra il potenziale quantomeccanico e quello calcolato da queste cariche in un certo numero di punti M opportunamente scelti. w_i sono degli opportuni pesi da attribuire ai vari punti in base alla loro importanza.

La carica sull'atomo N -esimo (q_N) è legata alle altre attraverso la relazione:

$$q_N = Z - \sum_{j=1}^{N-1} q_j \quad (2.24)$$

dove Z è la carica molcolare totale, pertanto il potenziale coulombiano descritto dalle N cariche si può esprimere come:

$$\phi(\mathbf{r}_i)^{Coul} = \sum_{j=1}^{N-1} \frac{q_j}{4\pi\epsilon_0 r_{ij}} + \frac{Z - \sum_{j=1}^{N-1} q_j}{4\pi\epsilon_0 r_{iN}} \quad (2.25)$$

derivando R in q_k si ottiene

$$\frac{\partial R}{\partial q_k} = -2 \sum_{i=1}^M w_i (\phi(\mathbf{r}_i)^{QM} - \phi(\mathbf{r}_i)^{Coul}) \left(\frac{\partial \phi(\mathbf{r}_i)^{Coul}}{\partial q_k} \right) = 0 \quad (2.26)$$

che può essere scritta come un set di M equazioni

$$\sum_{i=1}^M w_i \left(\phi(\mathbf{r}_i)^{QM} - \frac{Z}{r_{iN}} \right) \left(\frac{1}{r_{ik}} - \frac{1}{r_{iN}} \right) = \sum_{j=1}^{N-1} \left[\sum_{i=1}^M w_i \left(\frac{1}{r_{ik}} - \frac{1}{r_{iN}} \right) \left(\frac{1}{r_{ij}} - \frac{1}{r_{iN}} \right) \right] \frac{q_j}{4\pi\epsilon_0} \quad (2.27)$$

o, in forma matriciale $\mathbf{a} = \mathbf{A}\mathbf{q}$. Usando i metodi dell'algebra lineare, la precedente equazione può essere risolta ottenendo le cariche $\mathbf{q} = \mathbf{A}^{-1}\mathbf{a}$. La scelta degli M punti viene fatta in differenti modi, ma devono essere presi in una regione di spazio che sia significativa per definire un corretto modello di interazioni molecolari. Due sono le procedure più utilizzate per questo calcolo:

- CHELPG: sviluppato da Breneman e Wiberg nel 1990. In questo metodo, la molecola viene posta in una reticolo cubico di punti (con spaziatura compresa tra 0.3 e 0.8 Å). Tutti i punti che si trovano entro il raggio di van der waals e 2.8Å al di fuori di questo vengono scartati, e i punti rimanenti usati per il calcolo.
- RESP: sviluppato dal gruppo di P. Kolmann per calcolare le cariche parziali da usare nel campo di forza AMBER. In questo caso, i punti vengono presi da superfici molecolari costruite usando raggi di van der waals progressivamente crescenti, e imponendo dei vincoli su particolari atomi (per esempio carboni localizzati all'interno della molecola) che potrebbero altrimenti assumere cariche eccessivamente alte. A differenza del CHELPG, le cariche ottenute con questo metodo sono meno sensibili alla conformazione molecolare.

Per effettuare il calcolo delle cariche parziali si può ricorrere ai programmi per calcoli di QM, GAUSSIAN o GAMESS/US (quest'ultimo è un programma liberamente distribuito). Per il calcolo è necessario disporre di coordinate iniziali della molecola. Queste possono essere ottenute dal data base di strutture (come il Cambridge Structural Database) oppure costruite usando programmi di modellistica molecolare. Un programma molto diffuso di questo tipo è **MOLDEN** che può essere liberamente scaricato dal sito:

www.cmbi.kun.nl/schaft/molden/molden.html

Usando questo programma è possibile costruire la struttura iniziale della molecola e salvare su file le coordinate nel formato GAUSSIAN o GAMESS/US,

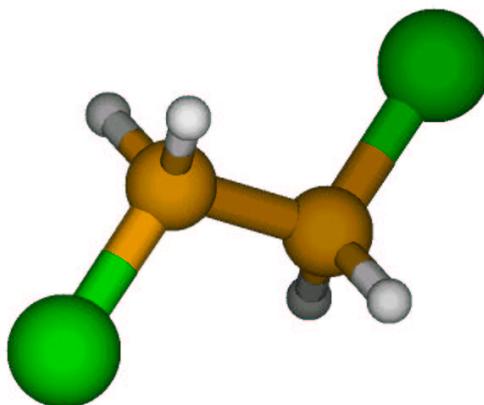


Figura 2.7: Molecola di 1,2-dicloroetano in conformazione *anti* .

per il calcolo delle cariche. Come esempio di calcolo delle cariche viene mostrato il contenuto del file di input per GAUSSIAN per al calcolo a punto singolo delle conformazioni *syn*, *gauge* e *anti* del 1,2,dicloroetano.

```
%Chk=DCLETANO
# RHF/6-31G(d) F0pt Test Pop=(CHelpG,dipole)
```

Calcolo delle cariche con il metodo di CHELPG

```
0 1
c
cl 1 clc2
c 1 cc3      2 cccl3
h 1 hc4      2 hcc14      3 dih4
h 1 hc5      2 hcc15      3 dih5
cl 3 clc6    1 clcc6      2 60.
h 3 hc7      1 hcc7      6 dih7
h 3 hc8      1 hcc8      6 dih8
```

```
clc2      1.750000
cc3       1.540000
cccl3     109.471
hc4       1.089000
hcc14     109.471
dih4      120.000
hc5       1.089000
hcc15     109.471
dih5     -120.000
clc6      1.750000
```

Tabella 2.1: Cariche parziali (in unità di carica atomica) calcolate su tre configurazioni del 1,2-dicloroetano. Tra parentesi vengono riportate le cariche di Mulliken.

Atomo	<i>syn</i>	<i>gauge</i>	<i>anti</i>
C	0.153	0.037 (-0.434)	0.029 (-0.433)
Cl	-0.215	-0.205 (-0.110)	-0.222 (-0.124)
C	0.149	0.037 (-0.434)	0.029 (-0.433)
H	0.031	0.094 (0.281)	0.096 (0.278)
H	0.031	0.073 (0.264)	0.096 (0.278)
Cl	-0.215	-0.205 (-0.110)	-0.222 (-0.124)
H	0.033	0.073 (0.264)	0.096 (0.278)
H	0.033	0.094 (0.281)	0.096 (0.278)
dipolo (Debye)	3.585	3.227	0.0000

```

clcc6      109.471
hc7        1.089000
hcc7       109.471
dih7       120.000
hc8        1.089000
hcc8       109.471
dih8       240.000
dih6        0.0

```

Come si può vedere dalla prima linea, i programmi effettuano dapprima una minimizzazione delle coordinate, e una volta che si è raggiunta la convergenza viene effettuato un calcolo ESP usando il metodo CHELPG.

Nella tabella 2.1 sono riportati i valori delle cariche ottenute per le varie configurazioni e il relativo momento di dipolo.

2.6.2 Calcolo di un profilo di energia potenziale per rotazione di un legame.

Quando in una molecola sono presenti vari angoli diedri, può essere utile calcolare il profilo dell'energia conformazionale ottenuta dalla rotazione intorno a questi. Questo profilo può essere usato per modellare l'energia potenziale classica dovuta alla rotazione intorno al suddetto diedro, usando il potenziale di Rycheman-Belleman ed escudendo le interazioni di non-legame intramolecolari tra gli atomi coinvolti.

In GAUSSIAN è possibile calcolare automaticamente un profilo di energia conformazionale. Nell'esempio che segue viene mostrato come ottenere, il profilo di energia conformazionale per la molecola di 1,2-dicloroetano. Di seguito viene mostrato il file di input per GAUSSIAN per il calcolo del profilo di energia.

```

%Chk=DCLETANO
# RHF/6-311G(D) Opt=Z-matrix NoSymm Test

```

```

Relaxed PES scan

0 1
  c
cl  1 clc2
.
.
.
dih6      0.    18    10.

```

Nell'ultima riga viene indicato l'angolo diedro Cl-C-C-Cl (dih6) interessato alla rotazione. La scansione viene effettuata ruotando 18 volte di 10 gradi l'angolo diedro. Per estrarre dal file di uscita il profilo di energia in funzione dell'angolo si può lo script per il programma **awk**, riportato di seguito.

```

#!/usr/bin/awk -f
BEGIN {
# Incremento in gradi della scansione
  step = 10.
# Fattore di conversione Hartree --> Kcal/mol
  conv = 627.5095
  cc = 0
  count = 0
}
$1 == "EIGENVALUES" { for (i=3;i<=NF;i++) {
  k=split($i,a,"-")
  for (ii=1;ii<=k;ii++) {
    if (a[ii] != "") {
      val = -a[ii]
      en[cc] = val
      if (val < min) {min = val}
      count+=step
      cc++
    }
  }
}
}
END {
  count = 0
  for (ii=0;ii<cc;ii++) {
    printf "% 4.1f   %8.3f\n", count, (en[ii]-min)*conv
    count+=step
  }
}

```

Nella Figura 2.8 viene riportato il grafico del profilo di energia ottenuto dalla scansione

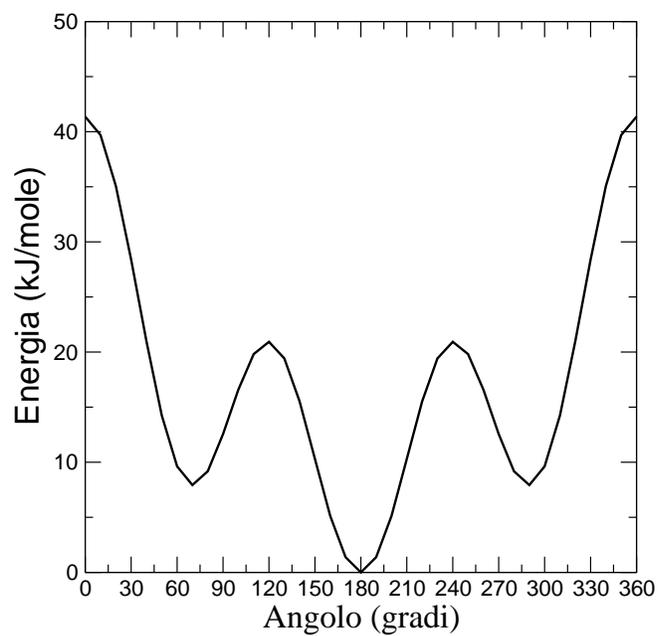


Figura 2.8: Andamento della profilo di energia potenziale ottenuto dalla rotazione intorno a un diedro Cl-C-C-Cl del 1,2-dicloroetano.

Capitolo 3

Equazioni del moto

Le equazioni del moto possono essere formulate in modi differenti. Per gli scopi della dinamica molecolare classica si usa la formulazione più semplice, cioè quella delle equazioni del moto di Newton. La formulazione Lagrangiana e Hamiltoniana vengono usate per descrivere in modo semplice ed elegante alcuni aspetti metodologici della DM (per esempio, vincoli, accoppiamento termico, etc.).

Il principio generale su cui si fondano le formulazioni non newtoniane della meccanica classica è il principio variazionale di Hamilton. Questo principio stabilisce che la traiettoria seguita da un sistema meccanico nello spazio delle fasi è quella per la quale l'integrale sul tempo $\int \mathcal{L} dt$, dove \mathcal{L} è il Lagrangiano del sistema, assume il valore minimo. In un campo di forze conservativo, dato un insieme di N coordinate e velocità generalizzate q_i, \dot{q}_i , e definita $\mathcal{L} = \mathcal{L}(\{q_i\}, \{\dot{q}_i\}, t)$, si ha che

$$\frac{d}{dt} \left(\frac{\partial \mathcal{L}}{\partial \dot{q}_i} \right) - \frac{\partial \mathcal{L}}{\partial q_i} = 0 \quad i = 1, \dots, N \quad (3.1)$$

Nel caso in cui q_i siano coordinate cartesiane il Lagrangiano diventa

$$\mathcal{L} = \frac{1}{2} m \sum_i [\dot{q}_i^2 - U(q_i)] \quad (3.2)$$

e dalla 3.1 si ottiene l'equazione di Newton:

$$m\ddot{q}_i = -\frac{\partial U}{\partial q_i} = F_i \quad (3.3)$$

dove F_i rappresenta la forza agente sulla particella i -esima. Il formalismo delle equazioni di Lagrange permette di trattare in modo semplice e diretto i vincoli. Il principio di Hamilton può essere esteso a sistemi con vincoli aventi la forma generale

$$\sum_k a_{lk} \dot{q}_k + a_l = 0 \quad l = 1, \dots, N \quad (3.4)$$

Nel caso di vincoli che dipendono solo dalle coordinate (olonomici) per i quali esiste una relazione tra le coordinate del tipo $g_l(\{q_k\}, t) = 0$, in cui

$$a_{lk} = \frac{\partial g_l}{\partial q_k} \quad a_l = \frac{\partial g_l}{\partial t} \quad (3.5)$$

l'equazioni di Lagrange possono essere scritte nella forma

$$\frac{d}{dt} \left(\frac{\partial \mathcal{L}}{\partial \dot{q}_i} \right) - \frac{\partial \mathcal{L}}{\partial q_i} = \sum_l \lambda_l a_{li} \quad i = 1, \dots, N \quad (3.6)$$

dove gli M moltiplicatori di Lagrange λ_i sono valutati sulle N coordinate. Il membro a destra della equazione 3.6 può essere considerato come una forza generalizzata che produce lo stesso effetto dei vincoli che sono stati imposti.

La formulazione di Hamilton delle equazioni del moto mette elegantemente in relazione la meccanica classica con quella quantistica. L' Hamiltoniano di un sistema viene definito in termini di velocità generalizzate $\{\dot{q}_i\}$ e momenti $p_i = \partial \mathcal{L} / \partial \dot{q}_i$ (che diventa $m\dot{q}_i$ nel caso di coordinate cartesiane) nel modo seguente

$$\mathcal{H}(\{q_i\}, \{p_i\}, t) = \sum_i \dot{q}_i p_i - \mathcal{L} \quad (3.7)$$

Le equazioni del moto di Hamilton sono allora definite:

$$\dot{q}_i = \frac{\partial \mathcal{H}}{\partial p_i}, \quad \dot{p}_i = -\frac{\partial \mathcal{H}}{\partial q_i} \quad (3.8)$$

Nel caso in cui si usi un potenziale U conservativo, le coordinate cartesiane x_i e le velocità $v_i = p_i/m_i$, l'Hamiltoniano 3.7 diventa

$$H(p_i, x_i) = \sum_i \frac{p_i^2}{2m_i} + U(x_i) \quad (3.9)$$

Dalle equazioni di Hamilton si ottengono le equazioni del moto di Newton:

$$\dot{\mathbf{r}}_i(t) = \mathbf{v}_i(t) \quad (3.10)$$

$$m_i \dot{\mathbf{v}}_i(t) = \mathbf{F}_i(t) \quad (3.11)$$

con \mathbf{r}_i e \mathbf{v}_i rispettivamente le coordinate cartesiane e le velocità di una particella i-esima con massa m_i e $\mathbf{F}_i(t)$ la forza sulla particella i-esima, che per sistemi conservativi è funzione delle sole coordinate:

$$\mathbf{F}_i = -\frac{\partial}{\partial x_i} V(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) \equiv V(\mathbf{r}) \quad (3.12)$$

Le equazioni 3.10 e 3.11 vengono integrate numericamente usando opportuni algoritmi (vedi paragrafo 3.1).

Nel trattare le forze che agiscono sul sistema simulato si può distinguere tra forze interne F_{int} e forze esterne F_{est} , le prime rappresentano le forze tra le particelle del sistema, mentre le seconde sono le forze originate dall'ambiente

esterno al sistema. Le forze esterne possono essere rappresentate attraverso l'effetto medio che esse producono. Si parla in tal caso di una forza media F_i^M definita come l'interazione media esercitata dalle M particelle esterne sulla particella i-esima. La forza media ha un corrispondente *potenziale di forza medio* V^M , definito come $F_i^M = -\partial V^M / \partial \mathbf{r}_i$, che rappresenta l'effetto medio esercitato da tutti i gradi di libertà non esplicitamente rappresentati nel sistema simulato. L'inclusione del termine di potenziale di forza medio e le sue fluttuazioni nel tempo porta alla formulazione della equazione del moto stocastica di Langevin, e i metodi simulativi che ne fanno uso sono chiamati genericamente metodi di dinamica stocastica (SD) [88, 76].

3.1 Metodi d'integrazione delle equazioni del moto

La Dinamica Molecolare simula l'evoluzione di un sistema finito di particelle attraverso l'integrazione delle equazioni del moto di Newton. Per integrare queste equazioni differenziali del secondo ordine si ricorre ai metodi numerici delle differenze finite, poichè non esiste soluzione analitica.

Gli algoritmi utilizzati si differenziano nella precisione di calcolo (ordine dell'algoritmo), nella occupazione di memoria e nel tipo di rappresentazione. La rappresentazione degli algoritmi usati si ottiene sviluppando in serie di Taylor le grandezze $r_i(t + \Delta t)$ e $v_i(t + \Delta t)$. In DM si usano generalmente algoritmi che usano sviluppi di Taylor fino al terzo ordine in quanto si è osservato che algoritmi di ordine inferiore sono poco accurati mentre quelli di ordine superiore sono troppo dispendiosi in termini di risorse computazionali.

L'algoritmo più usato è quello di Verlet [91, 92] che si ottiene da addizioni e sottrazioni di sviluppi in serie di Taylor delle coordinate $r_i(t + \Delta t)$ e $r_i(t - \Delta t)$ di particelle di massa m_i :

$$\mathbf{r}_i(t \pm \Delta t) = \mathbf{r}_i(t) \pm \Delta t \dot{\mathbf{r}}_i(t) + \frac{(\Delta t)^2}{2!} \ddot{\mathbf{r}}_i(t) \pm \frac{(\Delta t)^3}{3!} \dddot{\mathbf{r}}_i(t) + O(\Delta t^4) \quad (3.13)$$

da cui si ricava:

$$\mathbf{r}_i(t + \Delta t) = -\mathbf{r}_i(t - \Delta t) + 2\mathbf{r}_i(t) + \frac{(\Delta t)^2}{m_i} \mathbf{F}_i(t) \quad (3.14)$$

Dove $\mathbf{F}_i(t)$ è la forza esercitata sulla particella i al tempo t e dove i termini superiori al terzo ordine sono stati trascurati. Pertanto questo schema di integrazione fornisce una accuratezza del quarto ordine in Δt . Le velocità in questo schema sono date da:

$$\mathbf{v}_i(t) = \dot{\mathbf{r}}_i(t) = \frac{1}{2\Delta t} [\mathbf{r}_i(t + \Delta t) - \mathbf{r}_i(t - \Delta t)] \quad (3.15)$$

L'errore sulle velocità è dell'ordine di $(\Delta t)^3$.

Una variante di questo schema d'integrazione, molto diffuso nei programmi di DM, è il *leap-frog*:

$$\mathbf{v}_i(t + \Delta t) = \mathbf{v}_i\left(t - \frac{\Delta t}{2}\right) + \frac{\mathbf{F}_i(t)}{\Delta t} \quad (3.16)$$

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \mathbf{v}_i(t + \Delta t)\Delta t \quad (3.17)$$

Questo schema usa per le coordinate intervalli interi e per le velocità intervalli semi-interi. Nella Figura 3.1 viene rappresentato il modo in cui viene applicato questo algoritmo.

