Diss. ETH No. 16564

Algorithms and software for efficient biomolecular simulation

A dissertation submitted to the
SWISS FEDERAL INSTITUTE OF TECHNOLOGY
ZÜRICH

for the degree of
Doctor of Natural Sciences

presented by
MARKUS CHRISTEN
Dipl. Chem. ETH
born March 29, 1975
citizen of Bülach (ZH) and Wynau (BE), Switzerland

accepted on the recommendation of
Prof. Dr. Wilfred F. van Gunsteren, examiner
Prof. Dr. Andrew E. Torda and Prof. Dr. Matthias Troyer, co-examiners

2006

Berichte aus der Chemie

Markus Christen

Algorithms and software for efficient biomolecular simulation

Shaker Verlag
Aachen 2006

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.d-nb.de.

Zugl.: Zürich, ETH, Diss., 2006

Printed in Germany.

ISBN-10: 3-8322-5519-2
ISBN-13: 978-3-8322-5519-0
ISSN 0945-070X

Shaker Verlag GmbH • P.O. BOX 101818 • D-52018 Aachen
Phone: 0049/2407/9596-0 • Telefax: 0049/2407/9596-9
Internet: www.shaker.de • e-mail: info@shaker.de

Für Bea

Acknowledgements

It was with great pleasure that I spent my last five years with the group of IGC. This is for no small reason due to Wilfred, who, with his never ending enthusiasm, patience (especially when correcting the many sign errors I did - which may cancel if you just do enough of them) and humour, created a warm and friendly atmosphere in his group. His confidence and trust in me (be it concerning GROMOS, teaching, or the mere fact that I would finally produce a thesis), never got me bored. Together with Jolande, they were cheerful hosts the many times we were invited. Thank you.

Thanks also go to my co-examiners Andrew Torda and Matthias Troyer for reading this thesis and allowing me to defend it. Especially, I thank Andrew for the discussion about mixing fine-grained and coarse-grained simulations which brought me along towards *Chapter 5*.

I started working in this group a long time ago. In this time I enjoyed working together with many different people. Especially I remember: Xavier (for writing a Burg-Arras talk on the spot), Tomas (for his views on any topic), Heiko (for cleaning my apartment), Salomon (for his genius), Urs (for sleeping on the floor while awaiting a meeting with Wilfred), Roland (for leaving me with a mysterious server and now providing me with a - hopefully less mysterious - job), Dirk (for having to take the backseat in a Fiat cinquecento), Lukas (for showing me all about butane), Christian (for teaching me Informatik), Fred (for believing I quit and made a career taking care of elderly people), Jed (for not complaining about my endless discussions with Tomas), Indira (for taking care of us all), Regula (for her nice smile), Phil (for drinking a bowl of sangria after an already funny borrel) and Janez (for showing me Ljubljana).

Of course, most important were the people of my own generation that were here during most of my time within the group: Thereza and Roberto and Phaedra (because they introduced us to wine-tasting parties), Christine (because of her ability to see the funny side of everything), David (because of a nice game of tennis), Nico (because of inviting us all to Mainz), Alex (because everybody has his groundhog days), Alice (because she stayed sysadmin with me in good and in bad times), Haibo (because he never succumbed to the access of evil), Peter (because he just got the hang of everything), Mika (despite his attitude problem), Ulf (for his dry humour), Vincent (because he upheld the ideas of Mao), Michel (because of him showing us the desert) Tim (because of his teasing of Zrinka and sharing the last months with me) and especially Chris (because of all the 10 minutes of work we shared - *e.g.* “ene_ana”).

During the last months (and years), I enjoyed the presence of Daan (because he is helpful in every way), Valerie (because of sharing BIOMOS), Zrinka (because she keeps the group healthy with propolis), Jozi and Urban (because they bring joy to my home), Merijn (because of "it's Friday" discussions) and Annemarie (because I, too, like deep snow skiing), Bojan (because his guitar rules), Maria-Grazia (because of her one-of-a-kind tiramisu and her vivid opinions about politicians), Riccardo (because of his cooperativity and entropy), Cristina (because she brought Brazilian sunshine to us), Daniel (because "you know, it works for me"), Moritz (because of his Austrian charm), Claire (because of her french way of life), Lovorka (because of her refreshing attitude), Vreni (for sharing the GROMOS05 burden) and Maria (because of letting the Suns shine).

Almost all of the work presented in this thesis relies heavily on MD05 (GROMOSXX for us). I could never have written all of it alone. Some of the preparatory work was already started by Heiko, and brought to the next level by Roland with GROMOS++. But most of the long nights discussing design, new features or just hunting a tricky bug were shared with Chris who contributed a lot not only to the eventual success but also to a very pleasurable experience of getting there.

