ETH zürich

ETH Library

Enhanced conformational sampling in molecular dynamics simulations

Doctoral Thesis

Author(s):
Schrøder Hansen, Halvor

Publication date:
2010

Permanent link:
https://doi.org/https://doi.org/10.3929/ethz-a-006193666

Rights / license:
In Copyright - Non-Commercial Use Permitted

This page was generated automatically upon download from the ETH Zurich Research Collection.
For more information, please consult the Terms of use.

Diss. ETH No. 19141

Enhanced conformational sampling in molecular dynamics simulations

A dissertation submitted to
ETH ZURICH

for the degree of
Doctor of Sciences

presented by
HALVOR SCHRØDER HANSEN
Master i Teknologi, Norwegian University of Science and Technology (NTNU), Trondheim
born July 6, 1983
citizen of Norway

accepted on the recommendation of
Prof. Dr. Wilfred F. van Gunsteren, examiner
Dr. Philippe H. Hünenberger, Prof. Dr. Martin Kröger, PD Dr. Jürgen Schlitter, co–examiners

2010

Summary

A computational scientist without an algorithm is like a carpenter without a tool. It is only with the help of appropriate algorithms that he/she can address, attack, understand and (hopefully) ultimately solve problems. This thesis represents a journey through the development process of a set of enhanced-sampling algorithms for use in molecular simulations, that will hopefully once find their way into the standard simulation toolbox.

Chapter 1 provides a short introduction into the topic of molecular simulation. The basic theory and choices underlying molecular models are discussed, as well as their current problems and limitations. A brief introduction into statistical mechanics and free-energy calculations is given as well.

In Chapter 2, a method is proposed to combine the local elevation (LE) conformational searching and the umbrella sampling (US) conformational sampling approaches in to a single local elevation umbrella sampling (LEUS) scheme for molecular dynamics (MD) simulations. In this scheme an initial (relatively short) LE build-up (searching) phase is used to construct an optimized biasing potential within a subspace of conformationally relevant degrees of freedom, that is then used in a (comparatively longer) US sampling phase. This scheme dramatically enhances the sampling power of MD simulations, taking advantage of the fact that the preoptimized biasing potential represents a reasonable approximation to the negative of the free energy surface in the considered conformational subspace. The method is applied to the calculation of the relative free energies of β-D-glucopyranose ring conformers in water within the GROMOS 45A4 force field.

Chapter 3 presents a reoptimization of the GROMOS 53A6 force field for hexopyranose-based carbohydrates into a new version $56A_{CARBO}$. A central part of this reoptimization is the application of the LEUS method for enhancing the sampling along slow degrees of freedom, allowing for the inclusion of the corresponding conformational free energies into the parameterization process. This reoptimization repairs a number of shortcomings of the 53A6 parameter set (nearly identical to 45A4 in the context of carbohydrate compounds), and extends the scope of the force field to properties that had not been included previously into the parameterization procedure (ring conformational preferences, relative free energies of anomers and epimers). The new $56A_{CARBO}$ force field is characterized by : (*i*) the formulation of systematic build-up rules for the automatic generation of force-field topologies; (*ii*) the systematic use of enhanced-sampling methods into the parameterization procedure; (*iii*) an extensive validation against available experimental data in solution. The force field is shown to reproduce well the relative free energies of ring conformers, anomers, epimers, hydroxymethyl rotamers and glycosidic linkage conformers.

In Chapter 4 a new method, fragment-based local elevation umbrella sampling (FB-LEUS), is proposed as an attempt to expand the scope of the LEUS approach to solvated polymers with more than a few relevant degrees of freedom. The FB-LEUS scheme involves an US sampling phase that relies on a superposition of low-dimensionality biasing potentials optimized using LEUS at the fragment level. The feasibility of this approach is tested using polyalanine and polyvaline oligopeptides. Two types of fragment-based biasing potentials are distinguished: (*i*) the basin-filling (BF) potentials act so as to "fill" free-energy basins up to a prescribed free-energy level above the global minimum; (*ii*) the valley-digging (VD) potentials act so as to "dig" valleys between the free-energy minima, preserving barriers of a prescribed magnitude. For the systems considered, the application of FB-LEUS using potentials of the BF-type shows an impressive enhancement of the searching power, but this increase is largely offset by a deterioration of the statistical efficiency. For VD-type potentials, the searching enhancement is less significant, but the statistical efficiency is only slightly decreased, resulting in a more favorable trade-off. For the oligopeptide systems considered, the FB-LEUS(VD) approach still does not represent a significant improvement over plain MD. However,

for systems characterized by higher conformational barriers (*e.g.* oligosaccharides), the improvement is expected to be more significant.