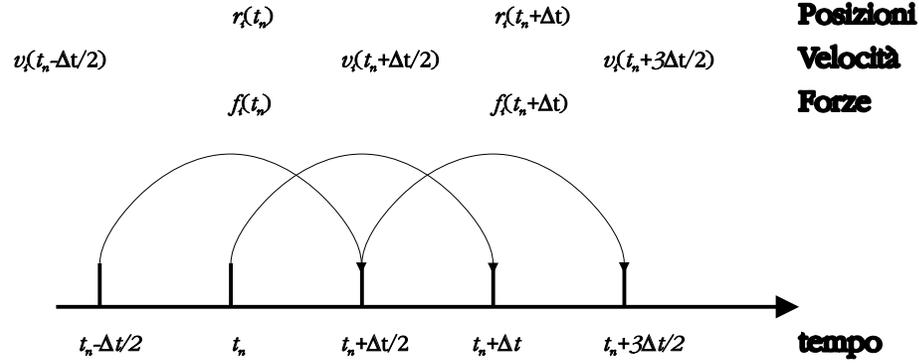


Figura 3.1: Schema *leap-frog* d'integrazione dell'equazione del moto di Newton.

Questo è molto efficiente da un punto di vista computazionale, essendo semplice e richiedendo meno tempo di CPU e memoria rispetto ad altri più avanzati. Uno schema di integrazione più accurato dei precedenti è quello di Beeman [7].

$$\mathbf{r}_i(t_n + \Delta t) = \mathbf{r}_i(t_n) + \mathbf{v}_i(t_n)\Delta t + \frac{[4\mathbf{F}_i(t_n) - \mathbf{F}_i(t_n - \Delta t)]\Delta t^2}{6m} \quad (3.18)$$

$$\mathbf{v}_i(t_n + \Delta t) = \mathbf{v}_i(t_n) + \frac{[2\mathbf{F}_i(t_n + \Delta t) + 5\mathbf{F}_i(t_n) - \mathbf{F}_i(t_n - \Delta t)]\Delta t}{6m} \quad (3.19)$$

L'algoritmo è poco usato in quanto la migliore accuratezza non compensa il maggior tempo di calcolo e la memoria impiegata.

Esiste un classe di algoritmi noti con il nome di *predictor-corrector*, di cui il più noto è quello di Gear [26]. Questi algoritmi funzionano nel modo che segue. Partendo dalle posizioni al tempo t_{n-1} vengono predette le posizioni al tempo t_n usando l'equazione (*predictor*):

$$\mathbf{r}'(t_n + \Delta t) = \mathbf{r}(t_n - \Delta t) + 2\mathbf{v}(t_n)\Delta t \quad (3.20)$$

Il valore previsto della posizione viene usato per prevedere l'accelerazione $\mathbf{a}'(t_n + \Delta t)$. Infine, usando l'accelerazione prevista, si calcolano i valori corretti

di $\mathbf{r}(t_n + \Delta t)$ e $\mathbf{v}(t_n + \Delta t)$ per mezzo delle equazioni (*correctors*):

$$\mathbf{v}(t_n + \Delta t) = \mathbf{v}(t_n) + \frac{1}{2}[\mathbf{a}'(t_n + \Delta t) + \mathbf{a}(t_n)]\Delta t \quad (3.21)$$

$$\mathbf{r}(t_n + \Delta t) = \mathbf{r}(t_n) + \frac{1}{2}[\mathbf{v}(t_n + \Delta t) + \mathbf{v}(t_n)]\Delta t \quad (3.22)$$

A questo punto il valore corretto di $\mathbf{r}(t_n + \Delta t)$ viene usato per calcolare un nuovo valore previsto di $\mathbf{a}'(t_n + \Delta t)$ e quindi un nuovo valore previsto per $\mathbf{v}(t_n + \Delta t)$ e $\mathbf{r}(t_n + \Delta t)$. Il processo viene ripetuto fino a quando la differenza tra il valore corretto e quello previsto di $\mathbf{r}(t_n + \Delta t)$ è minore di un fissato valore. Il metodo può essere generalizzato a ordini più elevati che usano non solo relazioni tra $\mathbf{r}(t_n + \Delta t)$, $\mathbf{v}(t_n + \Delta t)$ e $\mathbf{v}(t_n)$, ma anche i valori di $\mathbf{v}(t_n - \Delta t)$ e $\mathbf{v}(t_n - 2\Delta t)$. La scelta dell'algoritmo d'integrazione da usare in DM deve essere fatta sulla base di quattro differenti considerazioni.

1. L'algoritmo non deve richiedere più di una valutazione delle interazioni per passo d'integrazione.
2. La scelta dell'ordine dell'algoritmo è legata al tipo di modello scelto nella rappresentazione del sistema molecolare. In DM si usano algoritmi del 3° ordine poichè quelli di ordine inferiore non forniscono sufficiente accuratezza.
3. La rappresentazione dell'algoritmo è molto importante. Infatti se si vogliono applicare, a esempio, metodi di controllo della temperatura (vedi paragrafo successivo) che agiscono scalando le velocità, è necessario che queste vengano esplicitamente espresse nell'algoritmo.
4. La richiesta di risorse di calcolo e di memoria sono importanti fattori che influenzano le prestazioni del programma di DM.

Un' ottima rassegna degli algoritmi utilizzabili nell'integrazione delle equazioni del moto in Dinamica Molecolare è stata presentata da van Gunsteren e Berendsen. [86]

3.2 Passo d'integrazione e vincoli olonomici

In molecole flessibili la lunghezza del passo d'integrazione è limitata dalla frequenza di oscillazione dei moti più rapidi. Infatti, per integrare questi moti correttamente, e assicurare la conservazione dell'energia totale del sistema, è necessaria una scelta di valori del passo d'integrazione molto piccoli. Quando, non si è interessati allo studio di proprietà molecolari legate a queste vibrazioni, è preferibile bloccarle usando opportuni *vincoli olonomici*. L'uso di questi vincoli permette generalmente di aumentare fino a 10 volte il passo d'integrazione. Esistono in generale due modi di applicare vincoli alle equazioni del moto di Newton. Il primo modo è quello di far uso delle coordinate generalizzate, ma

è un metodo poco usato a causa della difficoltà di lavorare con tali coordinate con molecole complesse [25, 97]. Il secondo metodo consiste nell' usare le coordinate cartesiane. In questo caso il problema può essere affrontato usando due differenti tipi di metodi: quello matriciale [75, 37] e quello del rilassamento [75]. Il primo metodo è poco utilizzato in quanto applicabile solo a sistemi di piccole dimensioni. In genere si preferisce usare nei programmi di DM i metodi di rilassamento, tra questo il più diffuso è il cosiddetto SHAKE [75]. Lo SHAKE è una procedura iterativa, che dopo ogni passo d' integrazione varia le posizioni $\mathbf{r}'_i(t + \Delta t)$ degli atomi, finché tutte le condizioni di vincolo non vengano soddisfatte, nelle nuove posizioni vincolate $\mathbf{r}_i(t + \Delta t)$. Per fare in modo che le velocità degli atomi dopo l'applicazione dello SHAKE non abbiano componenti lungo la direzione dei vincoli queste vengono calcolate dalle posizioni ai tempi t e $t + \Delta t$ secondo l'espressione

$$v_i \left(t + \frac{\Delta t}{2} \right) = \frac{\mathbf{r}_i(t + \Delta t) - \mathbf{r}_i(t)}{\Delta t} \quad (3.23)$$

Infine le forze vincolari \mathbf{G}_i possono essere ottenute come

$$\mathbf{G}_i(t) = m_i \frac{\mathbf{r}_i(t + \Delta t) - \mathbf{r}'_i(t + \Delta t)}{\Delta t^2} \quad (3.24)$$

Lo SHAKE consente di ottenere un incremento del passo d'integrazione di simulazione di 2–4 volte, ma presenta comunque degli svantaggi che elenchiamo di seguito:

1. può dare dei problemi di convergenza quando viene applicato a grandi gruppi planari;
2. la sua applicazione costituisce generalmente un collo di bottiglia computazionale nella parallelizzazione dell'algoritmo;
3. in alcune applicazioni il vincolo dei legami può costituire un modello fisico non realistico.

Recentemente è stato messo a punto un nuovo metodo basato sulla soluzione algebrica locale di vincoli lineari [37], che risulta essere molto più veloce e stabile di SHAKE. Il metodo inoltre è molto più adatto ad essere usato su calcolatori paralleli.

Un metodo alternativo all'uso dei vincoli consiste nell'algoritmo del *Multiple Time Step* (MTS) [3, 80]. La lunghezza del passo d'integrazione è limitata dalla frequenza di oscillazione o dal tempo di rilassamento delle forze. Infatti le forze possono essere classificate in base al loro tempo di rilassamento. In base a tale criterio possiamo distinguere tre tipi fondamentali di forze:

1. forze (F_{af}) che determinano alte frequenze di oscillazioni: oscillatori armonici con elevati valori delle costanti di forza;

2. forze (F_{fi}) a frequenza intermedia: oscillazioni degli angoli di legame e torsionali, forze di van der Waals, forze elettrostatiche a corto raggio (es. dipolari).
3. forze (F_{bf}) a bassa frequenza (forze a lungo raggio);

Come si è detto, la presenza di forze ad alta frequenza limita l'impiego di passi d'integrazione lunghi. Con l'algoritmo del MTS si supera questo limite impiegando passi d'integrazione di differente lunghezza secondo lo schema che segue: si divide il passo d'integrazione Δt in m sottointervalli $\Delta t'$, le forze F_{af} vengono valutate ogni sottointervallo $\Delta t'$ mentre le forze $F_{bf} + F_{fi}$ ogni intervallo Δt . L'efficienza di questo metodo è paragonabile a quella dello SHA-KE, con il vantaggio di non avere le limitazioni di quest'ultimo. Tuckerman *et al.* [83] hanno proposto un nuovo algoritmo per il MTS. L'algoritmo è stato applicato con buoni risultati alla simulazione di sistemi molecolari di differenti dimensioni [93, 94].

Capitolo 4

Trattamento delle interazioni a lungo raggio

In una simulazione di DM la maggior parte del tempo di calcolo viene speso nel calcolo delle interazioni di non legame. Infatti mentre il tempo di calcolo delle interazioni di legame è proporzionale al numero N di atomi del sistema, il tempo di calcolo dell'interazioni tra atomi non legati è proporzionale a N^2 . Si comprende, quindi, come la lunghezza della simulazione e la grandezza del sistema simulato dipendano da quanto velocemente queste interazioni vengono calcolate. Esistono vari modi per velocizzare il calcolo delle interazioni a lungo raggio.

4.1 Il Cutoff

La procedura più semplice è quella di applicare il cosiddetto criterio di *cutoff* alle interazioni di non legame, che consiste nel calcolare solo le interazioni tra atomi che si trovano entro una sfera di raggio R_c . L'introduzione del raggio di *cutoff* si rivela una buona tecnica per trattare le interazioni di non legame a corta distanza (forze di van der Waals), ma può creare artefatti quando si devono trattare le interazioni coulombiane che sono a lunga distanza (essendo l'energia proporzionale a r^{-1}). Uno di questi problemi è rappresentato dal rumore prodotto dalla discontinuità delle interazioni alla distanza $r = R_c$. Tale discontinuità produce un aumento artificiale della energia cinetica e quindi della temperatura del sistema. Per ridurre gli effetti dovuti al *cutoff* si ricorre all'uso di funzioni che moltiplicate (funzioni di *switching*) o addizionate (funzioni di *shifting*) al potenziale d'interazione lo portano a zero in maniera continua in corrispondenza della distanza di *cutoff*. [8] L'uso di queste funzioni non si basa su nessun tipo di argomentazione fisica e il loro uso può produrre artefatti in sistemi particolari [65, 50]. Un altro modo molto semplice per ridurre la distanza di azione delle forze coulombiane è quello di agire sul valore della costante dielettrica relativa. In alcuni campi di forza [] viene introdotta, ma senza una valida

ragione fisica, una costante dielettrica in cui valore è proporzionale alla distanza r , questo produce una dipendenza dalla distanza, del potenziale elettrostatico, del tipo r^{-2} e quindi comporta una riduzione del raggio di azione della interazione. Il raggio di azione delle forze elettrostatiche può essere ulteriormente ridotto se si distribuiscono le cariche sugli atomi in modo da avere interazioni di tipo dipolare e multipolare. Quando la somma delle cariche atomiche (parziali) di un gruppo di atomi è zero, il termine di interazione risultante tra due gruppi di atomi di questo tipo è multipolare, ovvero, l'energia è proporzionale a r^{-3} o a potenze maggiori. Per grandi distanze la somma dei contributi di monopolo (r^{-1}) tra le varie coppie di atomi diventa nulla. Pertanto se si raggruppano gli atomi della molecola simulata in opportuni *gruppi di carica* aventi carica netta zero [5, 84], il raggio di azione delle forze coulombiane viene considerevolmente ridotto.

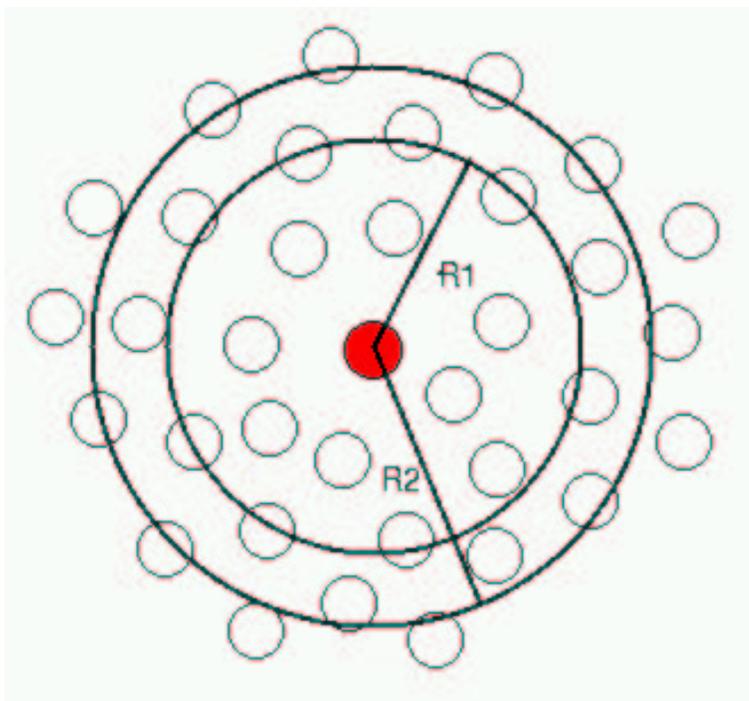


Figura 4.1: Nella figura viene mostrato l'uso del metodo del doppio cutoff (*twin range cutoff*). La particella centrale interagisce sempre con le particelle che si trovano dentro il raggio (R1) mentre con quelle entro il raggio R2 ogni n passi.

4.2 Metodi basati sulle somme di Ewalds

Un metodo, molto diffuso nel trattamento delle interazioni elettrostatiche in sistemi ionici, è quello delle somme di Ewald [3, 23, 38]. Il metodo nella sua formulazione è adatto a simulare sistemi periodici come cristalli ionici, ma viene anche applicato nella simulazione di sistemi liquidi ionici. In un sistema periodico il campo elettrico è ottenuto dalla formula:

$$U = \frac{1}{4\pi\epsilon_0} \frac{1}{2} \sum_{\mathbf{n}=0}^{\infty} \sum_{i=1}^N \sum_{j=1}^N \frac{q_i q_j}{|\mathbf{r}_{ij} + \mathbf{n}|} \quad (4.1)$$

dove i e j si riferiscono alle particelle ed \mathbf{n} ai punti $\mathbf{n} = (n_x L, n_y L, n_z L)$ (con n_x, n_y, n_z interi) di un reticolo cubico semplice di lato L (il metodo si può comunque applicare, con le dovute modifiche, a qualunque tipo di reticolo). Quando $\mathbf{n} = 0$ i termini con $i = j$ sono omessi in modo da prevenire l'interazione degli atomi con se stessi ma non con la loro immagine replicata. Le somme della equazione 4.1 possono essere riorganizzate in modo opportuno [23, 38, 3] ottenendo l'espressione (formula di Ewald):

$$\begin{aligned} U &= \frac{1}{4\pi\epsilon_0} \sum_{\mathbf{n}} \sum_{i=1}^N \sum_{j=i+1}^N q_i q_j \frac{\text{erfc}(\alpha|\mathbf{r}_{ij} + \mathbf{n}|)}{|\mathbf{r}_{ij} + \mathbf{n}|} \\ &+ \frac{1}{2\epsilon_0 V} \sum_{\mathbf{k} \neq 0} \frac{1}{k^2} e^{-\frac{k^2}{4\alpha^2}} \left\{ \left| \sum_{i=1}^N q_i \cos(\mathbf{k} \cdot \mathbf{r}_i) \right|^2 + \left| \sum_{i=1}^N q_i \sin(\mathbf{k} \cdot \mathbf{r}_i) \right|^2 \right\} \\ &- \frac{\alpha}{4\pi^{\frac{3}{2}} \epsilon_0} \sum_{i=1}^N q_i^2 \end{aligned} \quad (4.2)$$

dove si è usata la notazione

- \mathbf{n} Vettore sui punti del reticolo periodico.
- \mathbf{k} Vettore reciproco di \mathbf{n} .
- k Modulo di \mathbf{k} .
- i, j Indici assoluti dei siti carichi.
- N Numero totale di siti carichi.
- q_i Carica sul sito assoluto i -esimo.
- \mathbf{r}_i Coordinate cartesiane del sito i -esimo.
- \mathbf{r}_{ij} $\mathbf{r}_j - \mathbf{r}_i$.
- α Parametro di partizione dello spazio reciproco.
- V Volume della scatola di simulazione.

Il metodo per riformulare l'equazione 4.1 consiste essenzialmente nell'aggiungere alla distribuzione infinita di cariche puntiformi $\rho(\mathbf{r})$ (definita, usando la funzione delta, come $\rho_i(\mathbf{r}) = q_i \delta(\mathbf{r} - \mathbf{r}_i)$) una distribuzione isotropa di carica di stessa grandezza, ma di segno opposto.

$$\rho(\mathbf{r})_i^G = -q_i \left(\frac{\alpha}{\sqrt{\pi}} \right)^3 e^{-\alpha^2 |\mathbf{r} - \mathbf{r}_i|^2} \quad (4.3)$$

$$\rho_i^s(\mathbf{r}) \approx \rho_i(\mathbf{r}) + \rho_i^G(\mathbf{r}) \quad (4.4)$$

Le interazioni calcolate con questa distribuzione di carica schermata dal termine gaussiano diventano a corto raggio per la comparsa della funzione complementare di errore * come si vede nel primo termine della espressione 4.2. La prima sommatoria dovrebbe estendersi all'infinito, si può, comunque, con una buona approssimazione sostituirla con una sommatoria su un numero di termini finito. Naturalmente per riottenere la distribuzione di carica originale occorre aggiungere a $\rho_i^s(\mathbf{r})$ la distribuzione di carica $-\rho_i^G(\mathbf{r})$. L'energia di interazione di queste distribuzioni gaussiane è espressa come una somma sul reticolo dello spazio reciproco (secondo termine nella equazione 4.2) più un termine di auto-interazione. Anche in questo caso la somma sul vettore k dovrebbe essere estesa all'infinito; ma la presenza di un termine esponenziale permette di avere una buona approssimazione con un numero finito di termini. Il parametro α viene scelto in modo da garantire la convergenza delle proprietà delle due sommatorie che danno luogo a $U = U^s + U^G$. Il vantaggio delle somme di Ewald consiste nel ridurre la complessità del calcolo da N^2 a $\approx N \log N$, eliminando i problemi connessi all'uso del raggio di *cutoff* e mantenendo un buon livello di accuratezza nel calcolo del potenziale d'interazione. Oltre a questi due metodi ne esistono altri [87, 8] che non verranno descritti in questa sede .

