Diss. ETH No. 19479

Refinement of molecular simulation methodology: development of polarisable solvents and techniques to enhance configurational sampling

A dissertation submitted to the
ETH ZÜRICH

for the degree of
Doctor of Sciences

presented by
ANNA-PITSCHNA ENRICA KUNZ
Dipl. Sc. Nat. ETH
born 23. September 1980
citizen of Thun - Switzerland

accepted on the recommendation of
Prof. Dr. Wilfred F. van Gunsteren, examiner
Prof. Dr. Matthias Troyer, co-examiner
Prof. Dr. Xavier Daura, co-examiner

2011

Allen meinen Eltern gewidmet.

Acknowledgements

First I would like to express my gratitude towards Wilfred van Gunsteren, Matthias Troyer and Xavier Daura for agreeing to attend my thesis defence that late in the year.

Wilfred I would like to thank further for his way of finding positive aspects in all results even the ones that brought me close to insanity, for keeping the whole IGC group happy and generating a productive but relaxed working atmosphere. There is no point in thanking Wilfred without thanking Jolande van Gunsteren-Bolt to whom I am grateful for not only opening her house for farewells but also for keeping Wilfred alive and in good spirit.

Next I would like to thank the people that lured me into my PhD. Basically I would have to thank Wilfred again for his delightful lectures but now it is time to thank Daniel Trzesniak with whom I did my ICG3 project, Markus Christen who oversaw a semester project of mine, and Merijn Schenk who supervised my diploma thesis. I also have to thank Zrinka Gattin, who promised me a pair of stripey socks when I would join the group. I am still waiting for these socks, but Zrinka made up for that by being a good friend.

The next batch of people I would like to thank is the system administrators. The team taming Beaver was Merijn Schenk, Moritz Winger, Alexandra Choutko and Sereina Riniker. Daan Geerke, Clara Christ, Katharina Meier and Jožica Dolenc were taking care of Squirrel. The rest of the zoo of Animals was tended by Halvor Hansen and Nathan Schmid. Shining light on Ray were Markus Christen, Bettina Keller, Maria Reif, Denise Steiner, Andreas Eichenberger and Oliwia Szklarczyk. Since most of the simulations for this thesis were done on the BRUTUS cluster, I would like to acknowledge its support team as well.

The people that explicitly helped me with my scientific projects were Daan Geerke and Merijn Schenk for the polarisability, Christina Pereira, Lovorka Perić-Hassler and Halvor Hansen for projects containing carbohydrates, Irene Schober, Philippe Hünenberger and Stephan Bachmann for ions in water with carbohydrates, Denise Steiner for replica exchange, Sereina Riniker for dielectric permittivity, and Zhixiong Lin and Wei Huang for β-peptides. I feel much obliged to everybody who worked on GROMOS, which is basically the whole group.

I had the privilege to spend my whole PhD time in the office next to the common room and to share the room with Daniel Trzesniak, Daan Geerke, Clara Christ, Jane Allison, Jožica Dolenc, Katharina Meier, Andreas Eichenberger and Wei Huang. They made

my working days delightful.

I visited Groningen in February and March 2008 and had a swell time with Alex de Vries, Siewert Jan Marrink, Durba Sengupta and Jelger Risselada.

I am indebted to the good spirits of our group, our secretaries Daniela Kalbermatter, Anna Nedic Eigenmann and Carmen Ziegler without whom the whole group would fall apart. For all the lunch breaks shared, the entertaining coffee breaks, the amusing farewells and the joyful barbeques I would like to thank Dongqi Wang, Dirk Bakowies, Michel Cuendet, Bojan Žagrović, Claire Samson, Magdalena Siwko, Stefano Borini, Niels Hansen, Monika Laner and all other members of the IGC group. I enjoyed being part of such a inspiring environment.

My peer groups outside of IGC were on the one hand my Friday Physicists and on the other hand Möne Grütter. They put the necessary final pressure on me to finish my PhD. As did Peter Limacher, who had to endure a lot of my bureaucratic questions.

Bruno Horta I have to thank for not only helping me with scientific problems but also taking good care of me whenever I need it.

Last but most definitely not least I would like to thank my parents, Heinrich Kunz and Cornelia Schellenberg-Rauscher, for being loving and supportive and for bringing me this far.

Contents

Summary

Molecular dynamics (MD) simulations can help to understand and interpret experimental findings and to show a dynamic picture of the world at a microscopic level. However, how well this picture represents reality depends on the choice of the explicitly treated degrees of freedom, the interaction function governing the motion along them, the sampling time and the system size, which choices are limited by the available computing power. In Chapter 1 these issues are described and an introduction to the subjects of this thesis is given.

The steady growth in available computer power can be invested to improve each of these aspects. Chapters 2 and 3 present the introduction of electronic dipolar polarisability in molecules to obtain a better description of the non-bonded interaction between atoms. Chapter 2 describes a polarisable water model with a non-linear, damped polarisability, Chapter 3 a polarisable model of carbon tetrachloride. The models use a single charge-on-spring approach to include the dipolar polarisability to better reproduce experimental properties especially the dielectric ones, but are a factor 1.7 to 5 slower than the non-polarisable ones. The dielectric permittivity, which tended to be too large for other GROMOS polarisable water models, was reduced to 69.8 for COS/D compared to 78.4 as experimental value and was raised to the experimental value of 2.24 for CCl_4.

Since even the growth in available computer power does not and will not solve all problems of sampling relevant regions of the energy hypersurface, the development of methods to enhance sampling is still of interest. Chapters 4 to 7 describe investigations of the method of adiabatic decoupling with temperature or force scaling. The detailed theory and its application to a test system of liquid water can be found in Chapter 4. The sampling of the heavy mass degrees of freedom can be significantly enhanced depending on their number and the particular technique used. For certain water systems, the diffusion of a small number of water molecules could be increased up to 35 fold. An application of the method to enhance the sampling of ionic degrees of freedom in aqueous solution using adiabatic decoupling of ionic and water degrees of freedom with subsequent translational temperature scaling of the former is described in Chapter 5. The configurational sampling of ions in aqueous solution could be considerably enhanced as reflected by an increase in ion diffusion by a factor 15. An application to peptide folding in which the peptide and solvent degrees of freedom are adiabatically decoupled is described in Chapter 6. In this case the sampling of the folding equilibrium is only marginally enhanced compared to standard molecular dynamics simulation. In Chapter 7 the application of adiabatic decoupling to a much larger system, a protein in

water, is described. It appears that the sampling of the possible conformations of a 11-residue loop in a 124-residue protein can be much enhanced using either temperature or force scaling of the loop degrees of freedom. In all cases the effects of a variation of various scaling factors for the masses, temperature or forces of the "heavy" degrees of freedom were investigated.

Chapter 8 gives suggestions how the described methodology could be extended, further tested and developed.

Zusammenfassung

Molekulardynamische (MD) Simulationen können beim Verstehen und Interpretieren von experimentellen Resultaten helfen und sie zeigen ein dynamisches Bild der Welt im Kleinsten. Doch wie gut dieses Bild die Wirklichkeit wiedergibt, hängt von der Wahl der explizit behandelten Freiheitgrade, der Wechselwirkungsfunktion der Bewegung entlang diesen, der Beprobungsdauer und der Systemgrösse ab. Die Wahl im Bezug auf diese Punkte ist beschränkt durch die verfügbare Rechnerleistung. Im Kapitel 1 werden diese beschrieben und es wird eine Einführung in die Themen dieser Dissertation gegeben.

Das stete Wachstum der verfügbaren Rechnerleistung kann zur Verbesserung dieser Aspekte eingesetzt werden. In Kapitel 2 und 3 wird elektronische Dipolpolarisierbarkeit eingeführt um eine bessere Beschreibung der elektrostatischen Wechselwirkungen zwischen Atomen zu erhalten. Kapitel 2 beschreibt ein polarisierbares Wassermodell mit nicht linearer, gedämpfter Polarisierbarkeit, Kapitel 3 ein polarisierbares Modell für Tetrachlormethan. Die Modelle verwenden einen Ladung-an-Feder-Ansatz für die Dipolpolarisierbarkeit. Sie sind damit in der Lage experimentelle Werte, vor allem dielektrische, besser zu reproduzieren, sie sind aber auch 1.7- bis 5-mal langsamer als nicht polarisierbare Modelle. Die Dielektrizitätskonstante, die für andere polarisierbare Wassermodelle von GROMOS eher zu gross ist, wurde für das COS/D Modell auf 69.8 reduziert, verglichen mit 78.4 als experimentellem Wert, und wurde für CCl_4 auf den experimentellen Wert von 2.24 erhöht .

Da auch das Wachstum der verfügbaren Rechnerleistung nicht in der Lage ist — und es niemals sein wird — das Problem der Beprobung relevanter Regionen der Energiehyperfläche zu lösen, bleibt das Entwickeln von Methoden zur Verbesserung der Beprobung weiterhin interessant. Kapitel 4 bis 7 beschreiben die Überprüfung der Methode der adiabatischen Entkopplung mit Temperatur- oder Kraftanpassung. Die detaillierte Theorie und ihre Anwendung für ein Testsystem aus Wasser wird im Kapitel 4 gegeben. Die Beprobung der schweren Freiheitsgrade kann erheblich ver-bes-sert werden, abhängig von ihrer Zahl und welche Technik benutzt wurde. Für bestimmte Wassersysteme konnte die Diffusion der wenigen schweren Wassermoleküle 35-fach erhöht werden. Eine Anwendung der Methode der adiabatischen Entkopplung und folgender Anpassung der Translationstemperatur zur Verbesserung der Beprobung von ionischen Freiheitsgraden in wässeriger Lösung wird in Kapitel 5 beschrieben. Die Beprobung des Phasenraums für Ionen in wässeriger Lösung konnte wesentlich verbessert werden, was sich in einer Erhöhung der Ionendiffusion um einen Faktor

15 zeigte. Die Anwendung für Peptidfaltung, bei der die Peptid- und Lösungsmittelfreiheitsgrade adiabatisch entkoppelt wurden, wird in Kapitel 6 erörtert. In diesem Fall ist die Beprobung des Faltungsgleichgewichts nur leicht verbessert im Vergleich zur molekulardynamischen Standardsimulation. Kapitel 7 beschreibt die Anwendung der adiabatischen Entkopplung für ein grösseres System, ein Protein in Wasser. Es scheint, dass die Beprobung der möglichen Konformationen einer Schleife aus 11 Aminosäuren in einem 124 Aminosäuren-Protein signifikant verbessert werden kann, wenn die Temperatur oder die Kraft auf die Schleifenfreiheitsgrade angepasst wird. In allen Fällen wurde der Effekt, der das Ändern der unterschiedlichen Anpassungsgrössen für die Temperatur, die Kraft und die Masse der schweren Freiheitsgrade hat, untersucht.

Kapitel 8 gibt Vorschläge, wie die beschriebenen Methoden weiter getestet, entwickelt und erweitert werden könnten.

Resümee

Molekulardynamischi (MD) Simulazione chöi häufe bim Verstah u Interpretiere vo experimentelle Ärgäbnis u sy zeygenes bewegts Biud vo dr Wäut im Chlinschte. Wie guet das Biud isch, hanget aber vo dr Wau vo de usdrücklech behandlete Freyheytsgrad, vo dr Wächsuwürkigsfunktion, wo d Bewegig entlang disne bestimmt, vo dr Abtaschtigszyt und vo dr Systemgrössi ab. D Wau isch beschränkt dürdi vorhandeni Rächnerkapazität. Im Kapitu 1 wärde d Pünkt, wo zur Wau stöh, beschribe u es wird än lileytig zu dere Diss gäh.

Ds kontinuierliche Wachse vo dr Rächnerkapazität chamä zur Verbesserig vo dene Pünkt iisetze. Im Kapitu 2 u 3 wird die elektrischi Dipoupolarisierbarkeyt iigfüert für e besseri Beschribig vo de elektrostatische Wächsuwürkige zwüsche Atom. Kapitu 2 beschribt es polarisierbars Wassermodäu mit nid-linearer, dämpfter Polarisierbarkeyt, Kapitu 3 es polarisierbars Modäu für Tetrachlormethan. Die Modäu verwände ä Ladigare-Fädäre-Aasatz für d Dipoupolarisierbarkeyt. Si sy dermit i dr Lag experimenteui Wärt, vor auem dielektrischi, besser z reproduziere. Aber si sy o 1.7 bis 5 mau langsamer als nid polarisierbari Modäu. D Dielektrizitätskonstante, wo für anderi polarisierbari Wassermodäu vo GROMOS ender z gross sy, si mit em COS/D Modäu uf 69.8 reduziert worde, u uf dä experimenteui Wärt vo 2.24 erhöht für CCl_4.

Wüu o ds Wachstum vo dr Rächnerkapazität nid i dr Lag isch — us o nie wird sy — d Problem vo dr Abtaschtig vo de relevante Regione vo dr Energiihüperflächi z löse, blybt ds Entwickle vo Methode zur Verbesserig vo dr Abtaschtig intressant. Kapitu 4 bis 7 beschribe d Überprüefig vo dr Methode vom adiabatische Entkopple mit Tämpäratur- oder Chraftaapassig. Di detaillierti Theory und iri Aawändig uf es Teschtsyschtem us Wasser wird im Kapitu 4 gäh. D Abtaschtig vo de schwäre Freyheytsgrad cha dermit erheblich verbesseret wärde, je nach ihrere Zau und dr verwändete Technik. Für bestimmti Wassersystem het d Diffusion vo de schwärä Wassermolekül chönne um z 35 fache erhöht wärde. Ä Aawändig vo dr Methode zur adiabatische Entkopplig und dr Aapassig vo dr Translationstämpäratur zur Verbesserig vo dr Abtaschtig vo ionische Freyheytsgrad i wässriger Lösig wird im Kapitu 5 beschribe. D Abtaschtig vom Phaseruum für Ione i wässriger Lösig het dütlech chönne verbesseret wärde,was sech inere Erhöchig vo dr Ionädiffusion umene Faktor 15 zeygt. D Awändig fürd Peptidfauätig, wo d Peptid- ud Lösigsmittufreyheytsgrad adiabatisch entkopplet worde sy, wird im Kapitu 6 beschribe. I däm Fau isch d Abtaschtig vom Fautigsglychgwicht nume liecht verbesseret im Verglich zur molekulare Standartsimulation. Kapitu 7 beschribt d Awändig vo dr adiabatische Entkopplig für es grössers System, es Protein im Wasser.

Es schynt, dass d Abtaschtig vo dä mügleche Konformatione vo dr Schloufe us 11 Aminosürine imene 124 Aminosüri-Protein signifikant cha verbesseret wärde. I aune Fäu isch dr Effekt vo dä unterschidleche Apassigsgrössine für d Tämpäratur, d Charft u d Masse vo de schwäre Freyheytsgrad undersuecht worde.

Kapitu 8 git Vorschleg, wie di beschribene Methode chönte wiiter teschtet, entwicklet u erwiteret wärde.

Publications

This thesis led to the following publications:

Chapter 2

Anna-Pitschna E. Kunz and Wilfred F. van Gunsteren,
"Development of a non-linear classical polarisation model for liquid water and aqueous solutions: COS/D."
Journal of Physical Chemistry A, **113** (2009), 11570-11579

Chapter 3

Anna-Pitschna E. Kunz, Andreas E. Eichenberger and Wilfred F. van Gunsteren,
"A simple, efficient polarisable molecular model for liquid carbon tetrachloride"
Molecular physics, **109** (2010), 365-372

Chapter 4

Anna-Pitschna E. Kunz, Haiyan Liu and Wilfred F. van Gunsteren,
"Enhanced sampling of particular degrees of freedom in molecular systems based on adiabatic decoupling and temperature or force scaling"
Journal of Chemical Physics, (2011) accepted

Chapter 5

Anna-Pitschna E. Kunz and Wilfred F. van Gunsteren,
"Enhancing the configurational sampling of ions in aqueous solution using adiabatic decoupling with translational temperature scaling"
Journal of Physical Chemistry B, **115** (2011), 2931-2936

Chapter 6

Anna-Pitschna E. Kunz, Zhixiong Lin and Wilfred F. van Gunsteren,
"A method for sampling the internal degrees of freedom of a flexible solute molecule based on adiabatic decoupling and temperature or force scaling"
Molecular physics,(2011) submitted

Chapter 7

Anna-Pitschna E. Kunz and Wilfred F. van Gunsteren,
"A method for conformational sampling of loops in protein based on adiabatic decoupling and temperature or force scaling"
ChemPhysChem , (2011) accepted

Related Publications

Zhixiong Lin, Anna-Pitschna E. Kunz and Wilfred F. van Gunsteren,
"A one-site polarisable model for liquid chloroform: COS/C."
Molecular Physics, **108** (2010), 1749-1757

Sereina Riniker, Anna-Pitschna E. Kunz and Wilfred F. van Gunsteren,
"On the calculation of the dielectric permittivity of molecular models in the liquid phase"
Journal of Chemical Theory and Computation, **7** (2011), 1469-1475

Nathan Schmid, Jane R. Allison, Jozica Dolenc, Andreas P. Eichenberger, Anna-Pitschna E. Kunz und Wilfred F. van Gunsteren
"Biomolecular structure refinement using the GROMOS simulation software"
Journal of Biomolecular NMR , (2011), DOI: 10.1007/s10858-011-9534-0

Anna-Pitschna E. Kunz, Jane R. Allison, Daan P. Geerke, Bruno A. C. Horta, Philippe H. Hunenberger, Sereina Riniker, Nathan Schmid, and Wilfred F. van Gunsteren
"New functionalities in the GROMOS biomolecular simulation software"
Journal of Computational Chemistry , (2011) accepted

The cure for boredom is curiosity. There is no cure for curiosity.

Dorothy Parker

Chapter 1

Introduction

1.1 Simulation instead of experiment

The reasons to do simulations instead of experiments are manifold and of different character. The experiment may be impossible as it is the case for weather forecasting where measuring of the actual weather is always too late, or it might be too dangerous as the testing of nuclear bombs — at least by today's opinion. An experiment can also be too expensive where the building of real size aeroplane models for wind channel experiments can be given as an example. The motive for molecular simulation is of different nature: the experimental observation and interpretation of certain properties on very short time scales and very small space scales is problematic and so simulation may fill the gap by explaining experimental results — e.g. a dynamic ensemble can describe the measured NMR NOE intensities or distances better than a rigid model [1, 2] — or suggest new experiments — e.g. less experiments with a better chance of success need to be performed if a pre-screening of possible drug molecules is done [3, 4]. Where the experiment is blind for very short time scales and very small space scales, molecular simulation has problems with very long time scales and very large space scales. Molecular simulation and experiment are therefore complementary methods to study different aspects of nature.

When modelling a molecular system four choices have to be made [5].

1. Which degrees of freedom are explicitly modelled. Options range from the explicit treatment of electronic degrees of freedom over all-atom simulations including solvent or coarse-grained simulations that unite several atoms to a single particle up to continuum methods.
2. Which interaction function or force field is used for the description of the energy of the system using the explicitly treated degrees of freedom. For electronic degrees of freedom this may be a quantum Hamiltonian based on Coulomb's law, the Pauli principle and the Schrödinger equation, for all-atom or coarse-grained simulations it is generally a much more complex function of the coordinates of the particles and classical mechanics. An example is the Groningen Molecular Simulation (GROMOS) force field [6, 7].

3. How to generate a meaningful Boltzmann-weighted ensemble of structures by sampling the explicitly treated degrees of freedom, e.g. by Monte Carlo (MC), molecular dynamics (MD) or stochastic dynamics (SD) simulations.
4. How are the boundary conditions chosen that mimic the interaction of the explicitly treated degrees of freedom with those outside the system, including periodic boundary conditions, thermostats, barostats and other couplings.

The choices regarding these four aspects will depend on the system and its properties or processes of interest, the desired accuracy and the computer power available. This thesis addresses points 2 and 3 by proposing an improved treatment of polarisability in molecular simulation and improved methodology to enhance sampling of important degrees of freedom of biomolecular systems in the condensed phase. It treats on the one hand the inclusion of polarisability (Chapters 2 and 3) in the force field to improve its accuracy and widen its applicability and on the other hand adiabatic decoupling with temperature or force scaling (Chapters 4, 5, 6 and 7) to allow faster sampling of the appropriate ensemble.

