

Chapter 5

La Dinamica Molecolare nello Studio dei Biosistemi

Un approccio computazionale molto usato sia nella scienza dei materiali che nello studio di biomolecole è la *dinamica molecolare* (MD). Con questo metodo viene seguita passo per passo e simulata in dettaglio la dinamica del sistema. Di conseguenza, le simulazioni MD possono dare informazioni molto dettagliate su fenomeni quali cambiamenti conformazionali di proteine e acidi nucleici, assenti o difficili da estrarre con il metodo Monte Carlo. Il prezzo da pagare può essere elevato in termini di tempo di calcolo: simulazioni MD di sistemi complessi possono essere assai dispendiose. Tuttavia l'incremento di prestazioni dei computer da una parte, il raffinamento del metodo e degli algoritmi dall'altro, hanno reso possibili calcoli impensabili fino a pochi anni fa.

Con la MD possono essere studiati in particolare:

- stabilità di piccole proteine;
- cambi conformazionali
- protein folding (di piccole proteine!)
- riconoscimento molecolare: interazioni tra proteine, DNA, membrane, complessi
- trasporto ionico in sistemi biologici

Inoltre la tecnica fornisce anche un valido aiuto nel design di nuovi farmaci e nella determinazione di strutture di molecole complesse (in combinazione con tecniche sperimentali quali diffrazione di raggi X e NMR).

Il metodo fu introdotto inizialmente da Alder e Wainwright negli anni '50 per lo studio di sistemi composti da sfere rigide e portò rapidamente ottimi risultati anche nello studio di liquidi semplici. La prima simulazione di dinamica molecolare di un sistema realistico: l'acqua in fase liquida, risale al 1974 (Rahman e Stillinger). La prima simulazione di proteine (bovine pancreatic trypsin inhibitor, BPTI) fu fatta nel 1977 da McCammon e altri. Le simulazioni MD di proteine in acqua oggi raggiungono scale temporali fino ai nanosecondi; tuttavia, in assenza di solvente, sono riportati anche casi di simulazioni fino a millisecondi.

5.1 Le basi della dinamica molecolare

La MD si basa sulla seconda legge di Newton:

$$\mathbf{F} = m\mathbf{a} \quad (5.1)$$

dove \mathbf{F} è la forza esercitata sulla particella di massa m ed \mathbf{a} è la sua accelerazione. La conoscenza della forza su ogni atomo di un sistema permette di determinare le rispettive accelerazioni. Integrando le equazioni del moto si può ricavare la *traiettoria*, ovvero posizioni $\mathbf{r}(t)$ in funzione del tempo e velocità $\mathbf{r}(t)$. Il metodo è deterministico! la traiettoria è (in linea di principio) esattamente determinata dalle condizioni iniziali.

Per un sistema di molti atomi, il moto di ogni atomo del sistema è determinato dall'equazione

$$\mathbf{F}_i = m_i\mathbf{a}_i, \quad i = 1, 2, \dots, N. \quad (5.2)$$

Assunzione fondamentale è che la forza possa essere espressa come il gradiente di un'energia potenziale U :

$$\mathbf{F}_i = -\nabla_i U(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) = -\left(\frac{\partial U}{\partial x_i}, \frac{\partial U}{\partial y_i}, \frac{\partial U}{\partial z_i}\right) \quad (5.3)$$

che combinata con la precedente dà

$$-\nabla_i U = m_i\mathbf{a}_i \Rightarrow m_i \frac{d^2\mathbf{r}_i}{dt^2} = -\nabla_i U. \quad (5.4)$$

Per calcolare le traiettorie degli atomi si ha bisogno, oltre che della funzione energia potenziale U , delle posizioni iniziali (al tempo $t = 0$ diciamo) e delle velocità iniziali. Le posizioni iniziali vengono ricavate da dati sperimentali, provenienti per esempio da cristallografia a raggi X o da dati di struttura in soluzione ottenuta tramite spettroscopia NMR. Le velocità iniziali $\mathbf{v}_i(t = 0)$ sono spesso prese a caso da una distribuzione Gaussiana adatta alla temperatura di partenza:

$$p(\mathbf{v}) = \left(\frac{m_i}{2\pi k_b T}\right)^{1/2} e^{-m_i\mathbf{v}^2/2k_b T} \quad (5.5)$$

facendo in modo che la quantità di moto complessiva del sistema sia nulla: $\mathbf{P} = \sum_i m_i\mathbf{v}_i = 0$.

