

Doctoral Thesis ETH No. 16584

**Biomolecular simulation:
calculation of entropy and free energy,
polypeptide and carbopeptoid folding,
simplification of the force field for lipid simulations.**

A dissertation submitted to the

SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZURICH

for the degree of

Doctor of Sciences

presented by

RICCARDO BARON

Dr. Chem., Università degli Studi di Milano (IT)

born August 25, 1977

citizen of Italy

accepted on the recommendation of

Prof. Dr. Wilfred F. van Gunsteren, examiner

Prof. Dr. Ulrich W. Suter, Dr. Philippe H. Hünenberger, Dr. Dirk Bakowies, co-examiners

2006

Riassunto

Nunc age, quo motu genitalia materiai
 corpora res varias gignant genitasque resolvant
 et qua vi facere id cogantur quaeque sit ollis
 reddita mobilitas magnum per inane meandi,
 expeditam: ti te dictis praebere memento.
 Nam certe non inter se stipata cohaeret
 materies, quoniam minui rem quamque videmus
 et quasi longinquo fluere omnia cernimus aevo
 ex oculisque vetustatem subducere nostris,
 cum tamen incolumnis videatur summa manere
 propterea quia, quae decedunt corpora cuique,
 unde abeunt minuunt, quo venere augmine donant,
 illa senescere at haec contra florescere cogunt,
 nec remorantur ibi. Sic rerum summa novantur
 semper, et inter se mortales mutua vivunt.
 Augescunt aliae gentes, aliae minuuntur,
 inque brevi spatio mutantur saecula animantum
 et quasi cursores vitae lampada tradunt.

Ora, proseguendo, spiegherò con quale moto i corpuscoli / generatori della materia producano le varie cose, e create le dissolvano, / quale forza li costringa a operare ciò / e quale mobilità sia loro data nello spostarsi per il grande vuoto: / tu ricorda di prestare ascolto alle mie parole. / Infatti certamente la materia non si stringe in un unico blocco, / poichè tutti vediamo che ogni corpo si estenua, / e quasi scorgiamo il suo lento disfarsi nel tempo, / e la lunga durata sottrarlo alla nostra visione, / mentre la somma dell'universo rimane invariata, / perché le particelle elementari, infatti, che sfuggono a ciascuno dei corpi, / riducono quello che lasciano, o crescono quello cui approdano, quello fanno invecchiare, questo al contrario fiorire, né indugiano ivi: così l'universo si rinnova senza posa, / e le creature mortali vivono scambievoli vite. Certe specie si accrescono, altre a vicenda declinano, / in breve tempo mutano le stirpi animali, e come staffette si passano la fiaccola della vita.

(Tito Lucrezio Caro, in *De rerum natura*)

Le simulazioni di dinamica molecolare con campi di forza empirici sono un insostituibile strumento computazionale per lo studio, a risoluzione atomica, delle proprietà dei sistemi (bio)molecolari. Queste simulazioni, e i modelli di meccanica classica su cui si basano, sono di fondamentale importanza per la comprensione del comportamento di macromolecole complesse e aiutano a mettere in relazione la loro struttura, dinamica e funzione. Uno dei benefici primari di tali simulazioni è la loro complementarità ai risultati sperimentali, fornendo non soltanto valori medi, ma distribuzioni e serie temporali di qualsivoglia quantità definibile in termini matematici. Inoltre, la possibilità pratica di entrare in possesso di una visione meccanico-statistica dei sistemi studiati offre una chiave interpretativa che è, nella maggior parte dei casi, preclusa agli esperimenti. Il Capitolo 1 presenta, nella prima parte, una concisa introduzione sulla filosofia della simulazione (bio)molecolare. In seguito, sono descritte, con riferimenti ai capitoli seguenti, tre sfide per cui queste simulazioni sono di straordinaria importanza.

Una questione di interesse fondamentale in chimica e biologia è la comprensione del modo in cui le (bio)molecole si avvolgono in strutture tridimensionali relativamente stabili in condizioni fisiologiche. Il Capitolo 2 è uno studio dello stato denaturato di due peptidi costituiti da identiche catene peptidiche, ma da diverse catene laterali. Sperimentalmente è stato proposto che propendano verso un avvolgimento, nel primo caso, a struttura compatta di elica, nel secondo a foglietto β . Le simulazioni chiariscono la questione relativa all'ipotesi che questi due peptidi, caratterizzati da strutture secondarie differenti, possano condividere regioni comuni di parti popolate dei loro spazi configurazionali.

