

Computer simulations of peptides and proteins: a comparison of effects of solute and solvent variation upon simulated and measured observables

A dissertation submitted to the
ETH ZÜRICH

for the degree of
Doctor of Sciences

presented by
ZRINKA GATTIN
Dipl. Ing. Chem.
born September 07, 1980
citizen of Croatia

accepted on the recommendation of
Prof. Dr. Wilfred F. van Gunsteren, examiner
Prof. Dr. Alan E. Mark and Prof. Dr. Bernhard Jaun, co-examiners

2008

Kurzfassung

Da Biomoleküle keinen statischen, sondern vielmehr einen sehr dynamischen Charakter aufweisen, stellt die Untersuchung ihrer strukturellen und dynamischen Eigenschaften eine grosse Herausforderung für computergestützte und experimentelle Forschungsarbeiten dar.

Computergestützte Methoden, d.h. *in silico* Experimente, gewähren Einsicht auf einer derart detaillierten räumlichen und zeitlichen Ebene, die mit im Labor durchgeführten Experimenten nicht erreicht werden kann. Die technische Realisierung von computergestützten Studien ist jedoch von beträchtlichen Einschränkungen hinsichtlich der räumlichen Ausdehnung des zu untersuchenden Systems, sowie des Zeitraumes, über welchen dieses System betrachtet werden kann, gekennzeichnet. Molekulardynamik-Simulationen (MD-Simulationen), in welchen die Bewegung der Atome des biomolekularen Systems nach den Gesetzen der klassischen Mechanik berechnet wird, sind dabei eine der am meisten verwendeten computergestützten Methoden. Die ersten MD-Simulationen von Biomolekülen wurden in den späten 1970er Jahren des letzten Jahrhunderts durchgeführt. Seitdem sind sie zu einem bedeutenden und weit verbreiteten Hilfsmittel in der Erforschung biomolekularer Systeme geworden, wobei die Essenz ihrer Anwendung eine Darstellung der strukturellen und dynamischen Eigenschaften des zu untersuchenden Systems ist, um letztendlich die Beziehung zwischen Struktur, Dynamik und Funktion des Biomoleküls aufzuklären. Computersimulationen werden somit verwendet, um experimentelle Ergebnisse bezüglich Struktur und Dynamik auf atomarer Ebene zu interpretieren. Da mit (nassen) Experimenten Informationen über Zeitskalen von Nanosekunden (und weniger) nur schwer zugänglich sind, liegt bei den meisten computergestützten Studien die Betonung auf einem Vergleich mit Messdaten über strukturelle Eigenschaften. In dieser Dissertation werden MD-Simulationen verwendet, um eine Reihe von Peptiden und Proteinen zu untersuchen, wobei die Ergebnisse mit experimentellen Daten verglichen werden.

In Kapitel 1 wird ein Überblick über die theoretischen Grundlagen von MD-Simulationen gegeben. Dabei werden die Bedeutung einer korrekten Berechnung von Observablen, ausgehend von einem simulierten Ensemble, sowie die Probleme, die beim Vergleich eines aus einer herkömmlichen (auf experimentellen Daten basierenden) Strukturverfeinerung resultierenden, isolierten Strukturmodells mit einem Boltzmann-Ensemble von Strukturen aus MD-Simulationen auftreten können, herausgestellt. In diesem Zusammenhang werden drei Anwendungen von biomolekularen Simu-

lationen vorgestellt: i) Faltungsstudien für Peptide und Proteine; ii) die Interpretation von experimentellen Daten mit Hilfe von MD-Simulationen; iii) die Rolle von MD-Simulationen in der (auf experimentellen, räumlich und zeitlich gemittelten Daten basierenden) Strukturbestimmung und -verfeinerung von Biomolekülen.

Das Verständnis der Faltung von Biomolekülen in verhältnismässig stabile dreidimensionale Strukturen unter physiologischen Bedingungen ist für die Chemie und Biologie von grossem Interesse. In Kapitel 2 wird untersucht, wie das Konformationsgleichgewicht eines Peptides von der Fähigkeit der Ausbildung intramolekularer Wasserstoffbrücken beeinflusst wird. Die Interpretation von experimentellen Primärdaten aus der NMR-Strukturbestimmung (z.B. Signalintensitäten und -frequenzen) ist nicht trivial und wird durch das Mitteln über sowohl Molekülkonformationen als auch die Messzeit erschwert. Darüber hinaus liegt oftmals nur eine begrenzte Menge an verlässlichen Messdaten vor, wodurch die Gewissheit über eine wahrheitsgetreue Interpretation der Daten beeinträchtigt wird.