* La funzione complementare di errore è definita come segue:

$$erfc(x) = 2\pi^{-1/2} \int_x^\infty e^{-y^2} dy \quad (4.5)$$

Capitolo 5

Simulazione di sistemi in fase condensata

Nella Figura 5.1 vengono rappresentati diversi tipi di insiemi statistici usati nelle simulazioni di DM. Gli insiemi vengono caratterizzati in base alle variabili termodinamiche tenute vincolate e quelle lasciate libere di equilibrare con l'ambiente esterno:

- Insieme microcanonico (NVE). Un sistema NVE è caratterizzato da numero di particelle (N), il volume (V) e l'energia totale (E) costanti. Il sistema in queste condizioni è isolato dall'universo.
- Insieme canonico (NVT). In questo caso l'energia interna del sistema può variare, ma non la sua temperatura che è tenuta costante mediante lo scambio di calore con un bagno termico esterno (il sistema è in contatto con il resto dell'universo attraverso pareti diatermiche).
- Insieme NPT. In questo tipo di sistemi, il volume può variare in modo da mantenere la pressione costante e uguale a quella esterna.

I vari insiemi sono caratterizzati da proprietà termodinamiche descritte da opportune funzioni di stato. Si rimanda a un buon testo di termodinamica statistica l'approfondimento di questi concetti. In questo capitolo vedremo come è possibile definire e vincolare le variabili termodinamiche (N, V, T, P) per simulare sistemi molecolari in fase condensata.

5.1 Condizioni al contorno

La simulazione di un sistema molecolare con un numero limitato di atomi comporta il problema della valutazione delle condizioni al contorno. La scelta più semplice è quella di simulare la molecola nel vuoto, che dovrebbe descrivere una fase gassosa. Questa scelta comporta dei problemi, quali le deformazioni

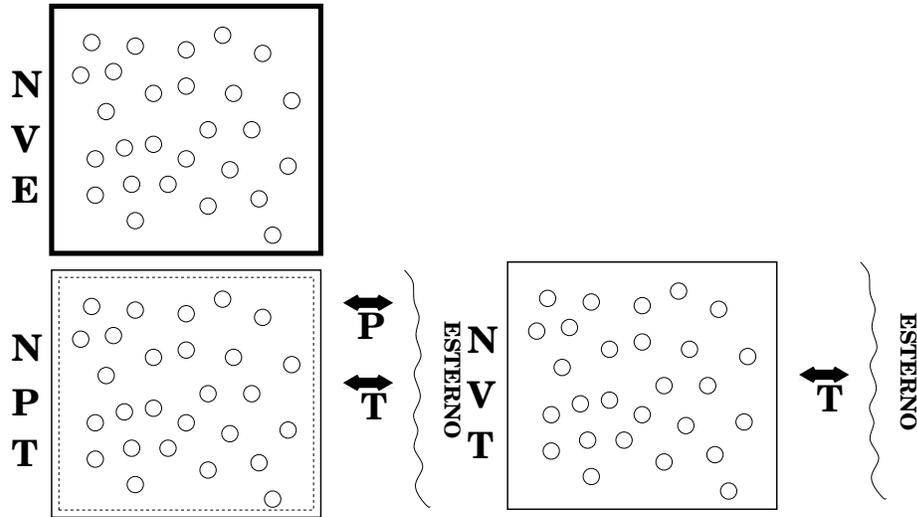


Figura 5.1: Insiemi statistici usati nelle simulazioni di DM.

superficiali, dovute alla tendenza della molecola a minimizzare la sua area superficiale. Questo effetto può essere ridotto se uno strato di atomi del sistema, a confine con il vuoto, viene tenuto bloccato (la cosiddetta *extended wall regione*) con dei potenziali armonici sulla posizione mentre la zona interna a questo strato subisce una normale dinamica [10] (vedi Figura 5.2):. In questo modo si riesce a limitare la deformazione superficiale e nello stesso tempo si riduce il tempo di calcolo, in quanto, la DM viene ristretta a una parte del sistema. Occorre, comunque, tener conto accuratamente dell'effetto che gli atomi bloccati possono avere sulla parte libera di muoversi per evitare artefatti.

Il metodo migliore per ridurre gli effetti ai bordi in un sistema finito è quello di usare condizioni periodiche. Le molecole del sistema che deve essere simulato sono poste in una scatola cubica (o più generalmente in un qualunque scatola avente una forma che consenta di riempire tutto lo spazio), che viene circondata da 26 immagini identiche di se stessa ottenute con una traslazione per una distanza pari al lato della scatola nelle direzioni x , y , z (fig. 5.3).

Vengono prese in considerazione solo le interazioni degli atomi nella scatola centrale con tutti gli atomi primi vicini. L'atomo nero nella scatola centrale in Figura 5.3, può interagire solo con gli atomi che si trovano entro la linea con tratto grosso (approssimazione della Minima Immagine, MI). L'anisotropia, dovuta alla forma cubica della MI, può essere evitata applicando un *cutoff* sferico di raggio R_c . Per evitare artefatti occorre che un atomo, i , della scatola, non interagisca simultaneamente con un atomo, j , e con l'immagine periodica di questo j' ; pertanto il lato più corto della scatola deve essere maggiore del doppio del raggio di *cutoff*. Oltre alla scatola cubica (o più in generale rettangolare) esistono altri tipi di scatole aventi forme più sferoidali [3] (per es. ottaedro troncato) che hanno il vantaggio di ridurre il numero di molecole simulate. L'uso

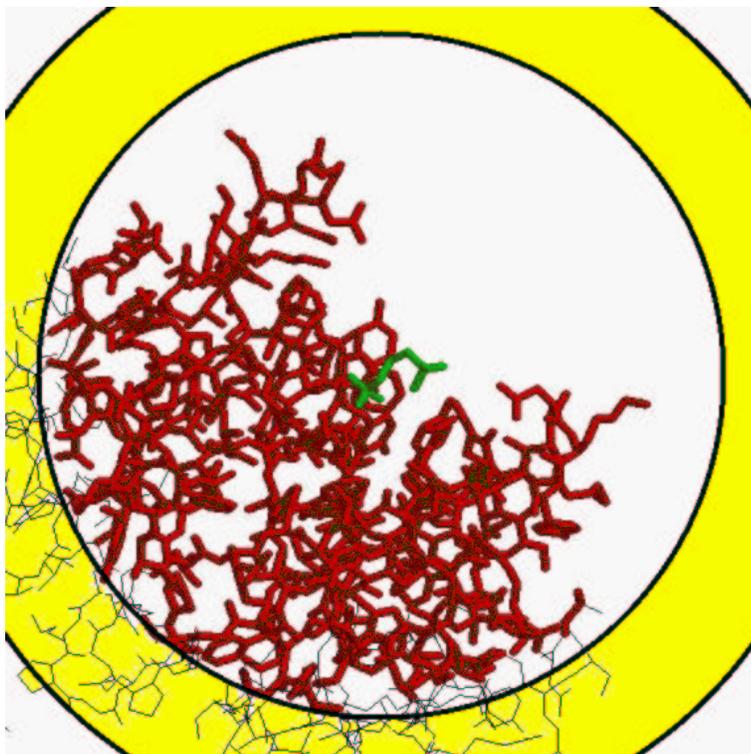


Figura 5.2: Esempio di applicazione del metodo dell' *Extended wall region* allo studio di una proteina. In giallo è indicata la regione che viene tenuta bloccata con dei vincoli sulle posizioni nella simulazione. In rossa e in verde le parti della proteina flessibili.

di queste scatole è utile quando si simulano sistemi in soluzione.

5.2 Calcolo della temperatura

La temperatura T del sistema simulato è ottenuta dalla media dell'energia cinetica secondo la nota espressione della equipartizione dell'energia cinetica:

$$\frac{1}{2} \sum_i^N m_i \mathbf{v}_i^2 = \frac{1}{2} N_f k_b T \quad (5.1)$$

da cui

$$T = \frac{1}{N_f k_b T} \sum_i^N m_i \mathbf{v}_i^2 \quad (5.2)$$

dove N è il numero di atomi, N_f il numero di gradi di libertà, k_b la costante di Boltzmann, m_i la massa e \mathbf{v}_i la velocità dell'atomi i -esimo. Per una molecola

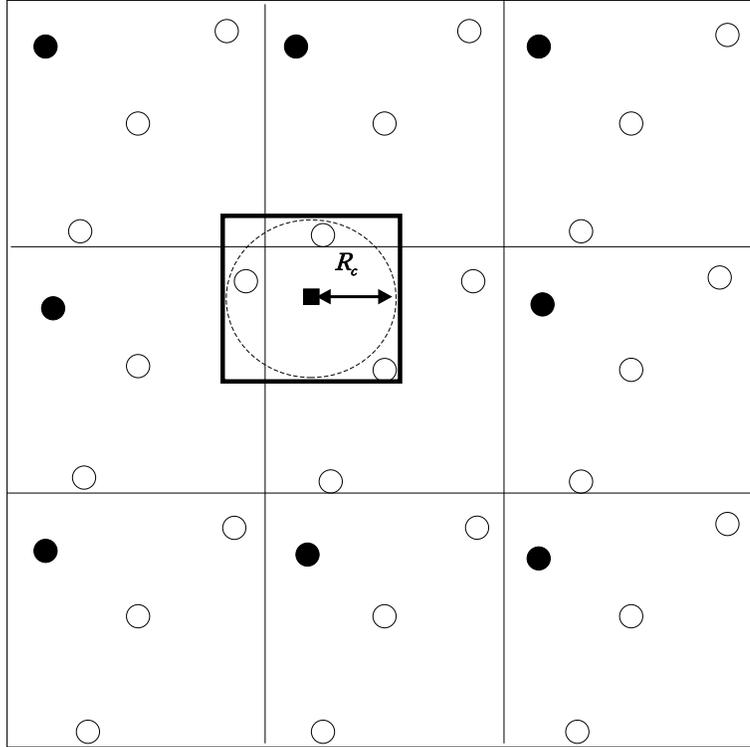


Figura 5.3: Esempio di scatola cubica periodica. I pallini rappresentano le particelle simulate. Il quadrato in tratto grosso rappresenta la MI. Il quadratino nero può interagire con le particelle che si trovano all'interno della circonferenza di raggio R_c . La periodicità fa in modo che quando un atomo lascia la scatola da una faccia, rientra, con la stessa velocità dalla faccia opposta

di N atomi, i gradi di libertà totali $N_f = 3N - N_b$ dove N_b indica il numero di legami.

5.3 Calcolo della pressione

La pressione di un sistema all'equilibrio può essere calcolata dal viriale delle forze che agiscono sul sistema:

$$\Xi = -\frac{1}{2} \sum_{i < j}^N \mathbf{r}_{ij} \cdot \mathbf{F}_{ij} \quad (5.3)$$

dove $\mathbf{r}_{ij} = \mathbf{t}_i - \mathbf{r}_j$ e \mathbf{F}_{ij} è la forza che agisce tra la particella i -esima e j -esima.

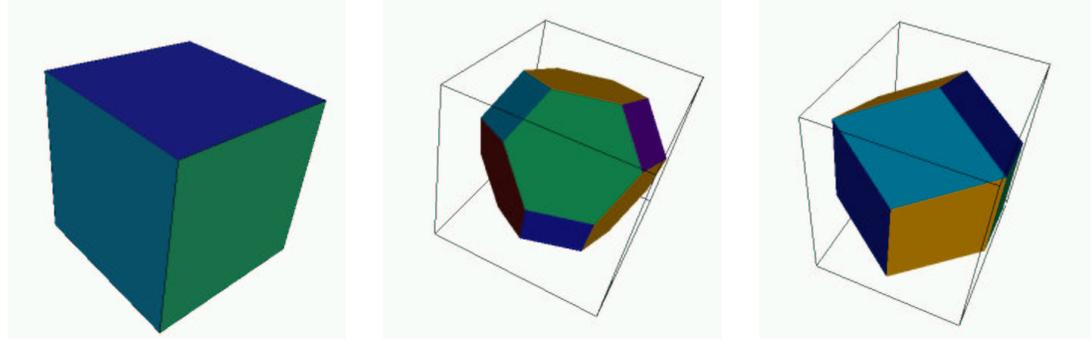


Figura 5.4: Differenti tipi di scatole periodiche, i volumi (V) si intendono relativi alla scatola cubica di volume unitario. a) Cubica, $V = 1$. b) Ottaedrica, $V=0.74$; c) Dodecaedrica, $V=0.71$.

Quindi la pressione di equilibrio è data da:

$$P = \frac{2}{3V} [\langle E_{cin} \rangle - \langle \Xi \rangle] \quad (5.4)$$

dove V è il volume del sistema e $\langle \rangle$ indica la media sulla traiettoria.

5.4 Metodi per il controllo della temperatura e della pressione

Quando si integrano le equazioni del moto l'energia totale del sistema è conservata, il sistema si trova in tal modo in condizioni adiabatiche, inoltre se il volume è tenuto costante il sistema simulato genererà un insieme statistico di tipo microcanonico. Un tale tipo di insieme non è adatto a riprodurre i dati ottenuti per via sperimentale poichè questi, generalmente, sono ottenuti a temperatura e pressione costante. Se poi si usa il raggio di *cutoff* nel calcolo delle interazioni a lungo raggio si introduce una perturbazione all'energia del sistema che porta a un leggero aumento della temperatura. Pertanto, per evitare questi problemi, si preferisce fare simulazioni in cui la temperatura e la pressione siano trattate come variabili indipendenti piuttosto che proprietà derivate.

I metodi generalmente impiegati per generare insiemi statistici isoterme e isobari si possono dividere in tre gruppi:

- *metodi del sistema esteso* [63]. Questi metodi fanno uso di meccanismi di retroazione per correggere le deviazioni di parametri controllati (temperatura e pressione) da valori medi prestabiliti, questo implica che il valore può fluttuare, ma l'ampiezza della fluttuazione può essere controllata.
- *Metodi di vincolo* [43, 22, 30]. In questi algoritmi le equazioni del moto vengono accoppiate a un opportuno vincolo (per esempio, nel caso della temperatura questo vincolo è l'energia cinetica costante).

- Nel caso della temperatura esiste un ulteriore gruppo costituito dai cosiddetti *metodi stocastici* [4, 39] che consistono nell'aggiustare la temperatura in modo da riprodurre l'effetto di un bagno termico esterno o, in modo ancora più semplice di generare a ogni step una distribuzione maxwelliana di velocità in modo da riportare il valore dell'energia cinetica totale a quello stabilito.

Nei paragrafi che seguono verranno descritti alcuni dei metodi più usati per il controllo della temperatura e della pressione.

5.4.1 Controllo della temperatura

5.4.2 Metodo isoterma

Questo metodo è stato sviluppato da Woodcock nel 1971 [98], ed è uno dei metodi più semplici che possono essere usati nelle simulazioni di Dinamica Molecolare. Esso consiste nel calcolare ad ogni passo d'integrazione un opportuno fattore che, moltiplicato per le velocità delle singole particelle, consente di ottenere il valore desiderato della temperatura del sistema. Il cambiamento di temperatura che si ottiene moltiplicando le velocità al tempo t per un fattore λ è pari a:

$$T_0 - T(t) = \Delta T = \frac{1}{2} \sum_j^N \frac{2}{3} \frac{m_i (\lambda \mathbf{v}_i)^2}{N k_b} - \frac{1}{2} \sum_j^N \frac{2}{3} \frac{m_i \mathbf{v}_i^2}{N k_b} \quad (5.5)$$

$$= (\lambda^2 - 1) \underbrace{\frac{1}{2} \sum_j^N \frac{2}{3} \frac{m_i \mathbf{v}_i^2}{N k_b}}_{T(t)} \quad (5.6)$$

da cui si ottiene che

$$\lambda = \sqrt{\frac{T_0}{T(t)}} \quad (5.7)$$

Quindi per ottenere ad ogni passo della simulazione una temperatura di riferimento T_0 , è sufficiente moltiplicare le componenti delle velocità di ogni particella per il fattore λ .

5.4.3 Metodo di Berendsen

Il metodo di Berendsen o dell'accoppiamento debole con il bagno termico [9] è uno dei metodi più usati nei programmi per simulazioni di DM. Uno dei pregi di questo metodo è la sua semplicità. L'algoritmo simula l'accoppiamento del sistema con un bagno termico esterno alla temperatura T_0 . L'entità dell'interazione tra questo bagno termostato e il sistema è determinata da una costante di tempo, τ , che può essere liberamente scelta. Schematicamente l'accoppiamento viene ottenuto modificando a ogni passo la velocità degli atomi. Questo

può essere fatto moltiplicando per una opportuna costante λ le velocità delle singole particelle. Il livellamento (*rescaling*) della temperatura T a un valore limite T_0 seguendo una legge esponenziale può essere ottenuto come segue:

$$\frac{dT}{dt} = \frac{1}{\tau}[T_0 - T(t)] \quad (5.8)$$

da cui, passando alle differenze finite, si ottiene:

$$\Delta T = \frac{\delta t}{\tau}[T_0 - T(t)] \quad (5.9)$$

dove δt è il passo d'integrazione. Sostituendo il ΔT della Equ. 5.9 con quello nella Equ. 5.6 si ottiene:

$$(\lambda^2 - 1)T(t) = \frac{\delta t}{\tau}[T_0 - T(t)] \quad (5.10)$$

da cui

$$\lambda^2 = 1 + \frac{\delta t}{\tau} \left(\frac{T_0}{T(t)} - 1 \right) \quad (5.11)$$

e quindi il nuovo valore della temperatura è dato da:

$$T'(t) = \sqrt{\left[1 + \frac{\Delta t}{\tau} \left(\frac{T_0}{T(t)} - 1 \right) \right]} T(t) \quad (5.12)$$

Per $\tau = \delta t$ si ottiene nuovamente l'Equ. 5.7, in tali condizioni il metodo di Berendsen diventa equivalente a quello isoterma.

5.4.4 Metodi di Nosé-Hoover

Un altro metodo largamente usato per il controllo della temperatura è quello di Nosé-Hoover [63, 42]. Per questo metodo si è dimostrato che produce conformazioni compatibili con un insieme canonico. Il metodo di Nosé, nella versione data da Hoover (1985), si basa sull'aggiunta alle usuali equazioni del moto di Newton di un termine (associato a una nuova variabile) che ha lo scopo di riportare (con un meccanismo a retroazione) l'energia cinetica del sistema a un valore medio predefinito.