1.2 Polarisability

In a classical treatment of molecular systems consisting of N particles the interactions between degrees of freedom are described by a so-called force field. Its Hamiltonian can be split into a kinetic $K(\mathbf{p}^N)$ and a potential energy term $V(\mathbf{r}^N)$,

$$H(\mathbf{p}^N, \mathbf{r}^N) = K(\mathbf{p}^N) + V(\mathbf{r}^N), \tag{1.1}$$

which are functions of the sets of particle coordinates $\mathbf{r}^N = (\mathbf{r}_1, \ldots, \mathbf{r}_N)$ and particle momenta $\mathbf{p}^N = (\mathbf{p}_1, \ldots, \mathbf{p}_N)$. The kinetic term

$$K(\mathbf{p}^N) = \sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2m_i} \tag{1.2}$$

depends generally on the momenta $\mathbf{p}^N$ and the masses m_i. This term is independent of the particle positions if no configurational constraints are applied. The potential energy term

$$\begin{aligned} V(\mathbf{r}^N) &= V^{phys}(\mathbf{r}^N) + V^{special}(\mathbf{r}^N) \\ &= V^{bonded}(\mathbf{r}^N) + V^{nonbonded}(\mathbf{r}^N) + V^{special}(\mathbf{r}^N) \end{aligned} \tag{1.3}$$

only depends on the particle coordinates $\mathbf{r}^N$ and can be split into a physical and a special contribution. The physical contribution can be further separated into bonded and nonbonded interactions. The bonded interactions — governing bond lengths, bond angles, torsional and improper dihedral angles — scale with the number of particles N,

while the nonbonded interactions — Lennard-Jones and electrostatic — scale with N^2 or with $N \log N$ for very large systems. For this reason, the bigger the system the more important the nonbonded interactions and the more important it is to describe their interaction properly. Adding polarisability is one possibility to improve the descriptions that are standardly used in biomolecular simulations. Polarisability describes the way a

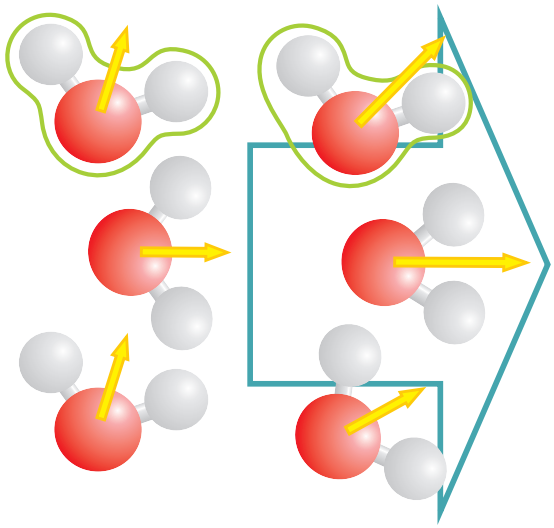

Figure 1.1: Types of change of the size and direction of a molecular dipole (yellow arrow) under the influence of an electric field (right hand side, direction given by the underlying blue arrow) compared to its size and direction in the absence of an electric field (left hand side). Top row: electron density (green) polarisability, middle row: geometric polarisability, bottom row: orientational polarisability.

system or molecule reacts to an electric field. For a real molecule this reaction can be explained by three different mechanisms [8] as illustrated in Fig. 1.1:

- electronic polarisability, i.e. the electron cloud moves or deforms according to the electric field
- geometric polarisability, i.e. the positions of the nuclei change according to the electric field leading to a possible change in bond lengths and angles
- orientational polarisability, i.e. reorientation of the whole molecule according to the electric field

In a real liquid all three processes will happen simultaneously and will lead to a change of the direction and/or strength of the dipoles of molecules. The electric field can be external or result from the other molecules.

In classical molecular dynamic simulation the electronic degrees of freedom are omitted foreclosing the possibility of electronic polarisability. Solvent molecules are usually rigid with all bond lengths and intramolecular distances constrained. In that case they

also do not have the possibility of geometric polarisability. If the solvent does not carry any partial charges, even orientational polarisability is not possible anymore as it is the case for the non-polarisable CCl_4 model described in Chapter 3. Polarisability can be included in an averaged way by increasing the permanent dipole moment of a model as it is done, e.g. in the SPC [9] water model that has a dipole moment of 2.27 D compared to the gas-phase one of 1.85 D of a water molecule, because it should mimic the average dipole of water in the bulk, which is considerably larger than the gas phase one. However, this representation may not be satisfactory in a different electric environment like the inside of a protein, a mixture with an other solvent or in the neighbourhood of an ion. This can lead to limited accuracy when trying to simultaneously reproduce thermodynamic properties in polar and apolar solvents as it was observed in the development of the GROMOS 53A5 and 53A6 force fields, where one set of van der Waals parameters and a fixed charge distribution failed to simultaneously reproduce for polar compounds the free enthalpy of hydration and the density and heat of vaporisation of the pure liquid [7]. Inclusion of dipolar polarisability could help solving this problem.

For classical simulation several ways to introduce dipolar polarisability are available: point polarisable dipole models [10–12], charge-on-spring (COS) models [13], also called Drude oscillator [14] or shell [15] models, and fluctuating charge [16] models.

Since the calculation of electric interactions using ideal dipoles is more tedious than the one for charges, GROMOS polarisable force fields [8] make use of the COS model to integrate polarisability as it leads to simple formulas that are easy to implement in biomolecular simulation software, is computationally efficient and can easily be combined with non-polarisable GROMOS force fields.

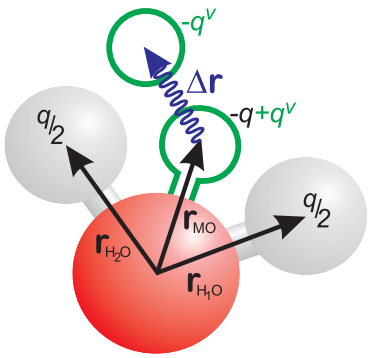

Figure 1.2: Off-site water model with four Coulomb interaction sites located at the two hydrogen atoms, the virtual-atom site M and the polarisation charge and one Lennard-Jones site located at the oxygen atom.

An ideal dipole is defined by

$$\boldsymbol{\mu} = \lim_{a\to\infty} q_a \mathbf{d}_a \text{ where } q_a = aq \text{ and } \mathbf{d}_a = \frac{\mathbf{d}}{a} \tag{1.4}$$

where q is a charge and $\mathbf{d}$ a distance. The dipole in the COS model is defined by

$$\boldsymbol{\mu} = q^{\nu}\Delta\mathbf{r} \tag{1.5}$$

where q^{ν} is the charge of the COS particle, called virtual charge, and $\Delta\mathbf{r}$ the length of the spring as shown in Fig.1.2. For $|\Delta\mathbf{r}| \ll |\mathbf{r}|$ where $\mathbf{r}$ is the distance to the COS, this acts as an ideal dipole. Since $|\Delta\mathbf{r}| \propto (q^{\nu})^{-1}$ for a given $|\boldsymbol{\mu}|$, q^{ν} has to be chosen big enough. However, choosing a too big q^{ν} leads to numerical precision problems in $\Delta\mathbf{r}$. An optimised value [13, 17] of $q^{\nu} = -8e$ is usually chosen in GROMOS.

In GROMOS, the position where the COS is attached can be chosen either on an atom or on a virtual site defined by the position of 3 atoms using, for the example of a water molecule [18],

$$\mathbf{r}_M = \mathbf{r}_O + \mathbf{r}_{MO} = \mathbf{r}_O + \frac{\gamma}{2}(\mathbf{r}_{H_1O} + \mathbf{r}_{H_2O}) \tag{1.6}$$

where $\mathbf{r}_{H_1O} = \mathbf{r}_{H_1} - \mathbf{r}_O$ and $\mathbf{r}_{H_2O} = \mathbf{r}_{H_1} - \mathbf{r}_O$, $\mathbf{r}_O$ is the position of the oxygen atom, $\mathbf{r}_{H_1}$ and $\mathbf{r}_{H_2}$ the positions of the hydrogen atoms and γ a scaling parameter. The addition of the massless site M does not introduce any extra degrees of freedom into the molecule. The force $\mathbf{f}_M$ acting on the virtual site M has to be redistributed to the atoms of the molecule using

$$\begin{aligned} \mathbf{F}_{H_1} &= \mathbf{f}_{H_1} + \frac{\gamma}{2}\mathbf{f}_M \\ \mathbf{F}_{H_2} &= \mathbf{f}_{H_2} + \frac{\gamma}{2}\mathbf{f}_M \\ \mathbf{F}_O &= \mathbf{f}_O + (1-\gamma)\mathbf{f}_M. \end{aligned} \tag{1.7}$$

The dipole $\boldsymbol{\mu}$ will be influenced [8] by the electric field $\mathbf{E}$

$$\boldsymbol{\mu} = \mathbf{f}(\mathbf{E}) \tag{1.8}$$

where f is a function of $\mathbf{E}$. Depending on the model, the dependence can be assumed linear

$$\boldsymbol{\mu} = \underline{\alpha}\mathbf{E} \tag{1.9}$$

where $\underline{\alpha}$ is the polarisability tensor. When the anisotropy of $\underline{\alpha}$ is small, $\underline{\alpha}$ can be replaced by a scalar α

$$\boldsymbol{\mu} = \alpha\mathbf{E}. \tag{1.10}$$

A way to obtain a scalar, simple but non-linear dependence of the dipole moment $\boldsymbol{\mu}$ on the electric field $\mathbf{E}$ is to make α a function of E

$$\boldsymbol{\mu} = \alpha(E)\mathbf{E}. \tag{1.11}$$

as described in Chapter 2.

For the COS model, the $\Delta\mathbf{r}_i$ of every spring can then be calculated using

$$\Delta\mathbf{r}_i = \frac{\mathbf{f}(\mathbf{E}_i)}{q^\nu} \tag{1.12}$$

where $\mathbf{f}(\mathbf{E}_i)$ can be $\underline{\alpha}\mathbf{E}$ or $\alpha\mathbf{E}$ or $\alpha(E)\mathbf{E}$ as described above. Since the electric field at site i, which is either a real atom or a virtual site M,

$$\mathbf{E}_i^\nu = \sum_{j=1,j\neq i}^{N} \left[\frac{q_j - q_j^\nu}{4\pi\epsilon_0|\mathbf{r}_i^\nu - \mathbf{r}_j|^2} \frac{\mathbf{r}_i^\nu - \mathbf{r}_j}{|\mathbf{r}_i^\nu - \mathbf{r}_j|} + \frac{q_j^\nu}{4\pi\epsilon_0|\mathbf{r}_i^\nu - \mathbf{r}_j^\nu|^2} \frac{\mathbf{r}_i^\nu - \mathbf{r}_j^\nu}{|\mathbf{r}_i^\nu - \mathbf{r}_j^\nu|} \right] \tag{1.13}$$

with $\mathbf{r}_i^\nu = \mathbf{r}_i + \Delta\mathbf{r}_i$, depends on the $\Delta\mathbf{r}_j$ values of all the other *COS* sites j, they have to be calculated iteratively.

When the $\Delta\mathbf{r}_i$s are converged, the electric energy

$$U^{el} = U^{coulomb} + U^{self} \tag{1.14}$$

consisting of a contribution from Coulombic interactions

$$U^{coulomb} = \sum_{i=1}^{N}\sum_{j>i}^{N} \left[\frac{(q_i - q_i^\nu)(q_j - q_j^\nu)}{4\pi\epsilon_0|\mathbf{r}_i - \mathbf{r}_j|} + \frac{(q_i - q_i^\nu)q_j^\nu}{4\pi\epsilon_0|\mathbf{r}_i - \mathbf{r}_j^\nu|} + \frac{q_i^\nu(q_j - q_j^\nu)}{4\pi\epsilon_0|\mathbf{r}_i^\nu - \mathbf{r}_j|} + \frac{q_i^\nu q_j^\nu}{4\pi\epsilon_0|\mathbf{r}_i^\nu - \mathbf{r}_j^\nu|} \right] \tag{1.15}$$

and a self-polarisation energy

$$U^{self} = \sum_{i=1}^{N} \int_0^{\boldsymbol{\mu}_i} \mathbf{E}_i \cdot d\boldsymbol{\mu}_i \tag{1.16}$$

can be determined. This leads to an equation for the electronic contribution to the force on particle i

$$\mathbf{f}_i^{el} = \sum_{j=1,j\neq i}^{N} \left[\frac{(q_i - q_i^\nu)(q_j - q_j^\nu)}{4\pi\epsilon_0|\mathbf{r}_i - \mathbf{r}_j|^2} \frac{\mathbf{r}_i - \mathbf{r}_j}{|\mathbf{r}_i - \mathbf{r}_j|} + \frac{(q_i - q_i^\nu)q_j^\nu}{4\pi\epsilon_0|\mathbf{r}_i - \mathbf{r}_j^\nu|^2} \frac{\mathbf{r}_i - \mathbf{r}_j^\nu}{|\mathbf{r}_i - \mathbf{r}_j^\nu|} + \frac{q_i^\nu(q_j - q_j^\nu)}{4\pi\epsilon_0|\mathbf{r}_i^\nu - \mathbf{r}_j|^2} \frac{\mathbf{r}_i^\nu - \mathbf{r}_j}{|\mathbf{r}_i^\nu - \mathbf{r}_j|} + \frac{q_i^\nu q_j^\nu}{4\pi\epsilon_0|\mathbf{r}_i^\nu - \mathbf{r}_j^\nu|^2} \frac{\mathbf{r}_i^\nu - \mathbf{r}_j^\nu}{|\mathbf{r}_i^\nu - \mathbf{r}_j^\nu|} \right]. \tag{1.17}$$

Note that the self-polarisation term does not contribute to the force since its minimum is found during the iteration procedure.

Inclusion of the polarisability makes each MD time step 5 times more expensive for the one-site polarisable model of water (Chapter 2) and 1.7 times for the one-site polarisable model of CCl_4 (Chapter 3). However, it improves the properties of the model, especially the dielectric ones. For the COS/D water model proposed in Chapter 2 the dielectric permittivity $\epsilon(0)$ is 69.8, which is higher than the non-polarisable SPC model ($\epsilon(0) = 61$), but, because a damped non-linear polarisability was used, closer to the experimental value of $\epsilon(0) = 78.4$ than the previous COS/G2 model, which had $\epsilon(0) = 87.8$. The non-polarisable model of CCl_4 had a dielectric permittivity $\epsilon(0)$ of 1 since the atoms did not carry any partial charges and therefore had no electrostatic interactions, while the polarisable model proposed in Chapter 3 was parametrised to reproduce the experimental value of 2.24.

1.3 Adiabatic decoupling with temperature or force scaling

Molecular dynamics simulation generates a new configuration of the complete system from the previous configuration using the force along and the inertia of a degree of freedom and integrating them over a time-step. In GROMOS the leap-frog algorithm is used to this end,

$$\begin{aligned} \mathbf{v}_i(t + \Delta t/2) &= \mathbf{v}_i(t - \Delta t/2) + m_i^{-1}\mathbf{f}_i(t)\Delta t + O(\Delta t^3) \\ \mathbf{r}_i(t + \Delta t) &= \mathbf{r}_i(t) + \mathbf{v}_i(t + \Delta t/2) + O(\Delta t^3), \end{aligned} \tag{1.18}$$

where Δt is the time step, $\mathbf{v}_i$ is the velocity of atom i and $\mathbf{f}_i$ the force acting on it.

According to statistical mechanics, the configurations $\mathbf{r}^N$ of a molecular system at a given temperature T must constitute a Boltzmann ensemble, i.e. they occur with a relative probability

$$exp(-V(\mathbf{r}^N)/k_BT) \tag{1.19}$$

in the ensemble of configurations, with k_B being the Boltzmann constant. In this ensemble low potential energy configurations have a high probability. Thus sampling can be viewed as the attempt to explore low energy regions of the energy hypersurface $V(\mathbf{r}^N)$, which usually has a multitude of high energy barriers between a multitude of local minima. This can make it difficult to visit all relevant regions of the energy hypersurface within a reasonable amount of sampling time. A great many methods to facilitate or enhance this search have been developed. They can be coarsely put into three major categories plus subgroups [19]

1. Deformation or smoothing of the potential energy surface $V(\mathbf{r}^N)$

   (a) Inclusion of restraints from experimental data (Xray, NMR) in $V(\mathbf{r}^N)$ [20]
   (b) Use of soft-core atoms in the non-bonded interactions [21]
   (c) Extending the spatial dimensionality from 3 to 4 dimensions [22]
   (d) Use of constraints to rigidify high frequency degrees of freedom [23]
   (e) Transformation to a coarse-grained (CG) model with less degrees of freedom [24]
   (f) Local potential energy elevation [25] or conformational flooding [26], a remedy against repeated sampling of the same part of configuration space
   (g) Scaling of the potential energy in order to make it a smoother function of $\mathbf{r}^N$ [27]

2. Scaling of system parameters

   (a) Temperature annealing [28]
   (b) Temperature scaling [29]
   (c) Mass scaling of a subset of degrees of freedom [30]
   (d) Mean-field approaches [31]
   (e) Force scaling for a subset of degrees of freedom, Chapter 4

3. Multi-copy searching and sampling

   (a) Genetic algorithms [32]
   (b) Replica exchange and multi canonical algorithms [33]
   (c) Cooperative search: SWARM simulation [34]

One method to improve sampling, adiabatic decoupling with temperature scaling [35], is described in detail in Chapter 4. It combines the scaling of the mass (2(c)) to decouple the frequency of a particular set of degrees of freedom from the rest of the system with either the scaling of the temperature (2(b)) to enhance the sampling of this particular set of decoupled degrees of freedom. Adiabatic decoupling was also used in combination with potential energy (1(g)) instead of temperature scaling [36]; see also Chapter 4. This chapter also contains a description of adiabatically decoupling in combination with force scaling, which avoids a particular disadvantage of potential energy scaling. This new method is tested and compared to temperature scaling using liquid water as test system in Chapter 4.

Adiabatic decoupling with temperature scaling is applied to investigate enhanced sampling of ionic degrees of freedom in aqueous solution in Chapter 5. It increases the ionic diffusion taken as a measure of sampling by a factor of 15, while no structural difference to the unchanged system is detected.

For enhanced sampling of the folding equilibrium of a peptide or of the various possible conformations of a loop of the amino-acid residue chain constituting a protein, force scaling is an interesting option as described in Chapter 6 and Chapter 7. Here it helps sampling the conformational space faster without noticeably distorting the distribution of conformations.

Chapter 2

Water is H_2O, hydrogen two parts, oxygen one, but there is also a third thing, that makes it water and nobody knows what that is.

D.H. Lawrence

Development of a non-linear classical polarisation model for liquid water and aqueous solutions: COS/D

2.1 Summary

A new charge-on-spring (COS) model for water is introduced (COS/D). It includes a sublinear dependence of the induced dipole on the electric field for large field strength to include the effect of hyperpolarisability by damping the polarisability. Only two new parameters were introduced to define the damping of the polarisability. In the parameterisation procedure these two damping parameters, the two Lennard-Jones parameters, the charge on the oxygen and the distance between the virtual site and the oxygen atom were varied to reproduce the density, the heat of vaporisation, the dielectric permittivity and the position of the first peak in the radial distribution function of liquid water at room temperature and pressure. In this way, a model was obtained that correctly describes a variety of thermodynamic, dynamic and dielectric properties of water, while still preserving the simplicity of the COS model which allows a straightforward introduction of explicit polarisation into (bio)molecular force fields.

2.2 Introduction

Faithful biomolecular simulation critically depends on the accuracy of the force field used. The most widely used general biomolecular force fields are all based on van der Waals and electrostatic non-bonded interaction terms using fixed (atomic) partial charges. Within this framework, polarisation is only accounted for by structural rearrangement of (fragments of) molecules. Improved accuracy is to be reached by variation and optimisation of the force field parameters. For example, the GROMOS (GROningen MOlecular Simulation) force field, based on reproducing the thermodynamic properties for small molecules, still improved its accuracy during the last decade. The newest pa-

rameter set 53A6 is able to reproduce the free energies of apolar (cyclohexane) and of polar (water) solvation for typical biomolecular compounds [7]. However, a limitation of a non-polarisable force field became eminent in this study. The attempt to simultaneously reproduce for polar compounds the free enthalpy of hydration *and* the density and heat of vaporisation of the pure liquids by a combination of a (fixed) charge distribution and a set of van der Waals parameters failed for most functional groups considered [7]. This is not surprising as one expects the degree of (electronic) polarisation of the solute to be larger when solvated in water compared to cyclohexane, which can only be achieved by using a polarisable molecular model or force field. Another example of the inadequacy of non-polarisable force fields is the absence of the experimentally known maximum in the solvation free enthalpy of argon at intermediate composition of ethylene glycol and water mixtures using non-polarisable models [37]. These and other findings reinforce the idea that a further improvement of biomolecular force fields should include polarisability [38–44]. There are several ways to introduce dipolar polarisability into classical simulations [8, 45]: the point polarisable dipole model (PPD) [10–12], the charge-on-spring (COS) [13], also called Drude-oscillator [14] or shell [15], model, and the fluctuating charge (FQ) [16] model. GROMOS polarisable force fields make use of the COS model [8] to integrate polarisability as it leads to simple formulas and is computationally efficient.

