

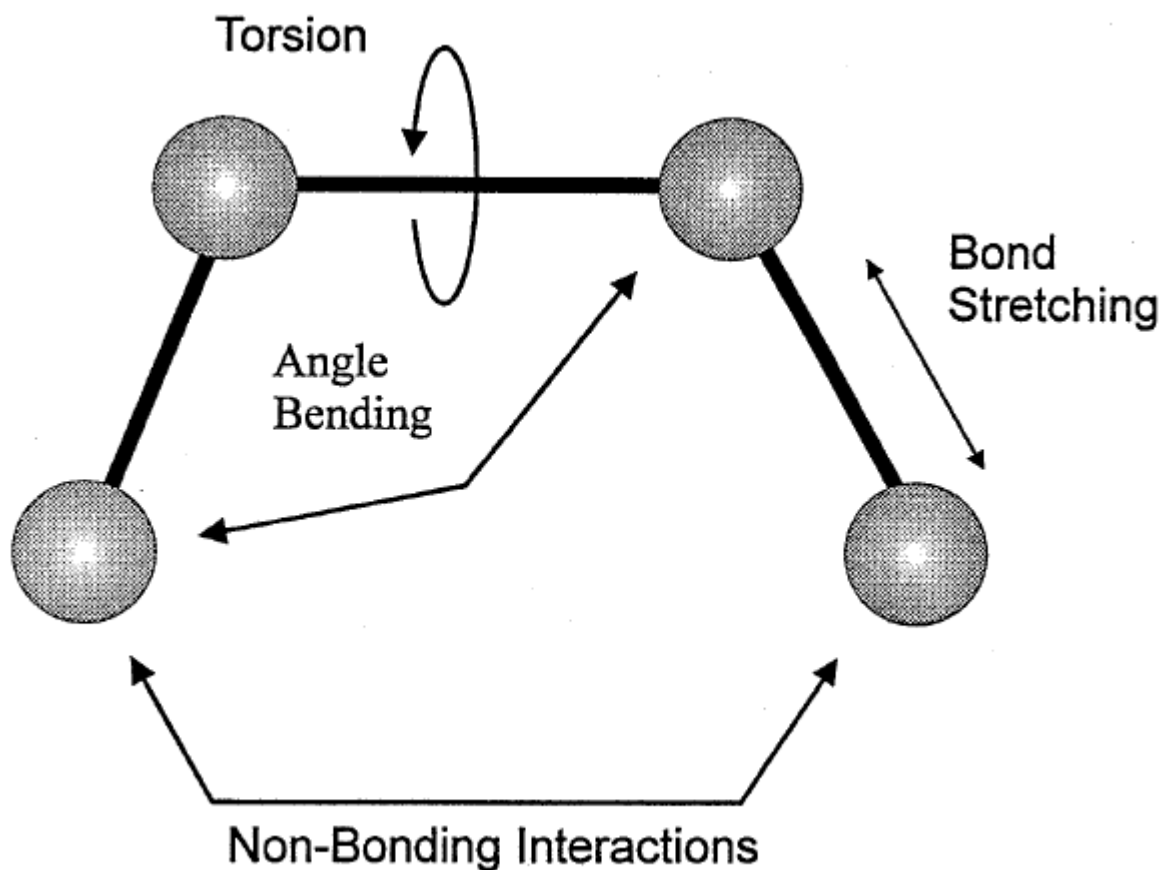
# METODOLOGIE COMPUTAZIONALI

## TUTORIAL DELLA Lezione DI MODELLISTICA MOLECOLARE

### MECCANICA MOLECOLARE

I modelli di meccanica molecolare (MM) sono stati sviluppati, nell'ambito della chimica computazionale, in base alla necessità di descrivere le strutture, e le proprietà molecolari, in un modo sufficientemente semplice tale da richiedere capacità di calcolo ragionevoli. Il campo d'applicazione della MM comprende sia piccole molecole di natura organica ed inorganica sia le macromolecole (proteine, acidi nucleici, polimeri). Normalmente la MM è applicata *in vacuo*, ma è anche possibile trattare gli effetti dovuti al solvente, che di solito è trattato in modo implicito. È prevedibile che presto la MM potrà fornire informazioni anche sui meccanismi di reazione, ma attualmente essa è applicabile esclusivamente a molecole allo stato fondamentale. Applicando la MM nel tempo (dinamica molecolare) è invece possibile descrivere l'evoluzione temporale di sistemi complessi quali proteine ed acidi nucleici. L'elevata velocità computazionale ne consente l'impiego in problemi che richiedono un alto numero di calcoli, quali la dinamica molecolare, l'analisi conformazionale e lo studio delle interazioni enzima substrato o ligandorecettore.

Obiettivo della MM è calcolare l'energia associata a ciascuna conformazione molecolare. Tali energie non vanno intese in senso assoluto poiché sono significative esclusivamente le differenze di energia calcolate fra due o più strutture. Gli elettroni non sono trattati distintamente dai loro nuclei (come invece avviene con i metodi quanto-meccanici), ma sono ad essi accorpati in strutture a simmetria sferica definibili *pseudoatomi*. Questo probabilmente costituisce il principale limite della MM, in quanto la simmetria di distribuzione elettronica per numerosi atomi non è affatto realistica. Con tale approccio, inoltre, non è possibile descrivere le situazioni di delocalizzazione elettronica o di coniugazione, tipiche dei sistemi aromatici. Il diametro delle strutture pseudoatomiche è ricavato da misurazioni sperimentali o da calcoli *ab initio*. In modo analogo si procede per la determinazione delle cariche atomiche parziali. Mentre gli atomi sono considerati delle sfere cariche, i legami atomici sono trattati alla stregua di molle. Le interazioni fra atomi sono quindi descritte secondo la teoria delle molle, basata sul calcolo delle energie potenziali proprie della meccanica classica. I legami atomici debbono essere forniti secondo uno schema predefinito. Non è quindi possibile analizzare la rottura o la formazione di nuovi legami. La trattazione matematica delle molle può essere impiegata per descrivere la capacità dei legami di accorciarsi allungarsi (stretching), flettersi (bending) e ruotare (torsion) come illustrato dalla figura.