5.4.5 Controllo della pressione

Il metodo dell'accoppiamento debole viene anche utilizzato per mantenere costante la pressione. Le equazioni del moto sono modificate in modo tale che il risultato netto sul sistema sia un rilassamento del primo ordine della pressione verso il valore prefissato P_0 :

$$\frac{dP(t)}{dt} = \frac{P_0 - P(t)}{\tau_p} \quad (5.13)$$

la pressione idrostatica P è definita per un sistema isotropico dalla relazione:

$$P(t) = \frac{2}{3} \frac{E_{cinetica}(t) - \Theta(t)}{V_{sistema}(t)} \quad (5.14)$$

in cui $V_{sistema}$ rappresenta il volume del sistema simulato e il viriale $\Theta(t)$ può essere ottenuto da

$$\Theta(t) = \frac{1}{2} \sum_{i < j} \mathbf{r}_{ij}(t) \cdot \mathbf{F}_{ij}(t) \quad (5.15)$$

in cui $\mathbf{F}_{ij}(t)$ è la forza che agisce tra gli atomi i e j . Per un sistema molecolare, le forze intramolecolari possono essere omesse dalla 5.15 insieme al contributo dell'energia cinetica. Le componenti della pressione lungo gli assi x , y e z è definito dalle componenti degli elementi diagonali del tensore P . La compressibilità isoterma (β_T) è legata al cambiamento di pressione a temperatura costante al cambiare del volume tramite la relazione:

$$\Delta P(t) = - \frac{\Delta V(t)}{\beta_T V(t)}. \quad (5.16)$$

Il cambiamento di volume può essere ottenuto moltiplicando le coordinate atomiche e gli spigoli della scatola con un fattore μ :

$$\Delta V(t) = (\mu(t)^3 - 1)V(t) \quad (5.17)$$

L'equazione 5.13 può essere discretizzata usando il passo di integrazione Δt e risolvendo rispetto a $\mu(t)$ si ottiene l'espressione:

$$\mu(t) = \left\{ 1 - \beta_T \frac{\delta t}{\tau_P} [P_0 - P(t)] \right\}^{1/3}. \quad (5.18)$$

Le equazioni modificate diventano:

$$\frac{d\mathbf{r}_i(t)}{dt} = \mathbf{v}_i(t) - \frac{\beta_T}{3\tau_P} [P_0 - P(t)] \mathbf{r}_i(t) \quad (5.19)$$

Il secondo termine nella equazione 5.19 determina il rilassamento della pressione $P(t)$ verso il il valore P_0 , la velocità di questo rilassamento viene controllata dal rapporto tra la compressibilità isoterma del sistema e il tempo di rilassamento della pressione (τ_P) che è stato scelto. Non è necessario conoscere accuratamente il valore di β_T poichè τ_P è un parametro che può essere agiustato. Tale valore dovrebbe essere sufficientemente piccolo (accoppiamento forte) per ottenere il valore di equilibrio della pressione, ma nello stesso tempo sufficientemente ampio (accoppiamento debole) da evitare di disturbare le proprietà del sistema.

Capitolo 6

Analisi delle traiettorie

L'analisi della traiettoria di una simulazione di dinamica molecolare consiste nel ricavare differenti proprietà macroscopiche di natura termodinamica o cinetica. Queste proprietà sono ottenute mediando lungo la traiettoria opportune grandezze. Come si è accennato nel Capitolo I, e come riprenderemo nel prossimo paragrafo, questa media è equivalente al valore meccanico statistico della grandezza desiderata in virtù dell'*ipotesi ergotica*. In questo capitolo, verranno descritte le proprietà che possono essere calcolate da simulazioni di DM. Le proprietà sono state divise in tre gruppi:

1. Proprietà termodinamiche.
2. Proprietà strutturali.
3. Proprietà dinamiche.

Come si vede dalla Tabella 6.1, queste proprietà possono essere confrontate con grandezze ottenibile da misure sperimentali dirette o indirette. La possibilità di avere questo confronto con misure sperimentali, consente di validare il modello microscopico del fenomeno che si sta studiando con le simulazioni.

6.1 Ipotesi Ergotica

Un sistema costituito da N particelle descriverà nel corso di una simulazione, una traiettoria nello spazio $6N$ dimensionale detto *spazio delle fasi*. Le $6N$ dimensioni sono costituite dalle $3N$ coordinate e dai $3N$ momenti del sistema. Da questa traiettoria è possibile (vedi Capitolo I) calcolare una stima della media della proprietà P attraverso l'integrale:

$$\langle P \rangle_t = \lim_{t \rightarrow \infty} \frac{1}{\tau} \int_0^t P(p^b, q^N) dt \quad (6.1)$$

In meccanica statistica si sostituisce alla media nel tempo, la media sulle possibili configurazioni che il sistema di N particelle può assumere nello spazio

delle fasi:

$$\langle A \rangle_e = \int \int dp^N dr^N A(p^N, q^N) \rho(p^N, q^N) dt \quad (6.2)$$

dove

$$\rho(p^N, q^N) A(p^N, q^N) = \frac{\exp \frac{-E(p^N, q^N)}{k_b T}}{Q} \quad (6.3)$$

è la densità di probabilità e Q è la funzione di partizione che, in un sistema in condizioni canoniche (NVT) è uguale a

$$Q_{NVT} = \frac{1}{N} \frac{1}{h^{3N}} \int \int dp^N dq^N \exp \frac{-E(p^N, q^N)}{k_b T} \quad (6.4)$$

L'ipotesi ergotica assume che per tempi molto lunghi valga l'equivalenza: $\langle P \rangle_e = \langle P \rangle_t$, ovvero che la traiettoria esplori tutte le possibili conformazioni del sistema nello spazio delle fasi. Purtroppo, anche con il sistema molecolare più semplice, l'esplorazione dello spazio delle fasi richiederebbe un tempo di calcolo troppo lungo per essere praticamente possibile. Pertanto, una simulazione di DM può fornire soltanto una stima delle proprietà termodinamiche del sistema simulato. Per aumentare l'accuratezza di questa stima è necessario effettuare diverse simulazioni partendo da punti diversi dello spazio delle fasi. Dalla stima della funzione di partizione è possibile ottenere tutte le proprietà termodinamiche macroscopiche del sistema quali, per esempio l'energia interna:

$$U = \frac{k_b T^2}{Q} \frac{\partial Q}{\partial T} \quad (6.5)$$

$$U = \int \int dp^N dq^N E(p^N, q^N) \rho(p^N, q^N) \quad (6.6)$$

e l'energia libera di Helmholtz:

$$A = -k_b \ln Q \quad (6.7)$$

$$A = k_b T \ln \left[\int \int dp^N dq^N \exp \frac{E(p^N, q^N)}{k_b T} \rho(p^N, q^N) \right] \quad (6.8)$$

Notare l'esponentiale positivo, questo fa sì che le configurazioni a più alta energia contribuiscano maggiormente al valore di A .

6.2 Determinazione delle proprietà termodinamiche

6.2.1 Entalpia di evaporazione

Da una simulazione in condizioni NPT è possibile calcolare il valore della entalpia di evaporazione (ΔH_{vap}) di una sostanza allo stato liquido. Questa grandezza è facilmente determinabile sperimentalmente da misure di tensione di vapore

e viene usata per valutare la qualità dei parametri del campo di forza usato per il modello della sostanza stessa. L'entalpia di evaporazione è il calore necessario a trasformare una mole di liquido nel suo vapore alla temperatura di 300 K e alla pressione di 1 bar. Il valore di ΔH_{vap} può essere ottenuto dall'energia interna molecolare nel seguente modo:

$$\Delta H_{vap} = \underbrace{[E_{inter}(g) - E_{inter}(l)]}_{\text{Contributo intermolecolare}} + \underbrace{[E_{intra}(g) - E_{intra}(l)]}_{\text{Contributo intramolecolare}} + \underbrace{RT}_{\text{Contributo p } \Delta V} \quad (6.9)$$

dove E_{inter} e E_{intra} sono le energie potenziali relative alle interazioni intermolecolari e intramolecolari, mentre (g) e (l) si riferiscono allo stato liquido e gassoso in cui la molecola si trova, quando i termini vengono calcolati. Il termine $E_{inter}(g)$ lo si considera nullo (di fatti è trascurabile).

6.2.2 Calcolo di funzioni di risposta termodinamiche

Una funzione di risposta indica come una semplice quantità termodinamica risponde al cambiamento di grandezze misurabili, quali la temperatura o la pressione. Un esempio di queste funzioni è la capacità termica a volume costante:

$$C_v = \left(\frac{\partial U}{\partial T} \right)_V \quad (6.10)$$

Le funzioni di risposta possono essere calcolate in due modi diversi. Nel primo metodo si effettuano diverse simulazioni per determinare i valori della grandezza termodinamica (l'energia interna, U , nel caso del C_v) in funzione della variabile indipendente (T per il C_v). Quindi la curva ottenuta viene derivata numericamente. Nel secondo metodo, la funzione di risposta viene derivata applicando la teoria delle fluttuazioni della meccanica statistica. In generale, il primo metodo è quello più accurato ma richiede l'uso di varie simulazioni. Il secondo consente di ottenere la funzione di risposta da una sola simulazione ma, a causa della lenta convergenza delle fluttuazioni della grandezza termodinamica da derivare, per migliorarne l'accuratezza occorre che questa simulazione sia molto lunga.

6.2.3 Capacità termica

La capacità termica si definisce come la quantità di calore assorbita da una sostanza per variazione di un grado di temperatura a volume (o pressione) costante. La capacità termica a volume costante, come si è detto nel precedente paragrafo, può essere determinata effettuando simulazioni della sostanza a due diverse temperature e approssimando con una differenza finita la derivata della energia interna:

$$C_v = \left(\frac{\partial U}{\partial T} \right)_V \approx \left(\frac{U_2 - U_1}{T_2 - T_1} \right)_V + 3R + C_V^{vibr} \quad (6.11)$$

In un insieme canonico (NVT) la capacità termica è anche legata alle fluttuazioni dell'energia interna dalla relazione:

$$C_v = \frac{\langle U^2 \rangle - \langle U \rangle^2}{NRT^2} + 3R + C_V^{vibr} \quad (6.12)$$

Questa espressione può essere derivata partendo dall'espressione della energia interna in termini di funzione di partizione nel seguente modo. L'energia interna U è uguale a

$$U = \frac{k_b T^2}{Q} \frac{\partial Q}{\partial T} \quad (6.13)$$

la derivata a volume costante è data da:

$$U = \frac{\partial}{\partial T} \left(\frac{k_b T^2}{Q} \frac{\partial Q}{\partial T} \right)_V = \frac{k_b T^2}{Q} \frac{\partial^2 Q^2}{\partial T^2} + 2 \frac{k_b T}{Q} \frac{\partial Q}{\partial T} - \frac{k_b T^2}{Q^2} \left(\frac{\partial Q}{\partial T} \right) \quad (6.14)$$

L'energia interna U non è niente altro che il valore atteso dell'energia $\langle E \rangle$ ovvero:

$$\langle E \rangle = \frac{k_b T^2}{Q} \frac{\partial Q}{\partial T} \quad (6.15)$$

Pertanto abbiamo che

$$2 \frac{k_b T^2}{Q} \frac{\partial Q}{\partial T} = 2 \frac{\langle E \rangle}{T} \quad (6.16)$$

e

$$\frac{k_b T^2}{Q^2} \left(\frac{\partial Q}{\partial T} \right) = k_b T \left(\frac{1}{Q} \frac{\partial Q}{\partial T} \right) = \frac{\langle E \rangle^2}{k_b T} \quad (6.17)$$

Il primo termine nella Equ. 6.14 si deriva in questo modo:

$$\frac{\partial}{\partial T} \left(\frac{\langle E \rangle}{k_B T^2} \right) = \frac{\partial}{\partial T} \left[\frac{1}{Q} \left(\frac{\partial Q}{\partial T} \right) \right] \quad (6.18)$$

da cui

$$-2 \frac{\langle E \rangle}{k_B T^3} = \frac{1}{Q} \frac{\partial^2 Q}{\partial T^2} + \frac{\partial Q}{\partial T} \frac{\partial}{\partial T} \left(\frac{1}{Q} \right) \quad (6.19)$$

usando la regola della catena si ottiene

$$\frac{\partial Q}{\partial T} \frac{\partial}{\partial T} \left(\frac{1}{Q} \right) = \frac{\partial Q}{\partial T} \frac{\partial Q}{\partial T} \frac{\partial}{\partial T} \frac{1}{Q} = - \left(\frac{\partial Q}{\partial T} \right)^2 \left(\frac{1}{Q} \right)^2 \quad (6.20)$$

da cui

$$\frac{k_b T^2}{Q} \frac{\partial^2 Q^2}{\partial T^2} = -2 \frac{\langle E \rangle}{k_b T^3} + \frac{\langle E^2 \rangle}{k_B^2 T^4} \quad (6.21)$$

quindi sostituendo le precedenti espressioni nella Equ. 6.14 si ottiene:

$$C_v = k_b T^2 \left\{ -2 \frac{\langle E \rangle}{k_b T^3} + \frac{\langle E^2 \rangle}{k_B^2 T^4} \right\} + 2 \frac{\langle E \rangle}{T} - \frac{\langle E \rangle^2}{k_b T^2} \quad (6.22)$$

da cui

$$C_v = \frac{(\langle E^2 \rangle - \langle E \rangle^2)}{k_b T^2} \quad (6.23)$$

La capacità termica a pressione costante può essere ottenuta usando la relazione di Mayer:

$$C_v = C_p + T \frac{\left(\frac{\partial V}{\partial T}\right)_P^2}{\left(\frac{\partial V}{\partial P}\right)_T} = C_p - TV \frac{\alpha^2}{\beta_T} \quad (6.24)$$

dove V è il volume molare, α il coefficiente di espandibilità termica e β_T il coefficiente di compressibilità isoterma.

6.2.4 Compressibilità isoterma

La compressibilità isoterma, β_T , misura come il volume del sistema risponde a un cambiamento isoterma reversibile di pressione. Usando le differenze finite si può ottenere la seguente relazione che permette di derivare β_T dalla variazione di densità ρ a due differenti temperature.

$$\beta_T = \frac{1}{V} \left(\frac{\partial V}{\partial P}\right)_T = \frac{1}{\rho} \left(\frac{\partial V}{\partial P}\right)_T = \left(\frac{\partial \ln \rho}{\partial P}\right)_T \approx \left(\frac{\ln(\rho_2/\rho_1)}{P_2 - P_1}\right)_T \quad (6.25)$$

6.2.5 Coefficiente di dilatazione termico

Il coefficiente di dilatazione termico, α_P , misura come il volume del sistema risponde a un cambiamento isobaro reversibile di temperatura. Usando le differenze finite si può ottenere la seguente relazione che permette di derivare α_P dalla variazione di densità ρ a due differenti temperature.

$$\alpha_P = \frac{1}{V} \left(\frac{\partial V}{\partial T}\right)_P = \frac{1}{\rho} \left(\frac{\partial V}{\partial T}\right)_P = \left(\frac{\partial \ln \rho}{\partial T}\right)_P \approx \left(\frac{\ln(\rho_2/\rho_1)}{T_2 - T_1}\right)_P \quad (6.26)$$

6.2.6 Calcolo delle differenze di energia libera

Il calcolo delle differenze di energia libera possono essere ottenute utilizzando due differenti metodi:

- Perturbazione termodinamica
- Integrazione termodinamica

Perturbazione termodinamica

Si considerino due stati A e B di un sistema, la differenza di energia libera è data da

$$A_A - A_B = -k_b T \ln Q_A + k_b T \ln Q_B = k_b T \ln \frac{Q_A}{Q_B} \quad (6.27)$$

dove Q_A e Q_B sono le funzioni di partizione per lo stato A e B. Per la definizione di Q , la precedente espressione diventa

$$A_A - A_B = -k_b T \ln \left[\sum \exp -\frac{E_A - E_B}{k_b T} \right] \quad (6.28)$$

ovvero

$$A_A - A_B = -k_b T \ln \left\langle \exp -\frac{E_A - E_B}{k_b T} \right\rangle_M \quad (6.29)$$

se la differenza tra i due stati A e B è maggiore di $k_b T$ allora si pu' ricorrere alla determinazione di stati intermedi per i quali $E_\lambda = \lambda E_A + (1 - \lambda) E_B$ da cui

$$A_A - A_B = -k_b T \sum_\lambda \ln \left\langle \exp -\frac{\partial E_\lambda}{k_b T} \right\rangle_M \quad (6.30)$$

Integrazione termodinamica

$$A = -k_b T \ln Q(\lambda) \quad (6.31)$$

$$\frac{\partial A}{\partial \lambda} = -\frac{k_b T}{Q} \frac{\partial Q}{\partial \lambda} \quad (6.32)$$

poichè

$$Q_{NVT} = \frac{1}{N} \frac{1}{h^{3N}} \int \int dp^N dq^N \exp \frac{-E(p^N, q^N)}{k_b T} \quad (6.33)$$

allora

$$\frac{\partial Q}{\partial \lambda} = \frac{1}{N} \frac{1}{h^{3N}} \int \int dp^N dq^N \frac{\partial}{\partial \lambda} \exp \frac{-E(p^N, q^N)}{k_b T} \quad (6.34)$$

$$= \frac{1}{N} \frac{1}{h^{3N}} \frac{1}{k_b T} \int \int dp^N dq^N \frac{\partial E(p^N, q^N)}{\partial \lambda} \exp \frac{-E(p^N, q^N)}{k_b T} \quad (6.35)$$

sostituendo questa espressione nella Equ. 6.32 si ottiene

$$\frac{\partial A(\lambda)}{\partial \lambda} = \frac{1}{N} \frac{1}{h^{3N}} \frac{1}{Q(\lambda)} \int \int dp^N dq^N \frac{\partial}{\partial \lambda} \exp \frac{-E(p^N, q^N)}{k_b T} \quad (6.36)$$

$$= \frac{1}{N} \frac{1}{h^{3N}} \int \int dp^N dq^N \frac{\partial E(p^N, q^N)}{\partial \lambda} \left[\frac{\exp \frac{-E(p^N, q^N)}{k_b T}}{Q(\lambda)} \right] \quad (6.37)$$

$$= \left\langle \frac{\partial E(p^N, q^N)}{\partial \lambda} \right\rangle \quad (6.38)$$

$$\Delta A = \int_0^1 \frac{\partial E(p^N, q^N)}{\partial \lambda} d\lambda \quad (6.39)$$

Il fattore lambda viene applicato moltiplicandolo per le funzioni di potenziale dello stato A e B in questo modo $V_\lambda = \lambda V_A + (1 - \lambda)V_B$. Per esempio nel caso di una funzione di LJ otteniamo:

$$V_{LJ}(r) = (1 - \lambda) \left(\frac{C_{12}^A}{r^{12}} - \frac{C_6^A}{r^6} \right) + \lambda \left(\frac{C_{12}^B}{r^{12}} - \frac{C_6^B}{r^6} \right) \quad (6.40)$$

Pertanto quando $\lambda = 0$ il sistema si troverà nello stato A mentre per $\lambda = 1$ nello stato B. Per lambda intermedi, il sistema sarà una miscela dei due stati. Nella Figura 6.1 viene riportato un esempio di calcolo dell'energia libera di idratazione dello ione bromuro in soluzione acquosa. In questo caso, l'integrazione termodinamica viene applicata facendo crescere l'interazioni dello ione con l'acqua aumentando progressivamente il valore di λ .

