

Diss. ETH No. 20461

# Methodology Development for Classical Molecular Dynamics Simulations

A dissertation submitted to the  
ETH Zürich

for the degree of  
Doctor of Sciences

presented by  
SEREINA ZOE RINIKER  
M. Sc. ETH  
born January 15th, 1985  
citizen of Habsburg – Switzerland

accepted on the recommendation of  
Prof. Dr. Wilfred F. van Gunsteren, examiner  
Prof. Dr. Gisbert Schneider, co-examiner  
Prof. Dr. Nico van der Vegt, co-examiner

2012

# Summary

The dynamics and structure of biomolecules can be studied at the atomic level using molecular dynamics (MD) computer simulations. The insight gained through such simulations may explain experimental results and complement experimental studies and in some cases even suggest experiments and predict experimental results. However, the aim of studying biomolecular systems of ever growing complexity and size with increasing accuracy necessitates a constant development and advancement of methodology.

In Chapter 1, a brief introduction to classical MD simulations and the three methodological areas covered in this thesis is given followed by detailed introductions to the methods in Chapters 2, 3 and 4. In Chapter 2, the basic ingredients of developing coarse-grained (CG) models are discussed. In coarse graining, selected degrees of freedom are eliminated leading to a reduction of the number of interactions to be calculated and thus to a decrease in computational effort. As coarse graining involves *per se* a loss in information, a set of conditions is given which must be satisfied for the CG model to reproduce the correct physical mechanisms. In Chapter 3, different methods to calculate the static relative dielectric permittivity and Debye relaxation time of a molecular model in simulations of the liquid phase are reviewed. The accurate reproduction of the dielectric properties of a compound is of central importance for its use in biomolecular simulations. The performance of the different methods is tested on a set of four commonly employed solvents. In Chapter 4, a general introduction to methods to compute free energy differences from configurational ensembles generated by MD simulations is given and their advantages and disadvantages are discussed based on a selection of examples. In addition, implementation details of the GROMOS software package are specified.

In Chapter 5, the development of a supra-molecular CG model for water which allows the explicit treatment of the electrostatic interactions is outlined. The calibration of the model parameters is based on experimental data and data from atomic level, fine-grained (FG) simulations. Thermodynamic, structural and dynamic properties of the CG model are compared with those obtained from experiment and from clusters of FG water molecules.

In Chapter 6, the combination of two resolution levels in a single system is described. This hybrid approach allows the detailed, atomic description for the region of interest without losing the computational speed-up of coarse graining. Only two parameters are needed to tune the interactions between particles of different resolution, which are calib-

rated and tested on mixtures of atomic FG and supra-molecular CG water.

In Chapter 7, the hybrid FG/CG approach is extended to solvate FG proteins in CG water. The influence of the CG water on the structural properties of the solute is investigated using four proteins with different combinations of secondary structure elements. The results are compared to those obtained from simulations using a FG water model, an implicit solvent model and vacuum boundary conditions. In addition, for one protein a comparison with experimental NMR data is made.

In Chapter 8, a study of the influence of introducing a layer of FG water molecules in hybrid protein/water FG/CG simulations is presented. A FG water layer allows the side chains of the protein to form hydrogen bonds with solvent without losing too much of the advantageous computational speed-up gained by coarse graining the solvent. The diffusion of the FG water molecules into the CG water is prevented by the application of attractive distance restraints to the centre of geometry of the protein.

In Chapter 9, the temperature dependence of the static relative dielectric permittivity of acetic acid, propionic acid and their methyl esters is studied using the external electric field method. A molecular explanation for the anomalous behaviour of the carboxylic acids is given. A set of intermolecular hydrogen-bonded structures is identified and their occurrence depending on the temperature and the electric field strength is investigated.

In Chapter 10, the performance of the enveloping distribution method (EDS) to calculate free energy differences is tested on a set of inhibitors of phenylethanolamine *N*-methyltransferase (PNMT). Two approaches to determine the parameters of the reference state that is sampled in EDS as well as two approaches to define the topology of the system are investigated. The results are compared to those obtained from thermodynamic integration (TI) simulations and with experimental binding affinities.

In Chapter 11, the substitution of water sites by apolar probes in a binding pocket of a protein is explored. The free energy difference between a water molecule and an apolar probe is calculated for a selection of water sites indicating which water molecules can favourably be replaced by a ligand. The binding pockets of two proteins tRNA-guanine transglycosylase (TGT) and cyclin-dependent kinase 2 (CDK2) are partly filled with a ligand and the water sites in a subpocket are investigated.

In Chapter 12, a brief outlook of future advancement of the discussed methods is given.

# Zusammenfassung

Die Dynamik und Struktur von Biomolekülen auf atomarem Level kann mittels Molekulardynamikcomputersimulationen (MD) studiert werden. Die Einsichten, welche durch solche Simulationen erhalten werden, können experimentelle Resultate erklären und experimentelle Studien vervollständigen sowie in manchen Fällen sogar Experimente vorschlagen und experimentelle Resultate vorhersagen. Der Trend zu immer grösseren und komplexeren Systemen mit zugleich immer grösserer Genauigkeit setzt jedoch eine fortlaufende Entwicklung und Verbesserung von entsprechenden Methoden voraus.