After Chris left Daniel took over that burden, but in addition we enjoyed the airport of Athens during the Olympic Games, wake boarding and sailing under the worst possible weather conditions and a lot of beach volley. This definitely brightens daily life.

Some of the most productive times were the months spent working together with semester students. Clara was the first one and she had therefore a bit of a rough time with finding and fixing bugs in the, at that time very fresh, program code. Without her I do not think GROMOSXX would have reached maturity. I am happy that we could finally write *Chapter 9*. Next to follow was Bettina. With her, the program was already much more stable, but as compensation the scientific problem seemed to be a bit evasive. Still, we somehow managed to get to *Chapter 7*, which is one of my favoured chapters of this thesis. During the last semester-project I was working with Pitschna. We were struggling to rotate rings, to bend helices and to successfully put glucose into chair conformations. This work resulted in *Chapter 6*. I had a lot of fun working on all these projects.

When I started Prisca made sure that everything went in a smooth way. After the move to Hönggerberg Daniela took over and I am very grateful for her ease in handling group matters and keeping Wilfred organized as well as for her friendship.

Obviously, a thesis is not written in one day but a rather longish process, building up on things one has done before. All my life I have been happy to receive the trust and support from my family: my parents Christina and Felice, my brothers Stefan and Daniel and my grand parents. Thank you. I also greatly appreciated the cordiality and easy acceptance I received as an additional member of the Rosenberger family.

Most important, especially also during this quite stressful last few months was the continuous support from Bea. Without her managing our household and often also my life next to a demand-

ing study and job of her own while still patiently listening to incomprehensible statistical mechanics problems getting this far would have been all but impossible. Even more important was the time we spent without thinking about work.

I would like to dedicate this thesis to her.

Contents

Kurzfassung

In den letzten fünfzig Jahren entwickelte sich die klassische Molekülsimulation in ein leistungsstarkes Werkzeug zur Untersuchung von biomolekularen Vorgängen auf atomarer Ebene. Dies wurde ermöglicht durch eine kontinuierliche Entwicklung der Simulationsverfahren und Programme. Im ersten Kapitel wird eine kurze Einführung in die klassische Simulation gegeben, mit speziellem Augenmerk auf Techniken um den durchsuchten Konfigurationsraum auf relevante Bereiche einzuschränken und auf die Berechnung von freien Energien aus den Simulationen. Als nächstes wird die neueste Version des Groningen Molecular Simulation Programmes GROMOS 05 vorgestellt. In dieser Version enthält das Programmpacket zwei Varianten des eigentlichen Simulationsprogrammes: eine erweiterte Version von PROMD, der traditionellen Simulationsmaschine von GROMOS, welche immer noch in FORTRAN geschrieben ist und das neu erstellte MD05 in C++. Dieses versucht durch Benutzen objektorientierter und generischer Programmiertechniken die Modularität und Lesbarkeit des Programmes zu erhöhen. Alle Simulationsverfahren, welche in den weiteren Kapiteln vorgestellt werden, sind in MD05 integriert.

In *Kapitel 3* wird ein kurzer Überblick über Methoden, welche den Konfigurationsraum effizient durchsuchen, gegeben. Danach folgt ein genauerer Blick auf die Berechnung von Entropien im Zusammenhang mit Kopie - Austausch in stochastisch dynamischen Simulationen. Dies geschieht anhand eines einfachen Testsystems, bei welchem vollständige Abdeckung des Konfigurationsraumes in Simulationen bei höheren Temperaturen erreicht werden kann. Bei diesen Temperaturen sollten somit auch die berechneten Eigenschaften unabhängig von der Simulationsmethode sein und erwartungsgemäss erfüllte Kopie - Austausch Simulation diese Bedingung. Bei tiefen Temperaturen ist keine vollständige Abdeckung mehr möglich. Kopie - Austausch Simulation kann unter diesen Umständen effizienter als die standard Simulationsmethode sein. Entropien von Simulationen werden häufig aufgeteilt in Rotations-, Translations- und Konfigurationsentropien. Dies wird erreicht durch eine Rotationsüberlagerung der Strukturen vor der Berechnung. Es konnte gezeigt werden, dass gewisse Überlagerungstechniken die Rotationsentropie bei tiefen Temperaturen gegenüber der Konfigurationsentropie stark bevorzugen.

Ein Verfahren, um eine fein-körnige (atomistische) und eine grob-körnige Representation eines Systems zugleich zu simulieren wird in *Kapitel 5* gegeben. Der momentane Zustand kann

durch einen körnigkeits Regler angegeben werden. Es ist möglich, die Körnigkeit während einer Simulation kontinuierlich von fein-körnig zu grob-körnig und zurück zu ändern, oder auch viele Kopien gleichzeitig bei unterschiedlicher Körnigkeit zu simulieren und durch Kopie - Austausch die fein-körnigen Kopien von dem schnelleren Absuchen des Konfigurationsraumes der grob-körnigen Kopien profitieren zu lassen.