One additional deficiency of most polarisable models apart from their larger demand for computing power than non-polarisable ones is their tendency to show overpolarisation leading to the polarisation catastrophe and a static dielectric permittivity, $\epsilon(0)$, that is too large [46]. There are several approaches to resolve these problems. The polarisation catastrophe can be avoided by a big enough repulsive Lennard-Jones term [47] leading to dipole-dipole distances larger than $(4\alpha^2)^{1/6}$, by spreading the polarisability over more sites [48] which lowers α and therefore the critical distance $(4\alpha^2)^{1/6}$, by introducing a distance-dependent damping factor for short distance dipole-dipole interactions to keep them finite [49], or by substituting the linear dependence of the induced dipole $\boldsymbol{\mu}^{ind}$ on the electric field $\mathbf{E}$ for all values of $\mathbf{E}$ for a sublinear dependence for large field strengths [8, 50, 51], which can be achieved by making the polarisability α electric field dependent. A remedy against a too large static dielectric permittivity $\epsilon(0)$ is, for example, use of a polarisable site that is off the charge site [18], or again a sublinear behaviour for $\boldsymbol{\mu}^{ind}$ for large field strengths. The method used in this paper, following an earlier described idea [8, 50], is the damping of the polarisability $\alpha_D = \alpha_D(E)$ for large E in an analytical way,

$$\alpha_D = \begin{cases} \alpha & \text{for } E \leq E_0 \\ \frac{\alpha E_0}{pE}\left[p + 1 - \left(\frac{E_0}{E}\right)^p\right] & \text{for } E > E_0 \end{cases} \tag{2.1}$$

where α is the original polarisability, E the value of the electric field $\mathbf{E}$, and p and E_0 are adjustable parameters of the model. The dependence of the induced dipole μ^{ind} on the electric field E ($\boldsymbol{\mu}_{ind}$ along $\mathbf{E}$) with a damped polarisability α_D is displayed in Fig. 2.1.

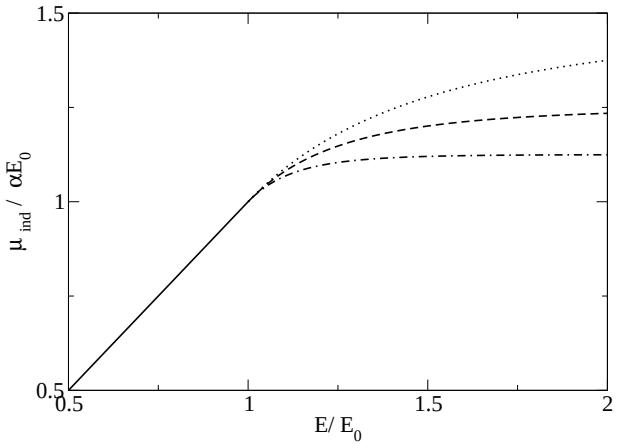

Figure 2.1: Dependence of the induced dipole μ^{ind} on the electric field strength E (Eq.(2.1)). The solid line is the linear dependence up to E_0, the dotted line is the damped part for $p = 2$, the dashed for $p = 4$ and the dash-dotted for $p = 8$.

The physical idea behind this approach of damping the polarisability α is to account for hyperpolarisability in a simplified scalar way as the linearity of $\boldsymbol{\mu}^{ind} = \alpha\mathbf{E}$ is breaking down beyond some value of E. Non-linear polarisation effects begin to become significant at a field strength of $20 - 30\,\mathrm{V/nm}$ [52–54] (about $150 - 250\,(\mathrm{kJmol^{-1}nm^{-3}})^{1/2}$) which is a strength that is comparable to the mean field strength in aqueous solution [55, 56]. The chosen damping in Equation (2.1) can straightforwardly be included in biomolecular force fields where many-atom molecules are to be treated likewise.

Since our main interest is in simulating biological systems at physiological temperature and pressure and since biomolecules are generally solvated in water, our major goal was to develop a water model that reproduces the bulk liquid phase properties of water at room temperature and pressure as well as the non-polarisable and previously developed polarisable models do. The gas phase properties of water, more representative for an isolated water molecule in a protein, are of less importance and the simulation of ice was merely done for completeness, as this phase is of little interest in biomolecular simulations.

The following properties were chosen for calibrating the parameters (p, E_0, q_H, d_{OM}, C_{12}, and C_6) of the model: the heat of vaporisation ΔH_{vap}, the density ρ, the static dielectric permittivity $\epsilon(0)$ estimated by the average molecular dipole $\langle\mu\rangle$ as predicted by Fig. 2.2 and the position of the first peak in the oxygen-oxygen radial distribution function $g_{OO}(r)$, all at about $298\,\mathrm{K}$ and $1\,\mathrm{atm}$. Using the optimised model parameters, other thermodynamic and dynamic properties and the solvation free energy of an argon probe were calculated to test the performance of the water model. To complete the study the gas phase dimer and the I_h ice were simulated.

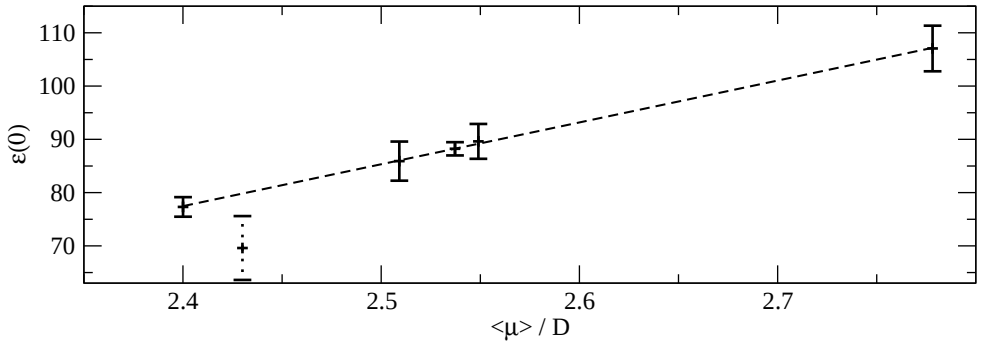

Figure 2.2: Relation between the static dielectric permittivity $\epsilon(0)$ and the average molecular dipole moment $\langle\mu\rangle$ as observed in simulations of liquid water. The dotted line is a linear regression on the solid lined error bars, which were obtained from the parametrisation (COS/D_A, COS/D_B, COS/D_C, COS/D_D, COS/D_E, COS/D_F) simulations. The error bar with the dashed line is obtained using the COS/D model.

2.3 Methods

2.3.1 Developing the model

The damped model is based on previous COS models [8] in which the electric field $\mathbf{E}_i$, which influences the damping and the induced dipole, is taken at the position of the COS charge of the virtual site of atom i [57]. The dependence of the polarisability of virtual site of atom i, α_i, on the electric field $\mathbf{E}_i$ at the COS of virtual site of atom i, $\alpha_i = \alpha_i(|\mathbf{E}_i|)$, was chosen such that the induced dipole $\boldsymbol{\mu}_i^{ind}$ depends linearly upon the electric field $\mathbf{E}_i$ up to a certain field strength $E_{0,i}$ and then levels off to a constant value as shown in Fig. 2.1. Additionally, the function and its first derivative are to be continuous which leads to the following formula for the induced dipole

$$\boldsymbol{\mu}_i^{ind} = \begin{cases} \alpha_i \mathbf{E}_i & \text{for } E_i \leq E_{0,i} \\ \frac{\alpha_i E_{0,i}}{p_i} \left[p_i + 1 - \left(\frac{E_{0,i}}{E_i} \right)^{p_i} \right] \frac{\mathbf{E}_i}{E_i} & \text{for } E_i > E_{0,i} \end{cases} \tag{2.2}$$

where p_i is a parameter that determines the damping. This is like replacing the harmonic force constant k_i^{HO} of the spring by a harmonic force parameter $k_i^{HO}(E)$ that depends on the strength of the electric field $\mathbf{E}_i$ and becomes larger above the truncation parameter $E_{0,i}$. The self-polarisation contribution to the potential energy, U_{self},

also becomes dependent on the electric field,

$$U_{self,i} = \begin{cases} \frac{1}{2}\alpha_i E_i^2 & \text{for } E_i \leq E_{0,i} \\ \frac{1}{2}\alpha_i {E_{0,i}}^2 + \frac{\alpha E_{0,i}^2}{p_i(p_i-1)}\left[-p_i^2 + (p_i^2-1)\left(\frac{E_i}{E_{0,i}}\right) + \left(\frac{E_{0,i}}{E_i}\right)^{p_i-1}\right] & \text{for } E_i > E_{0,i} \end{cases} \tag{2.3}$$

with $U_{self} = \sum_i U_{self,i}$, where i runs over all polarisable centers.

2.3.2 Simulation methods

A cubic box with an edge length of $3.1057\,\mathrm{nm}$ was filled with 1000 water molecules, resulting in a density of $997\,\mathrm{kg/m^3}$, corresponding to the density of liquid water at $298\,\mathrm{K}$ and $1\,\mathrm{atm}$ [58]. Molecular dynamics simulations were performed at constant pressure and temperature (*NpT* conditions) with the GROMOS96 package [59, 60], modified to incorporate the damped polarisable model. The geometry of the water molecules was constrained by applying the SHAKE algorithm [23] with a relative geometric tolerance of 10^{-4} on the OH bond length and on the intramolecular HH distance. The temperature was weakly coupled [29] to a bath of $298.15\,\mathrm{K}$ with a relaxation time of $0.1\,\mathrm{ps}$ and the pressure was weakly coupled [29] to a bath of $1\,\mathrm{atm}$ with a relaxation time of $0.5\,\mathrm{ps}$. The isothermal compressibility was set to the experimental value [58] of $7.513 \cdot 10^{-4}\,(\mathrm{kJmol^{-1}nm^{-3}})^{-1}$. The nonbonded van der Waals and electrostatic interactions were calculated using triple-range cutoff radii of $0.8/1.4\,\mathrm{nm}$. The short range interactions were calculated every time step by updating the molecular pair list for distances smaller than the first cutoff radius of $0.8\,\mathrm{nm}$. For the intermediate range of distances between 0.8 and $1.4\,\mathrm{nm}$ the pairlist was only updated every fifth timestep and at the same time the interaction was calculated and kept unchanged between these updates. The long range electrostatic interactions beyond the outer cutoff of $1.4\,\mathrm{nm}$ were represented by a reaction field [61,62] with $\epsilon_{RF} = 78.5$. The equations of motion were integrated using the leap frog algorithm with a time step of $2\,\mathrm{fs}$. The velocities of the atoms at the beginning of the simulation were assigned from a Maxwell distribution at $298\,\mathrm{K}$. During the runs, configurations of the system were saved every $0.5\,\mathrm{ps}$. The various properties for the COS/D water model were taken from a $2\,\mathrm{ns}$ simulation that followed a $50\,\mathrm{ps}$ equilibration period. To enhance convergence the dielectric properties were calculated from 10 separate independent runs of $5\,\mathrm{ns}$ at $298.15\,\mathrm{K}$. The starting structure of the ice I_h simulation was taken from the $3x2x2$ unit cell with 96 water molecules constructed by Hayward and Reimers [63] containing 12 copies of the smallest unit cell for ice I_h that contains eight water molecules. To ensure a big enough box for the triple range cutoff, this $3x2x2$ unit cell was copied three times along each of the $x-$, $y-$ and $z-$axes. This box then had edges with sizes of $4.056\,\mathrm{nm}$, $4.684\,\mathrm{nm}$, and $4.416\,\mathrm{nm}$. The structure was first equilibrated in 5 *NVT* simulations (each $5\,\mathrm{ps}$ long) with the temperature increasing from $1\,\mathrm{K}$ to $50\,\mathrm{K}$ followed by 3 *NpT* runs (each $5\,\mathrm{ps}$ long) with the temperature increasing from $50\,\mathrm{K}$ to $100\,\mathrm{K}$. The simulation was

performed at $100\,\mathrm{K}$ and $1\,\mathrm{atm}$ for $2\,\mathrm{ns}$ and the configurations of the system were saved every $0.5\,\mathrm{ps}$.

2.3.3 Parametrisation

The model parameters were fit to reproduce the experimental density ρ and heat of vaporisation ΔH_{vap} at room temperature and normal pressure. The heat of vaporisation was calculated using the following formula,

$$\Delta H_{vap}(T) = -U^{liquid}(T) + p\Delta V + Q^{int} + Q^{ext} = -U^{liquid}(T) + RT + Q, \qquad (2.4)$$

where ΔH_{vap} is the experimental molar heat of vaporisation, U_{liquid} is the computed intermolecular potential energy per mole, p the pressure, and ΔV the molar volume change between liquid and gas. R is the gas constant and T is the absolute temperature. Q^{int} and Q^{ext} are quantum corrections. Q^{int} accounts for the difference in vibration energy between water in the liquid and the gas phase. Q^{ext} is a correction due to the intermolecular interaction in the liquid and is the difference in vibrational energy calculated quantum-mechanically and classically. At $298\,\mathrm{K}$ this adds up to a total quantum correction [64] of $Q = -0.23\,\mathrm{kJmol}^{-1}$.

Additionally, the average molecular dipole moment $\langle\mu\rangle$ (as an approximation for the very slowly converging dielectric permittivity $\epsilon(0)$), and the first peak in the radial distribution function between the oxygens, g_{OO}, were chosen as reference points to be reproduced. The geometry was set to the experimental gas phase values and for α the experimental polarisability of water was used, first only the electronic contribution, then including the vibrational contribution as the geometry of the water molecules was kept rigid.

The parameters that were varied were the charge $q_O = -2q_H$ (and accordingly the distance between the oxygen and the virtual site d_{OM} to keep the dipole at the experimental value), the attractive van der Waals parameter C_6, the repulsive van der Waals parameter C_{12} (both for the $O-O$ interactions), the damping parameter p and the truncation parameter E_0.

2.3.4 Analysis

Radial distribution function g(r)

The structure of liquid water is characterized by a short-range order and a long-range disorder. This is reflected by the radial distribution function g(r), which is experimentally available, for instance, through neutron diffraction [65]. The pair distribution function g(r) gives the probability of finding another atom at a distance r from a given atom, relative to the probability expected for a completely uniform distribution at the same density, and

can be calculated by a simple histogram summation in radial shells over all molecules in the system.

Self-diffusion coefficient D

The diffusion coefficient is obtained from the long-time limit of the mean square displacement according to the Einstein relation, [66]

$$D = \lim_{t\to\infty} \frac{\langle(\mathbf{r}(t) - \mathbf{r}(0))^2\rangle}{6t} \tag{2.5}$$

where $\mathbf{r}(t)$ corresponds to the position vector of the center of mass at time t, and the averaging is performed over both time and water molecules. In a similar way we can calculate the $x-$, $y-$ and $z-$ components of the diffusion constant D.

Rotational correlation times τ_l^α

Reorientational correlation functions, $C_l^\alpha(t)$, are calculated for three different axes α: the H-H vector, the O-H vector and the molecular dipole vector μ, according to

$$C_l^\alpha(t) = \langle P_l(\mathbf{e}^\alpha(t) \cdot \mathbf{e}^\alpha(0)\rangle, \tag{2.6}$$

where P_l is the Legendre polynomial of order l and $\mathbf{e}^\alpha$ is a unit vector pointing along the α-axis in a molecular reference frame. $C_l^\alpha(t)$ shows in general an exponential decay, which can, therefore, be fitted using the following expression

$$C_l^\alpha(t) = A \exp\left(-\frac{t}{\tau_l^\alpha}\right), \tag{2.7}$$

where τ_l^α denotes the single-molecule correlation time and A is a constant. The H-H and O-H relaxation can be obtained from $^1H-^1H$ and $^{17}O-^1H$ dipolar relaxation NMR experiments, whereas the molecular dipolar orientational correlation function is experimentally obtained from optical measurements such as Raman scattering, fluorescence depolarisation or Kerr relaxation experiments. [67–69]

Dielectric permittivity $\epsilon(0)$

The static dielectric constant or permittivity $\epsilon(0)$ is calculated from the fluctuations in the total dipole of the simulation box according to a Kirkwood-Fröhlich-type equation derived by Neumann [70],

$$(\epsilon(0) - 1)\left(\frac{2\epsilon_{RF} + 1}{2\epsilon_{RF} + \epsilon(0)}\right) = \frac{\langle\mathbf{M}^2\rangle - \langle\mathbf{M}\rangle^2}{3\epsilon_0 V k_B T} \tag{2.8}$$

where ϵ_{RF} is the relative dielectric permittivity of the reaction field continuum that is used in the simulation, $\mathbf{M}$ is the total dipole moment of the system, V is the volume of the box, k_B is the Boltzmann constant, T is the absolute temperature, and ϵ_0 is the dielectric permittivity of vacuum.

Debye relaxation time τ_D and frequency-dependent permittivity $\epsilon(\omega)$

The Debye relaxation time τ_D can be obtained by calculation of the normalized auto-correlation function $\Phi(t)$ of the total dipole moment of the system,

$$\Phi(t) = \frac{\langle \mathbf{M}(0)\mathbf{M}(t)\rangle}{\langle \mathbf{M}^2\rangle}. \tag{2.9}$$

This function $\Phi(t)$ generally shows two decays, the first one being so fast that the used time resolution (a sampling rate of two per ps) is not able to capture it. It can, therefore, be approximated by a Heaviside function $\theta(t)$. The second decay is a single exponential decay. The $\Phi(t)$ function looks then as follows [71]:

$$\Phi(t) = (1-A)(1-\theta(t)) + A\exp(-t/\tau_s). \tag{2.10}$$

The frequency-dependent dielectric permittivity $\epsilon(\omega)$ of the system can be obtained from the normalized autocorrelation function $\Phi(t)$ of the total dipole moment $\mathbf{M}$ of the system, using its Fourier-Laplace transform [72],

$$\frac{(\epsilon(\omega)-1)}{(\epsilon(0)-1)}\frac{(2\epsilon_{RF}+\epsilon(0))}{(2\epsilon_{RF}+\epsilon(\omega))} = \int_0^\infty \left(-\frac{d\Phi}{dt}\right)\exp(-i\omega t)dt. \tag{2.11}$$

Assuming Debye dielectric behavior after the first initial phase, one has [71–74]

$$\frac{\epsilon(\omega)-\epsilon(\infty)}{\epsilon(0)-\epsilon(\infty)} = \frac{1}{1+i\omega\tau_D}. \tag{2.12}$$

Inserting Eq.(2.10) into Eq.(2.11) and using Eq.(2.12) one finds the infinite-frequency dielectric permittivity $\epsilon(\infty)$

$$\epsilon(\infty) = 1 + \frac{(1-A)(\epsilon(0)-1)}{1+A\lambda}, \tag{2.13}$$

and the Debye relaxation time τ_D

$$\tau_D = (1+A\lambda)\tau_s \tag{2.14}$$

with

$$\lambda = \frac{\epsilon(0)-1}{2\epsilon_{RF}+1}. \tag{2.15}$$

Finite and infinite system Kirkwood factors G_k and g_k

The finite system Kirkwood factor G_k measures the orientational correlation between a single dipole and all its peers. It is determined from

$$G_k = \frac{\langle M^2 \rangle - \langle M \rangle^2}{N \langle \mu^2 \rangle}, \tag{2.16}$$

where N is the number of molecules and μ is the dipole moment of a single molecule. The finite system Kirkwood factor G_k depends on the boundary conditions (ϵ_{RF}) and the box shape. For our simulation conditions, the relation to the infinite system Kirkwood factor g_k [73], which is available experimentally, is

$$g_k = \frac{(2\epsilon_{RF} + \epsilon(0))(2\epsilon(0) + 1)}{3\epsilon(0)(2\epsilon_{RF} + 1)} G_k. \tag{2.17}$$

Heat capacity C_p

The heat capacity at constant pressure can be approximated [64] according to the formula,

$$C_p \approx \frac{U_2^{tot} - U_1^{tot}}{T_2 - T_1} + \frac{\partial Q^{int}}{\partial T} + \frac{\partial Q^{ext}}{\partial T}, \tag{2.18}$$

where U^{tot} is the total energy per molecule and Q^{int} is the quantum contribution of the intramolecular vibrational modes to the specific heat, while Q^{ext} is the difference between the quantum-mechanical and classical intermolecular vibrational energy. These quantum contributions add up to about $-9.3\,\mathrm{J mol^{-1} K^{-1}}$ at $298\,\mathrm{K}$ and $1\,\mathrm{atm}$.

