

Molecular dynamics simulation of alkanes and proteins

methodology, prediction of properties and comparison to experimental data

Doctoral Thesis

Author(s):

Eichenberger, Andreas Paul

Publication date:

2012

Permanent link:

<https://doi.org/10.3929/ethz-a-009787706>

Rights / license:

[In Copyright - Non-Commercial Use Permitted](#)

Diss. ETH No. 20781

Molecular dynamics simulation of alkanes and proteins: methodology, prediction of properties and comparison to experimental data

A dissertation submitted to the
ETH Zürich

for the degree of
Doctor of Sciences

presented by
ANDREAS PAUL EICHENBERGER
M. Sc. ETH
born January 8, 1982
citizen of Beinwil a. See – Switzerland

accepted on the recommendation of
Prof. Dr. Wilfred F. van Gunsteren, examiner
Dr. Lorna J. Smith, co-examiner
Prof. Dr. Hans Jürgen Herrmann, co-examiner

2012

Summary

Molecular dynamics (MD) simulation is a powerful tool to investigate problems beyond the border of the area accessible to experiments and has been widely used to study the structure and dynamics of biomolecular systems. A reliable model and accurate algorithm is indispensable for a scientist to calculate and predict properties of the system under study, but treatment and post-processing of simulation trajectories is of no less importance.

Chapter 1 gives a brief introduction to MD simulations and how they may complement experimental studies, what basic decisions must be made prior to running an MD simulation, and mentions the two main limiting factors of MD simulations: insufficient configurational sampling and simulation speed.

Chapters 2 and 3 are addressing the simulation speed. In Chapter 2 a method to apply bond-angle or dihedral-angle constraints in MD simulations is presented. It constitutes an alternative to the use of distance constraints, which fail for linear molecule configurations, thereby enhancing the simulation time by allowing for a bigger integration time step.

In Chapter 3 another possibility to gain simulation speed is proposed which is based on reducing the number of degrees of freedom considered during a simulation. A supra-atomic coarse-grained (CG) model for linear alkanes in the liquid phase is presented, a possible precursor of a CG model for lipids to be used – a long-term goal – in combination with atomic-level, fine-grained (FG) protein or other biomolecules. It consists of six CG beads of sizes 2, 3 and 4 and distinguishes between beads at the chain ends (end beads) and beads between end beads (middle beads). The CG force field was derived based on experimental thermodynamic data and on mapped structural and energetic properties of corresponding FG alkane simulations.

Analysis of biomolecular simulation trajectories is considered in Chapter 4, where an overview of the GROMOS++ simulation software is given distinguishing between three types of programs: programs that calculate structural, dynamic or thermodynamic quantities from configurational simulation trajectories. GROMOS++ is a flexible and rich collection of analysis tools ready to be used for a variety of types of analysis regarding molecular simulation trajectories.

In Chapter 5, enveloping distribution sampling (EDS) is used to calculate folding free enthalpy differences between the wild-type Pin1 WW domain and 20 amide-to-ester mutants

in analogy to an experimental study [1]. It shows that a relation between the strength of the individual hydrogen bonds and the differences in folding free enthalpy differences, as done in the experimental study [1], is not straightforward due to structural changes within the folded protein between wild-type and mutants, indicated by different hydrogen-bonding patterns in the enveloping distribution sampling (EDS) MD simulations.

In Chapter 6 the effect of 222-trifluoroethanol (TFE) upon the protein structure of hen egg white lysozyme (HEWL) is studied. Experimental NMR studies report that in 70% TFE/30% water the regions corresponding to native β -strands in HEWL retain persistent helical conformation and increase in length, while the detailed and full protein structure in TFE/water is still not known. A possible configuration largely fulfilling the available experimental NMR data was generated using MD simulations, and is the currently best possible representation of this ensemble.

Chapters 5, 7 and 8 discuss proteins with backbone amide-to-ester substitutions. Amide-to-ester mutations are widely used in experimental and theoretical studies to influence the protein structure and stability by eliminating one backbone hydrogen-bond donor (NH versus O) and weakening one backbone hydrogen-bond acceptor (amide carbonyl versus ester carbonyl). The effect of amide-to-ester replacements on the protein structure of HEWL is described in Chapters 7 and 8. Fully ester-linked HEWL, i.e. containing as many amide-to-ester substitutions as possible, shows a slight compaction while losing its native structure. However, it does not unfold completely. Partly ester-linked HEWL, in which only 34 peptide linkages that are not involved in the helical or β -strand parts of native HEWL were replaced by ester linkages showed behaviour comparable to native HEWL. The conformational changes were analysed by comparing simulation averaged values of quantities that can be measured by NMR, such as ^1H - ^{15}N backbone order parameters, residual dipolar couplings, proton-proton NOE distances and ^3J -couplings with the corresponding values derived from experimental NMR data for native HEWL.

Possible further development and future work is mentioned in the outlook, Chapter 9.

Zusammenfassung

Molekulardynamikcomputersimulationen sind ein leistungsstarkes Werkzeug, um Problemstellungen jenseits der Möglichkeiten experimenteller Mittel zu betrachten. Sie werden vielfach zur Untersuchung der Struktur und Dynamik von biomolekularen Systemen verwendet. Ein vertrauenswürdiges Modell sowie ein präziser Algorithmus sind unabdingbar zur Berechnung und Vorhersage von Systemeigenschaften des simulierten Systems, aber auch der Umgang und die Auswertung von Simulationstrajektorien ist nicht von geringerer Wichtigkeit.