As shown in Chapter 4, sampling-enhancement approaches may open up the conformational space accessible to sampling to such an extent that the statistical efficiency is deteriorated, leading to slow convergence and low practical usefulness, especially when many relevant degrees of freedom are considered. Chapter 5 proposes a new method, ball-and-stick local elevation umbrella sampling (B&S-LEUS), applicable to subspaces of any dimensionality, promoting enhanced sampling within a restrained conformational subspace. The B&S-LEUS method consists of the following steps : (*A*) choice of a reduced conformational subspace permitting the definition of the relevant states; (*B*) representation of the relevant states by means of spheres ("balls"); (*C*) definition of a set of lines ("sticks") connecting these spheres; (*D*) unification of the biasing potentials associated with the union all the spheres and lines (active subspace) into a single biasing potential; (*E*) build-up of the memory using the local elevation (LE) procedure, leading to a biasing potential enabling a nearly uniform sampling of the active subspace; (*F*) generation of a biased ensemble of configurations using this preoptimized biasing potential, following an umbrella sampling (US) approach; (*G*) calculation of the relative free energies of the states *via* reweighting and state assignment. The main characteristics of this approach are : (*i*) a low internal dimensionality; (*ii*) a minimal irrelevant volume; (*iii*) a problem-adapted geometry. This B&S-LEUS method is applied to three systems : (*i*) a solvated (blocked) alanine monopeptide (two-dimensional conformational subspace), to illustrate the versatility of the method in promoting the sampling of arbitrary regions of the Ramachandran map; (*ii*) a solvated polyalanine decapeptide, to evaluate the relative free energies of three different types of helices (π, α and 3_{10}); (*iii*) an artifical hexopyranose, termed the ”mother” of all D-hexopyranoses, constructed as a hybrid of all D-hexopyranose stereoisomers, to calculate the relative free energies of the corresponding 32 isomers, anomers and chair conformers from a single simulation. The method provides the user with a very high level of control, *via* the accurate specification of the regions of the conformational subspace considered relevant (*i.e.* which should be sampled) or irrelevant (*i.e.* which should be avoided).

Finally, Chapter 6 provides concluding remarks and an outlook.

Kurzfassung

Ein Forscher des wissenschaftlichen Rechnens ohne Algorithmus ist wie ein Schreiner ohne Werkzeug. Nur mithilfe geeigneter Algorithmen kann er Probleme erkennen, verstehen, angehen und (hoffentlich) schließlich lösen. Diese Dissertation dokumentiert eine Reise durch die Entwicklungsprozesse eines Sets von Algorithmen, welche das Sampling bei Molekülsimulationen erweitern sollen, und die hoffentlich einmal den Weg in den Werkzeugkasten des Simulierens finden werden.

Kapitel 1 gibt eine kurze Einführung zum Thema Molekülsimulationen. Hierbei werden die Grundlagen der Theorie und die Wahlmöglichkeiten hinter einem Molekülmodell sowie aktuelle Probleme und Beschränkungen diskutiert. Zudem erfolgt eine knappe Einführung in die statistische Mechanik und die freie Energieberechnung.

In Kapitel 2 wird eine Methode vorgestellt, welche zwei bekannte Verfahren kombiniert. Zum einen die "local elevation"(LE)-Methode für das Searching und zum anderen die "umbrella sampling"(US)-Methode für das Sampling des Konfigurationsraumes. Die neue Methode wird "local elevation umbrella sampling" (LEUS) genannt. Sie beinhaltet eine (relativ kurze) LE-Phase, die ein optimiertes Verzerrungspotential in einem Unterraum von Konformations-relevanten Freiheitsgraden erzeugt, gefolgt von einer (vergleichsweise längeren) US-Phase. Da das voroptimierte Verzerrungspotential sich an die negative freie Energie in dem obengenannten Unterraum annähert, erweitert die LEUS-Methode das Sampling drastisch. Die Methode wird dazu benutzt, die relative freie Energie von den Ringkonformationen der β-D-Glukopyranose innerhalb des GROMOS 45A4-Kraftfeldes zu berechnen.