In assenza di accorgimenti, l'energia totale (cinetica più potenziale) in una simulazione MD è conservata. Si parla allora di dinamica *NVE*, per indicare che numero di particelle, volume ed energia sono conservate. Un'ovvia conseguenza è che la dinamica *NVE* "esplora" l'insieme microcanonico: per sistemi ergodici, la media temporale calcolata sulla traiettoria della MD equivale alla media statistica sull'insieme microcanonico. La temperatura del sistema varia durante la simulazione ed è determinata dall'energia cinetica istantanea:

$$\frac{3}{2}Nk_b T(t) = \sum_i \frac{1}{2}m_i\mathbf{v}_i^2(t). \quad (5.6)$$

Altri tipi di MD sono possibili: per esempio, l'insieme canonico può essere esplorato con la dinamica *NVT*, in cui si aggiunge un opportuno "termostato" che

tiene il sistema a temperatura data. E' anche possibile realizzare una dinamica *NPT*, ovvero in cui il sistema è tenuto a temperatura e pressione (anziché volume) dati. Queste tecniche, peraltro molto utili, sono tuttavia al di fuori degli scopi di questo corso.

5.1.1 L'approssimazione adiabatica

E' lecito chiedersi quanto sia vero che il moto degli atomi è determinato solo da una energia potenziale (o semplicemente, "potenziale") U che dipende solo dalle posizioni dei nuclei: che fine hanno fatto gli elettroni? In realtà gli elettroni sono ben presenti, ma un solido argomento teorico ci garantisce che possiamo incorporarne gli effetti in un potenziale interatomico effettivo. Gli elettroni sono molto più leggeri dei nuclei (almeno 1800 volte nel caso più sfavorevole) e di conseguenza il loro moto è assai più veloce di quello dei nuclei. Possiamo assumere che in ogni istante gli elettroni restino nel loro stato fondamentale, corrispondente alla posizione istantanea dei nuclei (*approssimazione adiabatica* o di *Born-Oppenheimer*). Si tratta di una eccellente approssimazione, con un intervallo di validità molto ampio che esclude solo una ristretta (ma importante!) classe di fenomeni *non-adiabatici*. L'approssimazione di Born-Oppenheimer garantisce quindi l'esistenza di un potenziale interatomico U e ne dà anche la formulazione in termini dell'energia di un sistema di elettroni sotto il campo dei nuclei. Nella MD *da principi primi*, il potenziale U è esplicitamente calcolato a partire dalle proprietà elettroniche. I risultati sono molto accurati ma il metodo è assai più pesante numericamente della MD "classica", in cui si usano forme semplici di potenziale (come vedremo nelle prossime sezioni).

5.1.2 Integrazione delle equazioni del moto

Il potenziale U è funzione delle posizioni di tutti gli atomi del sistema; ovviamente, a causa della complessità di tale funzione le equazioni del moto devono essere risolte numericamente. E' necessario *discretizzare* le equazioni, ovvero passare ad una descrizione in cui il tempo diventa una variabile discreta. Scegliamo un intervallo di tempo (*time step*) δt , piccolo (ma non più del necessario: il più grande possibile che ci permetta di descrivere il moto che ci aspettiamo). Usiamo lo sviluppo in serie di Taylor e scriviamo (per una particella)

$$\mathbf{r}(t + \delta t) = \mathbf{r}(t) + \mathbf{v}(t)\delta t + \frac{1}{2}\mathbf{a}(t)\delta t^2 + \frac{1}{6}\dot{\mathbf{a}}(t)\delta t^3 + \mathcal{O}(\delta t^4) \quad (5.7)$$

$$\mathbf{r}(t - \delta t) = \mathbf{r}(t) - \mathbf{v}(t)\delta t + \frac{1}{2}\mathbf{a}(t)\delta t^2 - \frac{1}{6}\dot{\mathbf{a}}(t)\delta t^3 + \mathcal{O}(\delta t^4). \quad (5.8)$$

