

Approfondimento sulla dinamica molecolare

1. La valutazione delle grandezze termodinamiche nella dinamica molecolare

Negli studi di dinamica molecolare siamo spesso interessati a esaminare l'andamento di **grandezze macroscopiche** del sistema, come la variazione di energia libera associata al legame di un inibitore ad un enzima o l'energia associata a differenti conformazioni di una proteina. Inoltre, per riprodurre nel miglior modo possibile le condizioni sperimentali, durante la simulazione è spesso necessario mantenere il controllo di grandezze termodinamiche macroscopiche, tra cui la temperatura, la pressione o il volume. Il legame tra i dati forniti a livello microscopico dalla dinamica molecolare (posizione e quantità di moto degli atomi facenti parte del sistema indagato) e le grandezze macroscopiche del sistema stesso si fonda sulle solide basi matematiche fornite dalla **meccanica statistica**. La meccanica statistica è quella branca della fisica che si occupa dello studio di sistemi macroscopici e di grandezze termodinamiche a partire dall'insieme degli stati microscopici del sistema. A livello microscopico, infatti, un sistema può essere definito dall'insieme delle posizioni (le tre coordinate spaziali x , y e z) e delle velocità (un vettore di componenti v_x , v_y e v_z) di ciascun elemento che ne fa parte. Si può dire quindi che da un punto di vista meccanico lo stato macroscopico di un sistema composto da N particelle può essere completamente definito da $6 \cdot N$ variabili indipendenti, date dall'insieme delle componenti di posizione e velocità dei suoi elementi. Si può quindi pensare all'esistenza di uno spazio multidimensionale, formato da $6 \cdot N$ dimensioni, nel quale ciascun punto dello spazio rappresenta uno stato del sistema. Tale spazio è chiamato **spazio delle fasi**. Pur descrivendo stati diversi del sistema, alcuni punti dello spazio delle fasi possono essere accomunati dalle medesime grandezze termodinamiche macroscopiche, per esempio temperatura, pressione, volume, e così via. L'insieme di tali punti costituisce il così detto **ensemble**, un concetto chiave in meccanica statistica e di conseguenza nella dinamica molecolare. È importante sottolineare che, sebbene gli stati che fanno parte di un ensemble condividano il medesimo valore delle grandezze termodinamiche da un punto di vista macroscopico, a livello microscopico ogni stato può discostarsi rispetto al valore suddetto. Si consideri, per esempio, la pressione dell'ensemble di un sistema per un certo valore di volume e temperatura. Ogni punto dell'ensemble, avendo diverse coordinate e velocità, non sarà completamente sincronizzato con gli altri punti, perciò in un dato istante saranno diversi il numero e la quantità di moto degli elementi del sistema che collidono sulle superfici che racchiudono il sistema stesso rispetto agli altri punti. Ne consegue che la pressione che viene misurata a livello macroscopico sarà data dalla media delle pressioni dei singoli punti dell'ensemble. Per definizione, tali punti condividono questo valore medio che concorrono a formare, pur **fluttuando** attorno a esso. Pertanto le grandezze termodinamiche di un sistema, a livello macroscopico, sono le **medie dei valori presenti in un ensemble** (*ensemble average*). Si può quindi pensare agli ensemble come a una popolazione di stati di un sistema che si distribuisce in maniera stocastica rispetto a un valore termodinamico medio, che può essere approssimato a quello sperimentalmente misurato. La dinamica molecolare genera, come abbiamo visto, la traiettoria di un sistema, cioè un insieme di stati (posizioni e velocità) di un sistema in un dato intervallo di tempo. Si può decidere di considerare gli stati di un sistema generati da un processo di dinamica molecolare come facenti parte di un dato ensemble, caratterizzato per esempio, oltre che da un numero fisso di particelle, da valori costanti di volume e temperatura (**ensemble NVT** o canonico) o di temperatura e pressione (**ensemble NPT** o isoterma-isobaro) che possono essere fissati in maniera arbitraria. Ne consegue che al termine della simulazione di dinamica molecolare è possibile

derivare grandezze termodinamiche macroscopiche (entalpia, energia libera di Gibbs e così via) attraverso la media dei valori mostrati dagli stati generati durante la simulazione. Così facendo, si sta assumendo implicitamente che la media dei valori istantanei di un ensemble coincida con la media dei valori presentati dal sistema nella sua evoluzione temporale (**ipotesi ergodica**), cioè che, se la simulazione di dinamica durasse un tempo sufficientemente lungo, essa finirebbe col generare tutti gli stati definiti dai punti di un ensemble dello spazio delle fasi. A rigor del vero, nonostante l'assunzione dell'ipotesi ergodica sembri plausibile se applicata ai sistemi generalmente studiati in dinamica molecolare, essa non è stata ancora completamente dimostrata da un punto di vista matematico. In ogni caso è importante tenere presente che la maggior parte delle grandezze termodinamiche che possono essere ottenute da uno studio di dinamica molecolare richiede il calcolo di un qualche tipo di *ensemble average*.