For this purpose, we carried out three additional NpT-simulations of $1\,\mathrm{ns}$ each (plus an initial $100\,\mathrm{ps}$ of equilibration) at 298, 318, and $338\,\mathrm{K}$, respectively.

Thermal expansion coefficient α

The thermal expansion coefficient α is calculated using a finite-difference expression, [75]

$$\alpha = \frac{1}{V}\left(\frac{\partial V}{\partial T}\right)_p \approx -\left(\frac{\ln(\rho_2/\rho_1)}{T_2 - T_1}\right)_p \tag{2.19}$$

Isothermal compressibility κ_T

The isothermal compressibility κ_T can be obtained by the following finite-difference expression, [76]

$$\kappa_T = -\frac{1}{V}\left(\frac{\partial V}{\partial p}\right)_T = \frac{1}{\rho}\left(\frac{\partial \rho}{\partial p}\right)_T = \left(\frac{\partial \ln(\rho)}{\partial p}\right)_T \approx \left(\frac{\ln(\rho_2/\rho_1)}{p_2 - p_1}\right)_T, \tag{2.20}$$

where ρ is the density of the system. For this purpose, we carried out three additional NVT simulations of 1ns each (plus an initial $100\,\mathrm{ps}$ of equilibration) at densities of 947.0, 997.0, and $1047.0\,\mathrm{kgm^{-3}}$, respectively.

Surface tension γ

The surface tension can be calculated using the following expression

$$\gamma = \frac{L_z}{2}\langle (p_{zz} - \frac{1}{2}(p_{xx} + p_{yy}))\rangle \tag{2.21}$$

where L_z is the length of the box in z direction, $\langle \ldots \rangle$ denotes a time averaging and p_{ii} are the diagonal elements of the pressure tensor. To calculate the surface tension an additional $1\,\mathrm{ns}$ simulation was performed with a box of 1000 particles where the box length in z direction was extended to $15\,\mathrm{nm}$.

Free enthalpy of solvation ΔG_S

The free enthalpy of solvation ΔG_S of argon as a case of a hydrophobic probe can be determined via Widom test-particle insertion [77] of an argon probe with van der Waals parameters $C_6 = 6.2647225\,10 \cdot 10^{-3}\mathrm{kJmol^{-1}nm^6}$ and $C_{12} = 9.84704\,10 \cdot 10^{-6}\mathrm{kJmol^{-1}nm^{12}}$ and a polarisability [78] $\frac{1}{4\pi\epsilon_0}\alpha$ of $1.586 \cdot 10^{-3}\,\mathrm{nm^3}$. For 600 water configurations of the simulation $2.1{\cdot}10^5$ test insertions were performed. From the volume- and Boltzmann-weighted average of the interaction energy U_{int} between the argon probe and the solvent molecules, ΔG_S was calculated from [77, 79]

$$\Delta G_S = -k_B T \ln\left(\frac{\langle V \exp(-U_{int}/k_B T)\rangle}{\langle V\rangle}\right) \tag{2.22}$$

where V is the volume of the box. U_{int} contains two terms [37] in case of a polarisable argon probe. One is the van der Waals term and the other one is the induced dipolar interaction energy U_{dip},

$$U_{dip} = -\frac{1}{2}\alpha E^2, \tag{2.23}$$

for simplicity taken from linear-response theory (without damping), where E is the size of the electric field at the argon probe due to all other partial charges (including charges-on-spring).

Argon solvation solute-solvent (uv) interaction enthalpies ΔH_{uv} and entropies $T\Delta S_{uv}$ were determined via

$$\Delta H_{uv} = \frac{\langle U_{int} V \exp(-U_{int}/k_B T)\rangle}{\langle V \exp(-U_{int}/k_B T)\rangle} \tag{2.24}$$

and

$$T\Delta S_{uv} = \Delta H_{uv} - \Delta G_S. \tag{2.25}$$

Note that ΔH_{uv} and $T\Delta S_{uv}$ are not directly available from experiment. Because there are no solute-solute interactions in a single argon probe and the solvent-solvent interaction enthalpy and entropy cancel exactly [79, 80] , the solute-solvent terms are the ones that determine the driving force of the solvation [37, 79].

2.4 Results

2.4.1 Liquid phase

Variation of model parameters

	ρ	ΔH_{vap}	R_{max}	$\langle\mu\rangle$
increasing q_H	↑	⇑	∽	↑
increasing α	↑	↑	∽	↑
increasing C_{12}	↓	⇓	↑	∽
increasing C_6	↑	↑	⇓	∽
increasing p	↓	↓	∽	↓
increasing E_0	↑	↑	∽	↑

Table 2.1: Effects of the COS/D model parameters on the properties of liquid water. One arrow: dependence visible, two arrows: stronger dependence, tilda: neglegable dependence. q_H: charge of the H atom, α: molecular polarisability, C_{12}: repulsive oxygen-oxygen Lennard-Jones parameter, C_6: attractive oxygen-oxygen Lennard-Jones parameter, p: damping parameter, E_0: truncation parameter, ρ: density, ΔH_{vap}: heat of vaporisation, R_{max}: first peak in the oxygen-oxygen radial distribution function, $\langle\mu\rangle$: averaged molecular dipole moment.

The impact of changing a single force-field parameter on the properties of the water model derived from observations made during the parametrisation process are summarised in Tab. 2.1, the details are given in Tabs.2.3 and 2.2 and the final COS/D model

	μ_0 [D]	$\alpha(4\pi\epsilon_0)^{-1}$ [$10^{-3} nm^3$]	$C_{12} \cdot 10^{-6}$ [$kJnm^{12}/mol$]	$C_6 \cdot 10^{-3}$ [$kJnm^6/mol$]	p	E_0 [$kJmol^{-1}nm^{-3}]^{1/2}$	ρ [kg/m^3]	ΔH_{vap} [kJ/mol]	U_{pol} [kJmol^{-1}]	U_{LJ} [kJmol^{-1}]	U_{cmb} [kJmol^{-1}]	R_{max} [nm]	$\langle\mu\rangle$ [D]
COS/DB$_1$	1.9	1.4	2.57	2.91	2	150	1027	40.1	24.4	14.3	-76.5	0.271	2.844
COS/DB$_2$	1.95	1.4	2.57	2.91	2	150	1052	42.6	28.1	16.6	-85.2	0.271	2.938
COS/DB$_3$	1.95	1.5	2.57	2.91	2	150	1081	43.1	34.6	19.0	-94.6	0.269	3.050
COS/DB$_4$	1.95	1.2	2.52	2.86	4	125	972	36.6	15.7	10.4	-60.5	0.275	2.604
COS/DB$_5$	1.95	1.2	2.62	2.86	4	125	985	37.2	16.0	10.2	-61.3	0.275	2.609
COS/DB$_6$	1.95	1.2	2.62	2.91	4	125	978	37.0	15.4	10.0	-60.2	0.273	2.603
COS/DB$_7$	1.95	1.15	2.8	2.85	2	125	1015	39.0	16.0	9.8	-62.6	0.273	2.611
COS/DB$_8$	1.95	1.15	2.8	2.85	2	130	1022	39.8	16.0	10.0	-63.7	0.273	2.628
COS/DB$_9$	1.95	1.15	2.8	2.85	3	125	1015	38.7	15.2	9.4	-61.2	0.273	2.592

Table 2.2: Collection of parameter sets and the influence of their variation on certain properties for the model with the charge on spring at the oxygen. μ_0: fixed molecular dipole moment, α: molecular static polarisability, C_{12}: repulsive oxygen-oxygen Lennard-Jones, C_6: attractive oxygen-oxygen Lennard-Jones coefficient, p: damping parameter, E_0: truncation parameter, ρ: density, ΔH_{vap}: heat of vaporisation, U^{self}: polarisation energy, U^{lj}: Lennard-Jones energy, U^{cmb}: Coulomb energy, R_{max}: position of thefirst peak in the oxygen-oxygen radial distribution function, $\langle\mu\rangle$: averaged molecular dipole moment, $\epsilon(0)$: dielectric permittivity.

	q_H [e]	$\alpha_{(4\pi\epsilon_0)^{-1}}$ [$10^{-3} nm^3$]	$C_{12} \cdot 10^{-6}$ [$kJnm^{12}mol^{-1}$]	$C_6 \cdot 10^{-3}$ [$kJmol^{-1}nm^6$]	p	E_0 [$kJmol^{-1}nm^{-3}]^{\frac{1}{2}}$	ρ [kg/m^3]	ΔH_{vap} [kJ/mol]	U_{pol} [kJ/mol]	U_{LJ} [kJ/mol]	U_{cmb} [kJ/mol]	R_{max} [nm]	$\langle\mu\rangle$ [D]	$\epsilon(0)$
COS/D	0.5863	1.4945	3.25	3.45	8	99	997	43.9	16.5	12.7	-71.0	0.275	2.482	69.8
COS/D$_1$	0.58	1.4945	3.1	3.35	8	99	991	43.3	16.6	13.0	-70.6	0.275	2.483	
COS/D$_2$	0.6055	1.4945	3.1	3.35	8	99	1000	46.1	17.4	15.0	-76.3	0.273	2.482	
COS/D$_3$	0.58	1.4945	3.3	3.35	8	99	1018	44.4	16.9	12.4	-71.5	0.273	2.483	
COS/D$_4$	0.58	1.4945	3.1	3.45	8	99	976	42.5	16.1	12.8	-69.2	0.277	2.482	
COS/D$_5$	0.6055	1.4945	3.24	3.46	4	100	1072	49.2	20.2	14.3	-82.0	0.273	2.539	
COS/D$_6$	0.6055	1.4945	3.24	3.46	4	200	1152	60.3	41.2	25.8	-125.0	0.275	3.055	
COS/D$_A$	0.5265	1.494	3.24	3.46	8	100	978	43.8	9.1	8.9	-59.6	0.275	2.400	77.3
COS/D$_B$	0.5265	1.494	3.24	3.46	8	110	987	44.9	10.8	9.9	-63.4	0.273	2.509	85.9
COS/D$_C$	0.5265	1.494	3.24	3.46	2	100	988	45.4	12.2	10.6	-66.0	0.275	2.537	88.2
COS/D$_D$	0.5265	1.494	3.24	3.46	8	1000	1021	47.4	21.5	14.4	-81.1	0.275	2.778	107.1

Table 2.3: Collection of parameter sets and the influence of their variation on certain properties for the model with the charge on spring at a virtual site M. q_H: point charge on the hydrogen, α: molecular static polarisability, C_{12}: repulsive oxygen-oxygen Lennard-Jones, C_6: attractive oxygen-oxygen Lennard-Jones coefficient, p: damping parameter, E_0: truncation parameter, ρ: density, ΔH_{vap}: heat of vaporisation, U^{self}: polarisation energy, U^{lj}: Lennard-Jones energy, U^{cmb}: Coulomb energy, R_{max}: position of the first peak in the oxygen-oxygen radial distribution function, $\langle\mu\rangle$: averaged molecular dipole moment, $\epsilon(0)$: dielectric permittivity.

Model	exp	SPC [9]	SPC/E [81]	COS/B2 [47]	COS/G2 [18]	COS/D
Number of force sites		3	3	4	5	5
d_{OH} [nm]	0.0957(3)(gas) [82], 0.0970(5)(liq) [83]	0.1	0.1	0.1	0.09572	0.09572
$\measuredangle HOH$ [deg]	104.52(5)(gas) [82], 106.1(18)(liq) [83]	109.47	109.47	109.47	104.52	104.52
d_{OM} [nm]					0.022	0.0257
p						8
E_0 $[kJmol^{-1}nm^{-3}]^{1/2}$						99
q_H [e]		0.41	0.4238	0.373	0.5265	0.5863
q_O [e]		-0.82	-0.8476	-0.746	0	0
q_M [e]					-1.0530	-1.1726
μ^0 [D]	1.855 [84]	2.27	2.35	2.07	1.85	1.855
q_{pol} [e]				-8.0	-8.0	-8.0
$\alpha(4\pi\epsilon_0)^{-1}$ $[10^{-3}nm^3]$	1.494(7)			0.930	1.255	1.4945
C_6 $[10^{-3}kJmol^{-1}nm^6]$		2.61735	2.61735	2.75691	3.24434	3.25
C_{12} $[10^{-6}kJmol^{-1}nm^{12}]$		2.63413	2.63413	3.01500	3.45765	3.45

Table 2.4: Parameters of three COS polarisable water models and the SPC and SPC/E models for comparison. d_{OH}: OH bond length, $\measuredangle HOH$: HOH bond angle, d_{OM}: oxygen-virtual M site distance, p: damping parameter, E_0: truncation parameter, q_H: partial charge on the hydrogen, q_O: partial charge on the oxygen, q_M: partial charge on the virtual M site, μ^0: fixed molecular dipole moment, q_{pol}: COS polarisation charge, α: molecular polarisability (which is the mean electronic polarisability (1.457(3)$(4\pi\epsilon_0)^{-1}$ $10^{-3}nm^3$) [85] plus the vibrational contribution (0.037$(4\pi\epsilon_0)^{-1}$ $10^{-3}nm^3$) [86] in the low-frequency limit) , C_6: attractive Lennard-Jones coefficient, C_{12}: repulsive Lennard-Jones coefficient, both for oxygen-oxygen interactions.

parameters are presented in Tab. 2.4. These properties are the heat of vaporisation (ΔH_{vap}), the density (ρ), the position of the first maximum in the oxygen-oxygen radial distribution function (R_{max}) and the average molecular dipole moment ($\langle\mu\rangle$).

A first approach starting from the COS/B2 model (Tab. 2.4), where the charge on spring is attached to the oxygen, did not lead to satisfying results for COS/DB type of models (Tab. 2.2). Fitting both the density ρ and the heat of vaporisation ΔH_{vap} to the experimental values while keeping the radial distribution function $g(r)$ and the dielectric permittivity $\epsilon(0)$ close to the experimental values turned out to be impossible for the COS/DB type of model. A fraction of the results of the parametrisation are shown in Tab. 2.2 to illustrate this.

Next the COS/G2 model (Tab. 2.4), which hast its charge-on-spring attached to a virtual site M at a distance d_{OM} from the oxygen on the molecular symmetry axis on the side of the H atoms, was used as starting point. For this COS/D type of model a satisfactory parameter set could be found (Tab. 2.3).

The density is increased when increasing the charge of the hydrogen atoms while also increasing the distance between the virtual site and the oxygen to keep the permanent dipole constant, the polarisability or the truncation parameter or when decreasing the damping parameter as all of this increases the Coulombic interaction leading to a denser packing. Obviously, the density will increase upon increasing the attractive van der Waals parameter or decreasing the repulsive oxygen-oxygen van der Waals parameter. The heat of vaporisation shows a qualitatively similar dependence on an increase or decrease of the parameters as the density, but a bit stronger one on a change of the charge on the hydrogen atoms and on increasing the repulsive oxygen-oxygen van der Waals parameter. An increase of the van der Waals well depth ($\epsilon_{LJ} = C_6^2/(4C_{12})$) leads to a decrease of the heat of vaporisation which is due to the fact that the Lennard-Jones potential energy is positive for liquid water. This complicates the parametrisation.

The position of the first peak in the oxygen-oxygen radial distribution is only dependent on the oxygen-oxygen van der Waals parameters, proportional to the repulsive one and a bit stronger reverse proportional to the attractive one. The latter shows a stronger effect.

The average molecular dipole is not sensitive to the oxygen-oxygen van der Waals parameter. As expected, it responds with increasing upon increasing the polarisability or the truncation parameter or upon decreasing the damping parameter.

The observed dependencies indicate that by varying these six parameters, the experimental values of the four observables should be reproducible.

The final parameters we chose for the COS/D model are given in Tab. 2.4 together with those of a few previous water models to which the COS/D model is compared in the next subsection.

Comparison of the COS/D model with other models

The non-polarisable models SPC [9] and SPC/E [81] and the polarisable models COS/B2 [47] and COS/G2 [18] were selected for comparison to the COS/D model, because a wide range of liquid state properties is available for these models [18, 47, 87]. For a comparison of such properties for yet other water models we refer to ref. [88–92].

Model	expt	SPC [47,87,93]	SPC/E [87,93]	COS/B2 [47,94]	COS/G2 [18]	COS/D
T [K]	298.15	300.7	301.0	302.5	302.8	296.8
p [atm]	1			5.5	0.93	0.8
ρ [kgm^{-3}]	997 [58]	972	994	992	997	997
ΔH_{vap} [kJmol^{-1}]	44.05 [46]	43.7	48.8	44.2		44.0
γ [mNm^{-1}]	71.6 [93]	53.4	61.3			72.9
U^{pot} [kJmol^{-1}]	-41.5 [64]	-41.3		-41.7	-41.3	-41.8
U^{cmb} [kJmol^{-1}]		-48.2		-64.0		-70.9
U^{self} [kJmol^{-1}]		-		11.5	15.4	16.5
U^{LJ} [kJmol^{-1}]		7.0		10.7		12.7

Table 2.5: Liquid state properties of different water models at $1\,\mathrm{atm}$ and $298.15\,\mathrm{K}$ (for expt, SPC, SPC/E, COS/D) and $300\,\mathrm{K}$ (for COS/B2, COS/G2), respectively. T: temperature, p: pressure, ρ: density, ΔH_{vap}: heat of vaporisation, ΔF_{exs}: excess free energy, γ: surface tension, U^{pot}: total potential energy, U^{cmb}: Coulomb energy, U^{self}: self-polarisation energy, U^{LJ}: Lennard-Jones energy.

Thermodynamic properties

In Tab. 2.5 the energetic properties and the densities of the SPC, SPC/E, COS/B2, COS/G2 and COS/D models are given. The density ρ of the SPC model is known to be too low. The densities of the polarisable models COS/B2, COS/G2 and COS/D and the SPC/E model are in better agreement with experiment. The heat of vaporisation ΔH_{vap} is in accordance with experiment for all models except SPC/E which has a larger value because of the additionally integrated polarisation energy [81].

The main contribution to the potential energy U^{pot} is the Coulombic energy U^{cmb}. The contribution from the Lennard-Jones energy U^{LJ} is 17-25%. And for the polarisable models the polarisation energy U^{self} is between a fourth and a third of the potential energy U^{pot}.

The surface tension γ shows too low values for SPC and SPC/E water. A tendency that is also observable with other water models [93]. The COS/D model has a surface tension close to the experimental value. The heat capacity C_p and the thermal expansion coefficient α are shown for three different models in Tab. 2.6. The heat capacity

T [K]	U^{tot} [kJmol^{-1}]	C_p [Jmol^{-1}K^{-1}]	ρ [kg m^{-3}]	α [10^{-4}K]
		expt		
318		75.31 [95]	990 [58]	4.22 [58]
		SPC [47]		
298	-33.72		972	
		74.2		8.13
318	-32.25		956	
		73.7		9.99
338	-30.59		937	
		COS/B2 [47]		
298	-34.65		997	
		86.7		10.1
318	-32.73		977	
		85.2		11.0
338	-30.84		956	
		COS/D		
298	-34.32		998	
		86.6		2.8
318	-33.04		992	
		85.6		2.8
338	-31.80		987	

Table 2.6: Heat capacity C_p and thermal expansion coefficient α at 1atm pressure. T: temperature, U^{tot}: total energy, ρ: density.

C_p is reasonably well reproduced by all models. The coefficients of thermal expansion α are overestimated by the old models which means that they all change their density ρ too much with changing temperature. While the COS/D model has a coefficient of thermal expansion α that is slightly too low.

The isothermal compressibility κ_T is listed in Tab. 2.7 for the same three water models. All models show results consistent with experiment.

Dynamic properties

The dynamic properties of the models are listed in Tab. 2.8. The SPC water model shows a too large diffusion coefficient D and the rotational relaxation times τ_l^α are too short, which points to a too large mobility of this model. All polarisable models lower the value for the diffusion coefficient D. For the rotational relaxation times τ_l^α the picture is not so clear. In the SPC/E, the COS/G2 and the COS/D model the rotational relaxation is slower than in experiment while the COS/B2 has almost the same values as the SPC model.