Sommando e sottraendo le due relazioni e ricordando $\mathbf{F} = m\mathbf{a}$ otteniamo l'*algoritmo di Verlet*:

$$\mathbf{r}(t + \delta t) = \mathbf{r}(t - \delta t) + 2\mathbf{r}(t) + \frac{\mathbf{F}(t)}{m}\delta t^2 + \mathcal{O}(\delta t^4) \quad (5.9)$$

$$\mathbf{v}(t) = \frac{\mathbf{r}(t + \delta t) - \mathbf{r}(t - \delta t)}{2\delta t} + \mathcal{O}(\delta t^3). \quad (5.10)$$

Da notare che le velocità non sono di fatto usate. L'algoritmo di Verlet è il prototipo di tutti gli algoritmi di integrazione ed è tuttora usato. E' efficiente perché richiede un solo calcolo di forze (la parte più costosa dal punto di vista computazionale) ad ogni passo. E' numericamente stabile ed ha una proprietà assai importante per una buona MD: conserva la *simmetria per inversione temporale* delle equazioni del moto. In altri termini: se si inverte la velocità e la direzione del tempo ad un certo istante, la traiettoria ritraccia all'inverso quella seguita fino ad allora all'inverso.

Combinando in modo diverso gli sviluppi in serie di Taylor si possono ricavare altri algoritmi di integrazione. Uno molto usato è il *velocity Verlet*:

$$\mathbf{r}(t + \delta t) = \mathbf{r}(t) + \mathbf{v}(t)\delta t + \frac{\mathbf{F}(t)}{m}\delta t^2 \quad (5.11)$$

$$\mathbf{v}(t + \delta t) = \mathbf{v}(t) + \frac{1}{2m}(\mathbf{F}(t + \delta t) + \mathbf{F}(t))\delta t. \quad (5.12)$$

Si può dimostrare che Verlet normale e velocity Verlet producono le stesse traiettorie (cosa per nulla evidente a prima vista). Un altro algoritmo molto usato è il *leap-frog* (a salto di rana):

$$\mathbf{r}(t + \delta t) = \mathbf{r}(t) + \mathbf{v}(t + \frac{\delta t}{2})\delta t \quad (5.13)$$

$$\mathbf{v}(t + \frac{\delta t}{2}) = \mathbf{v}(t - \frac{\delta t}{2}) + \frac{\mathbf{F}(t)}{m}\delta t, \quad (5.14)$$

così chiamato perché velocità e posizione non sono calcolati allo stesso tempo ma "saltellano" fra tempi sfalsati di $\delta t/2$. Esistono infine algoritmi più sofisticati, quale l'algoritmo di Beeman.

5.2 Potenziali Empirici

Lo studio teorico dei sistemi biologici permette di ricavare le relazioni tra struttura, funzione e dinamica a livello atomico. Molti dei problemi di interesse coinvolgono molti atomi per i quali non è ancora possibile adottare una trattazione quantistica. Tuttavia, i problemi diventano molto più trattabili utilizzando dei *potenziali empirici*, che sono molto meno dispendiosi (computazionalmente), anche se ciò porta ad una serie di approssimazioni e limitazioni. Le funzioni di energia potenziale (detti anche *force fields*) utilizzate portano ad un buon compromesso tra l'accuratezza fisica e l'efficienza computazionale. In effetti, tali funzioni sono in genere calibrate su dati sperimentali e calcoli quantistici attraverso cui vengono ricavati i parametri dai quali tali funzioni dipendono. Tra gli insiemi di funzioni di energia potenziale più usati, citiamo quelli usati dai più noti pacchetti di software per simulazioni MD: AMBER, CHARMM, GROMOS, più la variante OPLS alle funzioni AMBER. Una seria limitazione di tali approcci è la quasi impossibilità di descrivere drastici cambiamenti nella struttura elettronica del sistema: eventi come la formazione o la rottura di un legame non sono permessi, o più esattamente, sono male descritti. Una migliore descrizione richiederebbero approcci misti in cui parte delle interazioni