2. Preparazione e passaggi in una simulazione di dinamica molecolare

Un esperimento di dinamica molecolare consiste principalmente di quattro passaggi: preparazione, fase di equilibratura, fase di produzione, analisi dei risultati (Figura 1).

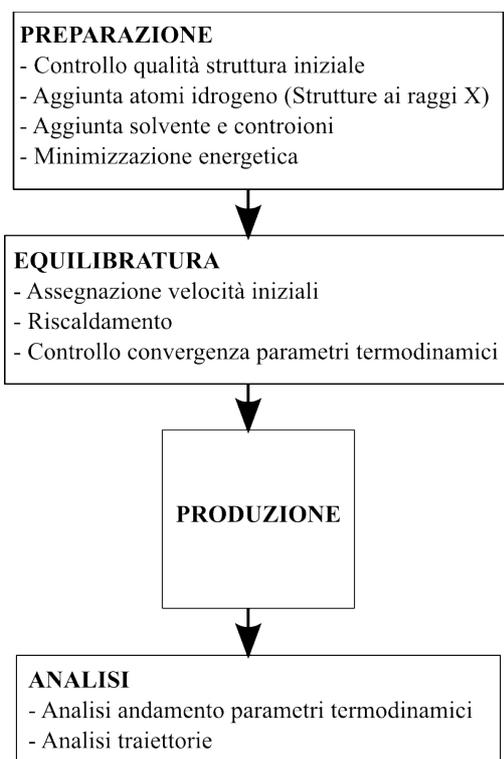


Figura 1 Passaggi comunemente eseguiti nella simulazione di dinamica molecolare di una proteina.

Nella prima fase, il sistema oggetto di studio va incontro a una serie di sottopassaggi di preparazione, la maggior parte dei quali consiste nel controllo della qualità e delle proprietà geometriche ed energetiche del sistema considerato, e nella creazione di una serie di condizioni al contorno necessarie affinché il sistema si accosti il più possibile alla realtà sperimentale. Si consideri, per esempio, la simulazione di dinamica molecolare di una proteina. Come accennato nel Paragrafo 12.3, le coordinate iniziali del sistema provengono da dati sperimentali (cristallografia ai raggi X o NMR) o da tecniche di modellistica molecolare (come la modellizzazione per omologia). Indipendentemente dalla metodologia, è sempre bene controllare la

qualità della struttura di partenza. Nel caso di una proteina risolta attraverso cristallografia a raggi X, si può far riferimento a parametri quali la risoluzione o l'*R-value* (si veda il Paragrafo 1.1.4); valori di risoluzione inferiori ai 2 Å con *R-value* al di sotto di 0,20 sono solitamente considerati indici di buona qualità strutturale. Sfortunatamente, non esistono valori di parametri universalmente accettati come indici di buona qualità nel caso di strutture risolte attraverso NMR (per avere una lista dei metodi più comunemente utilizzati per valutare la bontà di una struttura NMR si veda Nabuurs et al., 2004). Le metodologie impiegate per valutare la bontà di un modello proteico ottenuto da tecniche di modellistica molecolare sono illustrate nel Capitolo 11. Particolare attenzione va rivolta alle strutture in cui parti della catena polipeptidica, residui o atomi non sono presenti a causa di dati sperimentali incompleti. In questi casi si può procedere a una modellizzazione parziale delle regioni mancanti. Infine, è necessario ricordare che le strutture ottenute tramite diffrazione a raggi X non presentano gli atomi d'idrogeno, che sarà necessario ricostruire.