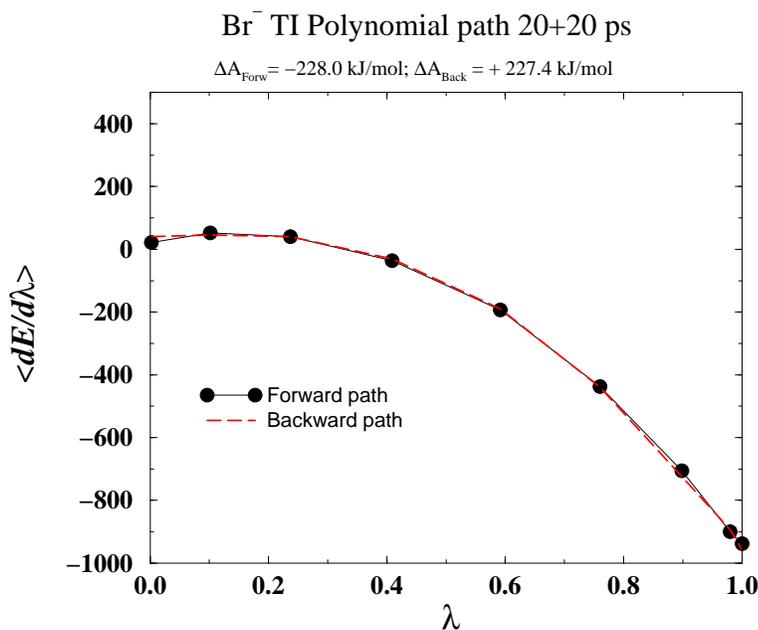


Figura 6.1: Esempio di calcolo della energia libera d'idratazione dello ione Br⁻ con il metodo della integrazione termodinamica. I due percorsi (Forward e Backward) sono stati ottenuti variando λ da 0 a 1 e viceversa. Per ogni lambda si è eseguita una simulazione di circa 100 ps.

6.3 Determinazione delle proprietà strutturali

6.3.1 Funzioni di distribuzione atomiche

Lo stato fluido è caratterizzato dalla assenza di una struttura permanente, questo significa che ogni molecola (o atomo) del fluido ha intorno a se dei continue riorganizzazioni delle altre molecole. Queste riorganizzazioni danno origine a ben definite correlazioni strutturali che possono essere messe in evidenza con opportune tecniche sperimentali [60, 31] Il trattamento delle correlazioni strutturali si ottiene dalla funzione di distribuzione canonica nella forma generale [60, 31]:

$$g(\mathbf{r}_1, \mathbf{r}_2) = \frac{N_a(N_a - 1) \int d\mathbf{r}_3 \dots \mathbf{r}_{N_a} \exp(-U(\mathbf{r}_1, \dots, \mathbf{r}_{N_a})/k_B T)}{\rho^2 \int d\mathbf{r}_1 \dots \mathbf{r}_{N_a} \exp(-U(\mathbf{r}_1, \dots, \mathbf{r}_{N_a})/k_B T)} \quad (6.41)$$

dove l'integrale al denominatore è la funzione di partizione, e l'integrale al denominatore differisce solo perchè \mathbf{r}_1 e \mathbf{r}_2 sono esclusi dall'integrazione. Nel caso di sistemi spazialmente omogenei, solo la distanza relativa tra le molecole è significativa, riducendo la 6.41 a una somma sulle coppie di molecole:

$$g(\mathbf{r}) = \frac{2V}{N^2} \left\langle \sum_{i < j} \delta(\mathbf{r} - \mathbf{r}_{ij}) \right\rangle \quad (6.42)$$

Se il sistema è anche isotropo la funzione può essere mediata sugli angoli senza perdita di informazione. La risultante funzione di distribuzione $g(r)$ descrive la organizzazione locale intorno a ogni dato atomo.

La $g(r)$ gioca un ruolo centrale nella fisica dello stato liquido e tutte le funzioni che dipendono dalla separazione a coppie (come la funzione di potenziale e la pressione) possono essere espresse in termini di integrali che coinvolgono la $g(r)$. La definizione di $g(r)$ implica che $\rho g(r) dr$ è proporzionale alla probabilità di trovare un atomo nell'elemento di volume $d\mathbf{r}$ alla distanza r da un dato atomo, e, nel caso di un fluido omogeneo ed isotropo, $4\pi\rho g(r)r^2 dr$ rappresenta il numero medio di atomi in un guscio di raggio r e spessore dr che circonda l'atomo. La $g(r)$ può essere misurata sperimentalmente con diverse tecniche. In generale essa è legata al fattore di struttura $S(\mathbf{k})$ attraverso la trasformata di Fourier. Il fattore di struttura può essere ottenuto o da misure di diffrazione di raggi X o di neutroni. In generale la relazione che lega $S(\mathbf{k})$ con $g(\mathbf{r})$ è

$$S(\mathbf{k}) = 1 + \rho \int g(\mathbf{r}) \exp(-i\mathbf{k} \cdot \mathbf{r}) d\mathbf{r} \quad (6.43)$$

che per fluidi isotropi si semplifica in

$$S(k) = 1 + 4\pi\rho \int \frac{\sin kr}{kr} g(r) r^2 dr \quad (6.44)$$

Un'altra tecnica sperimentale da cui è possibile determinare la funzione $g(r)$ intorno a particolari atomi sono le spettroscopie di assorbimento dei raggi X. Nel

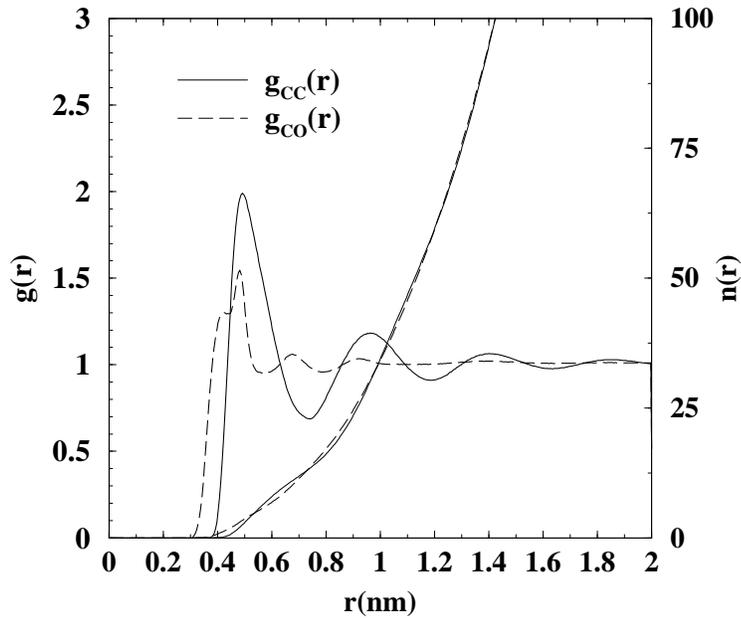


Figura 6.2: Esempio di $g(r)$.

capitolo V e VI questa tecnica verrà ampiamente illustrata. Queste relazioni costituiscono il legame tra le simulazioni di dinamica molecolare e il fenomeno fisico reale. Infatti la simulazione può fornire l'interpretazione di un dato sperimentale che altrimenti sarebbe difficile da spiegare.

6.4 Determinazione di proprietà dinamiche

6.4.1 Calcolo della costante di diffusione

La diffusione di un soluto in fluido è descritta dalle due leggi fenomenologiche di Fick:

$$J_z = -D \frac{\partial N}{\partial z} \quad (6.45)$$

$$\frac{\partial N}{\partial z} = D \frac{\partial^2 N(z, t)}{\partial z^2} \quad (6.46)$$

dove N è la concentrazione del soluto e z è la direzione spaziale lungo cui la diffusione avviene. La costante di diffusione D determina la velocità con cui il processo ha luogo. Nel 1905, A. Einstein ha mostrato che D è legato alla distanza media percorsa dalla particella al tempo t rispetto alla posizione iniziale, attraverso la relazione [3]:

$$6Dt = \frac{1}{N} \sum_{i=1}^N [\mathbf{r}_i(t) - \mathbf{r}_i(0)]^2 \quad (6.47)$$

dove $\mathbf{r}_i(t)$ è vettore delle coordinate della particella i -esima al tempo t , $\mathbf{r}_i(0)$ è il vettore delle coordinate della particella i -esima al tempo $t = 0$ e N numero totale di particelle.

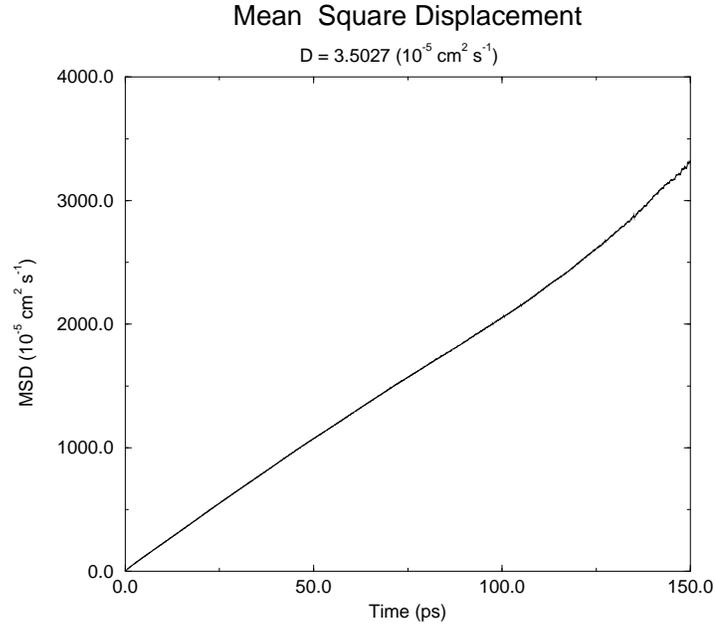


Figura 6.3: Esempio di $g(r)$.

Un' altro modo di calcolare D consiste nell'uso della funzione di autocorrelazione delle velocità del centro di massa della molecola, attraverso la formula di Green-Kubo [3]. Sia

$$|\mathbf{r}(t) - \mathbf{r}(0)| = \int_0^t \mathbf{v}e(t') dt' \quad (6.48)$$

elevando al quadrato entrambi i membri e calcolando la media si ottiene:

$$\langle |\mathbf{r}(t) - \mathbf{r}(0)|^2 \rangle = \int_0^t dt' \int_0^{t'} dt'' \langle \mathbf{v}(t') \cdot \mathbf{v}(t'') \rangle \quad (6.49)$$

poichè

$$\langle \mathbf{v}e(t') \cdot \mathbf{v}e(t'') \rangle = \langle \mathbf{v}e(t'' - t') \cdot \mathbf{0} \rangle \quad (6.50)$$

si ottiene la formula di Green-Kubo:

$$\langle |\mathbf{r}(t) - \mathbf{r}(0)|^2 \rangle = \int_0^t dt \langle \mathbf{v}(\tau) \cdot \mathbf{v}(0) \rangle \left(1 - \frac{\tau}{t}\right) \quad (6.51)$$

che al limite per t che tende ad infinito diventa:

$$\int_0^{T_{MD}} \left[\frac{1}{N} \sum_{i=1}^N C_i(t) \right] dt = 3D \quad (6.52)$$

dove

$$C_i(t) = \frac{1}{T_{MD} - t} \int_0^{T_{MD}-t} \mathbf{v}_i(t') \mathbf{v}_i(t' + t) dt' \quad (6.53)$$

è la funzione di autocorrelazione delle velocità e T_{MD} è il tempo totale della simulazione.

6.4.2 Il tempo di residenza e numero d'idratazione

Nello studio di processi di idratazione o solvatazione sono coinvolti fenomeni di scambio molecolare la cui scala di tempi non consente, generalmente, di avere informazioni sperimentali a livello microscopico. La funzione di distribuzione radiale fornisce le proprietà statiche di questo processo, ovvero la distanza media e il numero di coordinazione, cioè il numero medio di molecole intorno a un dato atomo o molecola. Se si osserva il processo su scale di tempo diversi, ci si accorge che il fenomeno della idratazione e solvatazione è in qualche modo legato al tempo di residenza medio delle molecole di solvente nella prima sfera di idratazione del soluto. Questo tempo di residenza può essere ottenuto molto semplicemente dalle simulazioni di DM e i valori ottenuti possono essere comparati con quelli ottenuti da tecniche sperimentali quali la Risonanza Magnetica Nucleare [64].

La definizione precisa di tempo di residenza la si può dare introducendo la funzione $P_j(t, t_n; t^*)$. Questa funzione può assumere due valori, 0 o 1, e si riferisce alla molecola di solvente j . Il valore uno viene assunto quando la molecola j si trova nella prima sfera d'idratazione del soluto al tempo t_n e al tempo $t + t_n$ e nel tempo intermedio non abbia lasciato la sfera di coordinazione per un periodo non superiore a t^* . Negli altri casi la funzione assume valore 0. Si può ora definire una quantità media caratteristica del soluto, definita dalla espressione:

$$n_{soluto} = \frac{1}{N} \sum_{n=1}^{N_t} \sum_j P_j(t, t_n; t^*) \quad (6.54)$$

Al tempo $t = 0$ segue dalla definizione che $n_{soluto}(0) = N_{coo}$, dove N_{coo} è il numero di coordinazione ottenuto dalla $g(r)$, mentre $n_{soluto}(t)$ rappresenta il numero di molecole di solvente che si trovano inizialmente nella prima sfera di coordinazione e rimangono ancora lì dopo che è trascorso un tempo t .

Il parametro t^* tiene conto delle molecole che lasciano solo temporaneamente la prima sfera di idratazione, senza entrare nella massa del solvente. Il comportamento di $n_{soluto}(t)$ su scale di tempi molto ampie è di tipo esponenziale, cioè $n_{soluto}(t) \approx n_{soluto} \exp(-t/\tau_{soluto}^s)$. La quantità τ_{soluto}^s rappresenta il tempo di correlazione per la persistenza della prima sfera di coordinazione intorno alla molecola di soluto. Lo stesso tipo di misura può essere effettuata per la massa del solvente ottenendo un tempo caratteristico $\tau_{solvente}^s$. Usando le definizioni dei tempi di τ_{soluto}^s e $\tau_{solvente}^s$ si può dare la definizione di numero di solvatazione n_{idr} di un soluto in un solvente come il numero medio di molecole di solvente che rimangono nella prima sfera di idratazione per il periodo di tempo in cui la massa del solvente della sfera di coordinazione è rinnovata. A causa della natura esponenziale delle funzioni $n_{soluto}(t)$ e $n_{solvente}(t)$, questo numero è dato da:

$$n_{idr} = n_{soluto} e^{\frac{\tau_{solvente}^s}{\tau_{soluto}^s}} \quad (6.55)$$

Tabella 6.1: Alcuni esempi di possibili confronti tra dati sperimentali e proprietà simulate dalla DM.

Proprietà atomiche	Metodi sperimentali	Gas	Soluzione	Membrana	Cristallo
Struttura					
-Posizioni	Diffrazione dei raggi X, neutroni o elettroni	✓			✓
-Distanza	Risonanza magnetica		✓		
-Orientazione	NMR				✓
Mobilità					
-fattori di Debye-Waller	Diffrazione dei raggi X, neutroni				✓
-fattori di occupanza	Diffrazione dei raggi X, neutroni				✓
Dinamiche					
-Frequenze di vibrazione	Spettroscopie nell'infrarosso	✓			✓
-Velocità di rilassamento	NMR e spett. ottiche		✓	✓	
-Diffusione	NO spin label		✓	✓	
Termodinamiche					
-densità	densitometria		✓	✓	✓
-Entalpia di evaporazione			✓		
-Viscosità e conduttanza	viscosimetria		✓	✓	

Capitolo 7

Calcolo parallelo e la DM

I recenti sviluppi tecnici nell'ambito dei calcolatori paralleli hanno reso disponibile notevoli potenze di calcolo a costi relativamente contenuti e hanno stimolato un intenso lavoro di programmazione atto realizzare per queste macchine gli stessi programmi per il calcolo scientifico e tecnico usati nei tradizionali calcolatori seriali. Uno dei settori scientifici che sta traendo beneficio da questi sviluppi è la chimica computazionale ed in particolare la Dinamica Molecolare. Il motivo di ciò è legato al crescente interesse nello studio, con questa tecnica, di sistemi sempre più grandi per tempi di simulazione molto lunghi con la conseguente necessità di elevate potenze di calcolo. Per questo motivo, negli ultimi anni sono stati fatti notevoli sforzi per ottimizzare i codici di DM su calcolatori paralleli aventi differenti architetture.

7.1 I calcolatori paralleli

Il termine *calcolo parallelo* si riferisce a un tipo di calcolo che utilizza contemporaneamente diversi processori. I calcolatori in cui si esegue questo tipo di calcolo vengono definiti in modi differenti: *multicomputer*, calcolatori paralleli, *cluster*, multiprocessori, etc. Ognuno di questi nomi suggerisce un particolare tipo di architettura. In generale sarà usato il termine *calcolatore parallelo* per indicare un sistema avente più processori. Ogni unità di processo verrà indicata come un *nodo* del calcolatore.

I calcolatori paralleli possono essere classificati, in base al modo con cui le istruzioni e i dati vengono elaborati, in due categorie fondamentali:

- *single instruction stream multiple data stream* (SIMD)
- *multiple instruction stream multiple data stream* (MIMD).

Un calcolatore SIMD è formato da numerosi nodi che eseguono nello stesso momento la stessa istruzione su dati differenti, distribuiti nella memoria dei singoli nodi, ottenendo in tal modo il parallelismo. I calcolatori MIMD sono

basati su una architettura parallela più generale con processori più complessi, in grado di eseguire, indipendentemente gli uni dagli altri, i propri programmi e operare sui propri dati.

In generale questi due tipi di architetture possono essere costituiti da centinaia a migliaia di processori ognuno dei quali dispone di una propria memoria locale.

La semplicità della architettura SIMD consente di avere, a parità di numero e potenza dei processori, prestazioni più elevate rispetto a macchine MIMD. D'altra parte comporta alcune limitazioni tecniche che ne rendono più difficile la programmazione quali per esempio il dover eseguire contemporaneamente su ogni processore lo stesso tipo di istruzione e la mancanza, in alcuni tipi di macchine SIMD, dell'indirizzamento locale delle variabili ovvero della possibilità di accedere localmente ad elementi di vettori.

Negli ultimi anni sono stati realizzati diversi programmi di dinamica molecolare per architetture di tipo MIMD con poche decine di processori [66, 34, 73] e, più recentemente, anche per macchine con centinaia o migliaia di processori [40, 67, 68].

Inoltre diversi programmi di DM, sviluppati per calcolatori seriali, per la simulazione di macromolecole biologiche, come CHARMM [11] e GROMOS [84], sono stati modificati per poter funzionare con calcolatori paralleli di tipo MIMD [68, 12, 55, 17].