sono trattate quantisticamente. Se tuttavia siamo interessati a seguire la dinamica di molecole biologiche in assenza di reazioni chimiche significative, tale limitazione non costituisce un problema.

Nei lucidi che seguono descriveremo le funzioni di energia potenziale implementate in CHARMM (Chemistry at HARvard Molecular Mechanics) e largamente utilizzate per simulare proteine, acidi nucleici, lipidi e carboidrati.

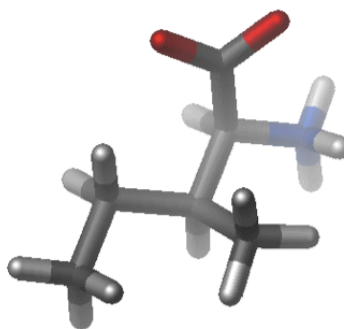
5.2.1 La Funzione Energia Potenziale in CHARMM

L'energia E , funzione delle posizioni $\{\mathbf{r}\}$ di tutti gli atomi del sistema, è data dalla somma di due termini:

$$E(\{\mathbf{r}\}) = E_{bonded} + E_{non-bonded} \quad (5.15)$$

dove

- E_{bonded} tiene conto delle interazioni tra gli atomi legati chimicamente da un legame covalente; queste dipendono dalle lunghezze di legame, dagli angoli e dalle rotazioni dei legami in una molecole;
- $E_{non-bonded}$ tiene conto delle interazioni tra gli atomi non legati chimicamente o tra gli atomi separati da 3 o più legami covalenti.



Il termine E_{bonded} può essere a sua volta scritto come somma di ulteriori termini:

$$E_{bonded} = E_{bond-stretch} + E_{bond-bend} + E_{rotate-along-bond}. \quad (5.16)$$

Nei prossimi paragrafi i vari termini sono esaminati in dettaglio.

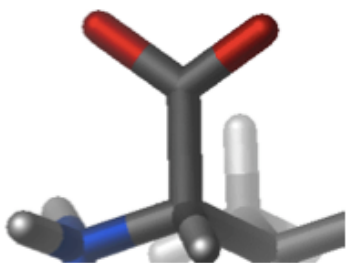
5.2.2 I vari termini di E_{bonded}

Il primo termine di tale energia corrisponde al potenziale armonico associato all'interazione tra la coppia di atomi separati da un legame covalente (1,2-pairs).



In questa espressione l'energia di un legame è funzione della variazione della lunghezza del legame, b , da quella ideale, b_0 . La costante K_b determina la forza del legame. Sia b_0 che K_b sono specifici per ogni coppia di atomi legati e quindi dipendono dal tipo di atomi che partecipano al legame.

$$E_{bond-stretch} = \sum_{1,2 \text{ pairs}} K_b (b - b_0)^2 \quad (5.17)$$

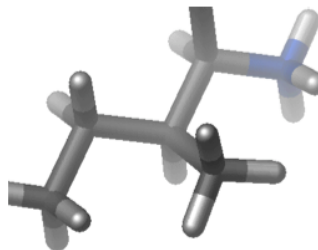


Il secondo termine è associato alle variazioni degli angoli di legame θ rispetto al suo valore ideale θ_0 , ed ha anch'esso una forma armonica:

$$E_{bond-bend} = \sum_{angles} K_{\theta}(\theta - \theta_0)^2. \quad (5.18)$$

Anche in questo caso θ_0 e K_{θ} dipendono dagli atomi coinvolti nel legame.