Il confronto tra risultati di simulazione e dati sperimentali è la chiave per giudicare l'attendibilità di simulazioni di dinamica molecolare. In seguito ad un controllo meticoloso, queste simulazioni possono essere impiegate per guidare gli esperimenti o per fornire informazioni preziose con risoluzioni spazio-temporali che agli esperimenti sono precluse. Nel Capitolo 3, questi concetti sono descritti nel caso di una famiglia di carbopeptoidi (peptidi costituiti da aminoacidi a base carboidrata) confrontando le traiettorie di simulazione con dati da esperimenti di risonanza magnetica nucleare. L'applicazione della dinamica molecolare mira alla comprensione del modo in cui differenti proprietà stereochimiche, estensioni delle catene carbopeptidiche e ambienti solvatanti possano influenzare il modo di avvolgersi di queste molecole.

Allo scopo di permettere calcoli più rapidi e di aumentare la capacità di campionamento delle simulazioni di dinamica molecolare, una parte dei gradi di libertà del sistema può essere eliminata dal modello, semplificandone il corrispondente campo di forze. L'importanza di tale rappresentazione, semplificata *a grana grossa*, è sinteticamente descritta nel Capitolo 1 e le applicazioni corrispondenti sono presentate nei capitoli 4, 5 e 6. In particolare, uno studio comparativo tra differenti risoluzioni del modello è il tema del Capitolo 4, che ha come soggetto molecole di idrocarburi liquidi. Il paragone si basa prevalentemente sulla stima di entropie configurazionali, che sono direttamente legate allo spazio configurazionale visitato da queste molecole. Questo permette, tra l'altro, di stimare la perdita di entropia configurazionale conseguente alla semplificazione del modello di simulazione. Nel Capitolo 5, questi concetti sono applicati al caso di membrane lipidiche in soluzione e i risultati sono confrontati con quelli ottenuti per catene alifatiche di idrocarburi liquidi. Queste analisi permettono di investigare la natura della fluidità delle membrane e del miscelamento di lipidi in bistrati lipidici, modelli di membrane cellulari. Inoltre, le entropie configurazionali sono calcolate allo scopo di completare i dati sperimentali relativi al grado di ordine delle catene aciliche in membrane di differente composizione.

I campi di forza a *grana grossa* sono tipicamente basati su una calibrazione *ad hoc* che permette di preservarne al meglio la semplicità intrinseca. Al contrario, il calcolo di quantità termodinamiche è utilizzato come procedura accurata per la parametrizzazione di campi di forza a livello atomico. Un'analisi dettagliata della termodinamica di evaporazione, solvatazione e separazione di fase è descritta nel Capitolo 6, con modelli a differente risoluzione. In questo caso, sono impiegati calcoli di energia libera per valutare la corrispondenza tra un campo di forze semplificato e dati sperimentali o simulazioni a risoluzione più alta. Nello stesso capitolo, sono proposti alcuni miglioramenti del campo di forze a *grana grossa*.

L'entropia è una proprietà chiave nella comprensione di una enorme varietà di fenomeni fisici, chimici e biochimici. Alcuni esempi sono gli effetti entropici riguardanti (i) l'avvolgimento e la stabilità dei sistemi proteici, (ii) i fenomeni di associazione molecolare e (iii) la catalisi enzimatica. Da un punto di vista pratico, nel corso degli ultimi decenni il calcolo di differenze di energia libera è diventato possibile utilizzando simulazioni molecolari. Una stima attendibile dell'entropia assoluta e le sue differenze da queste simulazioni é, invece, notoriamente un problema irrisolto. Il Capitolo 7 analizza criticamente

il metodo per la stima di entropie configurazionali da simulazioni di dinamica molecolare, che si basa sulla approssimazione quasi-armonica. Sono presentate la teoria, le assunzioni e le approssimazioni coinvolte. Inoltre, viene proposto un metodo per quantificare l'errore commesso nella stima dell'entropia. Nel complesso, il metodo presentato apre la possibilità di calcolare sia entropie assolute, sia differenze di entropia dovute a transizioni conformazionali di sistemi (macro)molecolari.