Una volta che il file contenente le coordinate tridimensionali della molecola oggetto di studio è stato controllato e, dove possibile, corretto e rifinito, è necessario aggiungere a questo file informazioni in grado di descrivere il sistema in termini di tipi di atomi, legami, cariche parziali e così via, affinché i potenziali possano essere assegnati correttamente (si veda il Paragrafo 12.1.2). I programmi più comunemente utilizzati nella dinamica molecolare, come **GROMACS**, sono in grado di riconoscere e assegnare automaticamente questi parametri per una vasta gamma di unità monomeriche costituenti il sistema (per esempio gli amminoacidi), utilizzando librerie di "unità di base" le cui regole sono applicate agli insiemi di atomi identificati (per esempio i residui di una proteina). Il risultato è un **file di topologia** (*topology file*) contenente, oltre alle coordinate, le informazioni necessarie a riconoscere un atomo e le sue proprietà. Naturalmente queste librerie non possono contenere tutti i diversi tipi di molecole esistenti in natura, per cui si rende spesso necessario, in fase di preparazione di una simulazione di dinamica molecolare, utilizzare sistemi alternativi (come il server **Dundee PRODRG2** o il pacchetto di programmi **Antechamber**) che siano in grado di ricavare la topologia di particolari molecole presenti nel sistema (cofattori, inibitori, ligandi e così via, necessari alla simulazione).

Il passaggio successivo nella preparazione del sistema, come visto in precedenza, consiste nell'effettuare uno o più cicli di minimizzazione energetica. Tali cicli si rendono necessari a causa della presenza di interazioni energeticamente sfavorevoli nelle configurazioni iniziali del sistema, in seguito all'aggiunta degli atomi di idrogeno o di altri atomi/molecole, alla bassa qualità delle strutture di partenza o alla ricostruzione di parti del sistema. La minimizzazione energetica è altresì necessaria ogniqualvolta il sistema sia soggetto a successive modifiche, per esempio dopo aver definito una *box* ed averla riempita di molecole di solvente (Paragrafo 12.3.1) o dopo aver aggiunto **controioni** al sistema per annullare la carica elettrostatica netta *e/o* per raggiungere un valore desiderato di forza ionica.

La fase di equilibratura del sistema consiste, come suggerisce il termine stesso, nel portare il sistema all'equilibrio termodinamico, in altre parole serve a evitare che il sistema sia soggetto a eccessive e irrealistiche fluttuazioni dei valori termodinamici (temperatura, pressione, energia cinetica e così via). Per far ciò, solitamente, all'inizio il sistema viene "riscaldato": sono assegnate le velocità iniziali del sistema; si procede poi a una simulazione nella quale tali velocità sono riscaldate per simulare l'**accoppiamento** (*coupling*) del sistema con un **bagno termico** (un bagno termico, o *heat bath*, è definito come un sistema a capacità termica illimitata, la cui temperatura rimane costante a contatto con il sistema di interesse), la cui temperatura viene progressivamente aumentata fino a raggiungere il valore desiderato e mantenuto durante la simulazione di dinamica vera e propria. In un ensemble di tipo NPT (Paragrafo 12.1e), all'accoppiamento con la temperatura segue l'accoppiamento con il valore desiderato di pressione. La fase di equilibratura procede finché tutti i parametri termodinamici di interesse non raggiungono un valore "costante", attorno al quale il sistema fluttua (Figura 12.2e).

A questo punto può iniziare la simulazione di dinamica vera e propria. Questo passaggio consiste di solito nell'assegnazione di valori a una serie di parametri sulla base del tipo di problema biologico che si sta affrontando e delle risposte che si desidera avere da un esperimento di dinamica molecolare. La decisione principale riguarda la durata della simulazione, che a sua volta deve prendere in considerazione:

- il tempo reale necessario affinché un dato evento biologico avvenga (Tabella 12.1);
- le dimensioni del sistema considerato;
- la potenza di calcolo di cui si dispone;

- il tempo reale di cui si dispone, approssimativamente, per effettuare l'intera simulazione.