Il terzo termine rappresenta la funzione potenziale associata agli angoli di torsione e modella la presenza di barriere steriche tra gli atomi separati 3 legami covalenti (1,4 pairs). Il moto associato con questi termini è una rotazione, descritta da un angolo diedrico e un coefficiente di simmetria ($n=1,2,3$), intorno al legame mediano. Questo potenziale si assume periodico ed è in genere espresso tramite una funzione coseno.

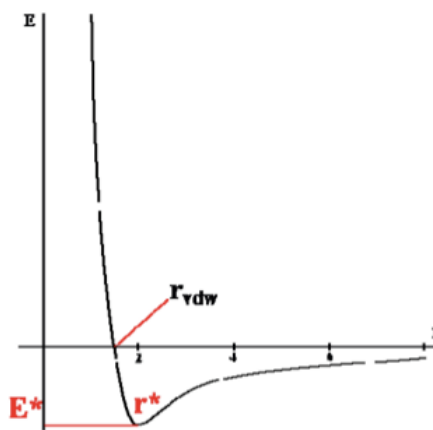


$$E_{rotate-along-bond} = \sum_{1,4 \text{ pairs}} K_{\phi}[1 - \cos(n\phi)]. \quad (5.19)$$

5.2.3 I vari termini di $E_{non-bonded}$

Tale termine dell'energia potenziale rappresenta il contributo delle interazioni tra atomi non legati e in CHARMM ha due componenti: l'energia di interazione Van der Waals e l'energia di interazione elettrostatica. Qualche altra funzione potenziale include un'ulteriore termine che tiene conto dei legami a idrogeno. In CHARMM queste interazioni sono inglobate nelle interazioni elettrostatiche e Van der Waals.

L'interazione van der Waals tra due atomi nasce dal bilanciamento tra una forza repulsiva ed una attrattiva: la forza repulsiva è quella tra gli elettroni dei due atomi e a piccole distanze è predominante; quella attrattiva è spesso denominata *forza dispersiva*, ed è associata alle fluttuazioni nelle distribuzioni di carica delle nuvole elettroniche. Le fluttuazioni nella distribuzione elettronica su di un atomo o molecola danno luogo a dipoli istantanei che a loro volta inducono un dipolo in un secondo atomo o molecola, dando luogo ad una forza attrattiva.



La forza attrattiva ha un raggio maggiore di quella repulsiva e quindi a piccole distanze la forza repulsiva domina. Ciò determina un minimo dell'energia potenziale. Il posizionamento degli atomi a tale distanza stabilizza il sistema. Il valore dell'energia al minimo, E^* , e la separazione ottimale tra gli atomi, r^*

(che corrispondono approssimativamente alla somma dei raggi di Van der Waals degli atomi) dipendono dal tipo di atomi.

Usualmente, l'interazione van der Waals è modellizzata utilizzando il potenziale 6-12 di Lennard-Jones che ha la forma qui sotto indicata:

$$E_{vdW} = \sum_{\text{nonbonded pairs}} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad (5.20)$$

dove ϵ_{ij} e σ_{ij} sono i parametri dell'interazione (ϵ_{ij} è la profondità della buca di potenziale, mentre σ_{ij} è dell'ordine del raggio di van der Waals per la coppia ij considerata). L'interazione di van der Waals è una delle più importanti per la stabilità delle macromolecole biologiche.

L'interazione elettrostatica tra una coppia di atomi è rappresentata dal potenziale coulombiano, che nel sistema MKSA ha la forma:

$$E_{electrostatic} = - \sum_{\text{nonbonded pairs}} \frac{q_i q_j}{4\pi\epsilon_0\epsilon r_{ij}} \quad (5.21)$$

dove ϵ è la costante dielettrica effettiva del mezzo e r_{ij} è la distanza tra i due atomi con carica q_i e q_j .