Eine kurze Einführung in die Methodik von Molekulardynamiksimulationen wird in Kapitel 1 gegeben. Es wird kurz aufgezeichnet, inwiefern Computersimulationen experimentelle Studien ergänzen können und welche grundsätzlichen Entscheidungen schon vor der Durchführung der Simulation getroffen werden müssen. Weiter werden die beiden Hauptgründe, welche die Aussagekraft von Computersimulationen einschränken, aufgezeichnet: mangelhafte Abtastung des Konfigurationsraums sowie die Simulationsgeschwindigkeit.

In Kapitel 2 und 3 wird die Problematik der Simulationsgeschwindigkeit angesprochen. Kapitel 2 beschreibt eine Methode zur Fixierung von Bindungs- und Diederwinkeln und stellt eine mögliche Alternative zur Fixierung von Bindungslängen dar, die im Gegensatz zu bisherigen Methoden auch für lineare Moleküle geeignet ist. Solche Methoden beschleunigen eine Computersimulation, da sie einen grösseren Integrationszeitschritt erlauben.

Eine andere Möglichkeit zur Beschleunigung von Computersimulationen wird in Kapitel 3 angesprochen und basiert auf der Reduktion der Freiheitsgrade, welche während der Simulation berücksichtigt werden. Es wird ein atomübergreifendes, grobkörniges Alkanmodell vorgestellt, welches als möglicher Vorläufer eines grobkörnigen Lipidmodells dienen soll, in Kombination mit Modellen für Proteine und anderen Biomolekülen auf feinkörnigem, atomarem Level. Das vorgeschlagene Modell besteht aus sechs grobkörnigen Kugeln, sogenannten "Beads" der Grösse 2, 3 und 4, und unterscheidet zwischen Beads am Ende der Alkankette (Endbeads) und solchen zwischen zwei Endbeads (Mittelbeads). Die Modellparameter des grobkörnigen Alkanmodells wurden basierend auf thermodynamischen, experimentellen Messwerten entworfen, wobei zusätzlich strukturelle und energetische Eigenschaften für das grobkörnige Modell aus feinkörnigen Alkansimulationen abgeleitet und benutzt wurden.

Das Analysieren von Simulationstrajektorien wird im Rahmen eines Überblicks über die

Computersoftware GROMOS++ in Kapitel 4 behandelt. Man unterscheidet drei Programmarten: Programme zur Berechnung von strukturellen, dynamischen oder thermodynamischen Grössen, alle basierend auf Simulationstrajektorien. GROMOS++ ist eine flexible und umfangreiche Sammlung von Auswertungsprogrammen für eine Vielfalt von Analysen basierend auf Molekularsimulationstrajektorien.

In Kapitel 5 wird die sogenannte “enveloping distribution sampling”, oder auch kurz EDS Methode benutzt, um Differenzen in der freien Faltungsenthalpie zwischen dem natürlichen und 20 mutierten Pin1 WW Teilproteinen zu berechnen, dies in Analogie zu einer experimentellen Studie [1]. Die Mutanten unterscheiden sich vom natürlichen Protein durch eine Esterbindung anstelle einer Peptidbindung. Die durchgeführten Simulationen zeigen, dass eine Verknüpfung zwischen der Stärke der einzelnen Wasserstoffbrücken und der Differenz in der freien Faltungsenthalpie, wie sie im Experiment [1] gemacht wird, schwierig ist. Der Grund dafür sind die strukturell unterschiedlichen Proteinfaltungen des natürlichen und der mutierten Proteine, angedeutet durch die verschiedenen Muster ausgebildeter Wasserstoffbrücken während der EDS Simulation.

Der Effekt von 222-Trifluoroethanol (TFE) auf die Struktur von Hühnereiweisslysozym (HEWL) wird in Kapitel 6 studiert. Experimentelle Studien beschreiben, dass in einer 70% TFE/30% Wasser Lösung β -Faltblätter in einem bestimmten Teilbereich des Proteins zu einer beständigen α -Helixstruktur wechseln, wobei die Gesamtstruktur des Proteins in diesem Faltungszustand unbekannt ist. Mit Hilfe von Molekulardynamiksimulationen wurde ein Ensemble von HEWL Proteinkonfigurationen generiert, welches verfügbaren Daten aus der NMR Spektroskopie grösstenteils entspricht. Es ist die zur Zeit beste Darstellung dieses Ensembles.

Kapitel 5, 7 und 8 behandeln Proteine, in welchen eine oder mehrere Peptidbindungen zu Esterbindungen mutiert sind. Solche Mutationen, welche einen Protonendonator eliminieren (NH gegenüber O) und einen Protonenakzeptor schwächen (Amidkarbonyl gegenüber Esterkarbonyl), werden in Theorie und Experiment häufig benutzt, um den Einfluss von Wasserstoffbrücken auf die Stabilität und Struktur des Proteins zu untersuchen. Kapitel 7 und 8 behandeln beide den Effekt solcher Mutationen auf die Proteinstruktur von HEWL. Vollständig verestertes HEWL zeigt eine leichte Verdichtung, verliert aber die natürliche Proteinfaltung. Ein teilweise verestertes HEWL, bei dem 34 nicht in die Sekundärstruktur des Proteins involvierte Aminosäuren mutiert sind, zeigt ein Verhalten ähnlich jenem von natürlichem Lysozym. Unterschiede in der Proteinkonfiguration wurden durch den Vergleich von gemittelten, berechneten NMR Grössen sowie experimentellen Grössen zugänglich durch NMR Spektroskopie (^1H - ^{15}N Ordnungs-

parameter, Dipol-Dipol Kopplungskonstanten, Proton-Protone NOE Distanzen und ^3J -Kopplungskonstanten) analysiert.

Abschliessend wird in Kapitel 9 ein Ausblick auf zukünftige, weiterführende Studien, Arbeiten und Entwicklungen gegeben.