Im Kapitel 1 wird eine kurze Einführung in die klassischen MD-Simulationen und die drei Methodikgebiete, welche in dieser Dissertation behandelt werden, gegeben, gefolgt von detaillierten Einführungen in die Methoden in den Kapiteln 2, 3 und 4. In Kapitel 2 werden die grundlegenden Faktoren für die Entwicklung von sogenannten "coarse-grained" (CG) Modellen diskutiert. Beim "coarse graining" werden ausgewählte Freiheitsgrade ausgelassen, was eine Reduzierung der Anzahl der zu berechnenden Wechselwirkungen und somit weniger Rechenaufwand zur Folge hat. Da dieser Prozess in jedem Fall zu einem Informationsverlust führt, wird eine Anzahl von Bedingungen angegeben, welche von einem CG-Modell erfüllt werden müssen, damit es die physikalischen Mechanismen korrekt nachbildet. In Kapitel 3 werden verschiedene Methoden zur Berechnung der relativen dielektrischen Permittivität und der Debye-Relaxationszeit eines molekularen Modells in Simulationen der flüssigen Phase vorgestellt. Die korrekte Nachbildung der dielektrischen Eigenschaften einer chemischen Verbindung ist von zentraler Bedeutung für den Einsatz eines Modells in biomolekularen Simulationen. Das Verhalten der verschiedenen Methoden wird anhand von vier üblichen Lösungsmitteln getestet. In Kapitel 4 wird eine allgemeine Einführung in verschiedene Methoden zur Berechnung von freien Energiedifferenzen aus Konfigurationskollektiven, welche durch MD Simulationen generiert werden, gegeben und deren Vor- und Nachteile werden basierend auf einer Auswahl von Beispielen diskutiert. Zudem werden die Implementationsdetails der GROMOS Software aufgeführt.

In Kapitel 5 wird die Entwicklung eines supramolekularen CG-Modells für Wasser geschildert, welches die explizite Behandlung der elektrostatischen Wechselwirkungen erlaubt. Die Kalibrierung der Modellparameter erfolgt basierend auf experimentellen Daten und solchen aus Simulationen mit atomarer, sogenannter "fine-grained" (FG) Auflösung. Die thermodynamischen, strukturellen und dynamischen Eigenschaften des CG-Modells

werden mit experimentellen Daten verglichen sowie mit solchen von FG-Wassermolekül-Gruppen.

In Kapitel 6 wird die Kombination zweier Auflösungsstufen in einem einzigen System beschrieben. Dieser hybride Ansatz erlaubt die detaillierte Beschreibung des Zielbereiches, ohne jedoch die Beschleunigung, die durch das "coarse graining" gewonnen wird, zu verlieren. Nur zwei Parameter sind vonnöten, um die Wechselwirkungen zwischen den Teilchen mit verschiedener Auflösung einzustellen, und diese werden anhand von Mischungen aus atomarem FG- und supramolekularem CG-Wasser bestimmt und getestet.

In Kapitel 7 wird der hybride FG/CG-Ansatz auf die Solvatisierung von FG-Proteinen in CG-Wasser erweitert. Der Einfluss des CG-Wassers auf die strukturellen Eigenschaften wird am Beispiel von vier Proteinen untersucht, welche unterschiedliche Kombinationen der sekundären Strukturelemente aufweisen. Die Resultate werden mit solchen aus Simulationen mit FG-Wasser, mit einem impliziten Lösungsmittelmodell und mit Vakuumrandbedingungen verglichen. Zudem wird für eines der Proteine ein Vergleich mit experimentellen NMR-Daten gemacht.

In Kapitel 8 wird eine Studie über den Einsatz einer Schicht aus FG-Wassermolekülen in den hybriden Protein/Wasser FG/CG-Simulationen präsentiert. Ein Schicht aus FG-Wassermolekülen erlaubt den Proteinseitenketten Wasserstoffbrücken mit dem Lösungsmittel zu formen, ohne dabei zu viel von der vorteilhaften Beschleunigung zu verlieren, welche durch das "coarse graining" des Lösungsmittel gewonnen wird. Die Diffusion der FG-Wassermoleküle in das CG-Wasser wird durch Anwendung von attraktiven Distanzbeschränkungen zum geometrischen Mittelpunkt des Proteins verhindert.

In Kapitel 9 wird die Temperaturabhängigkeit der relativen dielektrischen Permittivität von Essigsäure, Propansäure und deren Methylestern mithilfe der externen, elektrischen Feld-Methode (EEF) untersucht. Für das anomale Verhalten der Carbonsäuren wird eine molekulare Erklärung gegeben. Ein Satz von intermolekularen, über Wasserstoffbrücken verbundenen Strukturen wird identifiziert und deren Vorkommen abhängig von der Temperatur und der elektrischen Feldstärke angeschaut.

In Kapitel 10 wird die Leistung der sogenannten "enveloping distribution sampling" Methode (EDS) zur Berechnung von freien Energiedifferenzen an einem Satz von Inhibitoren für Phenylethanolamin *N*-Methyltransferase (PNMT) getestet. Zwei Ansätze zur Bestimmung der Parameter des Referenzzustandes, welcher in EDS simuliert wird, sowie zwei Ansätze zur Beschreibung der Topologie des Systems werden analysiert. Die Resultate werden mit solchen aus thermodynamischen Integrations-Simulationen (TI) gewonnenen

sowie mit experimentellen Bindungsaffinitäten verglichen.

In Kapitel 11 wird das Ersetzen von Wassermolekülen durch apolare Proben in der Bindungstasche eines Proteins untersucht. Die freie Energiedifferenz zwischen einem Wassermolekül und einer apolaren Probe wird für eine Auswahl von Wasser-Standorten berechnet. Damit wird aufgezeigt, welche Wassermoleküle geeignet sind, um durch einen Liganden ersetzt zu werden. Die Bindungstaschen der tRNA-Guanine Transglykosylase (TGT) und der cyclin-abhängigen Kinase 2 (CDK2) werden dazu durch einen Liganden teilweise gefüllt und die übrigbleibenden Wassermoleküle werden studiert.

In Kapitel 12 wird ein kurzer Ausblick in die zukünftige Entwicklung der vorgestellten Methoden gegeben.