Kapitel 3 stellt eine weitere Optimierung des GROMOS 53A6-Kraftfeldes für Hexopyranose-basierte Kohlenhydrate vor. Das hierbei entwickelte Kraftfeld wird $56A_{CARBO}$ genannt. Ein Hauptbestandteil dieser erneuten Optimierung ist die Anwendung der LEUS-Methode für das erweiterte Sampling entlang langsamer Freiheitsgrade, was eine Einbeziehung der dazugehörigen freien Konformationsenergien in dem Parametrisierungsprozess ermöglicht. Die erneute Optimierung verbessert verschiedene Mängel des 53A6-Kraftfeldes (welches übrigens fast identisch mit dem 45A4-Kraftfeld ist), und sie erweitert den Anwendungsbereich des Kraftfeldes um solche Eigenschaften, die in früheren Parametrisierungsprozessen nicht berücksichtigt worden sind (bevorzugte Ringkonformationen, relative freie Energien der Anomere und Epimere). Das neue $56A_{CARBO}$-Kraftfeld ist gekennzeichnet durch: (*i*) die Formulierung systematischer Aufbauregeln für die automatische Generierung von Kraftfeldtopologien; (*ii*) die systematische Anwendung von Methoden des erweiterten Samplings; (*iii*) eine extensive Validierung gegen verfügbare experimentelle Daten in Lösung. Das Kraftfeld reproduziert gut die relative freie Energie der Ringkonformere, Anomere, Epimere, Hydroxymethyl-Rotamere, und der glykosidischen Bindungskonformere.

In Kapitel 4 wird eine neue Methode, "fragment-based"-LEUS (FB-LEUS), vorgestellt als ein Versuch, den Anwendungsbereich der LEUS-Methode auf Polymere mit mehr als nur wenigen relevanten Freiheitsgraden zu erweitern. Die FB-LEUS-Methode beinhaltet eine US-Phase, die auf einer Superposition von niedrig-dimensionalen Verzerrungspotentialen basiert. Die Verzerrungspotentiale werden von LEUS-Simulationen abgeleitet, welche auf dem Level von Fragmenten durchgeführt werden. Die Durchführbarkeit dieser Methode wird an den Oligopeptiden Polyalanin und Polyvalin getestet. Es werden zwei Typen von Verzerrungspotentialen unterschieden: (*i*) das "basin-filling"(BF)-Potential, bei dem das Potential die freien Energieminima zu einer vorgeschriebenen Höhe auffüllt; (*ii*) das "valley-digging"(VD)-Potential, bei dem das Potential "Täler gräbt" zwischen freie Energieminima, wobei freie Energiebarrieren bis zu einer vorgeschriebenen Höhe erhalten bleiben. Für die betrachteten Systeme zeigt FB-LEUS(BF) eine beeindruckende Erhöhung des Searchings, jedoch wird diese Erhöhung gleichzeitig von einer Verminderung der statistischen Effizienz aufgehoben. Für VD-Potentiale ist die Erhöhung des Searchings

kleiner, aber die statistische Effizienz wird hier nur wenig vermindert, so dass diese Potentiale ein vergleichsweise vorteilhafteres Sampling vorweisen. Für die betrachteten Oligopeptide zeigen die FB-LEUS(VD)-Potentiale keine signifikante Verbesserung gegenüber einer normalen MD Simulation. Dennoch wird bei Systemen mit höheren Konformationsbarrieren (z.B. bei Oligosacchariden) ein erweitertes Sampling erwartet.

Wie in Kapitel 4 gezeigt worden ist, führt die Öffnung des zugänglichen Konformationsraumes durch erweitertes Sampling auch zu einer Verringerung der statistischen Effizienz. Dies kann zur Folge haben, dass die Simulationsresultate langsamer konvergieren und damit die praktische Anwendbarkeit des Algorithmus vermindert wird. Das ist vor allem dann der Fall, wenn viele Freiheitsgrade berücksichtigt werden. Kapitel 5 schlägt deshalb eine neue Methode vor, "ball-and-stick"-LEUS (B&S-LEUS), die auf beliebig viele Freiheitsgrade anwendbar ist und nur zu erweitertem Sampling in einem eingeschränkten Teil des Konformationsunterraumes führt. Die B&S-LEUS-Methode beinhaltet die folgenden Schritte: (*A*) Wahl eines reduzierten Konformationsunterraumes, der die Definition der relevanten Zustände erlaubt; (*B*) Darstellung der relevanten Zustände mit Kugeln ("balls"); (*C*) Definition von einem Set von Linien, welche die Kugeln verbinden; (*D*) Vereinigung der einzelnen Verzerrungspotentiale (einzelne Kugeln und Linien) zu einem gemeinsamen Verzerrungspotential (Union der Kugeln und Linien; aktiver Unterraum); (*E*) Aufbau des Gedächtnisses ("memory") mithilfe der LE-Methode zu einem Verzerrungspotential, das zu einer uniformen Verteilung in dem aktiven Unterraum führt; (*F*) Generierung eines verzerrten Ensembles durch dieses Verzerrungspotential mithilfe der US-Methode; (*G*) Berechnung der relativen freien Energien der Zustände mittels Rückgewichtung und Zustandszuweisung. Die Haupteigenschaften der Methode sind: (*i*) eine niedrige Eigendimensionalität; (*ii*) ein minimales irrelevantes Volumen; (*iii*) eine problemangepasste Geometrie. Die B&S-LEUS-Methode wird auf drei Systeme angewandt: (*i*) ein Alaninmonopeptid in Lösung, um die Vielseitigkeit der Methode im Sampling von willkürlichen Bereichen des Ramachandrandiagramms zu zeigen; (*ii*) ein Polyalanindekapeptid in Lösung, um die relative freie Energie von drei Helix-Typen (π, α und 3_{10}) zu berechnen; (*iii*) eine künstliche Hexopyranose, "die Mutter aller D-Hexopyranosen", welche als ein Hybrid von allen Stereoisomeren des D-Hexopyranoses konstruiert ist, um die relative freie Energie von den 32 Isomeren, Anomeren und Sesselkonformeren mit einer einzigen Simulation zu berechnen. Die Methode gibt dem Nutzer ein hohes Maß an Kontrolle, da genau festgelegt werden kann, welche Teile des Konformationsunterraums relevant (Sampling bevorzugt) und irrelevant (Sampling unerwünscht) sind.