I programmi di DM, soprattutto per la simulazione di sistemi molecolari complessi, per architetture di tipo SIMD sono meno numerosi. [96, 24, 62, 6]

7.2 Algoritmi per simulazioni di DM su calcolatori paralleli

In un programma di DM il calcolo delle forze di non legame è la parte che richiede il maggior tempo di calcolo, infatti, a seconda del metodo usato, può richiedere fino al 90% del tempo di calcolo totale. Per questo motivo nella realizzazione di un codice di un programma di DM che possa funzionare su calcolatori paralleli, occorre creare degli algoritmi efficienti in grado di parallelizzare il calcolo di queste interazioni. Due sono le possibili strategie.

7.2.1 Decomposizione geometrica.

Per effettuare la decomposizione geometrica si divide il sistema molecolare in $n_x \times n_y \times n_z$ scatole rettangolari (pari al numero di nodi che si stanno usando), contenenti un numero di atomi simile (Figura 7.1).

Gli atomi contenuti in ogni scatola vengono assegnati ad un processore del calcolatore, che si occuperà di calcolare le interazioni tra gli atomi contenuti in quella scatola. I processori adiacenti comunicano tra loro le informazioni sugli atomi che si trovano ai bordi delle rispettive scatole. Questa partizione fa in modo che le comunicazioni tra i processori sono molto limitate si riducono

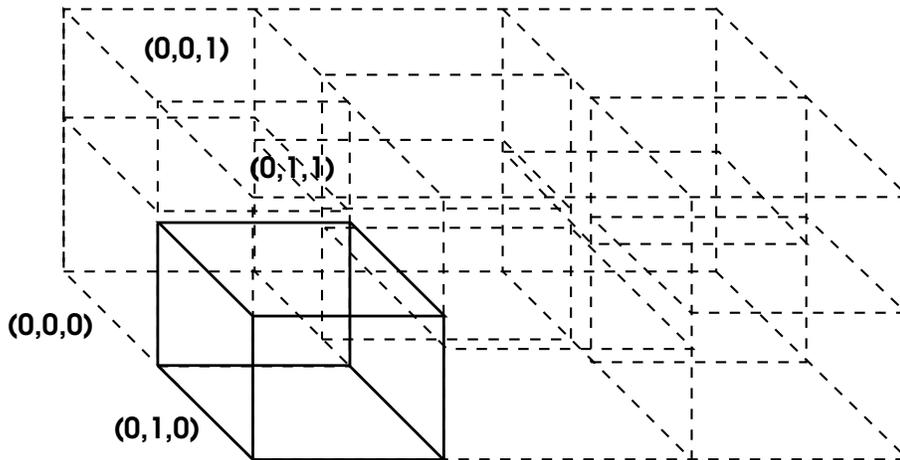


Figura 7.1: Suddivisione del sistema molecolare in scatole aventi un numero simile di atomi.

all'aggiornamento delle interazioni degli atomi che si trovano a confine tra le varie regioni.

7.2.2 Decomposizione dei dati e il *systolic loop*.

Il metodo del *systolic loop* (SL) è uno dei metodi più efficienti per calcolare le interazioni a due corpi su macchine SIMD e MIMD. [96, 62, 78, 45] Nella Figura 7.2 viene riportato schematicamente il principio del suo funzionamento. Il metodo consiste nell'assegnare a ogni nodo della macchina un gruppo di atomi (indicati come *residenti*). In ogni fase del ciclo ogni nodo calcola le interazioni tra il gruppo di atomi locali e un gruppo di atomi che gli è stato passato da un nodo adiacente (gruppo *transiente*). Completato il calcolo, il gruppo di atomi transienti viene poi passato a un altro nodo adiacente (seguendo una certa direzione) e ha luogo la fase successiva. Le interazioni calcolate vengono memorizzate sia localmente sia trasferite agli altri nodi insieme alle informazioni degli atomi transienti; in questo modo non è necessario effettuare la rotazione completa dei gruppi di atomi transienti in tutti i P nodi della macchina, ma, usando la III legge della dinamica, ci si ferma dopo $P/2$ fasi (se P è pari). Questa strategia di calcolo è stata adottata anche dal programma Gromacs.

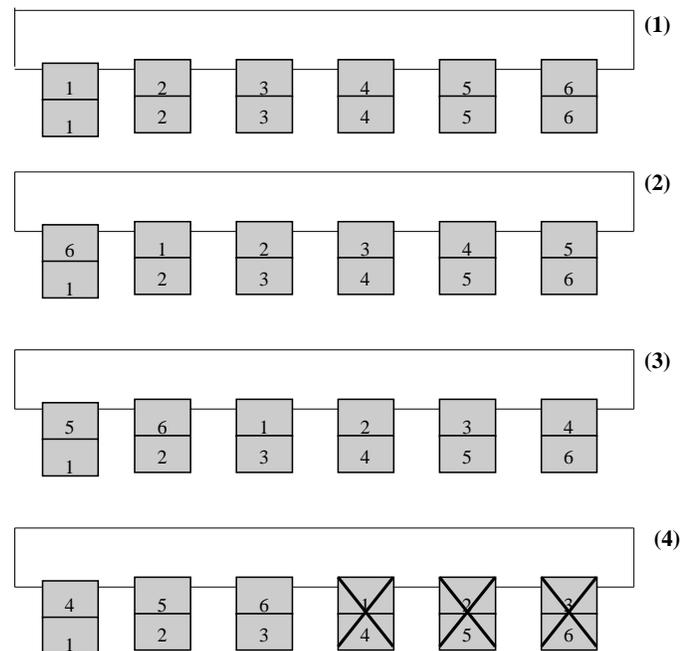


Figura 7.2: Il *systolic loop*. Ogni riquadro rappresenta un processore. Il numero nel riquadro in basso indica il gruppo di atomi residenti in quel processore mentre il numero nel riquadro in alto indica il gruppo di atomi transienti. Il numero tra parentesi a destra indica lo spostamento lungo l'anello. I processori segnati con X nel 4° spostamento sono forzati a rimanere inattivi per evitare di duplicare il calcolo delle interazioni già effettuato nei primi tre processori.

Capitolo 8

Simulazione di un sistema molecolare

In questo capitolo verrà fornita una descrizione di come preparare una simulazione di DM di un sistema molecolare. Per questo scopo, utilizzeremo il pacchetto di programmi per la DM, Gromacs. La scelta di questo pacchetto è motivata dal fatto che è uno dei programmi per simulazioni di DM più veloci e versatili liberamente distribuito sotto licenza GNU General Public Licence.

8.0.3 Il programma Gromacs

GroMaCS (Groningen MACHine for Chemical Simulations) è un pacchetto di programmi per simulazioni di Dinamica Molecolare concepito all'Università di Groningen dal Prof. H.J.C Berendsen e sviluppato in gran parte da alcuni studenti di PhD del suo gruppo. Inizialmente, Gromacs era il nome di un calcolatore parallelo i cui processori erano stati progettati appositamente per velocizzare i calcoli di simulazioni di Dinamica Molecolare. Per questo calcolatore, fu realizzato un apposito programma per simulazioni di DM, scritto in linguaggio C, basato sul campo di forze di GROMOS (vedi capitolo II) sviluppato anni prima nella stessa Università da H.J.C. Berendsen e W.F. van Gunsteren.

La rapida evoluzione dei microprocessori, rese rapidamente obsoleto il calcolatore Gromacs e i costi per un ulteriore sviluppo della macchina risultarono superiori all'utilizzo dei sempre più economici cluster di PC. Per questo motivo il perfezionamento del calcolatore venne fermato. Tuttavia, il progetto non venne completamente abbandonato, infatti il software utilizzato per effettuare le simulazioni era stato notevolmente sviluppato e reso funzionante su altri tipi di calcolatori seriali e paralleli. Il programma (che prese quindi il nome dal calcolatore) era stato fin dall'inizio concepito per essere usato in ambienti Unix e per il calcolo parallelo, e sono questi i due punti di forza che lo rendono sempre più diffuso negli ambienti di ricerca accademici e industriali. L'opera di incessante perfezionamento del suo creatore David van der Spoel affiancata

negli ultimi anni dai contributi di altrettanto validi collaboratori, ha reso Gromacs il più veloce programma di Dinamica Molecolare esistente. Nel 2001, Con l'uscita della versione 3.0, il programma è stato posto sotto licenza GNU GPL ed è così possibile contribuire liberamente al suo sviluppo. Il sito ufficiale di Gromacs è :

www.gromacs.org

Nel sito, oltre al codice sorgenti, gli eseguibili e la documentazione, è possibile trovare gli archivi di una *mail-list* molto attiva da cui attingere suggerimenti e informazioni sul programma e sulla tecnica della DM.

8.0.4 Struttura di Gromacs

A differenza di altri programmi di DM, Gromacs è altamente modularizzato in pieno accordo con la filosofia Unix. È infatti costituito da un insieme di programmi separati, suddivisi in varie categorie in base alla loro funzione:

- **Programmi per la generazione della topologia e delle coordinate:** pdb2gmx, editconf, genbox, genion, genbox, genconf, make_ndx
- **Programmi per la simulazione:** grompp, tpbconv, mdrun
- **Programmi di analisi delle traiettorie:** g_energy, g_analyze, g_rmsd, g_rmsf, etc.

Questi programmi sono collegati tra loro secondo il diagramma riportato in Figura 8.1.

8.0.5 Schema operativo per la simulazione di Dinamica Molecolare di un sistema molecolare

1. **Generazione delle coordinate atomiche iniziali della molecola da simulare.** Le coordinate iniziali della molecola possono essere ottenute in vari modi:
 - (a) Database di strutture ottenute da diffrazione di raggi X o NMR: Cambridge Structural Database (CSD) Protein Data Bank (PDB)
 - (b) Programmi di costruzione di strutture molecolari o database di strutture di piccole molecole.
 2. **Costruzione della topologia molecolare.** Si crea un file con la lista di tutti gli atomi con le relative proprietà, e dei legami, angoli di valenza, angoli diedri propri e impropri presenti nella molecola.
 3. **Generazione di una scatola di molecole.** Nel caso si voglia simulare un liquido è necessario replicare le coordinate di una molecola fino a riempire uniformemente un volume tale da ottenere la corretta densità del liquido ad una certa temperatura (per questo scopo si utilizzano i programmi: *editconf* e *genconf*). Nel caso si voglia studiare la molecola in un
-

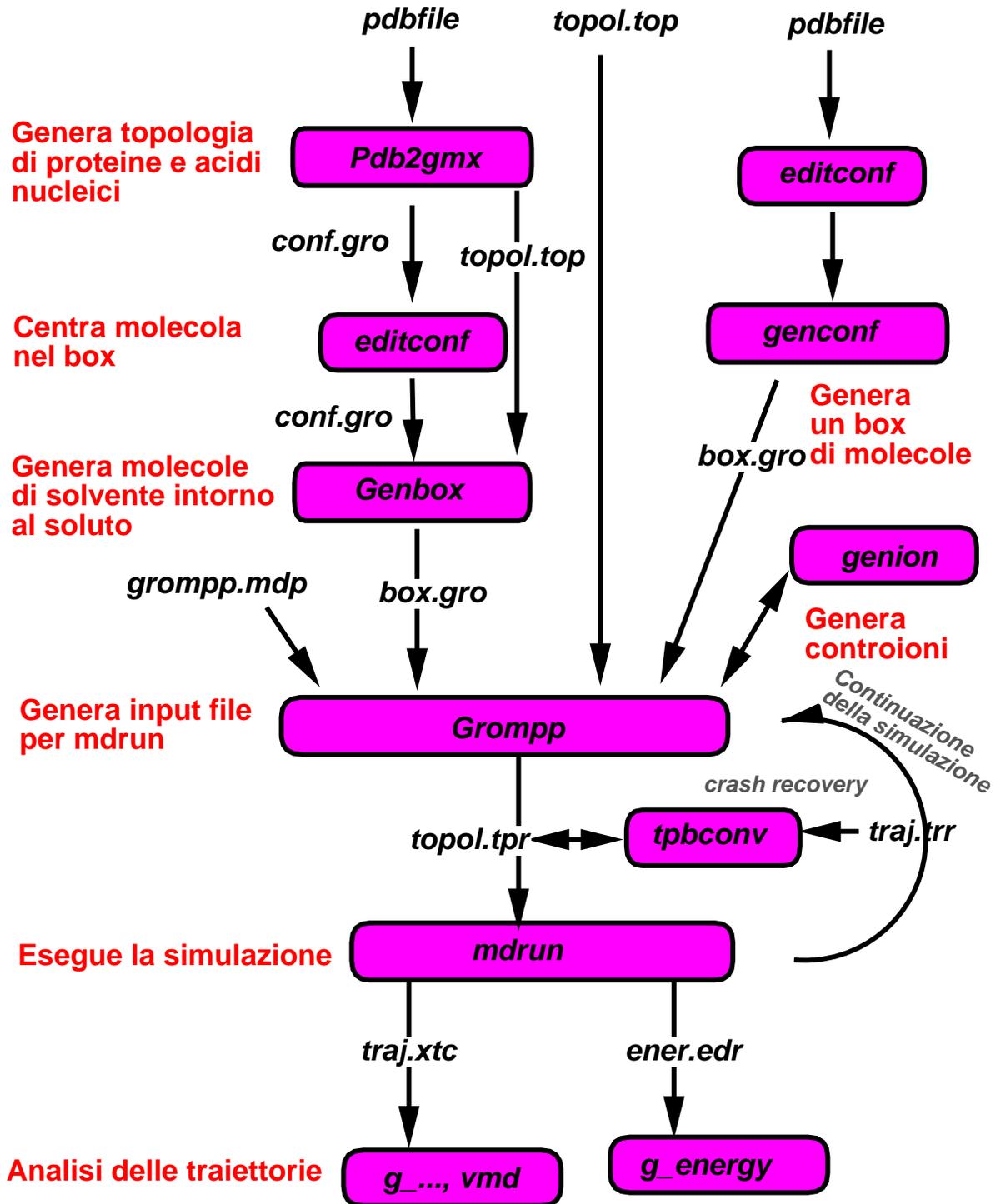


Figura 8.1: Diagramma di flusso del pacchetto di programmi GROMACS.

solvente allora occorre porre una (o più molecole) al centro di una scatola che viene poi uniformemente riempita con le molecole del solvente. Questo viene fatto replicando le coordinate equilibrate di molecole di solvente, fino a riempire la scatola di simulazione ed eliminando quelle molecole che si sovrappongono al soluto. Il programma usato per questo scopo è il *genbox*.

4. **Minimizzazione dell'energia potenziale totale.** Questa procedura serve a portare il sistema in un minimo locale di energia potenziale, in questo modo vengono eliminate eventuali tensioni che potrebbero dare problemi alla partenza della DM. Questa procedura viene effettuata con il programma di simulazione *mdrun* usando opportuni parametri per la procedura di minimizzazione.

5. **Simulazione di DM.** In questa fase vengono effettuate le simulazioni di DM. Una volta deciso il tipo di stato termodinamico con cui effettuare la simulazione (NVE,NVT,NPT) e scelti i parametri di controllo, in relazione al tipo di simulazione che si vuole effettuare, si può procedere all'esecuzione della simulazione stessa. Prima di iniziare la simulazione vera e propria è necessario equilibrare il sistema. Nel caso di un soluto immerso in una scatola di solvente, bisogna innanzitutto, equilibrare (alla temperatura ambiente e per qualche centinaio di picosecondi) le molecole di solvente intorno al soluto. Per far ciò, le coordinate del soluto vengono bloccate in una posizione di riferimento (conformazione iniziale) utilizzando i vincoli di posizione. Questo consente alle molecole di solvente di ridistribuirsi (generate con una procedura geometrica) intorno al soluto e nella scatola di simulazione omogeneamente eliminando possibili artefatti iniziali.

Quindi, per poter rilassare anche il soluto si procede ad un riscaldamento progressivo del sistema partendo da una temperatura molto bassa fino alla temperatura a cui si vuole effettuare la simulazione. Questa procedura può essere effettuata discontinuamente, ovvero con varie simulazioni di una certa lunghezza e a temperature progressivamente crescenti, oppure in modo continuo, lasciando al programma il compito di innalzare progressivamente la temperatura di riferimento del bagno termico fino al valore finale.

Una volta che il sistema è stato preparato in questo modo, si può procedere alla simulazione vera e propria. In genere vengono eliminati i primi 100 ps della traiettoria considerandoli di equilibratura.

6. **Analisi della traiettoria.** Le traiettorie finali vengono analizzate per studiare le proprietà strutturali, termodinamiche e cinetiche del sistema.

Nei prossimi paragrafi verranno descritti in dettaglio alcuni dei punti sopra elencati.

8.0.6 Le coordinate atomiche della molecola.

Le coordinate atomiche iniziali della molecola si possono ottenere in vari modi. Se è nota la struttura ai raggi X del cristallo del composto, allora si possono usare le coordinate cristallografiche reperibili dal:

- **CSD**, Cambridge Structural Database (<http://www.ccdc.cam.ac.uk>) per molecole organiche e inorganiche.
- **PDB**, Protein Data Bank (<http://www.rcsb.org/pdb/>), per le proteine.
- **NDB**, Nucleid Acid Database (<http://ndbserver.rutgers.edu>), per le strutture di acidi nucleici.

Se la struttura cristallografica non è nota si possono usare programmi di modellistica molecolare, come **MOLDEN**, per costruire geometricamente la struttura della molecola. Il formato del file di coordinate atomiche è importante poichè GROMACS è in grado di riconoscere solo i formati: GROMOS87, PDB e GROMOS96. Se si dovesse disporre di coordinate in formato diverso da quelli elencati si può procedere alla conversione usando programmi come, per esempio, OpenBabel (<http://openbabel.sourceforge.net>).

Una volta ottenuto il file di coordinate occorre verificare se il numero di atomi sia lo stesso di quello definito nella topologia (nelle strutture cristallografiche, potrebbero essere assenti gli idrogeni) e che l'ordine in cui gli atomi sono numerati sia lo stesso di quello nella topologia. Qualora non fossero nello stesso ordine occorre riordinare gli atomi nel file di coordinate nello stesso modo della topologia. All'uopo può essere utile usare un programma di grafica molecolare per identificare rapidamente gli atomi. Per trasformare le coordinate da formato GROMACS a PDB e viceversa si usa il programma:

```
editconf -f confout.gro -o confout.pdb
```

Di seguito viene riportato come esempio le coordinate della molecola del 1,2-dicloroetano nel formato PDB (nome file: coordinate.pdb):

```
HEADER      1,2-dicloroetano
REMARK      Esempio di coordinate in formato PDB
CRYST1     100.000  100.000  100.000  90.00  90.00  90.00 P 1      1
MODEL
HETATM     1  C      2      1      49.282  50.000  50.238  1.00  0.00
HETATM     2  CL     2      1      49.249  50.000  52.034  1.00  0.00
HETATM     4  H      2      1      48.756  49.120  49.910  1.00  0.00
HETATM     5  H      2      1      48.756  50.880  49.910  1.00  0.00
HETATM     3  C      2      1      50.718  50.000  49.762  1.00  0.00
HETATM     6  CL     2      1      50.751  50.000  47.966  1.00  0.00
HETATM     7  H      2      1      51.244  50.880  50.090  1.00  0.00
HETATM     8  H      2      1      51.244  49.120  50.090  1.00  0.00
TER
ENDMDL
```

e GROMOS87 (nome file: coordinate.gro):

```

Titolo: 1,2-dicloroetano
      8
      12      C      1      4.928      5.000      5.024
      12      CL     2      4.925      5.000      5.203
      12      C      3      5.072      5.000      4.976
      12      H      4      4.876      4.912      4.991
      12      H      5      4.876      5.088      4.991
      12      CL     6      5.075      5.000      4.797
      12      H      7      5.124      5.088      5.009
      12      H      8      5.124      4.912      5.009
10.00000 10.00000 10.00000

```

La riga contrassegnata con *CRYST1* nel file pdb, così l'ultima nel file GROMOS, contengono le dimensioni della scatola che contiene la molecola (posta nel suo centro), come vedremo queste informazioni saranno necessarie per le simulazioni in condizioni periodiche.