5.2.4 Limiti dei Potenziali Empirici

I potenziali empirici sono funzioni semplici e differenziabili delle coordinate atomiche, dalle quali è facile ricavare analiticamente la forza (vettoriale) che agisce su ogni atomo. Ciò permette di fare la simulazione della dinamica in modo efficiente. Le limitazioni imposte dai potenziali che abbiamo introdotto sono essenzialmente di due tipi:

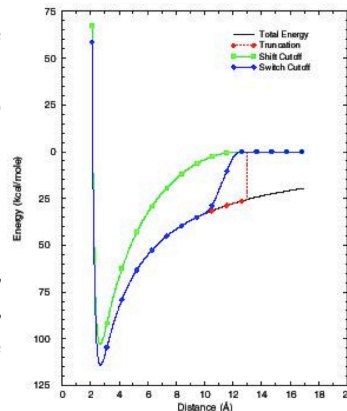
- I parametri che definiscono le funzioni di interazione fra ogni atomo del sistema e gli altri atomi vengono assegnate all'inizio e rimangono fissi per tutta la simulazione. Questo non permette di tenere conto del fatto che il comportamento di certi atomi (per esempio, per l'ossigeno e l'azoto) è influenzato dagli atomi vicini.
- I potenziali sono costruiti attraverso la somma di interazioni di coppia: non sono considerate interazioni a tre o più particelle. Questo non permette di tenere conto di effetti legati alla polarizzazione di certi gruppi e/o a ionizzazioni di certi residui aminoacidici indotti dal campo elettrico di tutta la proteina.

Infine, le funzioni di energia potenziale non includono effetti entropici. Quindi un minimo di energia non corrisponde necessariamente ad uno stato di equilibrio. Di conseguenza le simulazioni MD sono fatte in modo da riprodurre condizioni analoghe a quelle in cui normalmente sono condotti gli esperimenti, e cioè a temperatura e/o pressione costante.

Il termine energetico più dispendioso in una simulazione MD è quello associato alle interazioni non-bonded (elettrostatiche e van der Waals). Tali interazioni decrescono lentamente con la distanza fra gli atomi, per cui tale termine

dovrebbe essere valutato per ogni coppia di atomi del sistema. Ciò implica che il tempo necessario per la sua valutazione cresce con il quadrato del numero di atomi del sistema (N^2). Per velocizzare tale calcolo viene posto un cut-off alla distanza di interazione. Nel corso degli anni si sono sviluppati vari modi di attuazione del cut-off:

- *Troncamento*. L'energia d'interazione è semplicemente posta a zero per distanze maggiori di $r_{cut-off}$. Dato che questo metodo (brutale) porta spesso a grosse e innaturali fluttuazioni nell'energia, è poco utilizzato.
- *Shift*. Viene modificata l'intera funzione potenziale in modo che a $r_{cut-off}$ sia nulla. Tale metodo porta a distanze di equilibrio più piccole di quelle sperimentali.
- *Switch*. La funzione energia viene modificata solo in un piccolo intervallo di distanze (vedi figura a lato). Tale metodo ha il problema di generare forze notevoli nella regione di switch che possono perturbare l'equilibrio.



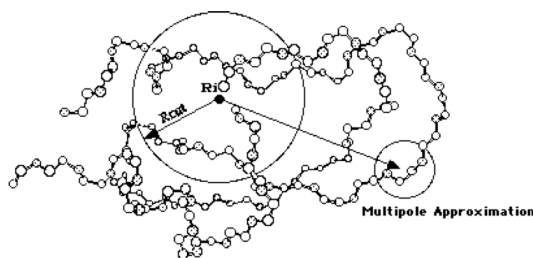
L'importanza delle interazioni elettrostatiche a lungo raggio nei sistemi biologici è dimostrato da un grande numero di studi sperimentali. In effetti, in una simulazione MD, l'inclusione di interazioni a raggio più elevato si ottiene semplicemente aumentando $r_{cut-off}$, con il risultato però di aumentare notevolmente il costo computazionale. A causa di ciò, spesso, tali interazioni vengono completamente ignorate, introducendo la severa approssimazione di non tenere conto delle proprietà dielettriche del mezzo. Più recentemente, sono stati introdotti vari modelli che permettono l'inclusione delle interazioni elettrostatiche a lungo range nelle simulazioni MD. Essi si basano su

- risommazione dell'energia di interazione;
- utilizzo dello sviluppo multipolare;
- partizione dell'interazione in termini a corto e lungo range.