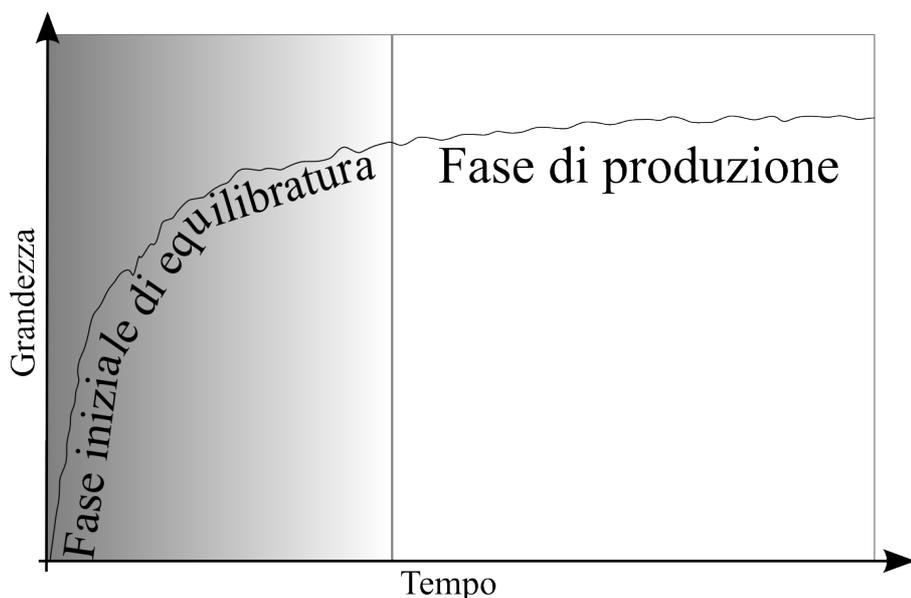


Figura 2 Fase di equilibratura e fase di produzione per una generica grandezza durante una simulazione di dinamica molecolare.

È da notare, a tal proposito, che anche i più potenti calcolatori oggi disponibili non sono in grado di simulare alcuni processi di notevole interesse biologico (primo tra tutti il ripiegamento di una proteina di medie dimensioni).

Tabella 1 Durata, in secondi, di alcuni eventi molecolari.

<i>Evento</i>	<i>Durata approssimativa (s)</i>
<i>Stiramento di un legame covalente</i>	$\sim 10^{-15}$
<i>Riorientamento di una molecola di H_2O</i>	$\sim 4 \cdot 10^{-12}$
<i>Chiusura/apertura inter-dominio proteico</i>	$10^{-11} - 10^{-7}$
<i>Cambiamenti conformazionali proteici</i>	$10^{-6} - \text{diversi secondi}$
<i>Denaturazione locale di una proteina</i>	$10^{-3} - \text{diversi secondi}$

Quando la simulazione è terminata si procede all'analisi dei dati ottenuti durante la fase di produzione. Questo passaggio, a sua volta, si suddivide in due fasi. Prima di tutto è necessario applicare alcune procedure di controllo per verificare la bontà del processo di dinamica. Queste procedure consistono principalmente nel valutare se l'energia del sistema e i parametri termodinamici abbiano mantenuto nel corso della simulazione un andamento stabile (si dice, in questo caso, che hanno raggiunto la convergenza) e se la struttura non abbia subito distorsioni eccessive o innaturali durante il processo (si valuta, per esempio, l'RMSD rispetto alla struttura di partenza, o l'andamento del raggio di girazione durante la simulazione). Un altro controllo che è necessario eseguire è che non ci siano state interazioni dirette tra le immagini delle celle periodiche durante l'intera simulazione; tale controllo avviene attraverso il monitoraggio della distanza minima tra le immagini.

Se tutto è andato bene, si può procedere all'analisi vera e propria. Il numero d'informazioni ricavabili da uno studio di dinamica molecolare è molto alto. A puro titolo di esempio, possono essere valutate proprietà come la variazione di area accessibile al solvente, la formazione e/o la rottura d'interazioni di tipo non covalente (come i legami idrogeno e i ponti salini), la strutturazione o destrutturazione di regioni a struttura secondaria regolare e le interazioni tra la proteina e il solvente. In ogni caso, la diretta osservazione della traiettoria ottenuta può forse fornire il maggior numero di informazioni sul rapporto struttura/funzione di una proteina, come riportato nell'esempio seguente.

3. Un esempio di simulazione di dinamica molecolare: accesso al sito attivo dell'enzima acetilcolinesterasi

L'acetilcolinesterasi (AChE) è una proteina di 75 kD appartenente alla famiglia delle idrolasi. Dal punto di vista filogenetico, l'AChE è un enzima ubiquitario negli organismi dotati di sistema nervoso ed evolutivamente molto conservato. L'AChE è coinvolto nei meccanismi di neurotrasmissione, dove catalizza l'idrolisi dell'acetilcolina (ACh), il neurotrasmettitore che media il passaggio degli impulsi nervosi attraverso le sinapsi colinergiche. A livello di tali sinapsi, la trasmissione termina quando l'acetilcolina rilasciata è idrolizzata a colina e acetato a opera dell'AChE, che è presente in abbondanza nella fessura sinaptica, in prossimità della superficie della membrana postsinaptica. La colina che rimane nella fessura sinaptica viene poi riassorbita dalla membrana presinaptica e riciclata nella reazione di condensazione con l'acetil-coenzima A, per formare nuove molecole di ACh.