8.0.7 Creazione della topologia per molecole semplici

La topologia di una molecola è la descrizione della sua struttura in termini di legami, angoli di legami, angoli di torsione e angoli impropri. Queste informazioni devono essere specificate in modo opportuno in un file. Il programma Gromacs ha delle librerie di topologie precostruite che comprendono amino acidi, acidi nucleici e svariate molecole organiche.

Nel caso la molecola da simulare non sia presente in queste librerie è necessario procedere alla costruzione della sua topologia.

Innanzitutto è necessario valutare se i campi di forza disponibili in Gromacs (GROMACS, GROMOS96, OPLS) contengano i parametri necessari per descrivere le interazioni presenti nella molecola da studiare, o se esistono lavori in letteratura sulla stessa molecola, o molecole simili, da cui ottenere i parametri del campo di forza. In caso contrario occorre procedere alla stima iniziale di questi, come è stato descritto nel Capitolo II. In particolare, occorre calcolare le cariche atomiche parziali e assegnare gli opportuni parametri di LJ.

Con questi dati si può procedere alla scrittura del file contenente le informazioni sulla topologia molecolare (nome file: topologia.top). Di seguito viene riportato un esempio relativo alla molecola di 1,2-dicloroetano, di cui abbiamo calcolato le cariche parziali atomiche nel capitolo II.

Iniziamo con lo scrivere delle linee di commento (iniziano con ;) in cui descriviamo la molecola con commenti utili a comprendere la struttura.

```

;           X
;           |
;      H11 - C1 - H12
;           |
;      H21 - C2 - H22

```

```

;           |
;           Y
;
;
;   X,Y : Alogen atoms

```

Inseriamo quindi un comando che richiama il file di libreria che contiene i parametri di LJ per i tipi di atomi che abbiamo nella nostro modello. In questo caso il campo di forza è quello di Gromacs.*

```
#include "ffgmx.itp"
```

Iniziamo quindi con la descrizione del tipo di molecola ([**moleculetype**])

```

[ moleculetype ]
;name nrexcl
DCLE      3

```

In corrispondenza del terzo rigo abbiamo assegnato il nome della molecola. Il numero 3 che vengono esclusi dalle interazioni di non legame intermolecolari atomi che distano non più di tre legami.

Il passo successivo è quello di definire la lista degli atomi della molecola (sezione [**atoms**]). L'ordine come vedremo è importante per la definizione dei gruppi di carica.

```

[ atoms ]
; nr  type resnr resid  atom  cgnr charge  mass
  1   C    1 DCE   C1     1    0.03 12.011
  2  Cl    1 DCE   CL     1   -0.23 35.453
  3   H    1 DCE  H11    1    0.10  1.008
  4   H    1 DCE  H12    1    0.10  1.008
  5   C    1 DCE   C2     2    0.03 12.011
  6  Cl    1 DCE   CL     2   -0.23 35.453
  7   H    1 DCE  H21    2    0.10  1.008
  8   H    1 DCE  H22    2    0.10  1.008

```

Nella prima colonna viene riportato il numero di sequenza dell'atomo, nella seconda il tipo atomico così come risulta nel file di libreria *ffgmxaa.atp*. La terza colonna indica il numero del residuo. Gromacs è un programma di DM, orientato allo studio di biopolimeri questo spiega il motivo di questa colonna che serve a indicare al programma la numerazione degli aminoacidi o nucleotidi che costituiscono l'intero biopolimero. Nella terza colonna viene riportato il codice che identifica il residuo, mentre nella quarta colonna viene richiesto un nome identificativo dell'atomo per renderlo facilmente distinguibile. Nella quinta colonna viene riportato il gruppo di carica, ovvero viene indicato con uno stesso

*Per questa molecola l'OPLS fornisce dei parametri più specifici (vedi Riferimento [47]) si lascia al lettore l'esercizio di modificare la topologia per questo tipo di campo di forza.

numero (e in modo consecutivo) gli atomi che hanno cariche parziali la cui somma è nulla. In questo caso, sono stati definiti due gruppi di cariche, usando le cariche parziali ottenute dal calcolo quantomeccanico sulla conformazione *anti*. Infine, nell'ultima colonna è riportata la massa dei singoli atomi.

```
[ bonds ]
;ai aj fu c0 c1
  1  2  1 0.178 3765600.0
  1  5  1 0.153 3347200.0
  1  3  1 0.100 3765600.0
  1  4  1 0.100 3765600.0
  5  6  1 0.178 3765600.0
  5  7  1 0.100 3765600.0
  5  8  1 0.100 3765600.0
```

Nella sezione **bonds** vengono elencati i legami presenti nella molecola. Nelle prime due colonne vengono riportati i numeri di sequenza degli atomi coinvolti nel legame. Nella terza colonna viene indicato il tipo di funzione di potenziale usata per descrivere il legame (1 = funzione armonica; 2 = funzione di Morse). Nelle ultime due colonne vengono indicati i parametri della funzione di potenziale. In questo caso viene usata una funzione armonica e, quindi vengono riportati il valore della distanza di equilibrio (in nanometri) e il valore della costante di forza (in kJ/mole).

```
[ pairs ]
;ai aj fu c0 c1
  2  6  1
  2  7  1
  2  8  1
  3  6  1
  3  7  1
  3  8  1
  4  6  1
  4  7  1
  4  8  1
```

Nella sezione [**pairs**] vengono elencate le coppie di atomi coinvolti nelle interazioni LJ tra atomi distanti 4 legami. Queste interazioni sono esplicitamente indicate poichè usano parametri opportunamente scalati (vedi capitolo II).

```
[ angles ]
;ai aj ak fu c0 c1
  2  1  3  1 109.0 460.2
  2  1  4  1 109.0 460.2
  2  1  5  1 109.0 460.2
  3  1  4  1 109.0 460.2
  3  1  5  1 109.0 460.2
```

```

4   1   5   1  109.0    460.2
1   5   6   1  109.0    460.2
1   5   7   1  109.0    460.2
1   5   8   1  109.0    460.2
6   5   7   1  109.0    460.2
6   5   8   1  109.0    460.2
7   5   8   1  109.0    460.2

```

Nella sezione [**angles**] vengono elencati gli angoli di valenza presenti nella molecola. Nelle prime tre colonne vengono riportati i numeri di sequenza dei tre atomi coinvolti nel formare l'angolo. Nella quarta colonna viene indicato il tipo di funzione di potenziale usata per descrivere il legame (1 = funzione armonica). Nelle ultime due colonne vengono indicati i parametri della funzione di potenziale. In questo caso viene usata una funzione armonica e, quindi vengono riportati il valore dell'angolo di equilibrio (in nanometri) e il valore della costante di forza (in kJ/mole).

```

[ dihedrals ]
;ai aj ak al fu      c0   c1  m
  6  5  1  2  1      0.0  5.9  3   ; Y - C2 - C1 - X

```

Nella sezione [**dihedrals**] vengono elencati gli angoli di torsione e gli angoli diedri impropri presenti nella molecola. Nel 1,2-dicloroetano non ci sono angoli di diedri impropri ma solo un angolo di torsione. Nelle prime quattro colonne vengono riportati i numeri di sequenza dei quattro atomi coinvolti nel formare il diedro. Nella quinta colonna viene indicato il tipo di funzione di potenziale usata per descrivere il legame (1 = angolo di torsione). Nelle ultime tre colonne vengono indicati i parametri della funzione di potenziale. In questo caso il primo valore indica la fase (ϕ_0), il secondo la costante di forza e il terzo la molteplicità (vedi Capitolo II). Nel caso dei diedri impropri il valore della quinta colonna è posto pari a due e vengono riportati due parametri nelle ultime due colonne, il primo indica l'angolo di riferimento e il secondo il valore della costante di forza.

```

[ system ]
1,2-dicloroetano

```

```

[ molecules ]
DCLE          1

```

La sezione [**system**] definisce il nome del sistema che si sta simulando. Infine, la sezione [**molecules**] fornisce la composizione del sistema, ovvero il numero di molecole di 1,2-dicloroetano che vengono simulate (in questo caso 1).

8.0.8 Creazione della topologia per biopolimeri

Contrariamente a quanto ci si può aspettare, la creazione di una topologia molecolare per una proteina o un acido nucleico è relativamente semplice. Infatti,

Gromacs, come la maggior parte dei programmi per simulazioni di DM, è stato concepito per essere usato principalmente per la simulazione di questi sistemi complessi. Pertanto, il pacchetto dispone di strumenti molto potenti per la creazione automatica della topologia di questi polimeri. La topologia viene creata assemblando insieme i vari monomeri (aminoacidi o basi puriniche e pirimidiniche) definiti come *building block* in opportuni files di libreria. Per preparare la topologia di una proteina si può quindi procedere in questo modo. Si cerca nel Protein Data Bank, la struttura della proteina da studiare. Si leggono attentamente le informazioni relative alla struttura nel *HEADER* del file pdb, per accertarsi della qualità e delle caratteristiche della struttura.[†]

A questo punto, si esegue il comando :

```
pdb2gmx -f pdbfile
```

Quindi, il programma chiede ad un certo punto quale campo di forza utilizzare:

```
Select the Force Field:
0: Gromacs Forcefield (see manual)
1: Gromacs Forcefield with all hydrogens (proteins only)
2: GROMOS96 43a1 Forcefield (official distribution)
3: GROMOS96 43b1 Vacuum Forcefield (official distribution)
4: GROMOS96 43a2 Forcefield (development) (improved alkane dihedrals)
5: OPLS-AA/L all-atom force field (2001 aminoacid dihedrals)
```

selezionato un campo di forze il programma procede alla creazione della topologia. Il programma assegna agli aminoacidi lo stato di protonazione relativo a pH=7, è comunque possibile indicare lo stato di protonazione dei vari aminoacidi o del gruppo carbossilico e aminico terminale con le seguenti opzioni:

```
-ter   Interactive termini selection, iso charged
-lys   Interactive Lysine selection, iso charged
-asp   Interactive Aspartic Acid selection, iso charged
-glu   Interactive Glutamic Acid selection, iso charged
-his   Interactive Histidine selection, iso checking
```

Per la selezione interattiva dei ponti disolfuro usare l'opzione *-ss*. Per attivare una sessione completamente interattiva usare il comando *-inter*.

In uscita il programma genera i files: *topol.top* con la topologia, *conf.gro* con le coordinate, *posres.itp* con la selezione degli atomi a cui applicare i vincoli sulla posizione (*position restraints*).

[†]Si deve considerare la risoluzione a cui questa è stata risolta, la presenza di tutti gli aminoacidi (potrebbero essere assenti alcune parti della proteina, non risolte per via del disordine molecolare, che devono essere modellate con opportuni programmi), la numerazione e così via.

8.0.9 Minimizzazione della struttura iniziale nel vuoto

Creato il file di topologia è necessario verificare che non siano stati commessi errori nella compilazione del file. Si deve inoltre verificare se la conformazione di minimo della molecola, per i parametri del campo di forza utilizzato, producano una rappresentazione geometrica della molecola simile a quella ottenuta dal calcolo quantomeccanico. Per far ciò dobbiamo preparare i files necessari alla minimizzazione energetica della molecola nel vuoto. Questi files sono:

1. File di coordinate iniziali in formato PDB o GROMOS.
2. File di topologia.
3. File con i parametri per la procedura di minimizzazione.

di questi dobbiamo ancora preparare solo il file di parametri.

File di parametri per la minimizzazione. Il file di parametri contiene una lista di comandi che servono a definire il tipo di minimizzazione e altre informazioni per la procedura di ottimizzazione della struttura molecolare. Di seguito viene riportato un esempio di file di parametri per effettuare una minimizzazione nel vuoto di una molecola. Di seguito viene riportato un esempio di file di parametri (nome file param.mdp) da usare nella minimizzazione nel vuoto di una molecola.

```
; RUN CONTROL PARAMETERS =
integrator           = steep
nsteps              = 5000

; ENERGY MINIMIZATION OPTIONS =
entol               = 1
emstep              = 0.01

; PBC SETTINGS
pbc                 = no
nstlist             = 0
rlist               = 0
rcoulomb            = 0
rvdw                = 0
```

Nella minimizzazione viene usato l'algoritmo dello *steepest descent* (*integrator=steep*), la tolleranza per la convergenza (*emtol*) è pari a 1 pN/mole mentre il numero di passi totali è pari a 5000. Non vengono usate condizioni periodiche al contorno (*pbc = no*) e cutoffs.

A questo punto abbiamo tutti i files necessari per procedere alla ottimizzazione della molecola. Per far questo useremo il programma *grompp* che combina insieme i vari files che abbiamo generato in un unico file binario. La sintassi del comando[‡] è la seguente:

[‡]GROMACS usa delle estensioni standard per i nomi dei file che non possono essere omesse. Abbiamo già visto vari esempi di estensioni:

```
grompp -f param.mdp -c coordinate.gro -p topologia.top -o topol.tpb
```

Quindi usando il comando *mdrun* possiamo lanciare la minimizzazione della molecola. La sintassi in questo caso è la seguente:

```
mdrun -v
```

Abbiamo usato solo l'opzione *-v* (verbose) che forza il programma a scrivere a video tutte le informazioni sui calcoli che sta eseguendo. Abbiamo ommesso la opzione *-s* che serve a specificare il file di input, poichè questo è stato chiamato con un nome predefinito (topol.tpb).

Durante l'esecuzione del programma si vedrà sullo schermo un susseguirsi di linee di questo tipo:

```
Step= 6751, Dmax= 1.2e-06 nm, Epot= -5.93153e+00 Fmax= 6.13962e+00, atom= 5
Step= 6752, Dmax= 1.5e-06 nm, Epot= -5.93155e+00 Fmax= 5.71714e+00, atom= 5
```

dove vengono indicati nell'ordine, il passo di minimizzazione, lo spostamento massimo delle coordinate, il valore dell'energia potenziale totale, la forza massima e qual'è il numero di sequenza dell'atomo su cui è stata valutata questo valore massimo della forza.

Mentre alla fine dell'esecuzione sul video rimarrà una scritta del tipo

```
Step= 6753, Dmax= 1.8e-06 nm, Epot= -5.93154e+00 Fmax= 1.30439e+01, atom= 5
Stepsize too small (8.9459e-07 nm)Converged to machine precision,
but not to the requested precision (1)
```

```
writing lowest energy coordinates.
```

```
Steepest Descents did not converge in 6754 steps
Potential Energy = -5.93155e+00
Maximum force: 5.71714e+00
```

La procedura di minimizzazione può essere resa più accurata eseguendo lo stesso calcolo in doppia precisione e diminuendo *emtol*. Se si vuole raggiungere il minimo, occorre dopo aver abbassato il più possibile la tolleranza con il metodo dello *steepest descent*, passare al metodo dei gradienti coniugati. Questa

- File della topologia : .top e in binario : .tpb.
- File coordinate: PDB=.pdb; gromacs (GROMOS87): .gro; GROMOS96: .g96.
- File di parametri: .mdp.
- File con le traiettorie: binario=.trr; binario compresso=.xtc.
- File con le energie: .edr.

Anche la radice di certi nomi dei files è predefinita. Infatti se viene specificata solo l'opzione di linea di comando senza essere seguita dal nome del file, il programma cerca nel direttorio corrente il nome predefinito del file. Per esempio nel caso del comando *grompp*, la specificazione del nome topol.tpb dopo l'opzione *-o* è ridondante poichè *topol.tpb* è un nome file predefinito per quella opzione. Le informazioni sui nomi predefiniti cosiccome sui comandi di linea, possono essere ottenute in forma sintetica da ogni comando di gromacs con l'opzione *-h*

minimizzazione completa è necessaria solo quando si vuole calcolare l'Hessiano del sistema per valutare i moti di vibrazione normali della molecola. In generale, l'ottimizzazione che si ottiene con lo *steepest descent* è sufficiente per il rilassamento del sistema molecolare prima della partenza di simulazioni di DM.

Il programma *mdrun* genera in uscita i seguenti files:

- *confout.gro*: contiene le coordinate finali minimizzate.
- *ener.edr*: contiene le energie durante la minimizzazione.
- *md.log*: contiene le informazioni relative al procedura di minimizzazione.

Per visualizzare la struttura finale si può procedere in due modi:

- Effettuare il calcolo usando l'opzione di *mdrun -c confout.pdb*. In questo modo viene indicato al programma di produrre in uscita un file di coordinate finali di tipo PDB.
- Trasformare le coordinate da formato GROMACS a PDB usando *editconf*.

Il file in formato PDB può essere letto da tutti i programmi di visualizzazione molecolare. Visualizzando il file di coordinate si può rapidamente vedere se la molecola ha subito deformazioni dovute a errori nella definizione di parametri della topologia. Per confrontare la nuova struttura con quella di riferimento si può usare il programma:

```
g_confrms -f1 iniziale.gro -f2 confout.gro
```

Il programma chiede quali gruppi di atomi usare per la sovrapposizione con una procedura di minimizzazione dei minimi quadrati, della seconda struttura (*confout.gro*) su quella di riferimento (*iniziale.gro*). Scegliendo 0 in entrambe le richieste si otterrà il seguente messaggio in uscita:

```
Root mean square deviation after lsq fit = 0.110199
```

ovvero il valore in nanometri della deviazione quadratica media delle coordinate nelle due strutture. Nel file *fit.pdb* sono contenute le due strutture sovrapposte in formato PDB. Caricando il file nel programma RASMOL e selezionando dal menu COLOR la voce STRUCTURE, le due conformazioni verranno colorate in modo diverso (blu e rosso), permettendo di valutare qualitativamente quali sono gli atomi che hanno subito le variazioni maggiori.

8.0.10 Simulazione della fase liquida

Creazione di un box di molecole dello stesso tipo.