## FORMALISMI DELLA MECCANICA MOLECOLARE

La più semplice equazione atta a descrivere il potenziale energetico molecolare è la seguente:

ENERGIA = Estretching + Ebending + Etorsion + Enon-bonding interactions

In MM è definito "**force field**" l'insieme delle equazioni e dei parametri che sono necessari a descrivere il comportamento dei vari tipi d'atomi e di legami. Nel corso degli anni sono stati sviluppati vari tipi di force field.

Alcuni di essi includono, oltre a quelli sopra riportati, dei termini addizionali, idonei a descrivere ulteriori tipi di interazioni. Altri force field, per migliorare l'accuratezza del modello, giustificano l'unificazione dei termini di stretching e di bending introducendo i cosiddetti "cross terms". La formulazione matematica dei termini dell'equazione energetica varia secondo il force field utilizzato. I termini più comuni sono i seguenti:

*Energia di stretching*

$$E = \sum_{\text{bonds}} k_b (r - r_0)^2$$

L'equazione dell'energia di stretching è basata sulla legge di Hook.  $k_b$  è il parametro che determina la costante di deformazione lineare della molla, rappresentante il legame atomico, mentre  $r_0$  ne costituisce la lunghezza di equilibrio. Per ciascun tipo di legame sono assegnati specifici valori di  $k_b$  e di  $r_0$ , secondo la natura del legame e degli atomi che lo formano (e.g. C-C, C-S, C=O).

L'equazione sopra riportata calcola l'energia associata alle vibrazioni attorno alla lunghezza d'equilibrio del legame, e conseguentemente, il modello tende a perdere la sua validità quando i legami si allungano verso i valori di dissociazione.

### Energia di bending

$$E = \underset{\text{angles}}{k_\theta} (\theta - \theta_0)^2$$

L'energia di bending è a sua volta basata sulla legge di Hook. Il parametro  $K_\theta$  regola la rigidità della deformazione angolare della molla, mentre  $\theta_0$  definisce il valore di equilibrio dell'angolo di legame. Anche in questo caso viene calcolata l'energia associata a piccole variazioni attorno all'angolo di equilibrio, e  $K_\theta$  e  $\theta_0$  assumono valori diversi a seconda degli atomi e della natura dei legami che formano un determinato angolo (O=C=O, C-O-C).

### Energia torsionale

$$E = \underset{\text{angles}}{k_\theta} (\theta - \theta_0)^2$$

In MM l'energia torsionale ha l'andamento di una funzione periodica e rappresenta la quantità di energia che deve essere algebricamente sommata alle energie di stretching, di bending e di non-interazione, affinché l'energia totale così calcolata sia in accordo con i valori determinati sperimentalmente o ottenuti da calcoli quanto-meccanici.  $A$  è il parametro che controlla l'ampiezza della curva,  $n$  ne determina la periodicità, mentre  $\theta$  trasla l'intera curva lungo l'asse di rotazione  $\tau$ . Per ciascun angolo diedro sono definiti valori diversi dei parametri secondo gli atomi e dei legami.

### Interazioni di non legame

$$E = \left( \underset{i \quad j}{- \frac{A_{ij}}{r_{ij}^6} + \frac{B_{ij}}{r_{ij}^{12}}} \right) + \left( \underset{i \quad j}{\frac{q_i q_j}{\epsilon r_{ij}}} \right)$$

(3)
(4)

Secondo la terminologia della MM, sono definiti atomi non-legati quelli tra loro separati da almeno due atomi legati covalentemente. L'energia di non legame dipende dalle interazioni fra tutte le possibili coppie di atomi non-legati  $i$  e  $j$ . Tali interazioni di

non-legame annoverano le forze di Van der Waals (descritte dal potenziale di Lennard-Jones o potenziale 6-12) e le forze di attrazione repulsione elettrostatica (definite dal potenziale coulombico).

L'equazione di Lennard-Jones (3) si articola in due componenti (attrattiva e repulsiva) regolate rispettivamente da funzioni inversamente proporzionali a  $r_6$  e a  $r_{12}$ , e dai parametri A e B. Il valore del parametro A "attrattivo" può essere ricavato da misure di polarizzabilità atomica, oppure mediante tecniche di meccanica quantistica. Il parametro B "repulsivo" è invece derivato da studi di diffrazione ai raggi X per riprodurre le distanze medie di

contatto fra diversi tipi di atomo nelle strutture cristalline di varie molecole. Le interazioni elettrostatiche sono descritte dal potenziale di Coulomb (4) e dipendono dalla carica atomica parziale presente sugli atomi non-legati, dalla loro distanza, e dalla natura del dielettrico interposto (che può essere sia un solvente sia una porzione della stessa molecola). Spesso al dielettrico si assegna un valore costante compreso tra 1 e 5 mentre valori linearmente dipendenti dalla distanza ( $\epsilon = r^{-1}$ ) vengono talvolta usati per simulare gli effetti dovuti al solvente quando esso è trattato implicitamente.

## MINIMIZZAZIONE DELL' ENERGIA

I metodi di minimizzazione energetica possono determinare con precisione la conformazione di minima energia di una molecola (minimo globale) localizzando, una alla volta, i minimi (locali) della funzione, che ne descrive il potenziale energetico. Obiettivo delle minimizzazioni è l'individuazione di un percorso che, attraverso la variazione dei gradi di libertà molecolari, conduca, con il minor numero possibile di calcoli, dalla conformazione iniziale al minimo locale più vicino. Tutti i gradi di libertà (al più  $3N - 6$  per una molecola ad N atomi) vengono di norma esaminati ma, per ridurre la mole dei calcoli, la maggior parte dei programmi consente di limitare i movimenti (tethering) di singoli atomi o di gruppi di atomi, oppure di bloccarli in posizione fissa introducendo i cosiddetti *restrains*.