Alcuni autori (Tai et al., 2001) hanno evidenziato, attraverso una simulazione della durata di circa 10 ns, caratteristiche dinamiche e strutturali dell'AChE non anticipate da dati sperimentali. Nel compiere questo studio, gli autori hanno inizialmente prelevato dalla banca dati PDB le coordinate della struttura tridimensionale dell'AChE da topo, in complesso con una proteina nota come fasciculina. Quest'ultima è stata rimossa dal file contenente le coordinate; in seguito sono stati aggiunti alla struttura, attraverso modellizzazione, sette residui mancanti. Dopo aver introdotto gli atomi di idrogeno, la proteina è stata solvatata in una *box* cubica (96 Å di lato). Al fine di neutralizzare la carica negativa netta dell'AChE, sono stati aggiunti nove ioni sodio. Al termine della preparazione, il sistema era composto da un totale di 8289 atomi di soluto e 75615 atomi di solvente. La fase di equilibratura è stata condotta in ensemble NPT. Il solvente e il soluto sono stati riscaldati alla temperatura di 298,15 °K in un tempo di equilibratura di 0,1 ps. Il *coupling* del sistema alla pressione di 1 atm è avvenuto in 0,4 ps. La fase di produzione di un tale sistema, nell'intervallo di 10,8 ns, è durata circa tre anni, utilizzando uno dei supercomputer più efficienti nel 2001 presso il centro di supercalcolo di San Diego, in California.

Lo studio ha evidenziato la presenza, in corrispondenza dell'accesso al sito attivo dell'enzima, di un "portale", la cui apertura e chiusura intermittente, mediata da due residui aromatici, sembra essere responsabile di una "selettività dinamica" dei substrati dell'AChE rispetto alle loro dimensioni. Il portale, infatti, si apre e si chiude spesso, creando un varco accessibile solo a piccoli substrati, per esempio ACh. Solo molto più raramente il varco si estende a tal punto da rendere accessibile al sito attivo anche ligandi (o inibitori) più voluminosi. In tale scenario, sussisterebbe una situazione nella quale le molecole di ACh che transitano nei pressi dell'entrata al sito attivo possono facilmente avere accesso, mentre molecole più ingombranti dovrebbero aspettare un tempo molto più lungo per poter entrare, con la conseguenza che sarebbe molto più probabile, per queste ultime, allontanarsi per diffusione dal varco (Figura 3).

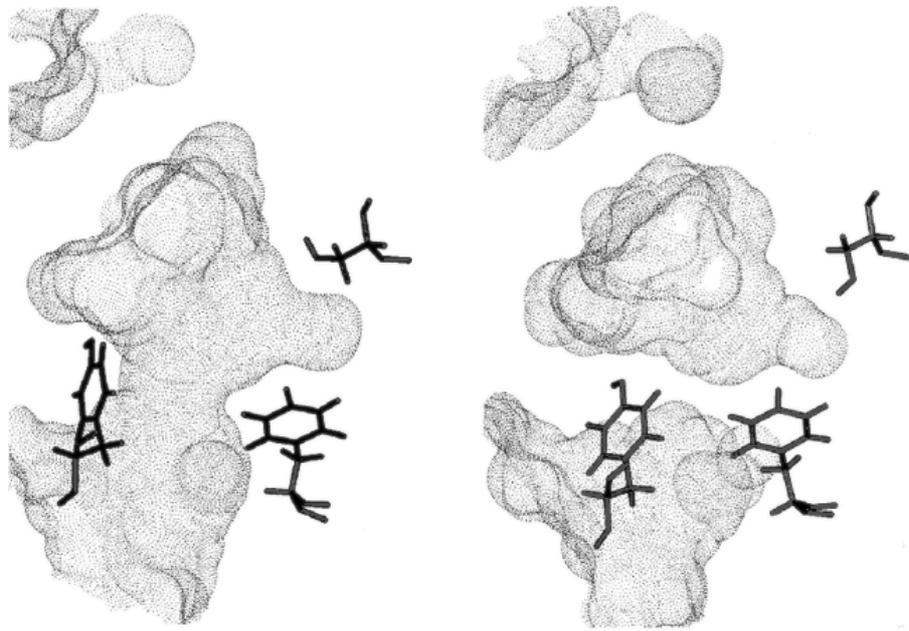


Figura 3 Apertura (sinistra) e chiusura (destra) del canale d'accesso al sito attivo dell'AChE. La superficie del canale è mostrata a punti. I residui che circondano il canale e che ne determinano l'apertura/chiusura sono mostrati come modelli a bastoncini. Riadattata da Karplus e McCammon, 2002.