In den Kapiteln 3, 4 und 5 wird der Einfluss verschiedener Faktoren auf die Übereinstimmung von experimentell ermittelten isolierten Strukturmodellen mit konformationellen Observablen aus MD-Ensembles untersucht und diskutiert. Dabei auftretende Diskrepanzen zwischen experimentellen Daten und den Ergebnissen von MD-Simulationen können erklärt werden, indem man z.B. MD-Simulationen mit zusätzlichen Potentialen, welche eine Annäherung an experimentelle Primärdaten (z.B. NOE Proton-Proton Abstände oder ^3J -Kopplungskonstanten) erzwingen, durchführt. Werden diese zusätzlichen Potentiale basierend auf einzelnen Zeitschritten angewendet, so kann man Aufschluss erhalten über mögliche inkorrekte, in der experimentellen Strukturverfeinerung benutzte, strukturelle Zwangsbedingungen. Zusammen mit der Analyse entsprechender MD-Simulationen mit zusätzlichen Potentialen basierend auf zeitlichen Durchschnittswerten kann somit dargelegt werden, ob die experimentell gemessenen Werte bestimmter Observablen aus einem Mitteln über viele Konformationen resultieren. In Kapitel 3 wird erläutert, wie eine derartige Studie zu einer Neuinterpretation von experimentellen Daten für ein β -Peptid führt. Kapitel 4 zeigt auf, dass es wichtig ist, ausreichende und gleichmässig verteilte strukturelle Zwangsbedingungen zu erhalten. Sind die Energiebarrieren zwischen einzelnen (hauptsächlich vorliegenden) Konformeren erheblich grösser als $k_B T$, so können MD-Simulationen kein korrektes Boltzmann-Ensemble erzeugen. In einem derartigen Fall kann man auf den sogenannten "local-elevation" Algorithmus zurückgreifen, um ein korrektes Ensemble von Konformationen zu erhalten. Dieser Algorithmus wurde, wie in Kapitel 5 beschrieben ist, auf zusätzliche Zwangspotentiale für die ^3J -Kopplungskonstanten der Rückgrat-Dihedralwinkel eines zyklischen β -Tetrapeptides angewendet, was zeigt, dass ^3J -Kopplungskonstanten direkt verwendet werden können um ein den experimentellen Daten entsprechendes Ensemble von Konformationen zu erhalten.

In den Kapiteln 6 und 7 wird die Stabilität von zwei kleinen Proteinen hinsichtlich Änderungen des pH-Wertes, der Temperatur und einem Zusatz von Harnstoff betrachtet. Die Ergebnisse (Kapitel 6) zeigen, dass die mit dem GROMOS Kraftfeld ermittelte Denaturierungs-Temperatur des Trp-cage Proteins besser mit dem experimentellen Wert übereinstimmt, als das für andere biomolekulare Kraftfelder der Fall ist. Desweiteren (Kapitel 7) wird anhand des Pysin Proteins dargestellt, dass nur kleine Änderungen im Protonierungszustand weniger Aminosäure-Seitenketten grosse strukturelle Änderungen bedingen können.

Schliesslich werden in Kapitel 8 die wichtigsten Schlussfolgerungen der in dieser Dissertation vorgestellten Studien zusammengefasst, sowie zukünftige Entwicklungen von MD-Simulationen biomolekularer Systeme umrissen.

Summary

The dynamic rather than static nature of biomolecules makes the investigation of their structural and dynamical properties a real challenge to theoretical, computational and experimental research. Computational methods, i.e. experiments *in silico*, provide elucidations on a detailed spatial and time level that wet (real-life) experiments often cannot reach, but they are severely limited regarding the spatial and time scales that can be covered. One of the most used computational techniques is molecular dynamics (MD) simulation, in which the motion of the atoms of a biomolecular system is simulated classically. From the first biomolecular dynamics simulations in late seventies of the last century, molecular simulations evolved into an important and widely used tool in studying biomolecular systems with particular emphasis on providing a structural and dynamical characterization of a particular biomolecular system aiming at an understanding of the relationship between its structure, dynamics and function. Computer simulation is thus used to interpret the results of experiments at an atomic level, both structurally and dynamically. Because dynamic information on a time scale of nanoseconds and below is experimentally not easily obtainable, most computational studies focus on the comparison with measured data on structural properties. In this thesis a number of small peptides and proteins is investigated computationally and the results are compared to experimental data.