Gli algoritmi di minimizzazione generalmente localizzano il minimo più vicino alla struttura di partenza, mentre i minimi lontani, separati da barriere energetiche dalla geometria iniziale, sono di difficile individuazione poiché la loro ricerca implica un aumento del gradiente della funzione energia. Conformazioni iniziali differenti possono quindi condurre ad un unico minimo od individuare minimi energetici distinti. Un minimo energetico può essere caratterizzato da una piccola variazione, tra due successivi passi di minimizzazione, del valore della funzione energia e/o da un suo gradiente nullo. Il gradiente è il miglior indice di ottimizzazione completa (convergenza), in quanto ridottissimi cambiamenti nel suo valore sono assai più significativi dei piccoli decrementi della funzione energia.

## ALGORITMI DI MINIMIZZAZIONE

Dal punto di vista matematico, le procedure di minimizzazione dell'energia di una molecola ricadono nell'area generale dei problemi d'ottimizzazione non lineare. Dato un insieme di variabili indipendenti  $\mathbf{x} = (x_1, x_2, x_3, \dots, x_n)$  ed una funzione  $V = V(\mathbf{x})$ , si tratta di trovare un sottoinsieme di valori di  $\mathbf{x}$ , denominato  $\mathbf{x}^*$ , per cui la funzione assuma il valore minimo:  
$$\min V(\mathbf{x}) = V(\mathbf{x}^*).$$

Intuitivamente il modo più semplice per localizzare il minimo di  $V(\mathbf{x})$  è compiere una scansione dello spazio vettoriale definito da  $\mathbf{x}$  con incrementi regolari di  $x_1, x_2, x_3, \dots, x_n$ . Concettualmente questo corrisponde a tracciare un reticolo sulla superficie della funzione di energia potenziale della molecola e ricercare quel punto della griglia per cui il potenziale ha il valore minimo. Si tratta di un metodo sostanzialmente simile all'analisi conformazionale sistematica, che si rivela però inapplicabile quando la superficie potenziale presenta molti minimi o quando vi sono troppi gradi di libertà. Per ovviare a questi inconvenienti, si ricorre a metodi di minimizzazione denominati di *primo* o di *secondo* ordine, con riferimento alla derivata da essi calcolata. Secondo questa classificazione, il metodo della griglia può essere definito d'ordine zero. I metodi che utilizzano il calcolo della derivata seconda della funzione di energia potenziale molecolare sono denominati di **Newton-Raphson (NR)**.

Essi si fondano su una proprietà matematica per la quale, in prossimità di un minimo, la funzione energia dipende in modo quadratico dalle variabili indipendenti. Nel caso di un problema monodimensionale, nei pressi di un minimo  $V(\mathbf{x})$  assumerà il valore di:

$$(1) V(x) = a + bx + cx^2$$

dove  $a, b, e c$  sono delle costanti. Le prime due derivate saranno rispettivamente:

$$(2) V' = b + 2cx$$

$$(3) V'' = 2c$$

In corrispondenza del minimo,  $V(x) = 0$ . Indicando l'ascissa del punto di minimo con  $x^*$ , dalla (2) si ricava che

$$(4) \quad x^* = -\frac{b}{2c} = -\frac{b}{V''(x)}$$

e quindi

$$(5) \quad V'(x) = b + V''(x)x$$

$$(6) \quad \frac{V'(x) - b}{V''(x)} = x$$

$$(7) \quad \frac{V'(x) - b}{V''(x)} - \frac{b}{V''(x)} = x$$

ricordando la (4) si ha che

$$(8) \quad \frac{V'(x) - b}{V''(x)} - x^* = x$$

da cui

$$(9) \quad x^* = x - \frac{V'(x)}{V''(x)}$$

Se ne deduce che, se il sistema si trova al punto  $x$  e sono note la sua derivata prima e seconda, la posizione del minimo può essere calcolata senza necessità di iterazioni. Tale metodica è però applicabile solo a piccole molecole, poiché richiede l'inversione di matrici ( $3N \times 3N$ ), dove  $N$  è il numero degli atomi.

I metodi del primo ordine, invece, trascurano il calcolo della derivata seconda e vengono classificati tra le *tecniche di discesa* (**Steepest Descent o SD**). Questi metodi prevedono il calcolo della derivata prima della funzione di energia potenziale, ovvero del calcolo della forza che agisce su ciascun atomo della molecola. Tale forza serve per determinare la direzione lungo la quale gli atomi saranno mossi di un fattore  $k$  secondo la formula

$$x_k = x_{k-1} \cdot \lambda \cdot s_k$$

dove  $s_k$  rappresenta un versore parallelo alla forza:  $\lambda$  assume un valore di 1.2 se l'iterazione  $k$ -esima porta alla diminuzione dell'energia globale della molecola, assume invece il valore di 0.5 nel caso l'energia aumenti. Una migliore efficienza dell'algoritmo può essere ottenuta *coniugando i gradienti* (**conjugate gradients o CG**) dell'energia, calcolati nelle interazioni  $k$  e  $(k-1)$ -esime. In questo caso la direzione da seguire è calcolata come:

$$s_k = \left( \frac{|s_k|}{|s_{k-1}|} \right)^2$$

Si è, infatti, dimostrato che, per superfici quadratiche, la direzione di ricerca così calcolata è molto più accurata di quella calcolata dal semplice gradiente. La scelta di un particolare algoritmo di minimizzazione dipende sia dalla complessità della molecola oggetto di studio, sia dall'accuratezza richiesta per le soluzioni (tabella 1.1). La procedura più usata consiste nell'iniziare la minimizzazione con un algoritmo relativamente poco preciso, ma efficace lontano dal minimo e nel sostituirlo con metodi più raffinati man mano che la diminuzione del gradiente energetico segnala l'approssimarsi della geometria di energia minima.