ρ [kgm^{-3}]	p [atm]	κ_T [10^{-6}atm^{-1}]
	Exp	
997	1	45.8 [58]
	SPC [47]	
947	-447.08	
		54.7
997	475.77	
		39.8
1047	1706.43	
	COS/B2 [47]	
947	-931.64	
		55.0
997	3.6	
		37.8
1047	1297.0	
	COS/D	
947	-1112.8	
		46.2
997	1.3	
		37.2
1047	1318.0	

Table 2.7: Isothermal compressibility κ_T of different water models at 298K. ρ: density, p: pressure.

Model	expt	SPC [87]	SPC/E [87]	COS/B2 [47]	COS/G2 [18]	COS/D
D [10^{-9}m^{2}s^{-1}]	2.3 [96]	4.2	2.4	2.6	2.3	2.5
τ_2^{HH} [ps]	2.0 [97]	1.7	2.9	1.7	2.4	3.3
τ_2^{OH} [ps]	1.95 [98]	1.6	2.6	1.6	2.2	3.1
τ_2^{μ} [ps]	1.92 [99]	1.4	2.5	1.6	2.0	3.0

Table 2.8: Dynamic properties of the different water models at $1\,\mathrm{atm}$ and $298.15\,\mathrm{K}$ (for expt, SPC, SPC/E, COS/D) and $300\,\mathrm{K}$ (for COS/B2, COS/G2), respectively. D: self-diffusion coefficient, τ_2^{α}: rotational relaxation times of different molecular axes.

Dielectric properties

Tab. 2.9 displays the dielectric properties for the different water models. The dielectric permittivity $\epsilon(0)$ is too low for the non-polarisable models and too high for the polarisable ones, except for the damped COS model (COS/D). The dielectric permittivity

Model	expt	SPC [87]	SPC/E [87]	COS/B2 [47]	COS/G2 [18]	COS/D
μ [D]		2.27	2.35	2.62	2.59	2.43
μ^{ind} [D]		0	0	0.58	0.78	0.56
$\epsilon(0)$	78.4 [100]	66.6	73.5	121.6	87.8	69.8
$\epsilon(\infty)$	5.2 [100], 1.79 [101]			2.67	3.18	2.37
τ_D [ps]	8.3 [100]	6.2	12.1	14.9	9.2	14.1
G_k		2.64	2.66	2.99		2.6
g_k	2.9 [102]	2.51	2.61	3.55	2.53	2.51

Table 2.9: Dielectric properties of the different water models at $1\,\mathrm{atm}$ and $298.15\,\mathrm{K}$ K (for expt, SPC, SPC/E, COS/D) and $300\,\mathrm{K}$ (for COS/B2, COS/G2), respectively. μ: average molecular dipole moment, μ^{ind}: average induced dipole moment per molecule, $\epsilon(0)$: static dielectric permittivity, $\epsilon(\infty)$: infinite frequency dielectric permittivity. τ_D: Debye dielectric relaxation time, G_k: finite system Kirkwood factor, g_k: infinite system Kirkwood factor.

$\epsilon(0)$ is directly related to the average molecular dipole $\langle\mu\rangle$ as shown in Fig. 2.2 and discussed by Guillot [46]. Guillot predicts a dipole moment of $2.4\,\mathrm{D}$ to reproduce the dielectric permittivity correctly, while Fig. 2.2 suggests a value of $2.41\,\mathrm{D}$. Experimentally, the value of the averaged dipole $\langle\mu\rangle$ in liquid water remains uncertain. The most recent experimental result suggests an average dipole moment $\langle\mu\rangle$ of water in the liquid phase of $2.9 \pm 0.6\,\mathrm{D}$ [56], while *ab initio* calculations predict a value between $2.33\,\mathrm{D}$ and $3.0\,\mathrm{D}$ [103–106].

The Debye dielectric relaxation time τ_D gives an approximation for the relaxation time of the hydrogen bond network. In contrast to the SPC model, the other models show larger Debye dielectric relaxation times τ_D than experiment. This overestimation is most probably due to the overestimation of the dielectric permittivity $\epsilon(0)$ as they are related, see Equation (2.14), and the lower mobility as seen in Tab. 2.8.

Structure

The radial distribution functions (RDF) $g(r)$ for the O-O, O-H, and H-H distances are shown in Fig. 2.3 for the COS/D model in comparison to the experimental data at $1\,\mathrm{atm}$ and $300\,\mathrm{K}$. The overall shape of the radial distribution functions g_{OO} of the COS/D model is comparable to the one derived from experiment [65]. As in other polarisable models [47], the first peak is overestimated indicating a slightly overstructured liquid. The coordination numbers of 4.5 for the experiment and 4.3 for the COS/D model are obtained by integrating $g(r)$ to the first minimum in the curve ($0.336\,\mathrm{nm}$). The second and third peaks for the COS/D are slightly more pronounced than in the curves derived from experimental data.

The g_{OH} curve shows the same type of agreement as the g_{OO} curve. The first peak is too pronounced and is shifted toward longer distances. The second peak has the

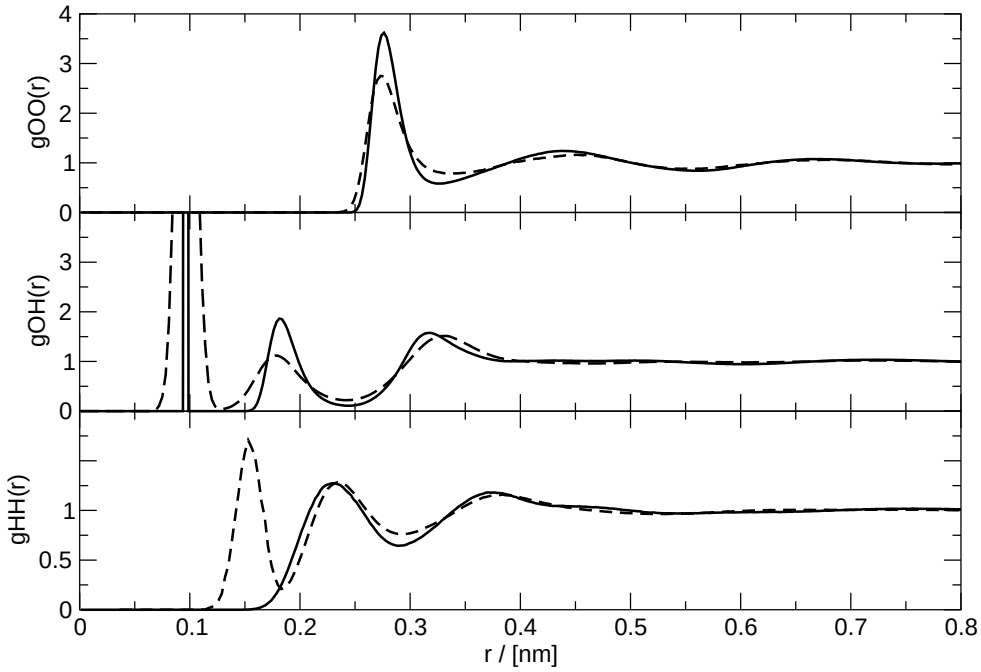

Figure 2.3: Liquid phase radial distribution function at room temperature and pressure for the oxygen-oxygen pair (top) , the oxygen-hydrogen pair (middle) and the hydrogen-hydrogen pair (bottom). The experimental results [65] are shown with a dashed line and the ones of the COS/D model with a solid line.

correct height but is shifted toward shorter distances.

For the g_{HH} curve the agreement with experimental data is good with an only marginally shifted first peak.

2.4.2 Gas phase

The dimer geometry as defined in Fig. 2.4 was optimised for COS/D. It was calculated performing a global conformational search in three dimensions (R_{OO}, θ, and ϕ) with the geometry of the monomers constrained to be rigid. The results are compared to the experimental findings in Tab. 2.10.

While all models give results close to the experimental one for the angle θ_{min}, the experimental angle ϕ_{min} is not properly reproduced by any of them. This less tetrahedral-like association of the monomers in the gas phase may lead to a deficiency in the description of the liquid structure such as a wrong or missing density maximum. The wrong optimal dimer geometry also influences the μ^{dimer} which is strongly depending on the orientation of the molecules and is therefore not well reproduced by any of the described models for liquid water.

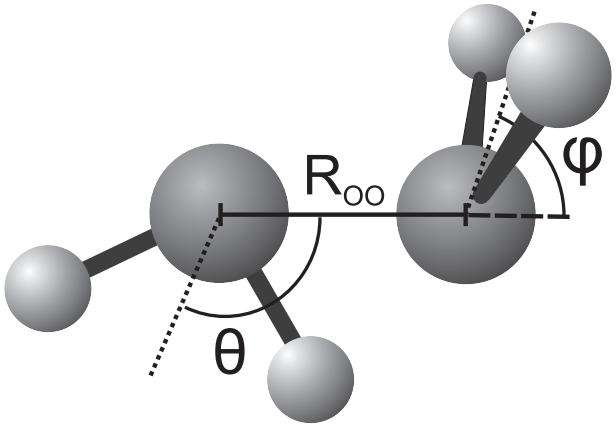

Figure 2.4: Definition of the water dimer geometry with the angles ϕ and θ and the distance R_{OO} that define the relative position and orientation of the monomers. Three atoms of the left monomer lie in one plane with the oxygen of the right monomer. The clockwise angle between the angular bisector of the left monomer and R_{OO} is θ. The vector between the hydrogen atoms of the other monomer is perpendicular to the mentioned plane. The anti-clockwise angle ϕ is between the angular bisector of the right monomer and R_{OO}.

Model	expt	SPC [18]	SPC/E [18]	COS/B2 [18]	COS/G2 [18]	COS/D
R_{OO}^{min} [nm]	0.297 [107]	0.275	0.274	0.279	0.281	0.269
θ_{min} [deg]	51 [107]	52	52	51	56	58
ϕ_{min} [deg]	57 [107]	23	22	20	74	100
U^{pot} [kJmol^{-1}]	-22.7(2.5) [108]	-27.65	-30.10	-23.29	-20.90	-26.16
μ^{dimer} [D]	2.6 [107]	3.59	3.76	3.76	2.08	3.16
μ^{mean} [D]		2.27	2.35	2.26	2.03	2.09

Table 2.10: Gas-phase dimer properties of different water models. Optimal (minimum energy) geometry defined by the O-O distance R(OO), angles θ and ϕ as defined in Fig. 2.4, U^{pot}: interaction energy, μ^{dimer}: total dipole moment, μ^{mean}: average molecular dipole moment.

The dimer separation distance R_{OO}^{min} is underestimated by all water models compared to the experimental value. Here the COS/B2 and COS/G2 models give some improvement compared to SPC and SPC/E, but the COS/D model shows an even smaller separation.

The binding strength that is overestimated by the non-polarisable models SPC and SPC/E, is better reproduced by the polarisable models COS/B2 and COS/G2. The damped polarisable model COS/D still gives a too strong binding strength which is, however, closer to the experimental one than that of the non-polarisable models.

The difference in potential energy between the optimesed dimer structure and the ideal dimer structure is, however, only $2.9\,\mathrm{kJmol}^{-1}$, a difference that can easily be overcome in a simulation at room temperature.

2.4.3 Ice

Model	expt	COS/G2 [18]	COS/D
ρ [kg m^{-3}]	931.0 [109]	958.0	946.7
μ [D]	2.6 [110], 3.09 [111]	2.81	2.53
μ^{ind} [D]		0.97	0.67
U [kJmol^{-1}]	-47.34 [112]	-49.05	-49.4

Table 2.11: Properties of the COS/G2 and COS/D models in the solid state (ice Ih): ρ: density, μ: average molecular dipole moment, μ^{ind}: average induced dipole moment, U: lattice energy.

To complete the investigation the results for the Ih ice are shown in Tab. 2.11. The COS/G2 and the COS/D models give only a slightly too high density ρ. The dipole is lower for the COS/D model as is expected as a direct result from the damping.

2.4.4 Hydrophobic solvation

Model	expt	SPC [79]	SPC/E [79]	COS/G2 [37]	COS/D
ΔG_S [kJ mol^{-1}]	8.4 [113]	8.5	8.9	8.9	8.6
$-\Delta H_{uv}$ [kJ mol^{-1}]		10.2	10.5	12.0	12.1
$-T\Delta S_{uv}$ [kJ mol^{-1}]		18.7	19.4	20.9	20.7

Table 2.12: Solvation properties for an argon probe for different water models at $300\,\mathrm{K}$ and $1\,\mathrm{atm}$. ΔG_S: solvation free enthalpy of a single argon molecule, ΔH_{uv}: argon solvation solute-solvent interaction enthalpy, ΔS_{uv}: argon solvation solute-solvent interaction entropy.

The results for the solvation of an argon probe in water are given in Tab. 2.12. The COS/D model does slightly better than COS/G2 and SPC/E. Compared to SPC it shows a 20% larger compansation of the argon solvation solute-solvent interaction enthalpy ΔH_{uv} and the entropic argon solvation solute-solvent interaction costs ΔS_{uv}. This leads to a slightly higher solvation free enthalpy ΔG_S. We note that the argon model was not parameterised using a COS/D type of polarisable model.

2.5 Conclusions

The presented polarisable water model is of the COS type which avoids complex evaluation of the dipole-dipole interactions and forces, as all electrostatic interactions are point charge interactions. The polarisable COS/D model has 5 interaction sites per molecule. Compared to the non-polarisable SPC model with 3 interaction sites per molecule the computational costs are approximatly a factor 5 higher. The introduced damping mimics the effect of hyperpolarisability in a scalar isotropic way.

The introduced damping of the polarisability α allows for a to reduction of the dielectric permittivity $\epsilon(0)$ of polarisable models to values significantly under the experimental value. It reduces the effect of overpolarisation, but it influences other quantities. Especially the radial distribution function $g_{OO}(r)$, which was rather insensitive to a variation of parameters in previous parametrisations [18, 87], showed a visible response. Using a COS model with its virtual charge site at the oxygen position, it turned out to be impossible to simultaneously obtain the heat of vaporisation ΔH_{vap}, the density ρ, the radial distribution function $g_{OO}(r)$, and the dielectric permittivity $\epsilon(0)$ close to their experimental values. Using a virtual site that is different from the oxygen positon a better model could be derived, the COS/D model.

The average dipole moment of the water molecules $\langle\mu\rangle$ turned out to be a fast computable first approximation for the slowly converging dielectric permittivity $\epsilon(0)$. Yet, the targeted dielectric permittivity $\epsilon(0)$ of 78.5 was not exactly reached for the COS/D model with a value of 69.8 and an average molecular dipole $\langle\mu\rangle$ of $2.43\,\mathrm{D}$.

For the pure liquid the inclusion of damped polarisability does not significantly improve the reproduction of thermodynamic properties. Yet, the developed COS/D model is expected to show more realistic behaviour in simulations in which single water molecules are in different sites experiencing different electric field strengths such as in proteins or within protein-protein or protein-DNA interfaces.

If we don't feel stupid it means we're not really trying.

Martin A. Schwartz

Chapter 3

A simple, efficient polarisable molecular model for liquid carbon tetrachloride

3.1 Summary

A one-site and a five-site polarisable model for liquid carbon tetrachloride CCl_4 is presented. They are based on a non-polarisable model consisting of five van der Waals sites not carrying any partial charges. In the one-site model, a charge on spring with a polarisability α_{CCl_4} of $11.1(4\pi\epsilon_0)10^{-3}\mathrm{nm}^3$ was attached to the carbon to make the model polarisable, while in the five-site model polarisabilities $\alpha_C = 1.848(4\pi\epsilon_0)10^{-3}\mathrm{nm}^3$ and $\alpha_{Cl} = 2.289(4\pi\epsilon_0)10^{-3}\mathrm{nm}^3$ were added to the carbon and chlorine atoms, respectively. Both models exactly reproduce the experimental static dielectric permittivity of 2.24 at $293\,\mathrm{K}$ and $1\,\mathrm{atm}$. This quantity was calculated by applying a homogeneous external electric field of varying strength. The one-site polarisable model is only about 1.7 times more computationally expensive than the non-polarisable one and is compatible with the GROMOS force field. A selected set of thermodynamic, dynamic and structural quantities was calculated and compared to experiment.

3.2 Introduction

The simulation of (bio)molecules has grown in importance over the past few decades as a means to understand their properties and behaviour [114–117]. The faithfulness of such simulations critically depends on the accuracy of the force field employed. The most widely used biomolecular force fields treat nonbonded interactions as pairwise additive, where many-body contributions are included in an average manner [59, 118–120]. This approach usually gives a satisfactory description for homogeneous systems, such as pure liquids or mixtures of compounds of similar polarity, but

may break down for spatially electronically inhomogeneous ones, such as a protein or membrane in water, interfaces or mixtures of different solvents. For this kind of systems the introduction of polarisability should and does lead to an improved description of the interactions [7, 37–44]. Polarisability can be included into classical simulations using different approaches [8, 45] such as the point polarisable dipole model (PPD) [10–12], the charge-on-spring (COS) [13] model, also called Drude oscillator [14] or shell [15] model, and the fluctuating charge (FQ) [16] model. The GROMOS polarisable force fields and software make use of the COS [8] model to integrate polarisability as it leads to simple formulas, is computationally efficient and compatible with the non-polarisable (GROMOS) force fields.

As the solvent plays an important role to describe the effect of polarisability, large efforts have been undertaken to find good polarisable models for a wide range of solvents, especially for water [18,47,121,122], which is the most common solvent in biomolecular simulations, but also for less polar solvents, such as chloroform [123–125], ethyleneglycol [37], dimethyl ether [94], methanol and ethanol [126–128]. These solvents have a higher polarisability than water and are therefore able to respond stronger to a change in the electric field of the environment.

The present paper describes the parametrisation of a polarisable model for carbon tetrachloride (CCl_4) for which the importance of polarisation has been shown earlier [129, 130]. The model introduced here is based on a non-polarisable model [131] for CCl_4. The only novelty is the addition of a charge on spring [8] which was added to obtain a polarisable model with the correct static dielectric permittivity $\epsilon(0)$. To keep the model as simple as possible and the computational cost low, a model with only one polarisable centre located at the carbon was developed. However, for comparison, a model with five polarisable centres, located at every atom, was also parametrised.

Since CCl_4 is a non-polar, symmetric molecule, most properties of the liquid as calculated for a non-polarisable model that does not contain partial charges, will not be changed by the introduction of polarisable sites in the molecule. Only the static dielectric permittivity will change from $\epsilon(0) = 1$ for the non-polarisable model. This would imply that we only would have to calculate this quantity, because an extensive set of properties was reported earlier [131]. However, to check the quality of the original parametrisation, a few quantities were recalculated using an increased system size and simulation time. Additionally, the total structure factor $I(Q)$ for neutron scattering was calculated, because it was not reported in [131]. It allows to compare to primary, observed experimental data and reflects the quality of the structural description of CCl_4 in the simulation of the liquid.

3.3 Methods

3.3.1 Developing the model and parametrisation

The static dielectric permittivity $\epsilon(0)$ is calculated applying a external homogeneous electric field $\mathbf{E}^{ex}$, e.g. along the z-axis $\mathbf{E}^{ex}$ = (0, 0, E^{ex}). Due to the response of the environment inside and outside the cut-off sphere the applied external field that acts on the charges of the system becomes

$$\mathbf{E}^{ex,cs} = \frac{3\epsilon_{RF}}{2\epsilon_{RF} + \epsilon_{cs}} \mathbf{E}^{ex} \tag{3.1}$$

where ϵ_{RF} is the dielectric constant of the reaction field outside and ϵ_{cs} within the cut-off sphere. For the system presented $\epsilon_{cs} = 1$. This field induces the charge on spring (COS) dipoles with a repulsive force $\mathbf{f}^{ex} = q^{v}\mathbf{E}^{ex,cs}$ between the two virtual (massless) charges q^{v} connected by a spring.