Chapter 1 gives a short overview of the molecular dynamics (MD) simulation technique. In particular it addresses the importance of a correct description of ensemble averages and the problems that may emerge when comparing a single model structure for a biomolecule derived using standard single-structure refinement procedures from the experimental data with a Boltzmann ensemble of structures generated using MD simulation. It discusses three applications of biomolecular simulation: i) the study of peptide and protein folding , ii) the interpretation of experimental data through molecular dynamics simulation and iii) molecular dynamics simulations as a tool for structure determination and refinement of biomolecules based on experimental (averaged) data.

A question of great interest in chemistry and biology is understanding why and how biomolecules fold into relatively stable three-dimensional structures under physiological conditions. Chapter 2 contains a study on how the ability to form an intramolecular hydrogen bonding network influences the conformational equilibrium of a peptide.

The structural interpretation of primary NMR experimental data such as signal intensities and frequencies is not straightforward and it is hampered by the occurrence of conformational and time averaging. Moreover, the number of measured data and their reliability may be limited, which enhances the uncertainty of the conformational interpretation of the data. In chapters 3, 4 and 5 the influence of several aspects regarding differences between the model structures derived from experimental data using standard procedures and MD conformational ensemble averages are investigated and discussed. One way to locate the origin of these differences is by performing MD simulation with restraints to the (primary) experimental data, e.g. NOE proton-proton distances or 3J -coupling constants. Application of instantaneous restraints may yield clues to incorrect restraints, and in combination with the analysis of the corresponding results from time-averaged restraining simulations it may clarify whether measured average values of particular observables are the result of extensive conformational averaging. This analysis leads in Chapter 3 to a re-interpretation of the experimental data of a β -peptide. The example of Chapter 4 illustrates the importance of obtaining sufficient and spatially well distributed restraints. When the energetic barriers separating well-populated conformers are much higher than $k_B T$, MD simulation will not produce a correct Boltzmann ensemble. In such a case the application of the local-elevation search algorithm may help to generate the appropriate conformational ensemble. In Chapter 5 we applied the local-elevation (LE) algorithm to 3J -value restraining of the backbone dihedral angles of a cyclic β -tetrapeptide and show that using 3J -value LE MD the measured 3J -coupling constant values can be directly used to derive the conformational ensemble.

Chapters 6 and 7 focus on predicting the stabilities of two small proteins, with respect to variation of pH, temperature and addition of urea. In Chapter 6 it is found that the denaturation temperature of a Trp-cage protein in the GROMOS force field is much closer to the measured one than those obtained using other biomolecular force fields. In Chapter 7 it is analysed how a small change in protonation states of a few amino acid side chains can result in a large change in the structure of the Pysin protein.

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

Sažetak

Dinamička (mnogo više nego statička) priroda biomolekula čini istraživanje njihovih strukturnih i dinamičkih svojstava pravim izazovom za teorijska, eksperimentalna i računalna istraživanja. Iako su jako ograničene s obzirom na prostornu i vremensku skalu koja im je dostupna, računalne metode, odnosno eksperimenti *in silico*, omogućavaju razlučivanje detalja na prostornoj i vremenskoj razini, koju eksperimenti često ne mogu dosegnuti. Jedna od najviše korištenih računalnih tehnika je molekulska dinamika (MD), u kojoj je kretanje atoma biomolekularnog sistema simulirano pomoću klasične fizike. Od prvih simulacija biomolekularne dinamike, krajem sedamdesetih godina prošlog stoljeća, molekularne simulacije su se razvile u važan i široko korišten alat u proučavanju biomolekulskih sustava s posebnim naglaskom na strukturnoj i dinamičkoj karakterizaciji određenog biomolekularnog sustava s ciljem razumijevanja odnosa strukture, dinamike i funkcije. Računalna simulacija se prema tome koristi za interpretaciju rezultata eksperimenata na atomarnoj razini, kako strukturno tako i dinamički. Pošto dinamička informacija na vremenskoj skali od nanosekunde i niže nije eksperimentalno lako dostižna, većina računalnih istraživanja usredotočuje se na usporedbu s izmjerenim podacima o strukturnim svojstvima. U ovoj disertaciji računalno je istražen niz peptida i proteina te su rezultati uspoređeni s eksperimentalnim podacima.