Infine, il Capitolo 8 abbozza le conclusioni principali tratte dagli studi presentati in questa tesi e suggerisce alcune prospettive per il futuro della simulazione di sistemi (bio)molecolari.

Summary

Molecular dynamics simulations with empirical force fields are an important computational tool to study the properties of (bio)molecular systems at atomic resolution. These simulations and their underlying models based on classical mechanics give fundamental insight into the behaviour of complex macromolecules and help to understand the relationship among their structure, dynamics, and function. One of the major benefits of such simulations is their complementarity to experimental results by providing not only averages, but also distributions and time series of any mathematically definable quantity. Additionally, the practical possibility to access a statistical-mechanical picture of the studied systems offers an interpretative key which is, in most of the cases, precluded to experiments. Chapter 1 presents a short introduction to the philosophy of (bio)molecular simulation. Three key challenging problems for which these simulations are of fundamental importance are also described with reference to the following chapters.

A question of great interest in chemistry and biology is the understanding of how (bio)molecules fold into relatively stable three-dimensional structures under physiological conditions. Chapter 2 is a study of the unfolded state of two small peptides with the same backbone, but different side chains. Experimentally they have been observed to fold into different secondary structures, one into a compact helix, the second into a β -hairpin fold. The simulations elucidate the question whether these two peptides of different secondary preferences share any common region in the populated parts of their configurational spaces. The comparison of simulation results to experimental data is the key for the validation of molecular dynamics simulations. After careful validation, such simulations can be used to guide experiments or gain insight at time and space resolutions inaccessible for experiment. In Chapter 3 these concepts are presented for a set of carbopeptoids (*i.e.* sugar-amino-acid based peptides) through the comparison of simulation trajectories to data from nuclear magnetic resonance experiments. The application of molecular dynamics is focused on the understanding of how different stereochemical properties, carbopeptoid length and solvent environment may influence the folding behaviour of this class of molecules.

In order to allow for faster computations and to increase the sampling possibilities of molecular dynamics simulations, some of the system degrees of freedom may be removed from the model, and the corresponding underlying force field simplified. The importance of such a coarse-grained representation is briefly described in Chapter 1 and corresponding applications are detailed in Chapters 4, 5, and 6. In particular, a comparison of different model resolutions for liquid hydrocarbons is given in Chapter 4 by estimating configurational entropies which directly relate to the configurational space accessed by these molecules. This allows to estimate the loss of configurational entropy upon coarse-graining the simulation model. In Chapter 5 these concepts are applied to fully hydrated lipid bilayers, and the results are compared to those for aliphatic hydrocarbon chains. This analysis gives insights into the nature of membrane fluidity and lipid mixing in bilayers. Configurational entropy calculations are also performed to complement experimental data on the ordering of lipid tails in lipid bilayers of different composition.

In order to keep coarse-grained force fields as simple as possible, they are typically based on *ad hoc* parametrizations. In contrast, the calculation of thermodynamic quantities is employed as an accurate procedure for parametrizing atomic-level force fields. A detailed picture of vaporization, solvation, and phase-partitioning thermodynamics at different model resolutions is presented in Chapter 6. There, free energy calculations are used to evaluate the correspondence between a simplified force field and experimental data and/or higher resolution simulations. Additionally, improvements of the coarse-grained force field are proposed.

Entropy is a key property to the understanding of a wide variety of physical, chemical and biochemical phenomena. For example, entropic effects play a fundamental role in protein folding and stability, molecular association, ligand binding, and enzyme catalysis. During the past decades, the calculation of accurate free-energy differences from molecular simulations has become feasible and is now almost a routine task. In contrast, reliable absolute entropies and entropy differences from these simulations are still notoriously difficult to obtain. Chapter 7 investigates critically the method to estimate configurational entropies from molecular dynamics simulations based on the quasi-harmonic approximation. The theory, assumptions, and approximations underlying this method are presented. Additionally, a method is proposed to characterize the magnitude of the error on the estimated entropy. This method opens the

possibility for an improved calculation of absolute entropies as well as of entropy differences for (macro)molecular systems which undergo conformational transitions.

Finally, Chapter 8 summarises the main conclusions drawn from the studies presented in this thesis and sketches some future perspectives for molecular dynamics simulation of (bio)molecular systems.