The polarisation of the system, that is the computational box,

$$\mathbf{P}(t) = V(t)^{-1}\mathbf{M}(t) = V(t)^{-1} \sum_{i=1}^{N_{mol}} \boldsymbol{\mu}_i(t) \tag{3.2}$$

can be related to the applied electric field by

$$\mathbf{P}(t) = \epsilon_0(\epsilon(0) - 1)\mathbf{E}^{ex} \tag{3.3}$$

with $V(t)$ the volume and $\mathbf{M}(t)$ the total dipole of the box at time t, ϵ_0 the permittivity of the vacuum, N_{mol} the number of molecules in the box and the dipole moment of a molecule given by the positions of the N_{COS} virtual charges,

$$\boldsymbol{\mu}_i(t) = \sum_{k=1}^{N_{COS}} q_{ik}^{v} \Delta\mathbf{r}_{ik}, \tag{3.4}$$

where $\Delta\mathbf{r}_{ik} = \mathbf{r}_{ik}^{+} - \mathbf{r}_{ik}^{-}$, the vector pointing from the negative virtual charge towards the positive virtual charge of the spring of the k^{th} polarisable site on molecule i. As the molecules have no permanent charges and because the massless COS charges adapt their positions instantaneously to the electric field, there is only polarisation in the field, i.e. z-direction. Since the induced dipoles of the different molecules interact with each other, the structure of the liquid will influence the polarisation. Within a molecule the induced dipoles do not interact with each other as this would lead to a possible polarisation catastrophe. So we have

$$\langle P_z \rangle_t = \epsilon_0(\epsilon(0) - 1)E_z^{ex} \tag{3.5}$$

where $\langle..\rangle_t$ stands for the average over time, which can be rewritten as

$$\epsilon(0) = 1 + \lim_{E_z^{ex}\to 0} \epsilon_0^{-1} \frac{\langle P_z\rangle_t}{E_z^{ex}}. \tag{3.6}$$

To get a linear response of the polarisation the applied electric field should be small enough to avoid saturation [132], which is the case if

$$\frac{\langle \mu_{i,z}\rangle_t E_z^{ex,cs}}{3k_B} \ll T \tag{3.7}$$

with k_B the Boltzmann constant and T the temperature, is fulfilled. Although this condition only applies when orientational polarisation of the molecules plays a role, which is only slightly possible for the five-site polarisable model considered here, in the present study with $T = 293\,\mathrm{K}$ only data with $\langle \mu_{i,z}\rangle_t E_z^{ex,cs}/3k_B < 50\,\mathrm{K}$ was considered in the calculation of the dielectric permittivity $\epsilon(0)$.

The COS model has three parameters that have to be decided upon: (i) the site of the charge on spring, (ii) the charge q^ν, and (iii) the polarisability α. In the case of one polarisable site the centre of the carbon atom was chosen as location of the COS and in the case of five sites, the centres of all atoms were a COS site. The charge q^ν was set to the standard value [13,17] of $-8e$. This leaves only the value of the polarisability α as an adjustable parameter. The starting value for one polarisable site was the molecular electronic polarisability α_{CCl_4} and for five sites the electronic polarisabilities α_C and α_{Cl} as derived from experiment.

3.3.2 Simulation methods

A cubic box with an edge length of $5.4305\,\mathrm{nm}$ was filled with 1000 CCl_4 molecules, resulting in a density of $1595\,\mathrm{kg/m^3}$, corresponding to the density of liquid carbon tetrachloride at $293\,\mathrm{K}$ and $1\,\mathrm{atm}$. [95]. Molecular dynamics simulations were performed at constant pressure and temperature (NpT conditions) with the GROMOS05 package [133], modified to incorporate polarisation and the possibility to apply an external electric field. The geometry of the rigid CCl_4 molecules was constrained by applying the SHAKE algorithm [23] with a relative geometric tolerance of 10^{-4} on the 4 CCl bond lengths and on 5 of the intramolecular ClCl distances. The temperature was weakly coupled [29] to a bath of $293\,\mathrm{K}$ with a relaxation time of $0.1\,\mathrm{ps}$ and the pressure was weakly coupled [29] to a bath of $1\,\mathrm{atm}$ with a relaxation time of $0.5\,\mathrm{ps}$. The isothermal compressibility was set to $7.513 \cdot 10^{-4}\,(\mathrm{kJmol^{-1}nm^{-3}})^{-1}$. The nonbonded van der Waals and electrostatic interactions were calculated using triple-range cutoff radii of $0.8/1.4\,\mathrm{nm}$. The short range interactions were calculated every time step by updating the molecular pair list for distances smaller than the first cutoff radius of $0.8\,\mathrm{nm}$. For the

intermediate range of distances between 0.8 and $1.4\,\mathrm{nm}$ the pairlist was only updated every fifth timestep and at the same time the interaction was calculated and kept unchanged between these updates. The long range electrostatic interactions beyond the outer cutoff of $1.4\,\mathrm{nm}$ were represented by a reaction field [61,62] with $\epsilon_{RF} = 2.2379$. The equations of motion were integrated using the leap frog algorithm with a time step of $2\,\mathrm{fs}$. The velocities of the atoms at the beginning of the simulation were assigned from a Maxwell distribution at $293\,\mathrm{K}$. In order to calculate the dielectric permittivity of the system, several simulations with applied electric fields of field strengths between 0.025 and 1 $(4\pi\epsilon_0)^{-1}\mathrm{enm}^2$ were performed. The position of the COS was determined iteratively with a convergence criterion of $2.4\,\mathrm{kJmol}^{-1}$ [8]. One simulation without an applied electric field was performed as well. The production runs were $1\,\mathrm{ns}$ long after $200\,\mathrm{ps}$ of equilibration. Coordinates were saved every $2\,\mathrm{ps}$ for analysis.

3.3.3 Analysis

Heat of vaporisation

The heat of vaporisation was calculated using the formula

$$\Delta H_{vap}(T) = -U^{liquid}(T) + p\Delta V = -U^{liquid}(T) + RT, \tag{3.8}$$

where ΔH_{vap} is the experimental molar heat of vaporisation, U_{liquid} is the computed intermolecular potential energy per mole, p the pressure, and ΔV the molar volume change between liquid and gas. R is the gas constant and T is the absolute temperature.

Self-diffusion coefficient D

The diffusion coefficient is obtained from the long-time limit of the mean-square displacement according to the Einstein relation, [66]

$$D = \lim_{t\to\infty} \frac{\langle(\mathbf{r}(\tau + t) - \mathbf{r}(\tau))^2\rangle_{\tau,molecules}}{6t} \tag{3.9}$$

where $\mathbf{r}(t)$ corresponds to the position vector of the centre of mass of a molecule at time$t$, and the averaging is performed over both time and carbon tetrachloride molecules.

Radial distribution function g(r)

The pair distribution function $g(r)$ represents the probability of finding another atom at a distance r from a given atom, relative to the probability expected for a completely

uniform distribution at the same density, and can be calculated by a simple histogram summation in radial shells over all molecules in the system using

$$g_{\alpha\beta}(r) = \frac{n(r)}{4\pi r^2 \Delta r \rho} \tag{3.10}$$

where $\alpha, \beta \in \{C, Cl\}$ are the atom types, $n(r)$ the number of atoms of type β around an atom of type α at distances between r and $r + \Delta r$ thereby excluding intra-molecular pairs (α, β), and ρ the density of the liquid.

Total structure factor I(Q)

Assuming an isotropic liquid, the time-dependent description of the scattering cross-section together with the theory of van Hove [134] allows to define partial structure factors $S(Q)$ to simplify the description of scattering effects. With these partial structure factors the total structure factor,

$$I(Q) = \frac{1}{N} \left\langle \sum_{i=1}^{N} \sum_{j=1}^{N} b_i b_j \exp(-i\mathbf{Q} \cdot (\mathbf{r}_i - \mathbf{r}_j)) \right\rangle, \tag{3.11}$$

where $\mathbf{Q} = 2\pi\mathbf{q}$ and $\mathbf{q}$ is a position vector in the reciprocal space, and b_i is the scattering length of atom i, for a system containing N atoms, can be described by

$$I(Q) = I_{self}(Q) + I_{distinct}(Q), \tag{3.12}$$

with

$$I_{self}(Q) = \sum_{\alpha}^{T_{atoms}} c_\alpha^2 \left\langle b_\alpha^2 \right\rangle, \tag{3.13}$$

$$I_{distinct}(Q) = \sum_{\alpha}^{T_{atoms}} \sum_{\beta \geq \alpha}^{T_{atoms}} (2 - \delta_{\alpha\beta}) c_\alpha c_\beta \langle b_\alpha \rangle \left\langle b_\beta \right\rangle S_{\alpha\beta}(Q), \tag{3.14}$$

where the δ-function $\delta_{\alpha\beta}$ is used to avoid double counting, $c_\alpha = \frac{N_\alpha}{N}$ is the atomic fraction of atomic species α with N_α the number of atoms of that species and T_{atoms} is the number of types of atoms. The partial structure factor $S_{\alpha\beta}(Q)$ is defined by

$$S_{\alpha\beta}(Q) = 4\pi\rho_\alpha \int_0^\infty r^2 (g_{\alpha\beta}(r) - 1) \frac{\sin(Qr)}{Qr} dr \tag{3.15}$$

where ρ_α is the atomic number density, and $g_{\alpha\beta}$ may or may not contain intra-molecular contributions.

To include the elasticity of the system, the distinct scattering intensity $I_{distinct}(Q)$ is split in two terms, an intra- and an inter-molecular term where the intra-molecular term includes the Debye-Waller factor [135, 136]

$$I_{distinct}(Q) = I_{intra}(Q) + I_{inter}(Q) \tag{3.16}$$

with

$$I_{intra}(Q) = \frac{1}{N}\sum_{i}^{N}\sum_{j\neq i}^{N}\langle b_i\rangle \left\langle b_j\right\rangle S_{ij}^{intra}(Q) \tag{3.17}$$

$$I_{inter}(Q) = \frac{1}{N}\sum_{\alpha}^{T_{atoms}}\sum_{\beta\geq\alpha}^{T_{atoms}}(2-\delta_{\alpha\beta})c_\alpha c_\beta \langle b_\alpha\rangle \left\langle b_\beta\right\rangle S_{\alpha\beta}^{inter}(Q) \tag{3.18}$$

where the summations in Eq.3.17 are over intra-molecular pairs and in Eq.3.18 over inter-molecular atom pairs, and

$$S_{ij}^{intra}(Q) = \frac{\sin(Qr_{ij})}{Qr_{ij}}\exp\left(-\frac{(\sigma_{ij}Q)^2}{2}\right), \tag{3.19}$$

$$S_{\alpha\beta}^{inter}(Q) = 4\pi\rho_a\int_0^\infty r^2(g_{\alpha\beta}(r)-1)\frac{\sin(Qr)}{Qr}\,dr, \tag{3.20}$$

where σ_{ij} is the root-mean-square deviation of the intra-molecular distance between atoms i and j.

For CCl_4 scattering lengths of $b_C = 6.646$ fm and $b_{Cl} = 9.58$ fm were used [100]. The σ_{CCl} and σ_{ClCl} were estimated from the ones from water [137] assuming a harmonic force with force constant k on the O-H bond in water and on the C-Cl bond in CCl_4: $\sigma_{CCl} = (k^{OH}/k^{CCl})^{1/2}\sigma_{OH}$ and likewise for σ_{ClCl}. The values derived in this manner are σ_{CCl} = 0.01 nm and $\sigma_{ClCl} = 0.02$ nm. The cut-off for the radial distribution function was 2.7 nm.

3.4 Results

3.4.1 Thermodynamic and dynamic properties

The results of the simulation without an applied electric field were compared to those for other models found in the literature (Table 3.1). Of these models only the one of Chang *et al.* [139] is polarisable. It has to be noted that the temperatures for the reported values are not identical for the different models.

The density is well reproduced by all the models. All of them underestimate it a bit, except the one of Tironi *et al.* [131] and this work which slightly overestimate it.

model	T [K]	ρ [kgm^{-3}]	ΔH_{vap} [kJmol^{-1}]	U^{pot} [kJmol^{-1}]	D [10^{-9}m^{2}s^{-1}]
this work	293.0	1600	32.40	-29.95	1.86
Tironi [131]	293	1601	32.45		1.8
McDonald [138]	288	1594		-32.6	
Chang [139]	298	1590	32.5	-29.4	1.8
Soetens [140]	298.4	1520		-30.88	2.2
Li [141]	293	1580			1.306
exp.	293	1594 [95]	32.4 [142]		1.15 [143]
exp.	298.15 [144]		32.4	-30.0	

Table 3.1: Properties of liquid carbon tetrachloride for different models. T: temperature, ρ: density, ΔH_{vap}: heat of vaporisation, U^{pot}: total potential energy, D: self-diffusion coefficient.

Since these two models are, apart from the polarisability, identical, the values should be identical, as long as the different system sizes, N_m=216 in [131] versus N_m=1000 here, and the simulation lengths, 50 ps in [131] versus 1 ns here, do not affect the values.

The heat of vaporisation is not reported for all models. The agreement for Chang *et al.* [139] , Tironi *et al.* [131] and this work are very good. For two other models only the potential energy was reported. It is too negative for Soetens *et al.* [140] and McDonald *et al.* [138].

The self-diffusion constant is overestimated by all models. Only Li *et al.* [141] reports a value close to the experimental one, the other models overestimate it by about 50%.

The inclusion of polarisability does not have an impact on the considered properties of the pure liquid, as expected.

3.4.2 Structural properties

The radial distribution functions for all possible pairs in the CCl_4 liquid are given in Fig. 3.1. The total structure factor $I(Q)$ was calculated using these radial distribution functions and compared to experiment, see Fig. 3.2. The agreement between the measured and the calculated total structure factors is quantitatively good, comparable to that reported by Li *et al.* [141] and Soetens *et al.* [140].

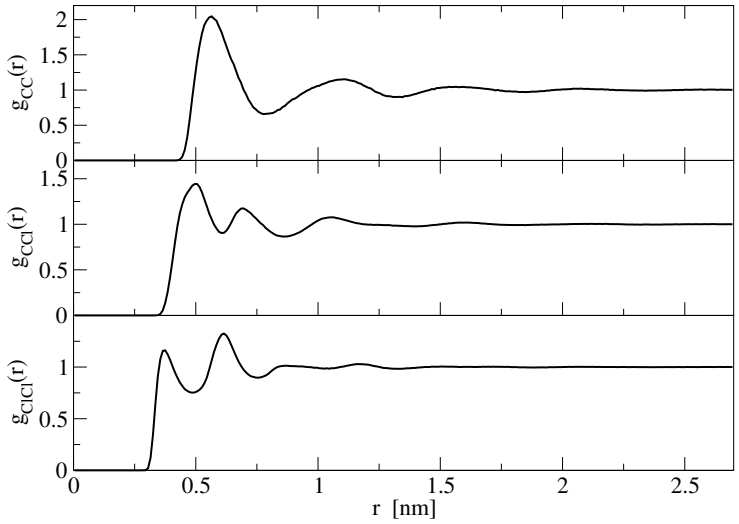

Figure 3.1: Radial distribution function for the carbon-carbon (top), the carbon-chloride (middle) and the chloride-chloride (bottom) pairs.

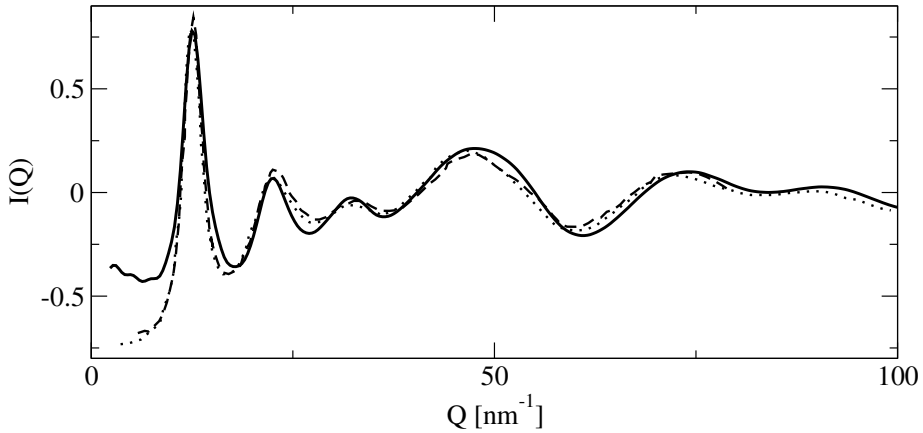

Figure 3.2: Comparison between experimentally measured structure factors (dotted line: Misawa [145], dashed line: Pusztai [146]) and the calculated one (solide line).

3.4.3 Static dielectric permittivity

For both models, the one with one polarisable site and the one with five polarisable sites, different values for the polarisability α were investigated, see Table 3.2. The models 5_3 and 1_3 reproduce the experimental static dielectric permittivity $\epsilon(0)$. To achieve that, the experimental electronic polarisability had to be increased by approximately 5%. As the model is rigid and does not have partial charges, the effect of intra-molecular vibrational polarisability, which is 0.69 $(4\pi\epsilon_0)10^{-3}\mathrm{nm}^3$ for CCl_4 in the low-frequency limit [86], justifies an increase of the polarisability to fit the dielectric per-

5 polarisable sites			
	α_C / $(4\pi\epsilon_0)10^{-3}$nm^3	α_{Cl} / $(4\pi\epsilon_0)10^{-3}$nm^3	$\epsilon(0)$
Exp. [100]	1.76	2.18	2.2379
model 5_1	1.76	2.18	2.17
model 5_2	1.936	2.398	2.31
model 5_3	1.848	2.289	2.24
1 polarisable site			
	α_{CCl_4} / $(4\pi\epsilon_0)10^{-3}$nm^3		$\epsilon(0)$
Exp. [100]	10.5, 11.2		2.2379
model 1_1	10.5		2.16
model 1_2	11.55		2.30
model 1_3	11.1		2.24

Table 3.2: Parametrisation of three polarisable models with five and one polarisable sites. Experimental and computed values at 293 K and 1 atm.

mittivity. Fig. 3.3 shows the linear relation between the polarisation and the applied electric field. Even data points above the chosen criterion of $\langle\mu_{i,z}\rangle_\tau E_z^{ex,cs}/3k_B < 50\,\mathrm{K}$ remain quite linear. In the case of model 1_3 the inclusion of these two additional points would have lowered the regression correlation factor R^2 from 0.999999991 (for values $\langle\mu_{i,z}\rangle_\tau E_z^{ex,cs}/3k_B 50\,\mathrm{K}$) to 0.9999992.

For the static dielectric permittivity the effect of introducing polarisability is eminent as the non-polarisable model by Tironi *et al.* [131] has an $\epsilon(0)$ of 1.

3.5 Conclusions

Two polarisable models for liquid CCl_4 were developed and investigated. They have exactly the same model parameters as those of the model by Tironi *et al.* [131], apart from the introduction of COS polarisability α on the carbon atom or on all 5 atoms of CCl_4. The results for the pure liquid without applying an electric field using an increased system size and simulation time are the same as reported before [131]. The calculated thermodynamic properties density, heat of vaporisation, potential energy at 293 K and 1 atm remain close to the experimental values, while the diffusion remains too high. The five-site model with α_C=1.848 $(4\pi\epsilon_0)10^{-3}$nm^3 and α_{Cl}=2.289 $(4\pi\epsilon_0)10^{-3}$nm^3 and the one-site model with α_{CCl_4}=11.1 $(4\pi\epsilon_0)10^{-3}$nm^3 exactly reproduce the experimental static dielectric permittivity of liquid CCl_4 at 293 K and 1 atm. The calculation of the total structure factor suggests that the structure of the simulated liquid matches closely the experimentally measured one. The computational cost of the one-site polarisable model is a factor 1.7 larger than for the non-polarisable model. For five polarisable sites this factor becomes 2.2. Calculation of the static dielectric permittivity using

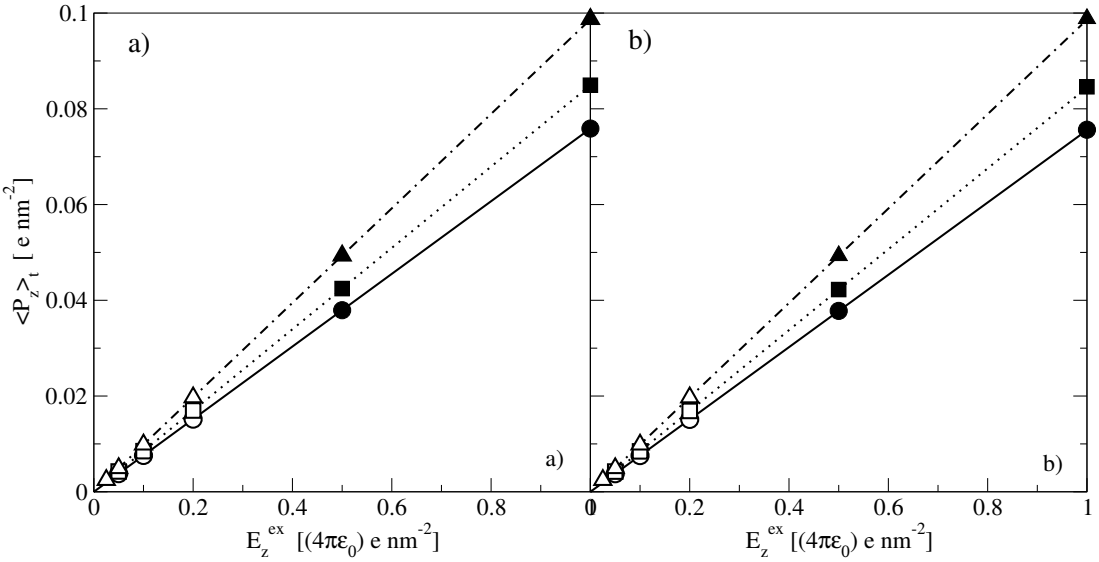

Figure 3.3: Dependence of the polarisation averaged over time $\langle P_z \rangle_t$ on the applied electric field E_z^{ex}. a) 1 polarisable site: circles: α_{CCl_4}=10.5 $(4\pi\epsilon_0)10^{-3}$nm^3, with full line from linear regression; squares: α_{CCl_4}=11.1 $(4\pi\epsilon_0)10^{-3}nm^3$, with dotted line from linear regression; triangle: α_{CCl_4}=11.55 $(4\pi\epsilon_0)10^{-3}$nm^3, with dash-dotted line from linear regression. Full symbols have a $\langle \mu_{i,z} \rangle_\tau E_z^{ex,cs}/3k_B > 50$ K and were excluded from the linear regression; b) 5 polarisable sites: circles: α_C=1.76 $(4\pi\epsilon_0)10^{-3}$nm^3, α_{Cl}=2.18 $(4\pi\epsilon_0)10^{-3}$nm^3, with full line from linear regression; squares: α_C=1.848 $(4\pi\epsilon_0)10^{-3}$nm^3, α_{Cl}=2.289 $(4\pi\epsilon_0)10^{-3}$nm^3, with dotted line from linear regression; triangle: α_C=1.936 $(4\pi\epsilon_0)10^{-3}$nm^3, α_{Cl}=2.398 $(4\pi\epsilon_0)10^{-3}$nm^3, with dash-dotted line from linear regression. Full symbols have a $\langle \mu_{i,z} \rangle_\tau E_z^{ex,cs}/3k_B > 50$ K and were excluded from the linear regression.

the described method of applying a homogeneous external electric field is a simple and efficient way to parametrise polarisable models against the static dielectric permittivity for molecular systems without permanent charges.