**Tabella 1.1: applicabilità dei metodi di minimizzazione**

	n. di atomi < 200		n. di atomi > 200	
	Lontano dal minimo	Vicino dal minimo	Lontano dal minimo	Vicino dal minimo
<i>Steepest Descent</i>	✓		✓	
<i>Conjugate Gradients</i>	✓	✓	✓	✓
<i>Newton-Raphson</i>		✓		

## ANALISI CONFORMAZIONALE

Sia le piccole molecole sufficientemente flessibili che, a maggior ragione, le macromolecole, costituiscono dei sistemi in grado di avere, alle normali condizioni di temperatura e pressione, vari isomeri conformazionali. Di conseguenza, se non si hanno precise informazioni sulla conformazione di una molecola, è necessario studiarne i vari conformeri in relazione all'energia posseduta.

Si definisce spazio conformazionale molecolare l'insieme di tutte le possibili conformazioni assumibili da una molecola. A condizioni normali, solo le conformazioni a bassa energia sono popolate e quando queste non sono preventivamente conosciute, si rende necessario esplorarle tutte o, quantomeno, studiarne un numero significativamente elevato da descrivere lo spazio conformazionale.

Nella maggior parte dei casi questo significa esaminare in modo quasi dettagliato, le possibili combinazioni degli angoli diedri che caratterizzano la geometria della molecola. Indipendentemente dalla tecnica adottata, tre sono i passi in cui si articola l'analisi conformazionale:

- 1) generazione di un insieme di geometrie iniziali che descrivano adeguatamente lo spazio conformazionale molecolare;
- 2) ottimizzazione delle strutture raccolte;
- 3) eliminazione, attraverso il calcolo degli RMS, dei duplicati energetici in modo da ottenere un insieme di strutture uniche e corrispondenti ai minimi della funzione energia potenziale.

Tra i metodi maggiormente impiegati nell'analisi conformazionale vanno ricordati:

- ☐= la ricerca sistematica (*grid search*);
- ☐= la dinamica ad alta temperatura seguita da raffreddamento ed equilibratura a 0 °K (*simulated annealing*);
- ☐= il campionamento conformazionale stocastico secondo il metodo di Monte Carlo.

La scelta di una particolare tecnica d'analisi conformazionale dipende dalla complessità del problema che si deve affrontare. In particolare *simulated annealing* e *metodo di Monte Carlo* vanno presi in considerazione quando i gradi di libertà sono molto numerosi o eccedono le capacità del campionamento sistematico.

## CAMPIONAMENTO SISTEMATICO

In linea di principio, il campionamento sistematico rappresenta il metodo di esplorazione più diretto. Le energie vengono campionate secondo i valori assunti, per regolari incrementi, da ciascuno dei gradi di libertà molecolare. Ad esempio, dovendo esplorare lo spazio conformazionale generato dalle rotazioni attorno a due angoli diedri ( $\alpha$  e  $\beta$ ), dapprima viene fissato, per ciascuno dei due angoli, l'intervallo di rotazione permesso ed il relativo passo di rotazione. Poi, al variare di  $\alpha$ , si fanno assumere a  $\beta$  tutti i valori consentiti. Si realizza in questo modo un *albero conformazionale* in cui ogni valore di  $\alpha$  è un nodo sul quale si innestano tutti i valori di  $\beta$ . Il numero complessivo di conformazioni da campionare è dato dalla relazione:

$$\prod_{i=1}^n \frac{R_i}{\delta_i}$$

dove  $n$  è il numero degli angoli diedri analizzati,  $R_i$  l'intervallo di rotazione ammesso e  $\delta_i$  il passo di rotazione.

Il principale problema del campionamento sistematico è che il numero delle conformazioni necessarie ad una descrizione adeguata dello spazio conformazionale molecolare diventa estremamente elevato per  $n > 5$  (con 6 angoli diedri ed un passo di  $30^\circ$  si dovrebbero raccogliere quasi  $3 \times 10^6$  strutture). Prima di optare per il campionamento sistematico è quindi necessario valutare con attenzione il numero degli atomi e dei legami attorno ai quali far avvenire le rotazioni, e fare una stima preventiva dei conformeri che saranno campionati. Per ridurre l'entità delle conformazioni da raccogliere sono disponibili varie strategie:

- 1) Limitare, in particolare per i sostituenti simmetrici, gli intervalli di rotazione al minimo indispensabile. Ad esempio, per un gruppo fenilico è sufficiente un valore compreso tra  $0^\circ$  e  $180^\circ$  anziché un intero angolo giro di  $360^\circ$ .
- 2) Aumentare il passo di rotazione: per un legame semplice C-C basta un passo di  $120^\circ$ . Va però tenuto presente che l'accuratezza di un'analisi conformazionale sistematica dipende pesantemente dalla finezza del passo di rotazione adottato.
- 3) Bloccare o non trattare del tutto quei gruppi la cui rotazione non fornisce informazioni utili. È il caso, ad esempio, dei gruppi metilici.
- 4) Suddividere la molecola in un certo numero di frammenti da analizzare poi separatamente.

Un altro approccio, per ridurre la mole dei dati da raccogliere, consiste nell'adottare degli opportuni filtri.



□ Il più semplice tipo di filtro è l'introduzione di un valore di soglia per l'energia (cutoff) in modo da scartare automaticamente tutti quei conformeri la cui energia superi il valore di cutoff. Se da un lato questo filtro riduce il numero delle conformazioni raccolte, dall'altro non diminuisce affatto quello dei calcoli che è necessario compiere perché, prima di decidere se un conformero debba o meno essere accettato, bisogna calcolarne l'energia. Questo tipo di selezione basata sull'energia fornisce i risultati migliori se è applicata al termine della ricerca conformazionale, quando cioè siano già note le conformazioni d'energia massima e minima in modo da poter stabilire un cutoff relativo.