Im abschließenden Kapitel 6 folgen Fazit und Ausblick.

Sammendrag

En forsker i beregningsvitenskap uten algoritmer er som en snekker uten verktøy. Det er først og fremst ved hjep av passende algoritmer at han kan undersøke, angripe, forstå og (forhåpentligvis) løse problemer. Denne doktoravhandlingen representerer en reise gjennom en utviklingsprosess for et sett med algoritmer for utvidet prøvetaking (”sampling”) for bruk i molekylsimuleringer, algoritmer som forhåpentligvis vil finne sin vei inn i simuleringenes verktøykasse.

Kapittel 1 gir en kort introduksjon til emnet molekylsimuleringer. Den grunnleggende teorien bak molekylære modeller diskuteres, i tillegg til aktuelle problemer og begrensinger. Kapittelet gir også en innledning til statistisk mekanikk og fri energi-beregninger.

I kapittel 2 foreslås en metode som kombinerer ”local elevation” (LE)-metoden, en metode for utforskning (”searching”) av konformasjoner, og ”umbrella sampling” (US)-metoden, en metode for prøvetaking av konformasjoner, til en felles metode kalt ”local elevation umbrella sampling” (LEUS). I denne metoden benyttes innledningsvis en (relativ kort) LE-fase for a bygge et optimert energibias innenfor et underrom bestående av konformasjonsrelevante frihetsgrader. Påfølgende blir dette energibiaset benyttet i en (forholdsvis lengre) US-fase. Siden det optimerte energibiaset er en rimelig tilnærming til det negative av den frie energien i det ovennevnte underrommet, gir LEUS-metoden en drastisk forbedring av prøvetakingsegenskapene til en MD-simulering. Metoden blir i dette kapittelet benyttet til å beregne den relative frie energien til ringkonformasjoner av β-D-glukopyranose ved hjelp av kraftfeltet GROMOS 45A4.

Kapittel 3 presenterer en reoptimering av kraftfeltet GROMOS 53A6 for heksopyranosebaserte karbohydrater, til et nytt kraftfelt med navnet $56A_{CARBO}$. En sentral del av denne reoptimeringen er bruken av LEUS-metoden for å utvide prøvetakingen langs trege frihetsgrader slik at den frie energien for konformasjonene langs disse frihetsgradene kan inkluderes i parametriseringsprosessen. Denne reoptimeringen reparerer en del mangler som er tilstede i kraftfeltet 53A6 (som forøvrig er nesten identisk med 45A4 for karbohydrater) og utvider bruksområdene til kraftfeltet til å også inkludere egenskaper som ikke har vært inkludert i tidligere parametriseringer (ringkonformasjonspreferanser og den relative frie energien til anomerer og epimerer). Det nye kraftfeltet $56A_{CARBO}$ kan karakteriseres av: (*i*) en formulering av systematiske oppbygningsregler for automatisk generering av kraftfelttopologier; (*ii*) en systematisk bruk av algorimer for utvidet prøvetaking i parametriseringsprosedyren; (*iii*) en omfattende validering mot tilgjengelige eksperimentelle data i løsningsfase. Kraftfeltet viser seg å kunne reprodusere den relative frie energien mellom ringkonformasjoner, anomerer, epimerer, hydroksymetyl-rotamerer og konformasjoner rundt glykosebindinger.