You have created a monster, and it will destroy you!

Doctor Waldmann in Frankenstein

Chapter 4

Enhanced sampling of particular degrees of freedom in molecular systems based on adiabatic decoupling and temperature or force scaling

4.1 Summary

A method to enhance sampling of a small subset of N^h particular degrees of freedom a system of $N^h + N^l$ degrees of freedom is presented. It makes use of adiabatically decoupling these degrees of freedom by increasing their mass followed by either increasing their temperature or reducing their interaction or the force acting on them. The appropriate statistical-mechanical expressions for use of these methods in simulation studies are derived. As long as the subset of mass-increased degrees of freedom is small compared to the total number of degrees of freedom of the system, sampling of this subset of degrees of freedom can be much enhanced at the cost of a slight perturbation of the configurational distribution. This is illustrated for a test system of 1000 SPC water molecules at $300\,\mathrm{K}$ and a density of $997\,kgm^{-3}$. Various fractions $N^h/(N^h + N^l)$ of water molecules were adiabatically decoupled. The size of the diffusion coefficient of these decoupled water molecules was used as a measure for how much the sampling was enhanced and the average potential energy per water molecule was used as a measure of how much the configurational distribution of the system gets distorted. A variety of parameter values was investigated and it was found that for $N^h/(N^h + N^l) \leq 0.1$ the diffusion of the N^h molecules could be enhanced by factors up to 35 depending on the method, the ratio $N^h/(N^h + N^l)$, the degree of decoupling, and the temperature or force scaling factors, at the cost of a slight perturbation of the configurational distribution.

4.2 Introduction

Computer simulation of molecular systems at the atomic level has contributed much to explanation of measured structural and thermodynamic properties of large solute molecules in aqueous or other solutions in terms of conformational or configurational distributions [5, 114–117]. Molecular dynamics (MD) or Monte Carlo (MC) simulation of mixtures of different molecules in the condensed phase yield a Boltzmann distributed conformational or configurational ensemble, from which a variety of properties that depend on the molecular coordinates, can be derived through averaging over the ensemble. The obtained accuracy of the averaged properties is limited on the one hand by the quality of the chosen molecular model and force field, and on the other hand by the attained extent of sampling of the, generally vast, conformational or configurational space that is accessible to the molecular system at a given temperature. Over the past decades, a plethora of methods and techniques to enhance the sampling power of MD and MC simulations has been proposed and applied to a diversity of systems, see e.g. [19, 147, 148] and references therein. Since the configurational part of the canonical distribution function of a molecular system does not depend on the masses of the atoms, as long as no configurational constraints are used in flexible molecules, a change of the masses of a subset of the atoms will not influence the configurational ensemble. This consideration was used to lengthen the time step for integrating the equations of motion in an MD simulation of liquid water [149,150] or a polypeptide [151–153] by increasing the masses of the light hydrogen atoms, the fast motion of which limited the MD time step. Using appropriate masses for the various atoms in a molecular system high-frequency motions can be slowed down and low-frequency ones sped up, thereby increasing the efficiency with which the configuration space can be explored in a given amount of computer time [154, 155]. For example, the sampling of the folding equilibrium of a polypeptide can be considerably enhanced by a reduction of the viscosity of the solvent, which is simply attained by a down scaling of its mass [156, 157].

In the mentioned studies, mass scaling was used to equalise the frequencies of the various motions in a molecular system, thereby increasing the coupling between these motions. It can, however also be used to decouple the motions of a particular subset of degrees of freedom of molecular system from the remaining degrees of freedom. By increasing the masses of the atoms defining the particular subset of degrees of freedom, their motion becomes adiabatically decoupled from the motion of the remaining, light particle degrees of freedom. This allows an enhanced sampling of the heavy particle degrees of freedom by raising their temperature without a rapid loss of energy from the high temperature, heavy particle degrees of freedom to the low temperature, light particle degrees of freedom [30, 35, 158]. Whether particular degrees of freedom can be easily decoupled from the remaining ones depends on their nature, e.g. whether they concern separate molecules or collective coordinates of a macromolecule [159–161]. In case the large mass degrees of freedom are a set of collective coordinates that are a function of the atom positions, two alternative approaches can be followed. One may

perform a coordinate transformation from Cartesian coordinates to generalised coordinates, which explicitly contain the collective coordinates [162]. This is, however, a cumbersome procedure in practice. Alternatively, one may treat the collective, large mass coordinates as extended phase space coordinates, thereby avoiding the need of coordinate transformations at the cost of the introduction of an additional parameter governing the harmonic coupling to the extended phase space coordinates which may induce spurious oscillations [163, 164]. Other schemes based on modifying the momentum distribution of the molecular system with the aim of enhanced configurational sampling have been proposed too [165–167].

When sampling from a canonical distribution, probability is proportional to $exp[-\beta V(\mathbf{r}^N)]$, where $\beta = (k_B T)^{-1}$, k_B is the Boltzmann constant, T the temperature, and $V(\mathbf{r}^N)$ the total potential energy of the system, which is a function of the Cartesian coordinates of the N atoms, $\mathbf{r}^N = (\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N)$. This means that the probability of sampling a given configuration $\mathbf{r}^N$ decreases exponentially as the potential energy increases. Sampling can be improved by changing the probability distribution such that the probability of high energy configurations is increased. Sampling from another distribution than the canonical one can be expressed in terms of canonical sampling in which an additional biasing potential energy term has been added to the Hamiltonian, $K(\mathbf{p}^N) + V(\mathbf{r}^N)$, where $K(\mathbf{p}^N)$ denotes the kinetic energy of the atoms. The ensemble that is sampled should, however, not deviate significantly from the original canonical distribution, as otherwise time will be spent sampling configurations which do not contribute to the property or quantity Q of interest,

$$\langle Q \rangle = \frac{\int Q(\mathbf{r}^N) exp[-\beta V(\mathbf{r}^N)] d\mathbf{r}^N}{\int exp[-\beta V(\mathbf{r}^N)] d\mathbf{r}^N}. \tag{4.1}$$

As mentioned, one approach to increase the probability of visiting high energy configurations is to simulate at high temperature. This corresponds to a simulation using an effective Hamiltonian in which all interactions are scaled down by a constant factor. This approach can, however, easily lead to insufficient sampling of low energy configurations, which contribute most to the ensemble average $\langle Q \rangle$ in Eq.(4.1). To alleviate this problem a variety of techniques has been proposed which are based on a down or upscaling of only particular regions of the energy hypersurface $V(\mathbf{r}^N)$, see [19, 147] and references therein. For example, solute-solvent interactions can be reduced in order to speed up the sampling of the folding equilibrium of a polypeptide [36, 168, 169].

Enhanced sampling can also be achieved by a direct modification of the forces acting on the atoms of the system. In an early application of this idea, the forces are modified such that a set of configurations with constant potential energy is generated [170]. This potential-energy annealing conformational search (PEACS) leads to enhanced sampling. Alternatively, the physical force can be augmented by a guiding force which ideally corresponds to the gradient of the local free-energy surface [171–173]. This guiding force is estimated from the physical forces experienced by the atoms using

a memory function. As the method uses time-averaged information, it generates irreversible trajectories which may lead to errors in the calculated canonical averages (Eq.(4.1)) [174, 175]. The method called self-guided Langevin dynamics uses yet another additional guiding force to enhance the sampling [176, 177].

In the presented article, we investigate three methods to enhance the sampling of a particular subset of the degrees of freedom of a molecular system that are all based on an adiabatic decoupling of the motion along the particular subset of degrees of freedom from that of the remaining degrees of freedom. The adiabatic decoupling is achieved by an up-scaling of the masses of the atoms determining the particular subset of degrees of freedom. The sampling of the latter degrees of freedom is then enhanced by i) raising their temperature, as in standard adiabatic decoupling simulations [30, 35, 158], ii) down-scaling of the interactions, i.e. potential energy, involving the atoms of the particular degrees of freedom, as in [36, 168, 169], or iii) modification of the forces on the atoms of the particular degrees of freedom.

As a test system to evaluate the performance of the different methods we use liquid water at room temperature and pressure, because its configurational distribution can be analysed using different radial distribution functions and is governed by translational as well as rotational motion.

4.3 Methods

4.3.1 Adiabatic decoupling

The basic idea of adiabatic decoupling is to prevent energy exchange between two sets of degrees of freedom of a system, while maintaining their mutual interaction. This can be achieved by scaling the masses of one set of degrees of freedom such that they differ so much from the masses governing the motion along the other set of degrees of freedom that the transfer of kinetic energy between them, through momentum conserving collisions, becomes very small. Because the masses of the particles only appear in the momentum part or factor in the partition function, at least if no constraints are present in molecules with internal degrees of freedom, the momentum part can be analytically calculated and the configurational part of the partition function is independent of the masses. Thus when applying adiabatic decoupling, the non-dynamic equilibrium properties of a system remain unaltered. If we are not interested in the dynamical properties of a system, but only in the structural and thermodynamic ones, we may apply sampling enhancement techniques to speed up the sampling of one of the two sets of degrees of freedom.

We consider a system in which we distinguish between two sets of particles or degrees of freedom, N^h particles of type h and N^l particles of type l, for which we wish to obtain the canonical ensemble at temperature T^l or $\beta^l = (k_B T^l)^{-1}$ by simulation.

Extension of the formalism to constant temperature and pressure is straightforward and will not be considered here. We assume that the motion of the two sets of particles is adiabatically decoupled by increasing the mass of all particles of type h by a factor of $s_m \gg 1$, assuming that the particles of type h and l have originally similar masses. For simplicity of notation we use

$$m^h = s_m m^l \tag{4.2}$$

with $s_m \geq 1$.

In practical simulations we are often interested in a speed-up of the sampling of a particular subset of the degrees of freedom of a system, e.g. the distribution of counter ions around a polymer, or a protein in aqueous solution, or the (un)folding conformational equilibrium distribution of a polypeptide in solution, or the conformational distribution of a part of a protein such as a loop of a protein in solution. In these cases we may identify the mentioned degrees of freedom with the adiabatically decoupled N^h degrees of freedom of the particles of type h and the other, i.e. solvent or remaining part of the solute degrees of freedom with the N^l l type degrees of freedom.

The Hamiltonian of the system S with N particles that we are interested in, can be written as

$$H(p^h, p^l, x^h, x^l) = K^h(p^h) + V^{hh}(x^h) + V^{hl}(x^h, x^l) + K^l(p^l) + V^{ll}(x^l), \tag{4.3}$$

where p and x denote the momentum and position vectors $\mathbf{p}^N$ and $\mathbf{r}^N$ respectively, and the superscripts h and l are used to distinguish the h- and l type particles. $K(p)$ is the kinetic energy, split into one for the h type particles and one for the l type particles, and $V(x)$ the potential energy of the system, which is separated into three terms, one describing the interaction between the h type particles, one describing the interaction between h type and l type particles, and one describing the interaction between the l type particles. The two sets of degrees of freedom differ in their dynamics: the changes in the positions of the l type particles are much faster than those of the h type ones.

The sampling of the h type degrees of freedom can now be enhanced by application of various sampling enhancement techniques. Here we consider three of them:

(a) Raising the temperature T^h of the h type degrees of freedom, which is part of the original adiabatic decoupling technique [30, 35, 158].

(b) Down-scaling of the interaction function of the h type particles, either only V^{hh}, or V^{hl}, or both V^{hh} and V^{hl}.

(c) Down-scaling of the forces $\mathbf{f}_{hh}$ and $\mathbf{f}_{hl}$ exerted on the h type particles by the h type and l type particles, respectively.

4.3.2 High T^h to enhance sampling

In classical adiabatic decoupling the temperature T^h of the h type particles is scaled up, e.g.

$$T^h = s_T T^l \quad or \quad \beta^h = \beta^l / s_T \tag{4.4}$$

with $s_T > 1$ in order to enhance sampling of the h type degrees of freedom. The system S we are interested in is characterized by

N^h	N^l
β^l	β^l

(4.5)

while the system S^T that is simulated is characterised by

N^h	N^l
$\beta^h = \beta^l / s_T$	β^l

. (4.6)

The task is now to express the probability distribution of system S in terms of the probability distribution of system S^T, or more simply, to express ensemble averages Eq.(4.1) of system S, $\langle Q \rangle$, in terms of ensemble averages of system S^T, $\langle Q \rangle_{\beta^h}$.

In the adiabatically decoupled limit, the partition function of the l type particles for a particular configuration x^h and momentum p^h of the h type particles is, apart from a normalization constant,

$$Z^l(x^h;\beta^l) = \int e^{-\beta^l K^l(p^l)} dp^l \int e^{-\beta^l V^{ll}(x^l)} e^{-\beta^l V^{hl}(x^h,x^l)} dx^l. \tag{4.7}$$

In the adiabatically decoupled limit, the Hamiltonian for the h type degrees of freedom becomes

$$H^h(p^h,x^h;\beta^l) = K^h(p^h) + V^{hh}(x^h) - (\beta^l)^{-1} \ln Z^l(x^h;\beta^l). \tag{4.8}$$

The partition function or the h type degrees of freedom of system S^T, which are at temperature T^h, is then

$$\begin{aligned} Z^h(\beta^h;\beta^l) &= \int\int e^{-\beta^h H^h(p^h,x^h;\beta^l)} dp^h dx^h \\ &= \int e^{-\beta^h K^h(p^h)} dp^h \int e^{-\beta^h V^{hh}(x^h)} [Z^l(x^h;\beta^l)]^{\beta^h/\beta^l} dx^h. \end{aligned} \tag{4.9}$$

The configurational probability of x^h at T^h is then

$$P^h(x^h;\beta^h;\beta^l) = \frac{\int e^{-\beta^h K^h(p^h)} dp^h e^{-\beta^h V^{hh}(x^h)} [Z^l(x^h;\beta^l)]^{\beta^h/\beta^l}}{Z^h(\beta^h;\beta^l)}, \tag{4.10}$$

or in other words

$$P^h(x^h;\beta^h;\beta^l) \propto e^{-\beta^h V^{hh}(x^h)} [Z^l(x^h;\beta^l)]^{\beta^h/\beta^l}. \tag{4.11}$$

Likewise we find for the partition function of the h type degrees of freedom of system S, which are at temperature T^l,

$$Z^h(\beta^l;\beta^l) = \int e^{-\beta^l K^h(p^h)} dp^h \int e^{-\beta^l V^{hh}(x^h)} [Z^l(x^h;\beta^l)] dx^h \tag{4.12}$$

and for the configurational probability of x^h at T^l

$$P^h(x^h;\beta^l;\beta^l) \propto e^{-\beta^l V^{hh}(x^h)} [Z^l(x^h;\beta^l)]. \tag{4.13}$$

The configurational distribution of the h type degrees of freedom at T^l, i.e. of system S, can now be expressed in terms of the configurational distribution of the h type degrees of freedom at T^h, i.e. of system S^T. Using Eq.(4.11) we have

$$Z^l(x^h;\beta^l) \propto e^{+\beta^l V^{hh}(x^h)} [P^h(x^h;\beta^h;\beta^l)]^{\beta^l/\beta^h} \tag{4.14}$$

which can be inserted into Eq.(4.13) to obtain

$$P^h(x^h;\beta^l;\beta^l) \propto [P^h(x^h;\beta^h;\beta^l)]^{\beta^l/\beta^h} = [P^h(x^h;\beta^h;\beta^l)]^{s_T}. \tag{4.15}$$

Thus the two distributions $P^h(x^h;\beta^h;\beta^l)$ and $P^h(x^h;\beta^l;\beta^l)$ are not identical, as was suggested in [30, 35, 158], if $s_T \neq 1$.

An ensemble average $\langle Q\rangle_{\beta^l}$ of a quantity $Q(x^h)$ at T^l can be expressed in terms of an ensemble average $\langle Q\rangle_{\beta^h}$ of $Q(x^h)$ at T^h using the standard unbiasing formula,

$$\langle Q\rangle_{\beta^l} = \frac{\int\int Q(x^h) e^{-\beta^l H^h(p^h,x^h;\beta^l)} dp^h dx^h}{\int\int e^{-\beta^l H^h(p^h,x^h;\beta^l)} dp^h dx^h} = \frac{\langle Q e^{-(\beta^l-\beta^h)H^h}\rangle_{\beta^h}}{\langle e^{-(\beta^l-\beta^h)H^h}\rangle_{\beta^h}}. \tag{4.16}$$

Since $Q(x^h)$ does not depend on p^h and the kinetic part $K^h(p^h)$ in the Hamiltonian $H^h(p^h,x^h;\beta^l)$ of Eq.(4.8) only depends on p^h, the kinetic part will not contribute to Eq.(4.16). Using Eq.(4.8) we get

$$\langle Q\rangle_{\beta^l} = \frac{\langle Q e^{-(\beta^l-\beta^h)[V^{hh}-(\beta^l)^{-1}\ln Z^l]}\rangle_{\beta^h}}{\langle e^{-(\beta^l-\beta^h)[V^{hh}-(\beta^l)^{-1}\ln Z^l]}\rangle_{\beta^h}}. \tag{4.17}$$

This expression implies that each configuration x^h sampled at T^h must be reweighted by the factor

$$e^{-(\beta^l-\beta^h)[V^{hh}(x^h)-(\beta^l)^{-1}\ln Z^l(x^h;\beta^l)]}. \tag{4.18}$$

The partition function $Z^l(x^h;\beta^l)$ defined in Eq.(4.7) can be written as an ensemble average over the l type degrees of freedom,

$$Z^l(x^h;\beta^l) = \frac{\int e^{-\beta^l K^l(p^l)}dp^l \int e^{-\beta^l V^{ll}(x^l)}dx^l}{\langle e^{+\beta^l V^{hl}(x^h)}\rangle}. \tag{4.19}$$

Since the two factors in the numerator do not depend on x^h, the reweighting factor Eq.(4.18) can be simplified, using Eq.(4.19), to

$$e^{-(\beta^l-\beta^h)[V^{hh}(x^h)+(\beta^l)^{-1}\ln\langle e^{+\beta^l V^{hl}(x^h)}\rangle]}, \tag{4.20}$$

where the ensemble average is over the l type degrees of freedom at T^l for a given x^h. In the adiabatic limit, x^h is changing slowly compared to x^l, which suggests the approximation [1, 178]

$$\langle e^{+\beta^l V^{hl}(x^h)}\rangle_t = [\tau_{MF}(1-exp(-t/\tau_{MF})]^{-1}\int_0^t e^{-(t-t')/\tau_{MF}} e^{+\beta^l V^{hl}(x^h(t'),x^l(t'))}dt', \tag{4.21}$$

which time-average can be obtained as running average from the trajectory of system S^T, while τ_{MF} determines the averaging period to obtain the mean force due to the l type particles. Expression (4.21) in discretised form becomes for $t_n \gg \tau_{MF}$

$$\langle e^{+\beta^l V^{hl}(x^h)}\rangle_{t_n} = [1-e^{-\Delta t/\tau_{MF}}]e^{+\beta^l V^{hl}(x^h(t_n))} + e^{-\Delta t/\tau_{MF}}\langle e^{+\beta^l V^{hl}(x^h)}\rangle_{t_n-\Delta t}. \tag{4.22}$$

For particular types of degrees of freedom, such as the translational degrees of freedom of ions in water or of water molecules themselves, the configurational distribution may be rather insensitive to the temperature, which would allow a neglect of reweighting using Eq.(4.20).