□ La cosiddetta potatura dell'albero conformazionale è una tecnica di filtraggio più sofisticata, che si basa sul fatto che se una conformazione è ad alta energia lo sono pure tutte quelle che da essa derivano. Una volta individuato un nodo ad elevata energia non è più necessario proseguire lungo quel ramo della ricerca e si possono così ridurre sia i calcoli sia il numero delle conformazioni da raccogliere.

□ Si possono infine imporre delle restrizioni strutturali (constrains) decidendo di accettare esclusivamente quelle geometrie che soddisfano a determinati requisiti, quali una certa lunghezza o un dato angolo di legame.

## DINAMICA MOLECOLARE

La dinamica molecolare (Molecular Dynamics o MD) studia, nella sua accezione più ampia, qualsiasi tipo di movimento a livello molecolare.

Per simulare il moto degli atomi la MD applica la legge di Newton

$$(1) \quad \frac{\overline{F}}{m} = \overline{a} = \frac{d\overline{v}}{dt} = \frac{d^2\overline{r}}{dt^2}$$

L'azione fondamentale della MD è il calcolo della risultante delle forze di interazione reciproca che agiscono su ciascun atomo  $i$ -esimo, in quanto, conoscendo anche la massa atomica, è possibile ricavare il modulo e la direzione della velocità che lo caratterizza. Tale risultante può essere a sua volta ottenuta dalla variazione dell'energia misurata nel passaggio dell'atomo tra due successive posizioni.

$$(2) \quad -\frac{dE}{dr_i} = \overline{F}_i$$

Confrontando le equazioni (1) e (2) è possibile esprimere in altro modo l'equazione di Newton

$$(3) \quad -\frac{dE}{dr_i} = m_i \frac{d^2 \bar{r}_i}{dt^2}$$

Dalla variazione dell'energia e dalla massa atomica è quindi possibile conoscere in ogni istante la posizione occupata nel tempo da ciascun atomo. Si ottiene così una successione di *frames* che fotografano, di norma ogni  $10^{-12}$  secondi, i cambiamenti della geometria molecolare. L'insieme di queste "istantanee" è detto traiettoria molecolare.

In pratica, però, per mancanza di soluzioni analitiche, non è possibile risolvere direttamente l'equazione (3) ed è necessario procedere per tappe. Ricavata la accelerazione dalla variazione di energia e dalla massa atomica è possibile risalire alla velocità integrando l'equazione

$$(4) \quad \bar{a}_i = \frac{d\bar{v}_i}{dt}$$

e da questa è infine possibile ottenere la posizione occupata dal singolo atomo.

$$(5) \quad \bar{v}_i = \frac{d\bar{r}_i}{dt}$$

La dinamica molecolare trova applicazione sostanzialmente in due settori principali:

□ l'analisi conformazionale realizzata attraverso dinamiche ad alta temperatura (900 o 1000 °K);†

□ lo studio delle interazioni tra molecole diverse o fra differenti regioni di una macromolecola, condotto a temperatura ambiente (300 °K), per simulare l'energia cinetica propria dei sistemi biologici alle normali condizioni sperimentali. Il principale vantaggio dell'analisi conformazionale ad alta temperatura risiede nella sua applicabilità a sistemi comprendenti un elevato numero di atomi (proteine ed acidi nucleici), per i quali non sono utilizzabili né l'analisi sequenziale né il metodo di Monte Carlo. Poiché l'elevata temperatura alla quale la simulazione viene condotta si traduce in una notevole energia cinetica a livello dei singoli atomi, questo consente l'accesso virtuale a tutte le regioni dello spazio conformazionale molecolare mediante il superamento di barriere torsionali altrimenti invalicabili. Va però osservato che l'energia cinetica potrebbe essere talmente elevata da portare a situazioni indesiderate, quali, ad esempio, le inversioni di chiralità o la trasformazione di domini proteici ad  $\alpha$  elica in regioni  $\beta$  sheet. Per ovviare al problema è possibile agire all'origine della simulazione, introducendo preventivamente delle opportune restrizioni per la conservazione della chiralità, o a valle, rigettando tutte quelle strutture

---

† Poiché le conformazioni campionate ad alta temperatura vengono successivamente minimizzate e le minimizzazioni avvengono a 0 °K l'analisi conformazionale ad alta temperatura viene sovente indicata come *simulated annealing*.

che siano palesemente incompatibili con i dati strutturali noti o con le ipotesi di lavoro formulate. Una dinamica ad alta temperatura è più impegnativa dal punto di vista computazionale di una dinamica a temperatura ambiente, giacché richiede un controllo estremamente attento e rigoroso della temperatura e dell'energia cinetica, per evitare che i legami atomici (soprattutto quelli che vedono coinvolti gli idrogeni) si allunghino a dismisura determinando il cosiddetto "blow up" molecolare. Tale controllo è realizzato tramite una ridistribuzione a frequenza estremamente ravvicinata (ogni pochi femtosecondi) dell'energia cinetica tra tutti gli atomi della molecola. Sia le dinamiche ad alta temperatura che quelle a temperatura ambiente si articolano in due fasi, dette d'equilibratura termica e di simulazione. La fase di simulazione è generalmente condotta o a temperatura costante o ad energia cinetica costante. Durante la fase di equilibratura una struttura precedentemente minimizzata, e quindi a 0 °K, viene portata per tappe successive alla temperatura richiesta per la simulazione assegnando a ciascun atomo una certa velocità iniziale (secondo una distribuzione di Gauss o di Boltzmann), in modo che la somma delle loro singole energie cinetiche corrisponda all'energia totale del sistema a quella temperatura. Nella fase di simulazione le energie cinetiche, e quindi le velocità possedute dai singoli atomi, determinano la variazione delle loro posizioni relative modificando le lunghezze di legame, le ampiezze degli angoli di legame e degli angoli diedri e le interazioni di non legame. L'energia cinetica viene così parzialmente convertita in energia potenziale molecolare. L'evoluzione del sistema nel tempo deriva da questo continuo equilibrio dinamico fra velocità e forze di richiamo esercitate dalle "molle" della meccanica molecolare. Questo processo viene, per così dire, fotografato ad intervalli regolari attraverso le strutture campionate che vengono poi successivamente minimizzate, onde associare ad ogni struttura un preciso contenuto energetico.