Poglavlje 1 daje kratki pregled MD simulacijske tehnike, a posebice spominje važnost ispravnog opisa srednjih vrijednosti ansambla i probleme koji se mogu pojaviti kada se uspoređuje "single model" struktura biomolekule izvedena iz eksperimentalnih podataka koristeći standardne ("single structure") postupke utočnjavanja s Boltzmannovim ansamblom struktura dobivenim korištenjem MD simulacija. U njemu se raspravlja o tri primjene biomolekulskih simulacija: i) proučavanje peptida i smatanje ("folding") proteina, ii) interpretacija eksperimentalnih podataka pomoću MD simulacija te iii) MD simulacije kao alat za određivanje i utočnavanje strukture biomolekula na temelju eksperimentalnih (usrednjenih) podataka. Kako i zašto biomolekule tvore relativno stabilne trodimenzionalne strukture pri fiziološkim uvjetima, pitanje je od velikog interesa u kemiji i biologiji. Poglavlje 2 sadrži istraživanje o tome kako sposobnost da se formira intramolekulska mreža vodikovih veza utječe na konformacijsku ravnotežu peptida.

Strukturna interpretacija primarnih eksperimentalnih NMR podataka kao što su intenziteti signala i frekvencije nije jednostavna, a k tome je ometa i pojava konformacijskog i vremenskog usrednjavanja. Povrh toga, broj izmjerenih podataka i njihova pouzdanost mogu biti ograničeni, što povećava nesigurnost konformacijske interpretacije podataka. U poglavljima 3, 4 i 5 istražen je i diskutiran utjecaj nekoliko aspekata koji se tiču razlika između modelne strukture dobivene pomoću standardnih postupaka iz eksperimentalnih podataka i iz usrednjenih konformacijskih MD ansambla. Jedan način za pronalaženje podrijetla tih razlika je izvođenje MD simulacije s konformacijskim ograničenjima izvedenima iz (primarnih) eksperimentalnih podataka, npr. NOE proton-proton udaljenosti ili konstante sprege, 3J . Primjena trenutačnih ograničenja može ukazati na nekorektna eksperimentalna ograničenja te u kombinaciji s analizom odgovarajućih rezultata dobivenih pomoću simulacija koje koriste vremenski usrednjena ograničenja ("time-averaged restrained") može razjasniti da li su izmjerene usrednjene vrijednosti pojedinih observabli posljedica opsežnog konformacijskog usrednjavanja. Analiza u Poglavlju 3 dovodi do reinterpretacije eksperimentalnih podataka vezanih za β -peptide. Poglavlje 4 ilustrira važnost postizanja dostatnih i prostorno dobro distribuiranih konformacijskih ograničenja. Kada su energetske barijere koje razdvajaju zastupljenije konformere mnogo više no $k_B T$, MD simulacija neće dati ispravan Boltzmannov ansambl. U takvom slučaju, primjena "local-elevation" (LE) pretražnog algoritma može pomoći pri generiranju odgovarajućih konformacijskih ansambla. U Poglavlju 5 primjenjen je LE algoritam na ograničenje dihedralnih kuteva okosnice cikličkog β -tetrapeptida pomoću 3J -vrijednosti i pokazano je da se izmjerene vrijednosti konstante sprege mogu pomoću 3J LE MD izravno koristiti za određivanje konformacijskog ansambla.

Poglavlja 6 i 7 usredotočuju se na predviđanje stabilnosti dva mala proteina, s obzirom na promjenu pH, temperature i dodatak uree. U Poglavlju 6 je utvrđeno kako je temperatura denaturacije proteina Trp-cage u GROMOS polju sila ("force field") mnogo bliža izmjerenoj vrijednosti od onih dobivenih pomoću drugih biomolekulskih polja sila. U Poglavlju 7 je analizirano kako mala promjena u protonaciji nekoliko aminokiselinskih ogranaka može dovesti do velike promjene u strukturi Pysin proteina. Konačno, Poglavlje 8 sažima glavne zaključke izvučene iz istraživanja iznesenih u ovoj disertaciji te daje kratak pregled nekih perspektiva za MD simulacije biomolekulskih sistema.