Tali modelli sono stati inclusi in CHARMM e le simulazioni mostrano la loro importanza in vari casi.

Il modello *Extended Electrostatics* approssima l'interazione elettrostatica complessiva partizionando il potenziale elettrico in la forza risultante su un atomo in un contributo *vicino* ed uno *esteso*.

Il contributo vicino è dovuto alle cariche presenti in una sfera di raggio $r_{cut-off}$, ed è calcolato tramite termini di interazione a coppie; Il contributo esteso è associato alle particelle all'esterno della sfera stessa ed è calcolato utilizzando un'approssimazione multipolare.



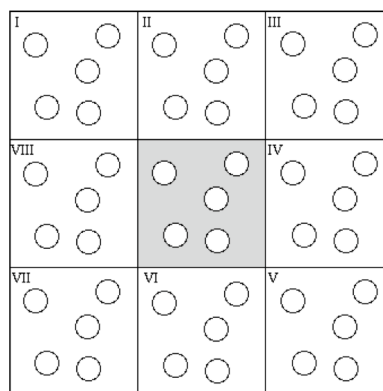
5.3 Inclusione del Solvente in una Simulazione MD

L'acqua ha un ruolo fondamentale su struttura, dinamica e termodinamica delle molecole biologiche. Uno dei più importanti effetti del solvente è quello di schermo per le interazioni elettrostatiche

L'inclusione del solvente nelle simulazioni può essere fatta in vari modi. Il modo *implicito* più semplice è quello di includere una costante elettrostatica nel termine elettrostatico dell'energia potenziale, usando se necessario una costante dielettrica effettiva dipendente dalla distanza: $\epsilon_{eff} \equiv \epsilon(r_{ij})$. Il modo *esplicito* consiste invece nell'includere molecole d'acqua nel sistema. Le molecole forniscono direttamente l'effetto di schermo (o almeno la parte legata al riorientamento delle stesse).

In questa trattazione più dettagliata bisogna utilizzare convenienti condizioni al contorno, sia per prevenire la diffusione delle molecole d'acqua in zone troppo lontane dalla proteina, che per permettere la simulazione e il calcolo delle proprietà macroscopiche utilizzando un numero limitato di molecole di solvente. Le condizioni al contorno periodiche permettono di fare la simulazione con un numero relativamente piccolo di molecole di solvente e in modo che le particelle risentano di forze analoghe a quelle che risentirebbero nel bulk della soluzione.

Qui a lato è mostrata la situazione in due dimensioni. Il quadrato centrale è circondato dagli otto quadrati vicini e le coordinate delle particelle immagine in tali quadrati sono ottenute mediante semplici traslazioni delle posizioni delle corrispondenti particelle nel quadrato principale. Le forze nel quadrato principale sono calcolate utilizzando tutte le particelle (anche quelle dei quadrati secondari). Il raggio di cut-off è in genere scelto in modo che una particella del quadrato primario non veda la sua immagine nei quadrati secondari.



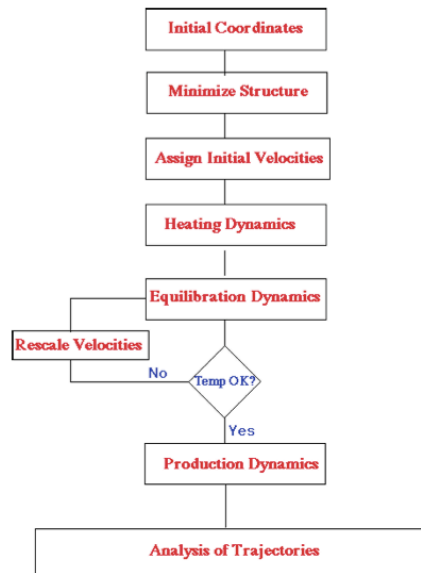
Ci sono casi in cui non è utile considerare le condizioni al contorno periodiche, e in cui, anche sfruttando la notevole potenza dei computer, le simulazioni vengono fatte incorporando un grande numero di molecole di solvente. In questi casi, la proteina viene circondata da una sfera d'acqua e ottenuta utilizzando un potenziale al bordo che confina le molecole d'acqua in tale volume

(vedi figura in basso a sinistra). In altri casi, l'acqua viene confinata solo nella parte interessante della simulazione (vedi figura in basso a destra).