## METODO DI MONTE CARLO

Il metodo detto di Monte Carlo è una tecnica d'esplorazione stocastica dello spazio conformazionale molecolare. I conformeri sono generati applicando delle variazioni casuali o pseudo-casuali alla struttura della molecola, per descrivere in modo statisticamente significativo l'insieme delle conformazioni assumibili. Questa tecnica si basa sull'assunto (euristico) che i conformeri a bassa energia di una molecola siano tra loro più strettamente correlati di quanto non lo siano quelli ad alta energia. Per campionare le popolazioni di conformeri localizzate nelle regioni a bassa energia dello spazio conformazionale è quindi sufficiente studiare la variazione di un numero limitato di gradi di libertà. Poiché le distanze e gli angoli di legame possono avere soltanto limitate deformazioni, sono le variazioni degli angoli diedri che principalmente differenziano i vari conformeri. Indicando con  $R$  il numero di conformazioni che si desidera campionare per ciascun grado di libertà, per una molecola con  $N$  atomi (e  $3N-6$  gradi di libertà) si dovrebbero determinare  $R^{3N-6}$  strutture. Se invece, dal punto di vista conformazionale, sono importanti solo  $N'$  angoli diedri indipendenti, allora il numero di conformeri da determinare è pari a  $R^{N'}$ . L'analisi conformazionale secondo il metodo Monte Carlo è generalmente implementata come una procedura iterativa che si articola in vari passi:

a) *Scelta della geometria iniziale*: la scelta della struttura di partenza rappresenta il passaggio più delicato, giacché influenza direttamente l'efficacia complessiva del metodo nell'esplorare le diverse regioni dello spazio conformazionale. I criteri di scelta sono sostanzialmente due:

1. Mediante la procedura detta *Random Walk* (RW) si sceglie come struttura di partenza l'ultima struttura salvata, purché la sua energia non sia troppo lontana da quella dell'attuale minimo globale. In altre parole: un passaggio che parte dalla struttura  $m$  conduce alla struttura  $m+1$ . Se l'energia della struttura  $m+1$  è entro, ad es., 10 kJ/mol dal minimo assoluto temporaneo allora viene utilizzata come struttura di partenza per il passaggio successivo. In caso contrario per giungere alla struttura  $m+2$  si riparte dalla struttura  $m$ .

2. Con il criterio denominato *Usage Directed* (UD) si opera invece nel modo seguente: dopo il passaggio che dalla struttura  $m$  porta alla struttura  $m+1$ , tra le  $m+1$  strutture già raccolte si isola il sottoinsieme delle  $M'$  conformazioni che distano meno di 10 kJ/mol dal minimo assoluto temporaneo. Nell'ambito di  $M'$  si sceglie, quale struttura di partenza per un nuovo ciclo, quella che è stata usata per questo scopo il minor numero di volte. In caso di parità si preferisce per quella ad energia più bassa. Attualmente si tende a privilegiare il criterio UD poiché con RW si ha la tendenza a ritrovare più volte le medesime strutture. E' stato inoltre dimostrato che UD aumenta l'efficienza del metodo Monte Carlo nel localizzare un maggior numero di conformeri a bassa energia.

b) *variazione casuale delle coordinate molecolari*: consiste nell'applicare una traslazione casuale in un intervallo predeterminato (0.5÷3.5 Å), ad uno o più atomi della molecola e/o nel variare, sempre casualmente e in un dato intervallo, uno o più angoli diedri. Talvolta, soprattutto per molecole asimmetriche, viene variato, in modo aleatorio, anche il numero degli angoli diedri di volta in volta modificati.

c) *ottimizzazione*: la struttura generata al passo precedente è minimizzata (usualmente in meccanica molecolare) ed eliminata qualora la sua energia non rientri in una finestra di ampiezza prefissata e riferita al minimo globale temporaneo.

d) *comparazione con le strutture precedentemente raccolte*: la nuova struttura è sovrapposta a tutte le altre precedentemente raccolte per stabilirne, mediante RMS, l'unicità. In caso positivo essa viene salvata in quanto rappresenta effettivamente un nuovo minimo energetico, e rigettata come duplicato in caso contrario. Va infatti ricordato che con la minimizzazione si può giungere allo stesso minimo anche partendo da strutture tra loro diverse. La procedura ha termine una volta raggiunto il numero stabilito di conformazioni salvate oppure quando non si trovano più nuovi minimi, bensì solo duplicati. Prima di avviare una simulazione di Monte Carlo è quindi necessario stabilire il numero massimo di conformazioni di minima energia da campionare ed entro quale intervallo energetico dal minimo globale debbano cadere.