I kapittel 4 foreslås en ny metode, fragment-basert-LEUS (FB-LEUS), som et forsøk på å utvide LEUS-metoden til polymerer (i løsningsfase) med flere enn et fåtall frihetsgrader. FB-LEUS-metoden involverer en US-fase, som er konstruert som en superposisjon av lav-dimensjonale energibias, optimert ved bruk av LEUS-metoden på fragmenter. Gjennomførbarheten til denne metoden blir testet på oligopeptider av polyalanin og polyvalin. To typer fragment-baserte energibias blir presentert: (*i*) ”basin-filling” (BF)-bias, konstruert for å fylle opp fri energi-minima til et forhåndsbestemt nivå over det globale minimumet; (*ii*) ”valley-digging” (VD)-bias, konstruert for å grave daler mellom fri energi-minima, og samtidig begrense høyden på barrierene til et forhåndsbestemt nivå. For systemene som betraktes, viser BF-biasene en imponerende forbedring av utforskningsegenskaper, men denne forbedringen blir spist opp av en samtidig forringelse av statistisk effektivitet. For VD-biasene derimot, er forbedringen av utforskningsegenskaper mindre. Samtidig er forringelsen av statistisk effektivitet også mindre slik at den totale effektiviteten i dette tilfellet er bedre. For systemene som betraktes viser FB-LEUS(VD)-metoden fortsatt ikke en signifikant forbedring i forhold til en standard MD simulering. Likevel forventes metoden å være fordelaktig

for systemer som er karakterisert av høyere konformasjonsbarrierer, slik som f.eks. oligosakkarider.

Som vist i Kapittel 4, kan en algoritme for utvidet prøvetaking åpne opp et konformasjonsrom for prøvetaking i en slik grad at den statistiske effektiviteten forringes drastisk, med det resultat at simuleringsresultatene konvergerer sakte og algoritmen blir mindre anvendbar i praksis. Dette viser seg spesielt å være tilfelle når mange relevante frihetsgrader taes i betraktning. Kapittel 5 foreslår en ny metode "ball-and-stick"(kule-og-pinne)-LEUS (B&S-LEUS) som kan benyttes på systemer av hvilken som helst dimensjon, og fremmer utvidet prøvetaking innenfor et begrenset konformasjonsunderrom. B&S-LEUS-metoden består av følgende steg: (*A*) valg av et redusert konformasjonsunderrom som muliggjør en definasjon av relevante tilstander; (*B*) representasjon av de relevante tilstandene ved hjelp av kuler (”balls”); (*C*) definering av et sett med linjer (”sticks”) som forbinder disse kulene; (*D*) sammenslåing av energibiasene som unionen av alle kuler og linjer (aktivt underrom) til ett enkelt energibias; (*E*) oppbygning av et minne, ved hjelp av LE-metoden, som fører til et energibias som gir tilnærmet enhetlig prøvetaking i det aktive underrommet; (*F*) generering av en populasjon (”ensemble”) av konfigurasjoner med dette biaset (US-metoden); (*G*) beregning av relative frie energier til tilstandene ved hjelp av en tilstandstilordning og en omvekting. Hovedegenskapene til denne metoden er: (*i*) en lav egendimensjonalitet; (*ii*) et minimalt irrelevant volum; (*iii*) en problemtilpasset geometri. B&S-LEUS metoden blir anvendt på tre systemer: (*i*) et (blokkert) alaninmonopeptid i løsningsfase (todimensjonalt konformasjonsunderrom), for å illustrere allsidigheten til metoden, med hensyn på å fremme prøvetaking av vilkårlige deler av Ramachandrandiagrammet; (*ii*) et polyalanindekapeptid i løsningsfase, for å beregne relativ free energi av tre forskjellige typer helikser (π, α og 3_{10}); (*iii*) en kunstig heksopyranose, med det passende tilnavnet ”alle D-heksopyranosers mor”, konstruert som en hybrid av alle stereoisomerer av D-heksopyranose, for å beregne de relative frie energiene til de 32 mulige isomerer, anomerer og stolkonformasjoner fra en enkelt simulering. Metoden gir brukeren en stor grad av kontroll, via en nøyaktig spesifikasjon av hvilke deler av konformasjonsunderrommet som er relevante (e.g. skal utsettes for prøvetaking) og hvilke som er irrelevante (e.g. skal ikke utsettes for prøvetaking).

Til slutt gir kapittel 6 konklusjoner, anbefalinger og fremtidsvisjoner.