Ein unterschiedlicher Ansatz, um die Effizienz von Simulationen zu steigern, wird in *Kapitel 6* aufgezeigt. Um eine freie Energie Differenz zwischen zwei Zuständen zu berechnen, muss die Simulation diese beiden Zustände verbinden. Damit dies schneller geschieht, kann man den Konfigurationsraum, welcher der Simulation zur Verfügung steht, einschränken. Wenn diese Zwänge so formuliert werden können, dass sie nichts zur potentiellen Energie und den Kräften in den Endzuständen beitragen, dann bleibt die berechnete freie Energie unabhängig von dem erzwungenen Pfad, der die Zustände verbindet. Diese Methode ist für Distanz- und Dihedralwinkelbeschränkungen ausgearbeitet und wird an der Ionen - Bindung eines zyklischen Peptides und an der Berechnung der freien Energie Differenz von zwei Zucker Konformationen gezeigt. Im zweiten Beispiel wird die Methode verglichen mit Resultaten erhalten aus einem Potential der mittleren Kraft.

Der Simulation auferlegte Zwänge können auch dazu benutzt werden, um experimentell bestimmte Eigenschaften zu reproduzieren. Eine Simulation kann von experimentellen Eigenschaften abweichen, wenn das verwendete Kraftfeld nicht für das Problem geeignet ist, oder wenn die Simulationszeit nicht ausreicht, um alle notwendigen Konfigurationen, die zum experimentellen Resultat beitragen, zu besuchen. Beide Probleme können durch geschickt gewählte Zwänge verkleinert werden. In *Kapitel 7* werden Zwänge vorgestellt, die sich während der Simulation anpassen. Dies wird durch eine Kombination mit der Technik der örtlichen Erhebung erreicht, in welcher die potentielle Energie während der Simulation für bestimmte Konfigurationen angehoben wird. Diese Kombination ergibt zwei Hauptsächliche Vorteile. Erstens ergibt sich durch die langsame Anpassung der potentiellen Energiefunktion eine minimale Beeinflussung der Simulation durch die zusätzlichen Zwänge. Und zweitens wird lokal effizienter nach einer Konfiguration gesucht, welche die experimentellen Eigenschaften wiedergibt.

In den zwei nächsten Kapiteln werden nicht mehr Zwänge angeschaut, sondern der Simulation Nebenbedingungen hinzugefügt, welche exakt erfüllt sein müssen. Diese Nebenbedingungen werden beispielsweise dazu gebraucht, um Bindungen starr zu machen. Es ist nun möglich, eine zusätzliche Flexibilität einzufügen, mit welcher die starren Bindungslängen sich in einem gewissen Mass an Veränderungen in der Umgebung anpassen können. Die Methode der flexiblen Nebenbedingungen wird angewendet in einer Simulation von Neopentan unter hohem Druck und in einer vergleichenden Studie von Modellen mit unterschiedlicher Behandlung von Bindungen.

Zum Schluss wird in *Kapitel 10* auf mögliche zukünftige Entwicklungen von Simulationsprogrammen und effizienten Algorithmen hingewiesen.

Summary

Over the last fifty years, continuous development of simulation algorithms and software has made classical molecular dynamics simulation into a powerful tool to investigate biomolecular processes in atomistic detail. After a short introduction of classical simulation with special attention to biased sampling of configurational space and some techniques to calculate free energies, the latest version of the Groningen Molecular Simulation package is introduced:
GROMOS05. In this version, the heart of the package is delivered in two variants. An enhanced version of PROMD, the FORTRAN simulation engine, and MD05, written in C++. MD05 strives for higher modularity and readability by making use of object oriented features and generic programming techniques. All algorithms presented in the other chapters are integrated into MD05.

A brief overview and classification of searching methods is given in *Chapter 3* before a closer look is taken at entropies calculated by replica-exchange stochastic dynamics simulations. For the simple test system complete sampling of configurational space can be achieved. Therefore the entropies calculated from standard stochastic dynamics simulations should match the ones obtained from the replica-exchange simulations. This was found to be true. At low temperature and incomplete sampling of the configurational space, replica-exchange simulation can be more efficient if the simulation parameters are carefully selected. The procedure of rotationally fitting structures to decompose entropy into configurational, translational and rotational entropies has a significant impact on this decomposition, often disfavouring configurational entropy at low temperatures.

An algorithm to combine fine-grained (atomistic) simulations with coarse-grained ones is introduced in *Chapter 5*. The algorithm allows to either continuously vary the grain-level of the simulation from fully fine-grained to coarse-grained and back, or, using replica exchange, simultaneously simulate a system at fine-grained, at coarse-grained and at some intermediate grain levels. The higher the grain level the bigger the simulation time-step may be, therefore increasing sampling efficiency. Through replica exchange, the replicas at lower grain-level can profit from the faster sampling available at the higher levels.