## **INTERAZIONE LEGANTE-RECETTORE: MOLECULAR DOCKING**

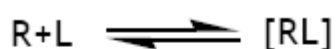
L'applicazione di metodologie di docking molecolare ha come obiettivo quello di poter predire la struttura, o le strutture, con la quale due o più molecole interagiscono nel formare complessi intermolecolari. Questa metodologia trova un'importante applicazione quando non siano disponibili informazioni strutturali, quali per esempio la struttura cristallografica od ottenuta per risonanza magnetica nucleare, del complesso intermolecolare d'interesse. In particolare, questa tecnica computazionale è spesso utilizzata nel predire se, e come, un legante possa interagire a livello molecolare con il suo bersaglio biologico, quale un enzima, un recettore od un segmento d'acido nucleico, di cui

non disponiamo informazioni strutturali sufficientemente accurate. Molti algoritmi di docking molecolare sono in grado di generare un gran numero di possibili strutture e sono in oltre in grado di fornire un criterio, o *score*, basato sull'energia di interazione legante-recettore, con il quale è possibile stabilire quale potrebbe essere il complesso più stabile. Lo scopo di un algoritmo di docking molecolare è quindi quello di generare e valutare in termini energetici possibili strutture di complessi intermolecolari. Il primo problema computazionale importante è che, nel processo di docking molecolare, devono essere presi in considerazione un elevato numero di gradi di libertà conformazionali. Infatti considerando la formazione di un complesso intermolecolare di tipo 1:1, ci sono sei gradi di libertà traslazionali e rotazionali di ogni molecola rispetto all'altra, oltre ai gradi di libertà vibrazionali di ogni singola molecola ( $3N-6$ ). Diversi algoritmi sono stati sviluppati per risolvere in maniera interamente automatizzata il problema del docking molecolare, e questi algoritmi possono essere classificati in funzione del numero dei gradi di libertà conformazionali che ignorano durante il processo di generazione dei diversi complessi intermolecolari. La prima fase di un qualsiasi protocollo di docking molecolare prevede di individuare la regione di riconoscimento tra le due specie interagenti. La zona di riconoscimento può essere localizzata a livello della superficie molecolare delle specie interagenti (come sovente succede nella formazione di complessi intermolecolari di sistemi macromolecolari quali proteina-proteina, proteina-DNA); oppure, come nel caso di riconoscimento di piccole molecole organiche e sistemi macromolecolari, può essere localizzata in speciali cavità (chiamate anche sito attivo o di riconoscimento), che vengono a formarsi dal ripiegamento della superficie della struttura macromolecolare, e all'interno delle quali il legante può essere ospitato esplicando a sua volta l'attività che gli compete. Se la zona di riconoscimento non fosse nota a priori, esistono alcuni algoritmi matematici che sono in grado di analizzare la struttura superficiale di una macromolecola, evidenziando tutti i possibili ripiegamenti della sua superficie (*site search analysis*) che potrebbero rappresentare, almeno in prima analisi, possibili siti di riconoscimento. La geometria con la quale il legante verrà a collocarsi all'interno della cavità sarà funzione, sia dei suoi requisiti sterici sia della natura delle interazioni elettrostatiche che esso instaurerà all'interno della cavità. Prima di iniziare il protocollo di docking molecolare il legante deve essere collocato in prossimità della zona di riconoscimento. Questa situazione rappresenterà la geometria di partenza per la generazione automatizzata dei possibili complessi intermolecolari. Il più semplice algoritmo utilizzabile considera le due molecole interagenti come dei corpi rigidi, e di conseguenza, sono presi in considerazione esclusivamente sei gradi di libertà traslazionali e rotazionali di una molecola (solitamente quella di dimensioni inferiori ovvero il legante) rispetto all'altra che è considerata conformazionalmente fissa nello spazio durante la generazione dei vari complessi intermolecolari. Questi algoritmi prendono il nome di *docking molecolare rigido*. In quest'approccio la scelta della conformazione del legante è cruciale. Infatti, essendo il legante considerato come corpo rigido, si deve immaginare, prima d'iniziare il protocollo di docking, in quale conformazione il legante andrà a riconoscere l'altra molecola (anch'essa in ogni caso rigida) nel complesso intermolecolare.

Per effettuare un *docking molecolare conformazionalmente flessibile*, tutti i gradi di libertà devono essere presi in considerazione durante la fase di generazione dei diversi complessi intermolecolari. Attualmente, molti degli algoritmi utilizzati prevedono l'esplorazione dello spazio conformazionale del legante mantenendo la struttura del recettore rigorosamente rigida nello spazio. Quest'approccio è noto come *docking molecolare semirigido*. Tutti i comuni metodi d'analisi conformazionale sono stati inclusi nei diversi algoritmi semirigidi. In particolare, i metodi Monte Carlo, spesso in combinazione con la tecnica di *simulated annealing*, trovano particolare impiego nell'esplorazione dello spazio conformazionale del legante durante il protocollo di docking

semirigido. Ad ogni interazione della procedura Monte Carlo la conformazione propria del legante è variata (solitamente per variazioni dei valori degli angoli diedri) e l'intera molecola è poi casualmente traslata, e/o ruotata, all'interno della cavità di riconoscimento. Come già anticipato, una volta generate le possibili conformazioni con le quali il legante può allocarsi all'interno della cavità di riconoscimento, abbiamo bisogno di un criterio guida, o *score*, che ci permetta di stabilire quale di queste sia energeticamente più stabile e quindi possa essere candidata a rappresentare in maniera più accurata la reale situazione di legame tra le due specie interagenti. Esistono due metodi principali che possono essere utilizzati per ottenere lo score durante il protocollo di docking molecolare:

1. ENERGY SCORE: l'energia del legante all'interno della cavità può essere calcolata utilizzando un campo di forze di meccanica molecolare, e quindi ricavare l'energia d'interazione mediante la seguente relazione:



$$\Delta E = E_{[RL]} - (E_L + E_R)$$

I valori delle energie d'interazioni sono solitamente negativi ad indicare una stabilizzazione del complesso intermolecolare rispetto alle due specie isolate, e il valore assoluto indica il grado di stabilizzazione del sistema. Questi valori non hanno significato termodinamico, derivando, come già detto, da un protocollo di meccanica molecolare, anche se in alcuni casi possono essere correlati con il valore termodinamico delle costanti di formazione del complesso.