5.4 Set-up e running di una Simulazione MD

Come per le simulazioni Monte Carlo, il risultato di una simulazione MD è una serie temporale di conformazioni corrispondenti alle traiettorie o cammini seguiti da ogni atomo. La maggior parte delle simulazioni MD sono effettuate a N , V ed E costanti (insieme microcanonico), ma è possibile fare simulazioni anche a N , T e P costanti che meglio riproducono le condizioni sperimentali. Qui a fianco vediamo un diagramma a blocchi (schematico) che riassume le varie tappe di una di queste simulazioni.



- *Inizializzazione.* Per cominciare si sceglie la configurazione iniziale del sistema $t = 0$. Nella maggior parte dei casi, per la simulazione di biomolecole si parte dalla struttura (cristallografica o NMR) ottenuta dal *Brookhaven Protein Databank*. La scelta della configurazione iniziale deve essere fatta con attenzione perchè essa può influenzare la qualità della simulazione. Prima di partire con la vera e propria simulazione è spesso bene procedere ad una minimizzazione dell'energia della struttura. A questo punto vengono addizionate le molecole d'acqua: spesso le molecole d'acqua sono prese da una precedente simulazione fatta in un box di dimensioni opportune; la proteina viene inserita in tale box eliminando le molecole d'acqua ad essa sovrapposte. A questo punto si effettua una nuova minimizzazione dell'energia per permettere all'acqua ed alla proteina di assestarsi nella nuova situazione.
- *Riscaldamento.* Vengono assegnate le velocità iniziali (in modo da avere una bassa temperatura) e si avvia la simulazione. Nel caso di solvente

esplicito, prima si blocca la proteina e si lascia termalizzare le molecole d'acqua, poi si sblocca la proteina e si comincia l'evoluzione del sistema proteina + acqua insieme. Durante la fase di riscaldamento periodicamente le velocità vengono riscaldate in modo da via via aumentare la temperatura fino ad arrivare a quella prescritta.

- *Equilibrizzazione.* Raggiunta la temperatura prescritta la simulazione continua in modo che possano essere monitorate quantità come pressione, temperatura ed energia. In questa fase, talvolta è necessario, per ovviare a innaturali fluttuazioni della temperatura, riscaldare le velocità opportunamente. La fase continua fino a quando l'equilibrizzazione dei parametri suddetti è raggiunta.
- *Fase di Produzione.* Nella fase finale, dopo l'equilibrizzazione, lasciando andare la simulazione, si possono le medie dei parametri termodinamici di interesse. Tale fase può corrispondere a varie centinaia di picosecondi a nanosecondi.

Durante una simulazione MD, le coordinate e le velocità degli atomi vengono memorizzate e successivamente utilizzate per l'analisi. Si possono calcolare medie di quantità come l'energia (cinetica e/o potenziale):

$$\langle E \rangle = \frac{1}{M} \sum_{i=1}^M E_i \quad (5.22)$$

dove M è il numero di passi di MD, $E_i \equiv E(t_i)$ al tempo $t_i = t_0 + (i - 1)\delta t$ (assumendo che $t = t_0$ sia l'istante in cui si inizia a raccogliere statistica). Si possono calcolare altri tipi di grandezze quali il raggio di girazione in funzione del tempo e la sua media; o le cosiddette *funzioni di correlazioni di coppia*, $g_{AB}(R)$, che danno la probabilità di trovare un atomo di tipo B a distanza R da un atomo di tipo A .