Se la topologia non presenta problemi, ovvero la geometria della molecola si mantiene nei limiti della accuratezza prevista, si può procedere alla costruzione di una scatola di molecole per la simulazione della fase liquida del composto. Per realizzare la scatola si deve conoscere la densità del liquido ad una certa

temperatura. Questa ci permette di calcolare il volume medio occupato da una singola molecola. Se d è la densità in kg m^{-3} , allora il volume in nanometri occupato da una molecola è uguale a:

$$V_m = \frac{1000M}{dN_a} 10^{21} \quad (8.1)$$

dove N_a è il numero di Avogadro e M il peso molecolare del composto. Calcolando la radice cubica del volume (V_M) si ottiene il lato (l) del volumetto che contiene la molecola. Nel caso del 1,2-dicloroetano abbiamo che $M = 98.96g$ e $d = 1246.3 \text{ kg m}^{-3}$, da cui si ottiene $l = \sqrt[3]{(0.132)} = 0.509 \text{ nm}$. Per costruire questo volumetto intorno alla molecola si utilizza il programma:

```
editconf -f confout.gro -c -box 0.509 0.509 0.509 -o OneMolBox.gro
```

il comando di linea `-box 0.509 0.509 0.509` serve a definire i lati della scatola cubica, mentre `-c` centra le coordinate della molecola rispetto alla scatola. Usando il nuovo file di coordinate possiamo generare una scatola di molecole avente la densità del liquido con il programma `genconf`. La sintassi del programma è la seguente:

```
genconf -f OneMolBox -nbox 5 5 5
```

Il programma replica il volumetto e la molecola in essa contenuta su un reticolo di $5 \times 5 \times 5$ nodi. Usando l'opzione `-rot -maxrot 20` la molecola viene ruotata di massimo 20 gradi (`-maxrot 20`). Nella Figura 8.2 vengono riportate le conformazioni delle molecole nella scatola ottenute con e senza l'opzione `-rot`. Una

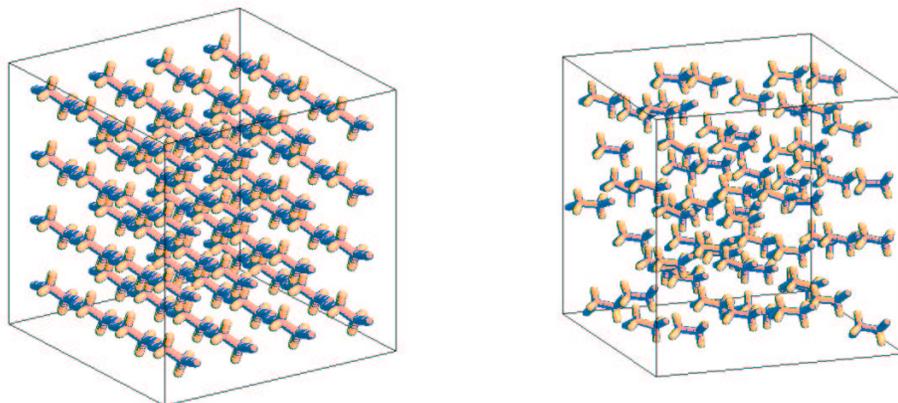


Figura 8.2: Conformazioni ottenute da `genconf` senza opzione `-rot` (a sinistra) e con (a destra).

volta generate le coordinate di partenza, prima di procedere alla vera e propria

simulazione è necessario minimizzare nuovamente l'intera scatola di simulazione per eliminare i falsi contatti. Per far questo occorre prima di tutto modificare l'ultima riga della topologia ponendo il valore del numero delle molecole pari a 125 (5x5x5). Si deve quindi modificare il file di parametri per tener conto delle condizioni periodiche al contorno e dell'uso del cutoff nel modo che segue:

```
; RUN CONTROL PARAMETERS =
integrator           = steep
nsteps              = 5000

; ENERGY MINIMIZATION OPTIONS =
; Force tolerance and initial step-size =
emtol               = 100
emstep              = 0.01

; NEIGHBORSEARCHING PARAMETERS =
; nblast update frequency =
nblast              = 5
; ns algorithm (simple or grid) =
ns-type             = grid
; Periodic boundary conditions: xyz or no =
pbc                 = xyz
; nblast cut-off
rlist               = 0.8

; OPTIONS FOR ELECTROSTATICS AND VDW =
; Method for doing electrostatics =
coulombtype         = Cut-off
rcoulomb            = 1.0
; Dielectric constant (DC) for cut-off or DC of reaction field =
epsilon-r           = 1
; Method for doing Van der Waals =
vdw-type            = Cut-off
; cut-off lengths
rvdw                = 0.8
```

Le interazioni di non legame vengono calcolate con il metodo delle liste di Verlet, usando il metodo del doppio cutoff. Ricordare che nel caso di simulazioni in condizioni periodiche il cutoff deve essere sempre minore della metà del lato della scatola. Pertanto si può procedere alla minimizzazione come descritto in precedenza. L'analisi visiva della configurazione finale, nonché i valori di energia potenziale totale consentiranno di valutare l'assenza di contatti sfavorevoli.

La solvatazione di un soluto.

Se si vuole simulare un soluto in una scatola di solvente allora si può usare il programma *genbox* per preparare la scatola di simulazione. Per far ciò si stabi-

lisce il tipo di scatola (rettangolare, ottaedrico etc..) e le sue dimensioni, quindi si centra la molecola di soluto in questa nuova scatola, usando il programma *editconf*:

```
editconf -f oneMolBox.gro -o Box.gro -c -box lx ly lz
```

Il programma *editconf* può generare la scatola in vari modi, quello che finora abbiamo usato consiste nel fissare le dimensioni di un box rettangolare (o cubico se le dimensioni sono uguali). Si può anche definire la distanza minima del soluto dalle pareti con l'opzione *-d distanza (in nm)*. Inoltre se si vuole definire un box più sferico per ridurre le dimensioni del sistema simulato, si pu' usare l'opzione *-bt tric (o cubic o dodecahedron o octahedron)*. Si rimanda la manuale di GROMACS per ulteriori dettagli.

Il nuovo file di coordinate viene quindi usato per creare la scatola del solvente:

```
genbox -cp Box.gro -cs $GMXLIB/spc.gro -p topol.top
```

Le coordinate delle molecole di acqua che sono state generate vengono scritte nel file *out.gro* insieme a quelle del soluto, e in fondo al file della topologia viene aggiunta un nuovo rigo:

```
SOL                209
```

indicante il numero di molecole di solvente presenti nel sistema.

Se non sono già presenti (come nel caso del file di topologia creato a mano), nel file di topologia vanno inoltre aggiunte le seguenti linee prima di *[system]*:

```
; Include water topology
#ifdef FLEX_SPC
#include "flexspc.itp"
#else
#include "spc.itp"
#endif
```

Come si può leggere dai commenti, queste linee servono ad inserire la topologia dell'acqua.

Bibliografia

- [1] Jr. A. D. MacKerell and al., *An all-atom empirical energy function for the simulation of nucleic acids*, J. Am. Chem. Soc. **117** (1995), 11946–11975.
- [2] B. J. Alder and T. E. Wainwright, J. Chem. Phys. **27** (1957), 1208.
- [3] M. P. Allen and D. J. Tildesly, *Computer simulation of liquids*, Oxford University Press, Oxford, 1989.
- [4] H. C. Andersen, J. Chem. Phys. **72** (1980), 2384.
- [5] J. Åqvist, W. F. van Gunsteren, M. Leijonmarck, and O. Tapia, J.Mol.Biol. **183** (1985), 461.
- [6] P. Ballestrero, P. Baglietto, and C. Ruggero, J. Comp. Chem. **17** (1996), 469.
- [7] D. J. Beeman, J. compt. Phys. **20** (1976), 130.
- [8] H. J. C. Berendsen, Computer Simulation of Biomolecular Systems (W. F. van Gunsteren and P. K. Weiner, eds.), vol. 2, ESCOM, Leiden, 1993, p. 161.
- [9] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. Di Nola, and J. R. Haak, *Molecular dynamics with coupling to an external bath*, J. Chem. Phys. **81** (1984), 3684.
- [10] M. Berkowitz and J. A. McCammon, Chem. Phys. Lett. **90** (1982), 215.
- [11] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, *Charmm: A program for macromolecular energy, minimization, and dynamics calculations*, J. Comp. Chem. **4** (1983), 187–217.
- [12] C. L. Brooks, S. Yong, and D. J. Tobias, The Int. J. of Sup. Appl. **5** (1991), 98.
- [13] D. Brown and J. H. R. Clarke, J. Chem. Phys. **84** (1986), 2858.
- [14] A. T. Brunger, Annu. Rev. Chem. **42** (1991), 197.

-
- [15] A. T. Brunger and J. Kuriyan M. Karplus, *Science* **235** (1987), 458.
- [16] R. A. Buckingham, *Proc. Roy. Soc. Ser. A* **168** (1938), 264.
- [17] W. T. Clark, R. Hanxleden, J. A. McCammon, and L. R. Scott, *Technical report crpctr 93356-s*, Tech. report, Center for Research on Parallel Computation, Houston, November 1993.
- [18] W. D. Cornell and al., *A second generation force field for the simulation of proteins, nucleic acids, and organic molecules*, *J. Am. Chem. Soc.* **117** (1995), 5179–5197.
- [19] P. Dauber-Osguthorpe, V. A. Roberts, D. J. Osguthorpe, J. Wolff, M. Genest, and A. T. Hagler, *Proteins* **4** (1988), 31.
- [20] A. Di Nola and E. Giglio, *Acta Crystallogr.* **A 26** (1970), 144.
- [21] E. Egberts, *Thesis*, Master's thesis, University of Groningen, 1988.
- [22] D. J. Evans, W. G. Hoover, B. H. Failor, B. Moran, and A. J. C. Ladd, *Nonequilibrium molecular dynamics via Gauss's principle of least constraint*, *Phys. Rev. A* **28** (1983), 1016.
- [23] P. Ewald, **64** (1921), 253.
- [24] D. Fincham, *Mol. Simul.* **1** (1987), 1.
- [25] M. Fixman, *J. Chem. Phys.* **69** (1978), 1527.
- [26] C. W. Gear, *Numerical initial value problems in ordinary differential equations*, Englewood Cliffs, NJ, USA.
- [27] E. Giglio, A. M. Liquori, and L. Mazzarella, *Nuovo Cim.* **1** (1969), 135.
- [28] A. T. Hagler, A. T. Huler, and S. Lifson, *J. Am. Chem. Soc.* **96** (1989), 5319.
- [29] J. M. Haile, *Molecular dynamics simulation: elementary methods*, John Wiley and Sons, Inc., New York, 1992.
- [30] J. M. Haile and S. Gupta, *J. Chem. Phys.* **79** (1983), 3067.
- [31] J. P. Hansen and I. R. McDonald, *Theory of simple liquids*, 2nd ed., Academic Press, London, 1986.
- [32] S. C. Harvey, M. Prabhakaran, and B. Mao J. A. McCammon, *Science* **223** (1984), 1189.
- [33] T. S. Harvey and W. F. van Gunsteren, *Techniques in Protein Chemistry* (R.H. Angeletti, ed.), vol. IV, Academic Press, New York, 1993, p. 615.
- [34] H. Heller, H. Grubmüller, and K. Shulten, *Mol. Simul.* **5** (1990), 133.
-

-
- [35] J. Hermans (ed.), *Molecular dynamics and protein structure*, Polycrystal Bookservice P.O. Box 27, Western Springs, Ill., USA, 1985.
- [36] J. Hermans, H. J. C. Berendsen, W. F. van Gunsteren, and J.P.M. Postma, *Biopolymers* **1984**, 1513.
- [37] B. Hess, H. Bekker, J.G.E.M. Fraaije, and H.J.C. Berendsen, *A linear constraint solver for molecular simulations*, *J.Comp.Chem.* **18** (1997), 1463–1472.
- [38] D. M. Heyes, *J. Chem. Phys.* **74** (1981), 1924.
- [39] ———, *Chem. Phys.* **82** (1981), 285.
- [40] P. A. J. Hilberts and K. Esselink, *Computer simulations in chemical physics*, M.P. Allen and D.J Tildesley (eds.), 1993.
- [41] Z. I. Hodes, G. Nemethy, and H. A. Scheraga, *Biopolymers* **18** (1979), 1565.
- [42] W. G. Hoover, *Canonical dynamics: Equilibrium phase-space distributions*, *Phys. Rev. A* **31** (1985), 1695.
- [43] W. G. Hoover, D. J. Evans, R. B. Hickman, A. J. C. Ladd, W. T. Ashurst, and B. Moran, *Phys. Rev. A* **22** (1980), 1960.
- [44] J. N. Hsraelachvili, *Intermolecular and surphace forces*, Acad. Press, London, New York, 1985.
- [45] J. F. Janak and P. C. Pattnaik, *J.Comp.Chem.* **13** (1992), 533.
- [46] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, and R. W. Impey and M. L. Klein, *J. Chem. Phys.* **79** (1983), 962.
- [47] W. L. Jorgensen, N. A. McDonald, M. Selmi, and P.R. Raben, *Importance of polarization for dipola solutes in low dielectric media: 1,2-dichloroethane and water in cyclohesane*, *J. Am. Chem. Soc.* **117** (1995), 11809–11810.
- [48] W. L. Jorgensen and J. Tirado-Rives, *The oplis potential functions for proteins. energy minimizations for crystals of cyclic peptides and crambin.*, *J. Am. Chem. Soc.* **110** (1988), 1657–1666.
- [49] R. Kaptein, E. R. P. Zuiderweg, R. M. Scheek, and R. Boelens W. F. van Gunsteren, *J.Mol.Biol.* **182** (1985), 179.
- [50] D. H. Kitson, F. Avbelj, J. Moult, D. T. Nguyen, J. E. Mertz, D. Hadzi, and A. T. Hagler, *Proc. Natl. Acad. Sci. USA* **90** (1993), 8920.
- [51] J. E. H. Koehler, W. Saenger, and W. F. van Gunsteren, *J. Mol. Biol.* **203** (1988), 241.
- [52] M. Levitt, *J. Mol. Biol.* **168** (1983), 595.
-

-
- [53] S. Lifson, C. E. Felder, and A. Shanzer, *J. Am. Chem. Soc.* **105** (1983), 3866.
- [54] S. Lifson, A. T. Hagler, and P. Dauber, *J. Am. Chem. Soc.* **101** (1979), 5111.
- [55] S. L. Lin, J. Mellor-Crummey, B. M. Pettit, and G. N. Phillips, *J. Comp. Chem.* **13** (1992), 1022.
- [56] E. Lippert, *Hydrogen bond*, vol. 1, P. Schuster and G. Zudel and C. Sandorfy, North-Holland, Amsterdam, New York, 1976.
- [57] F. London, *Z. Physik* **63** (1930), 245.
- [58] H. Margenau and N. R. Kestner, *Theory of intermolecular forces*, New York, 1969.
- [59] J. A. McCammon, B. R. Gelin, and M. Karplus, *Nature* **267** (1977), 585.
- [60] D. A. McQuarrie, *Statistical mechanics*, Harper & Row, New York, 1976.
- [61] F. A. Mommany, R. F. McGuire, A. W. Burgess, and H. A. Scheraga, *J. Phys. Chem.* **79** (1975), 2361.
- [62] K. M. Nelson, C. F. Cornwell, and L. T. Wille, *Comput. Mat. Sci.* **2** (1984), 525.
- [63] S. Nosé, *A molecular dynamics method for simulations in the canonical ensemble*, *Mol. Phys.* **52** (1984), 255.
- [64] H. Ohtaki and T. Radnai, *Chem. Rev.* **93** (1993), 1157.
- [65] L. Perera, U. Essmann, and M. L. Berkowitz, *J. Chem. Phys.* **102** (1995), 450.
- [66] M. R. S. Pinches, D. J. Tildesley, and W. Smith, *Mol. Simul.* **6** (1991), 51.
- [67] S. P. Plimpton, *J. Comput. Phys.* **117** (1995), 1.
- [68] S. P. Plimpton and B. Hendrickson, *J. Comp. Chem.* **17** (1996), 326.
- [69] J. P. M. Postma, *A molecular dynamics study of water*, Ph.D. thesis, University of Groningen. the Netherlands, 1985.
- [70] A. Rahman, *Phys. Rev. A* **405** (1964), 136.
- [71] A. Rahman and F. H. Stillinger, *J. Chem. Phys.* **55** (1971), 3336.
- [72] S. N. Rao, U. C. Singh, and P. A. Bash P. A. Kollman, *Nature* **328** (1987), 551.
- [73] D. C. Rapaport, *Comput. Phys. Commun.* **62** (1991), 217.
-

-
- [74] J. P. Ryckaert and A. Bellemans, *Chem. Phys. Lett.* **30** (1975), 123.
- [75] J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, *Numerical integration of the cartesian equations of motion of a system with constraints; Molecular dynamics of n-alkanes*, *J. Comp. Phys* **23** (1977), 327.
- [76] Y.-Y. Shi, L. Wang, and W. F. van Gunsteren, *Mol.Simul.* **1** (1988), 369.
- [77] U. C. Singh, S. C. Weiner, and P. Kollman, *Procl. Natl. Acad. Sci. USA* **82** (1985), 755.
- [78] W. Smith, *Comp. Phys. Comm.* **62** (1991), 229.
- [79] W. Soppe, *Thesis*, Master's thesis, University of Groningen, 1989.
- [80] W. B. Street, D. J. Tildesley, and G. Saville, *Mol. Phys.* **35** (1978), 639.
- [81] A. E. Torda, R. M. Brunne, T. Huber, H. Kessler, and W. F. van Gunsteren, *J. Biomol. NMR* **3** (1993), 55.
- [82] A. E. Torda, R. M. Scheek, and W. F. van Gunsteren, *J. Mol. Biol.* **214** (1990), 223.
- [83] M. Tuckerman, B. J. Berne, and G. J. Martyna, *J.Chem.Phys.* **97** (1992), 1990.
- [84] W. F. van Gunsteren and H. J. C. Berendsen, *Gromos-87 manual*, Biomos BV, Nijenborgh 4, 9747 AG Groningen, The Netherlands.
- [85] W. F. van Gunsteren and H. J. C. Berendsen, *J. Mol. Biol.* **176** (1984), 559.
- [86] ———, *Molecular dynamics simulations: techniques and applications to protein*, *Molecular Dynamics and Protein Structure* (J. Hermans, ed.), P.O. Box 27, Western Springs, Ill., USA: Polycrystal Bookservice, 1985.
- [87] ———, *Angew. Chem. Int. Engl.* **29** (1990), 992.
- [88] W. F. van Gunsteren, H. J. C. Berendsen, and J. A. C. Rullmann, *Mol. Phys.* **44** (1981), 69.
- [89] W. F. van Gunsteren, X. Daura, and A. E. Mark, *GROMOS force field*, *Encyclopedia of Computational Chemistry* **2** (1998), 1211–1216.
- [90] W.F. van Gunsteren, *Methods for calculation of free energies and binding constants: successes and problems*, *Computation of free energies for Biomolecular Systems* (W. F. van Gunsteren and P. K. Weiner, eds.), Escom Science Publishers, Leiden, 1989, pp. 27 – 59.
- [91] L. Verlet, *Phis. Rev.* **159** (1967), 98.
- [92] ———, *Phis. Rev.* **165** (1968), 201.
-

- [93] M. Watanabe and M. Karplus, *J.Chem.Phys.* **99** (1993), 8063.
 - [94] ———, *J.Chem.Phys.* **99** (1995), 5681.
 - [95] S. J. Weiner, P. A. Kollman, D. T. Nguyen, and D. A. Case, *An all atom force field for simulations of proteins and nucleic acids*, *J. Comp. Chem.* **7** (1986), 230–252.
 - [96] A. Windemuth and K. Schulten, *Mol. Simul.* **5** (1991), 353.
 - [97] J Wittenburg, *Dynamics of systems of rigid bodies*, Teubner, Stuttgart, 1977.
 - [98] L. V. Woodcock, *Chem. Phys. Lett.* **10** (1971), 257.
-