A different way to improve sampling is shown in the next chapter, where the simulation is restrained to follow a reaction coordinate from a given state *A* to a state *B* for obtaining the free energy difference between these two states. Opposed to standard methods, the restraining

functions are chosen to have zero potential energy and forces in the end states (*A* and *B*). This permits the calculation of a path independent free energy difference. The method, implemented in terms of distance and dihedral restraints is applied to ion-complexation in a cyclic peptide and the calculation of the relative stability of two different chair conformations of a glucopyranoside. For the second case a comparison with a potential of mean force calculated using dihedral-angle constraints is given.

Further on, restraints can be used to force a simulation to reproduce experimental data. If the force field is not well suited to represent a biomolecule or if the simulation time is not long enough to sample the relevant part of the configurational space, experimental properties might not be reproduced by simulations. Both problems may be overcome by adding restraints. In *Chapter 7* adaptive restraints based on local-elevation simulation are introduced. Adaptive restraints have two main advantages: First, as the restraint force slowly builds up over the simulation time, and only if the restraint is not fulfilled, a minimum of force is added, which leads to the least possible disturbance of the simulation. Second, as the adaptive restraint is based on local-elevation simulation, locally enhanced sampling is achieved, until the restraint is fulfilled.

In the next two chapters the focus is no longer on restraints but on constraints. But these are flexibilized using an approximate, but fast algorithm, which enables simulations to use a comparatively large time-step but still have the constraints slowly adapt to changes in the environment. Using this technique, fast bond vibrational frequencies can be avoided. The method is applied to neopentane simulations under high pressure and to a comparative study of different models of flexibility. In this study, the free energy difference of water and methanol is calculated by thermodynamic integration.

Finally, in *Chapter 10* an outlook of future development with regard to simulation software and configurational sampling is provided.

Publications

This thesis has led to the following publications:

Chapter 2:

Markus Christen, Philippe H. Hünenberger, Dirk Bakowies, Riccardo Baron, Roland Bürgi, Daan P. Geerke, Tim N. Heinz, Mika A. Kastenholz, Vincent Kräutler, Chris Oostenbrink, Christine Peter, Daniel Trzesniak, and Wilfred F. van Gunsteren,
" The GROMOS Software for Biomolecular Simulation: GROMOS05 "
J. Comput. Chem. **26** (2005), 1719–1751

Chapter 3:

Markus Christen and Wilfred F. van Gunsteren,
" On searching in, sampling of, and dynamically moving through conformational space of biomolecular systems: a review "
J. Comput. Chem. (2006), accepted

Chapter 5:

Markus Christen and Wilfred F. van Gunsteren,
" Multigraining: an algorithm for simultaneous fine-grained and coarse-grained simulation of molecular systems "
J. Chem. Phys. **124** (2006), 154106

Chapter 6:

Markus Christen, Anna-Pitschna E. Kunz, and Wilfred F. van Gunsteren,
" Sampling rare events using hidden restraints "
J. Phys. Chem. B **110** (2006), 8488–8498

Chapter 8:

Markus Christen and Wilfred F. van Gunsteren,
" An approximate but fast method to impose flexible distance constraints in molecular dynamics simulations "
J. Chem. Phys. **122** (2005), Art. No. 144106

Related publications:

Wilfred F. van Gunsteren, Dirk Bakowies, Roland Bürgi, Indira Chandrasekhar, Markus Christen, Xavier Daura, Peter Gee, Alice Glättli, Tomas Hansson, Chris Oostenbrink, Christine Peter, Jed Pitera, Lukas Schuler, Thereza Soares, and Haibo Yu,
"Molecular dynamics simulation of biomolecular systems"
CHIMIA **55** (2001), 856–860

Thereza A. Soares, Markus Christen, Kai F. Hu, and Wilfred F. van Gunsteren ,
" Alpha- and beta-polypeptides show a different stability of helical secondary structure "
Tetrahedron **60** (2004), 7775–7780

Wilfred F. van Gunsteren, Dirk Bakowies, Riccardo Baron, Indira Chandrasekhar, Markus Christen, Xavier Daura, Peter Gee, Daan P. Geerke, Alice Glättli, Philippe H. Hünenberger, Mika A. Kastenholz, Chris Oostenbrink, Merijn Schenk, Daniel Trzesniak, Nico F. A. van der Vegt, and Haibo B. Yu,
" Biomolecular modelling: goals, problems perspectives "
Angew. Chem. Int. Ed. **45** (2006), 4064–4092

Bettina Keller, Markus Christen, Chris Oostenbrink, and Wilfred F. van Gunsteren ,
" On using oscillating time-dependent restraints in MD simulation "
J. Biomol. NMR (2006), accepted