2. GRID SCORE: questo metodo si basa sul calcolo dell'energia d'interazione elettrostatica tra il legante e la regione di riconoscimento. Per il calcolo del potenziale elettrostatico della regione di riconoscimento, si racchiude la cavità all'interno di una griglia tridimensionale a maglie regolari (GRID). Una sonda chimica, quale per esempio un protone od una molecola d'acqua, è fatta scorrere attraverso tutti i vertici della griglia tridimensionale, e l'energia d'interazione tra la sonda e la proteina che costituisce per esempio la cavità è calcolata utilizzando una funzione empirica d'energia. Il risultato è una griglia tridimensionale con un valore d'energia ad ogni vertice. A questo punto è possibile calcolare l'energia d'interazione tra la griglia tridimensionale, che da questo momento costituisce da un punto di vista energetico la nostra cavità, e il potenziale elettrostatico generato dalle cariche atomiche del legante all'interno della griglia.

### **Metodo per il calcolo approssimato delle costanti di binding di complessi proteina-legante (SCORE)**

Come anticipato nei paragrafi precedenti il disporre della struttura tridimensionale di una proteina d'interesse biologico, sia essa determinata mediante tecnica di diffrazione dei raggi X oppure mediante tecniche di spettroscopia di risonanza magnetica nucleare, assume sempre di più un ruolo cruciale nella progettazione razionale di nuovi farmaci. Dalla struttura tridimensionale è, infatti, possibile analizzare nel dettaglio la topologia della nostra proteina e quindi estrapolare i possibili requisiti strutturali essenziali per il riconoscimento di potenziali leganti. Questo processo rappresenta la prima fase della

tecnica per la progettazione di nuovi farmaci conosciuta come *structure-based drug design*. A seguito dello studio della struttura proteica e dell'identificazione delle possibili interazioni efficaci per il riconoscimento del legante, nuove strutture possono essere proposte per uno screening di attività utilizzando sia la ricerca su banche date di strutture tridimensionali che rispondano ai diversi prerequisiti sterici ed elettronici imposti dal sito di riconoscimento della proteina, oppure mediante tecniche di costruzione molecolare chiamate *de novo design* che verrà descritta in dettaglio di seguito. Ognuna di queste due metodologie sarà in grado di proporre una serie di nuove strutture da sottoporre alla successiva fase di sintesi prima, e della valutazione della loro attività biologica poi. Ovviamente il successo nell'identificare nuovi potenziali farmaci sarà strettamente correlato alla capacità di predire in maniera quantitativa le costanti di affinità dei diversi complessi proteina-legante. In questo modo solo quei composti che presentano una costante di binding accettabile saranno diretti alla successiva fase di sintesi. La predizione delle costanti di affinità di un legante nei confronti di un suo target proteico è uno degli argomenti di ricerca biocomputazionale maggiormente interessanti. Da un punto di vista del rigore matematico, il metodo della "perturbazione dell'energia libera" (*free energy perturbation*, *FEB*) rappresenta ad oggi l'unico approccio che affronta in maniera termodinamicamente pertinente il calcolo della previsione delle costanti di affinità di un complesso proteina/legante. Questo metodo prevede di calcolare la variazione di energia libera in seguito alla formazione del complesso binario usando metodi di dinamica molecolare. Può essere dimostrato che la variazione di energia libera causata di piccole variazioni (perturbazioni) dello stato iniziale di un sistema è uguale alla sua variazione entalpia quando si considera la media di molti stati del sistema stesso:

$$\Delta G = -RT \ln \left\langle \exp \left( -\frac{H_{AB}}{RT} \right) \right\rangle_A$$

Dove  $\langle \rangle_A$  rappresenta la media rispetto ai diversi stati. La conseguenza importante di quest'approccio è che risulta matematicamente possibile calcolare la variazione di energia libera del processo di formazione del complesso binario senza esplicitamente calcolarne la corrispondente variazione entropica, ma solo calcolando ripetutamente la variazione entalpia su un numero molto elevato di stati. Nonostante l'incredibile incremento della velocità di calcolo dei nuovi computers, questo metodo rimane computazionalmente costoso ed applicabile al momento a sistemi molecolari di piccole/medie dimensioni. Questo approccio quindi è difficilmente applicabile in tandem con i metodi di drug design che prevedono un veloce feedback tra il chimico di sintesi ed il modellista.

Recentemente, l'approccio empirico per il calcolo delle costanti d'affinità ha riscosso un nuovo interesse. L'assunzione di base che caratterizza quest'approccio è che il contributo totale dell'energia libera di binding possa essere suddivisa in diverse componenti la cui espressione matematica è riportata qui a seguito:

$$\Delta G_{binding} = \Delta G_{vdw} + \Delta G_{metal} + \Delta G_{Hbond} + \Delta G_{desolvation} + \Delta G_{deformation} + \Delta G_0$$

I contributi riportati nell'equazione precedente possono essere così descritti:  $\Delta G_{vdw}$  rappresenta il contributo delle interazioni di van der Waals tra la proteina ed il suo legante;

$\Delta G_{metal}$  il contributo derivante dall'interazione metallo-legante;  $\Delta G_{Hbond}$  il contributo dovuto alla formazione dei legami ad idrogeno;  $\Delta G_{desolvation}$  il contributo della desolvatazione a seguito della formazione del complesso binario;  $\Delta G_{deformation}$  il contributo derivante dalla variazione conformazionale del legante durante il processo di binding;  $\Delta G_0$  rappresenta la costante di regressione e può contenere tutti i contributi entropici legati alla diminuzione dei gradi di libertà rotazionale e traslazionali che occorrono durante il processo di binding. In questo lavoro di tesi è stato utilizzato un nuovo metodo empirico conosciuto come SCORE, sviluppato da Wang et al (J. Mol. Model. 1998, 4, 379-394), che permette di calcolare la costante di affinità assoluta conoscendo la struttura tridimensionale di un complesso proteina-legante.