

Progetto di simulazione molecolare per il corso di “Complementi di algoritmi” A.A. 2004-05

25 novembre 2004

1 Descrizione

Si vogliono applicare tecniche del campo degli algoritmi per modellare l'interazione fra molecole organiche in soluzione acquosa. I dati sono costituiti dalla sequenza di posizioni occupate istante per istante da tutti gli atomi di un dato insieme di molecole organiche e di molecole d'acqua. Derivano da simulazioni basate sulle leggi della Meccanica e dell'Elettrodinamica.

Si dispone quindi, per ogni istante di tempo, di una lista di atomi, per ognuno dei quali è noto:

- un numero d'ordine
- l'elemento chimico
- la molecola di appartenenza (sempre la stessa: non avvengono reazioni chimiche)
- la posizione (l'atomo è modellato come un punto in uno spazio a tre dimensioni)

Particolare interesse assumono i *legami a idrogeno*: ogni atomo di idrogeno del sistema, originariamente legato a un atomo di un altro elemento (detto *donore*) può stabilire un legame con un atomo di un altro elemento chimico (detto *accettore*) appartenente ad una molecola differente. Nei casi più comuni, donore e accettore sono atomi di ossigeno, ma possono essere anche atomi di azoto, zolfo, ecc. . .

Per convenzione, si ritiene che si stabilisca un legame a idrogeno quando

1. la distanza fra donore e accettore è inferiore a una data soglia d_m
2. l'angolo formato dalla terna donore-idrogeno-accettore supera una data soglia α_m

È quindi possibile costruire un (multi)grafo ausiliario non orientato $G(V, E)$, i cui vertici sono le molecole del sistema, mentre i lati sono i legami a idrogeno. Si tratta di un multigrafo perché possono stabilirsi più legami a idrogeno fra le stesse due molecole, e addirittura fra gli stessi due atomi di molecole diverse.

La descrizione puntuale del sistema racchiude l'informazione sulle interazioni fra soluto e solvente in modo troppo implicito. L'ipotesi di fondo del progetto è che la topologia del grafo ausiliario possa invece fornire informazioni più leggibili.

In particolare, sarebbe interessante indagare i seguenti temi:

1. dato un sistema con più molecole di soluto e di solvente, definire il sottografo costituito dai soli legami diretti fra molecole di soluto (trascurando quelli fra soluto e solvente); si sospetta, infatti, che le molecole che tendono a interagire direttamente fra loro quando immerse in soluzione acquosa (ad es., l'urea) abbiano proprietà diverse da quelle che interagiscono soprattutto indirettamente attraverso il solvente (ad es., il TMAO). In particolare, l'urea ha un'azione "distruttiva" nei confronti della struttura delle proteine (non nel senso che ne rompa la catena polipeptidica, ma nel senso che ne altera la forma, e quindi la funzionalità). Al contrario, il TMAO, oltre a non modificare la struttura delle proteine, svolge un'azione "protettiva", contrastando gli effetti dell'urea.
2. qualora le molecole di soluto formino *cluster*, legati da legami a idrogeno, è interessante valutarne le dimensioni (ed eventualmente la struttura): si sospetta che molecole che formano cluster di dimensioni più ampie, con legami multipli fra le stesse molecole, piuttosto che con topologia ciclica, ecc... possano avere effetti macroscopici specifici
3. è interessante valutare se il soluto si distribuisca in modo omogeneo nel solvente o tenda a concentrarsi, se non addirittura a formare aggregati; questo suggerirebbe che alcune molecole (ad es., l'urea) possano avere blande proprietà idrofobiche, sinora non chiaramente riconosciute per via sperimentale
4. dato un sistema con una sola molecola di soluto, è interessante valutare l'estensione della sfera di influenza di tale molecola sul solvente, cioè quanto lunghe siano le catene di molecole di acqua legate direttamente al soluto

5. analogamente, è interessante valutare come cambiano le interazioni fra molecole d'acqua passando da un sistema puramente acquoso a un sistema con una o più molecole di soluto
6. in particolare, i legami a idrogeno fra molecole di acqua creano una struttura fittamente reticolata lungo il sistema, con veri e propri cammini; il fenomeno viene definito *percolation*: l'acqua è in effetti, più che un fluido, un *gel*, cioè un sistema con una struttura che sulla scala temporale dei picosecondi appare stabile, mentre si rompe e ricompone continuamente su scale più lunghe; potrebbe essere interessante (o almeno un pretesto per fare algoritmi) valutare l'entità di questa connessione attraverso algoritmi di massimo flusso
7. riguardo tutti i punti descritti precedentemente, è interessante l'analisi comparata di *framework* successivi, per valutare la durata dei legami a idrogeno e delle strutture che essi inducono

2 Come procedere

Il primo passo necessariamente sarà descrittivo e limitato a un solo *framework*. Dobbiamo trovare strutture dati che riflettano la complessità del problema fisico, ma consentano al tempo stesso di evidenziare i soli aspetti che a noi interessano.

Il punto fondamentale sono i legami a idrogeno. Questi sono legami fra atomi (terne di atomi, di cui due in una molecola e uno in un'altra), ma svolgono il loro ruolo in quanto legami fra molecole. Il modello che testeremo vedrà le molecole come vertici di un (multi)grafo, i legami a idrogeno come lati. Converrà mantenere la semplice rappresentazione a lista di lati, ma anche le informazioni legate all'incidenza fra legami e molecole (quindi la rappresentazione a liste di incidenza). Converrà anche conservare l'indicazione di appartenenza degli atomi alle molecole.

In breve, la prima fase del progetto consiste nel:

1. concepire una struttura dati per il multigrafo
2. caricare da file i dati relativi ai vertici
3. determinare i lati del multigrafo
4. stamparli in uscita per controllarne la validità