4.3.3 Downscaling of $V(x^h)$ to enhance sampling

When scaling down the interaction function of the h type particles, three cases can be distinguished

1. Scaling both V^{hh} and V^{hl} in Eq.(4.3), which would apply if both potential energy terms impede a rapid sampling of the h type degrees of freedom.
2. Scaling of only V^{hl}, which would apply when only the interaction between h type and l type particles impedes a rapid sampling of the h type degrees of freedom.

3. Scaling of only V^{hh}, which would apply when only the interaction between the h type particles impedes the sampling.

Regarding the interaction of the l type particles, three cases can be distinguished too.

1. The potential energy of the l type particles is dominated both by V^{ll} and V^{hl},

$$V^{ll} \approx V^{hl}. \tag{4.23}$$

2. The potential energy of the l type of particles is dominated by interactions between themselves, which is much stronger than that with the h type particles,

$$V^{ll} \ll V^{hl}. \tag{4.24}$$

3. Their potential energy is dominated by interactions with the h type particles

$$V^{ll} \gg V^{hl}. \tag{4.25}$$

We first consider scaling down both V^{hh} and V^{hl} using the scaling factor s_V^{hh} and s_V^{hl}, respectively. The system that is simulated, S^V, looks like

N^h $\quad s_V^{hh} < 1$ and/or $\beta^l \quad s_V^{hl} < 1$	N^l $\beta^l \quad s_V^{hl} < 1$

(4.26)

while the system S we are interested in is

N^h $\quad s_V^{hh} = 1$ and $\beta^l \quad s_V^{hl} = 1$	N^l $\beta^l \quad s_V^{hl} = 1$

(4.27)

Their Hamiltonians are different. For system S it is given by Eq.(4.3). For system S^V it becomes

$$H^V(p^h, p^l, x^h, x^l) = K^h(p^h) + s_V^{hh} V^{hh}(x^h) + s_V^{hl} V^{hl}(x^h, x^l) + K^l(p^l) + V^{ll}(x^l). \tag{4.28}$$

In the adiabatically decoupling limit, we get as in the previous section on temperature scaling

$$Z^l(x^h;\beta^l,\beta^{hl}) = \int e^{-\beta^l K^l(p^l)}dp^l \int e^{-\beta^l V^{ll}(x^l)} e^{-\beta^l s_V^{hl} V^{hl}(x^h,x^l)}dx^l \tag{4.29}$$

with $\beta^{hl} = \beta^l s_V^{hl}$. In the adiabatically decoupled limit, the Hamiltonian for the h type degrees of freedom of system S^V becomes

$$H^h(p^h,x^h;\beta^l,\beta^{hl}) = K^h(p^h) + s_V^{hh}V^{hh}(x^h) - (\beta^l)^{-1}\ln Z^l(x^h;\beta^l,\beta^{hl}). \tag{4.30}$$

The partition function of the h type degrees of freedom of system S^V at T^l is then

$$Z^h(\beta^l;\beta^{hl},\beta^{hh}) = \int e^{-\beta^l K^h(p^h)}dp^h \int e^{-\beta^{hh}V^{hh}(x^h)}[Z^l(x^h;\beta^l,\beta^{hl})]dx^h, \tag{4.31}$$

with $\beta^{hh} = \beta^l s_V^{hh}$. The configurational probability of x^h of system S^V at T^l is then

$$P^h(x^h;\beta^l;\beta^{hl},\beta^{hh}) \propto e^{-\beta^{hh}V^{hh}(x^h)}[Z^l(x^h;\beta^l,\beta^{hl})], \tag{4.32}$$

while for system S we find

$$P^h(x^h;\beta^l;\beta^l,\beta^l) \propto e^{-\beta^l V^{hh}(x^h)}[Z^l(x^h;\beta^l,\beta^l)]. \tag{4.33}$$

So we cannot easily express P^h of system S in terms of P^h of system S^V.

Yet, as in the T-scaling case, we may express an ensemble average $\langle Q\rangle_{\beta^l}$ of a quantity $Q(x^h)$ at T^l for system S in terms of an ensemble average $\langle Q\rangle_{\beta^l,\beta^{hl},\beta^{hh}}$ of $Q(x^h)$ at T^l for system S^V,

$$\langle Q\rangle_{\beta^l} = \frac{\langle Q e^{-[(\beta^l-\beta^{hh})V^{hh}+(\beta^l-\beta^{hl})(\beta^l)^{-1}\ln\langle e^{+\beta^{hl}V^{hl}}\rangle]}\rangle_{\beta^l,\beta^{hl},\beta^{hh}}}{\langle e^{-[(\beta^l-\beta^{hh})V^{hh}+(\beta^l-\beta^{hl})(\beta^l)^{-1}\ln\langle e^{+\beta^{hl}V^{hl}}\rangle]}]\rangle_{\beta^h,\beta^{hl},\beta^{hh}}}, \tag{4.34}$$

where the ensemble average in the exponential function is over the l type degrees of freedom of system S^V and the other two ensemble averages over the h type of degrees of freedom of system S^V. The reweighting factor thus becomes,

$$e^{-[(\beta^l-\beta^{hh})V^{hh}(x^h)+(\beta^l-\beta^{hl})(\beta^l)^{-1}\ln\langle e^{+\beta^{hl}V^{hl}(x^h)}\rangle]}, \tag{4.35}$$

which can be evaluated using the approximation

$$\langle e^{+\beta^{hl}V^{hl}(x^h)}\rangle_t = [\tau_{MF}(1-exp(-t/\tau_{MF})]^{-1}\int_0^t e^{-(t-t')/\tau_{MF}}e^{+\beta^{hl}V^{hl}(x^h(t'),x^l(t'))}dt'. \tag{4.36}$$

The differences between the probability distributions P^h for systems S and S^V will depend on whether the V^{ll} term or the V^{hl} term dominates the potential energy of the l type particles. If V^{ll} dominates the l type degrees of freedom, e.g. if $N^l \gg N^h$, we have

$$Z^l(x^h;\beta^l,\beta^{hl}) \approx Z^l(x^h;\beta^l,\beta^l) \tag{4.37}$$

and thus

$$P^h(x^h;\beta^l,\beta^l,\beta^l) \propto e^{-(\beta^l-\beta^{hh})V^{hh}(x^h)}P^h(x^h;\beta^l,\beta^{hl},\beta^{hh}). \tag{4.38}$$

The ensemble average of $Q(x^h)$ becomes

$$\langle Q\rangle_{\beta^l} = \langle Qe^{-(\beta^l-\beta^{hh})V^{hh}}\rangle_{\beta^l,\beta^{hl},\beta^{hh}}, \tag{4.39}$$

where the ensemble average is over configurations obtained using the scaling factors s_V^{hl} and s_V^{hh}. If the V^{hl} term dominates the l type of degrees of freedom, e.g. if $N^l \ll N^h$, we have

$$Z^l(x^h;\beta^l,\beta^{hl}) \approx Z^l(x^h;\beta^{hl},\beta^{hl}), \tag{4.40}$$

which does not lead to an expression of P^h for system S in terms of the P^h for system S^V.

If the h type degrees of freedom are dominated by V^{hl} compared to V^{hh} and the l type degrees of freedom are dominated by V^{ll} compared to V^{hl}, we may simulate system S^V at T^l enhancing the sampling using $s_V^{hl} < 1$ and $s_V^{hh} < 1$ and compute ensemble averages from Eq.(4.39). One would choose $s_V^{hl} = s_V^{hh}$ in order not to destroy the balance between the forces on the h type of particles. If one would use $s_V^{hh} = 1$, the generated ensemble need not be reweighted using Eq.(4.39) because $\beta^{hh} = \beta^l$ in that case. This would only apply if the interaction $V^{hh}(x^h)$ is a smooth function.

If the term V^{hh} dominates the h type degrees of freedom compared to V^{hl}, as may be the case for a protein (h) in water (l), one could do with only scaling down V^{hh} leaving V^{hl} as is. In that case we have $s_V^{hl} = 1$ and

$$P^h(x^h;\beta^l,\beta^l,\beta^{hh}) \propto e^{-\beta^{hh}V^{hh}(x^h)}Z^l(x^h;\beta^l,\beta^l) \tag{4.41}$$

and

$$P^h(x^h;\beta^l,\beta^l,\beta^l) \propto e^{-(\beta^l-\beta^{hh})V^{hh}(x^h)}P^h(x^h;\beta^l,\beta^l,\beta^{hh}). \tag{4.42}$$

The ensemble average of $Q(x^h)$ becomes

$$\langle Q\rangle_{\beta^l} = \langle Qe^{-(\beta^l-\beta^{hh})V^{hh}}\rangle_{\beta^l,\beta^l,\beta^{hh}} \tag{4.43}$$

as before.

Down-scaling of $V(x^h)$ will only yield usable results if an imbalance between V^{hh} and V^{hl} can be avoided. Since this is in practice not easily done, we propose in the following subsection a technique that enhances the sampling of the h type degrees of freedom by differentially scaling the forces on the l type and h type of particles without altering the balance between the forces on the h type particles.

4.3.4 Force-scaling to enhance sampling

The previous treatment of down-scaling the potential energy to enhance the sampling of the h type degrees of freedom suffers from the disadvantage that only in particular limiting cases the configurational distribution of the unscaled h type degrees of freedom at T^l, system S, is obtained, and that in other cases it is only approximated. This is due to the formulation of the problem in terms of a joint system of h type and l type particles, Eqs.(4.26) and (4.27). Alternatively we may consider the h type and l type degrees of freedom as separate systems S^h and S^l which are subject to external forces originating in the other system: S^h feels forces originating from S^l and vice versa, but these are not symmetric. By relaxing the condition of deriving these forces from the same potential energy term $V^{hl}(x^h, x^l)$ for the h type and l type particles, i.e. by allowing

$$\mathbf{f}_{hl} \neq -\mathbf{f}_{lh}, \tag{4.44}$$

where $\mathbf{f}_{hl}$ is the force on particle h exerted by particle l, a violation of Newton's third law

$$\mathbf{f}_{hl} = -\mathbf{f}_{lh}, \tag{4.45}$$

we may keep the correct balance between $\mathbf{f}_{hh}$ and $\mathbf{f}_{hl}$ for the h type particles and also between $\mathbf{f}_{ll}$ and $\mathbf{f}_{lh}$ for the l type particles. The h type particle system S^h is then Hamiltonian, and kinetic or potential energy scaling techniques to enhance sampling can be used. The l type particle system S^l is also Hamiltonian. Only the combined system $S^h + S^l$ is not Hamiltonian, which is no problem because we are only interested in the h type particle system S^h.

The derivation of the expressions for the ensemble averages follows the one of the previous section. The Hamiltonian for the h type particles is in analogy to Eqs.(4.29) and (4.30),

$$H^h(p^h, x^h) = K^h(p^h) + s_V^{hh} V^{hh}(x^h) - s_V^{hl}(\beta^l)^{-1} \ln Z^l(x^h; \beta^l) \tag{4.46}$$

with

$$Z^l(x^h; \beta^l) = \int e^{-\beta^l K^l(p^l)} dp^l \int e^{-\beta^l V^{ll}(x^l)} e^{-\beta^l V^{hl}(x^h, x^l)} dx^l. \tag{4.47}$$

If

$$s_V^{hh} = s_V^{hl}, \tag{4.48}$$

the forces on the h type particles exerted by the h type and l type particles are kept balanced, and can be scaled down to enhance sampling. Eq.(4.47) implies that the l type particles are subject to the full forces exerted by the h type and l type particles. The Hamiltonian for the l-type particles is thus

$$H^l(p^l, x^l) = K^l(p^l) + V^{ll}(x^l) + V^{hl}(x^h, x^l). \tag{4.49}$$

The ensemble average $\langle Q \rangle_{\beta^l}$ of a quantity $Q(x^h)$ at T^l for system S^h can be expressed in terms of an ensemble average at T^l for system S^h with the potential energy scaled Hamiltonian Eq.(4.46).

$$\langle Q \rangle_{\beta^l} = \frac{\langle Q e^{-[(\beta^l - \beta^{hh})V^{hh} + (\beta^l - \beta^{hl})(\beta^l)^{-1} \ln \langle e^{+\beta^l V^{hl}} \rangle]} \rangle_{\beta^l, \beta^{hl}, \beta^{hh}}}{\langle e^{-[(\beta^l - \beta^{hh})V^{hh} + (\beta^l - \beta^{hl})(\beta^l)^{-1} \ln \langle e^{+\beta^l V^{hl}} \rangle]} \rangle_{\beta^l, \beta^{hl}, \beta^{hh}}}, \tag{4.50}$$

as in Eq.(4.34), but with the second occurrence of β^{hl}, i.e. the one in the ensemble average over the l type degrees of freedom, replaced by β^l, both in the numerator and denominator. The reweighting factor thus becomes

$$e^{-[(\beta^l - \beta^{hh})V^{hh}(x^h) + (\beta^l - \beta^{hl})(\beta^l)^{-1} \ln \langle e^{+\beta^l V^{hl}(x^h)} \rangle]} \tag{4.51}$$

which can be evaluated using the approximation

$$\langle e^{+\beta^l V^{hl}(x^h)} \rangle_t = [\tau_{MF}(1 - exp(-t/\tau_{MF})]^{-1} \int_0^t e^{-(t-t')/\tau_{MF}} e^{+\beta^l V^{hl}(x^h(t'), x^l(t'))} dt'. \tag{4.52}$$

4.3.5 Replica-exchange formulation

In the three previous subsections expressions for the adiabatically decoupled Hamiltonians for the h type degrees of freedom at scaled temperatures, potential energies and forces were given. Trajectories generated at elevated temperature or down scaled potential energies or forces are then to be reweighted in order to obtain the ensemble averages of interest, at T^l and unscaled interactions or forces, see Eqs.(4.20), (4.35) and (4.51). Alternatively, one could use the replica-exchange technique [33, 179–184] with replicas at different temperatures or differently scaled interactions or forces to generate a generalised canonical ensemble, in which the trajectories at the temperature T^l or with the unscaled interaction or force, $s_V^{hh} = s_V^{hl} = 1$, can be used to obtain ensemble averages without reweighting, but at the cost of generating the additional replica trajectories.

In replica-exchange simulations, M independent MD simulations, replicas, are performed each having a different temperature, $\{T_1^h, T_2^h, \ldots, T_M^h\}$ or $\{\beta_1^h, \beta_2^h, \ldots, \beta_M^h\}$ in the case of temperature replica exchange, or having different h type particle interaction or force scaling factors, $\{s_{V_1}^{hh}, s_{V_1}^{hl}, s_{V_2}^{hh}, s_{V_2}^{hl}, \ldots, s_{V_M}^{hh}, s_{V_M}^{hl}\}$ or $\{\beta_1^{hh}, \beta_1^{hl}, \beta_2^{hh}, \beta_2^{hl}, \ldots, \beta_M^{hh}, \beta_M^{hl}\}$. After a chosen number of MD steps, exchanges of the configurations x_m^h and x_n^h of replicas with adjacent temperatures or interaction or force scale factors, i.e. β_m^h and β_n^h or $\{\beta_m^{hh}, \beta_m^{hl}\}$ and $\{\beta_n^{hh}, \beta_n^{hl}\}$, are attempted using the detailed balance condition

$$P_{S'} \cdot t(S' \to S'') = P_{S''} \cdot t(S'' \to S'), \tag{4.53}$$

where S' denotes the state of the system before the exchange, i.e. the replica m with β_m^h or $\{\beta_m^{hh}, \beta_m^{hl}\}$ has configuration x_m^h and the replica n with β_n^h or $\{\beta_n^{hh}, \beta_n^{hl}\}$ has configuration x_n^h, and S'' denotes the state after the exchange, i.e. the replica m has configuration x_n^h and the replica n has configuration x_m^h. P_S denotes the configurational probability for state S and $t(S \to S')$ the transition probability from state S to state S'. The relative configurational probability of states S'' and S' follows directly from the reweighting factors derived in the previous sections.

For temperature replica-exchange we find using Eq.(4.20)

$$\frac{P_{S''}}{P_{S'}} = \frac{e^{-\beta_m^h[(V^{hh}(x_n^h)-V^{hh}(x_m^h))+(\beta^l)^{-1}ln(\frac{\langle e^{+\beta^l V^{hl}(x_n^h)}\rangle}{\langle e^{+\beta^l V^{hl}(x_m^h)}\rangle})]}}{e^{-\beta_n^h[(V^{hh}(x_n^h)-V^{hh}(x_m^h))+(\beta^l)^{-1}ln(\frac{\langle e^{+\beta^l V^{hl}(x_n^h)}\rangle}{\langle e^{+\beta^l V^{hl}(x_m^h)}\rangle})]}} \tag{4.54}$$

Combining Eqs.(4.53)and (4.54) we find for the probability p of exchange of replicas m and n,

$$p(m \leftrightarrow n) = t(S' \to S'') = \begin{cases} 1 & \text{if } \Delta_{mn} \leq 0 \\ P_{S''}/P_{S'} = e^{-\Delta_{mn}} & \text{if } \Delta_{mn} \geq 0 \end{cases} \tag{4.55}$$

$$= min(1, e^{-\Delta_{mn}}) \tag{4.56}$$

with

$$\begin{aligned} \Delta_{mn} = &\Big\{\beta_m^h[V^{hh}(x_n^h) + (\beta^l)^{-1}\ln\langle e^{+\beta^l V^{hl}(x_n^h)}\rangle] \\ &+ \beta_n^h[V^{hh}(x_m^h) + (\beta^l)^{-1}\ln\langle e^{+\beta^l V^{hl}(x_m^h)}\rangle]\Big\} \\ &- \Big\{\beta_m^h[V^{hh}(x_m^h) + (\beta^l)^{-1}\ln\langle e^{+\beta^l V^{hl}(x_m^h)}\rangle] \\ &+ \beta_n^h[V^{hh}(x_n^h) + (\beta^l)^{-1}\ln\langle e^{+\beta^l V^{hl}(x_n^h)}\rangle]\Big\} \end{aligned} \tag{4.57}$$

where the ensemble averages over the l type degrees of freedom at T^l and for x_m^h and x_n^h can be obtained using Eq.(4.21) for replicas m and n.

For interaction replica-exchange Eq.(4.35) yields

$$\begin{aligned} \Delta_{mn} = &\Big\{\beta_m^{hh}V^{hh}(x_n^h) + \beta_m^{hl}(\beta^l)^{-1}\ln\langle e^{+\beta_m^{hl}V^{hl}(x_n^h)}\rangle \\ &+ \beta_n^{hh}V^{hh}(x_m^h) + \beta_n^{hl}(\beta^l)^{-1}\ln\langle e^{+\beta_n^{hl}V^{hl}(x_m^h)}\rangle\Big\} \\ &- \Big\{\beta_m^{hh}V^{hh}(x_m^h) + \beta_m^{hl}(\beta^l)^{-1}\ln\langle e^{+\beta_m^{hl}V^{hl}(x_m^h)}\rangle \\ &+ \beta_n^{hh}V^{hh}(x_n^h) + \beta_n^{hl}(\beta^l)^{-1}\ln\langle e^{+\beta_n^{hl}V^{hl}(x_n^h)}\rangle\Big\} \end{aligned} \tag{4.58}$$

where the ensemble averages over the l type degrees of freedom at T^l and scaling factor β^{hl}, and for x_m^h and x_n^h can be obtained using Eq.(4.36) for replicas m and n.

For force-scaling replica-exchange Eq.(4.51) yields

$$\begin{aligned}\Delta_{mn} = &\left\{\beta_m^{hh}V^{hh}(x_n^h) + \beta_m^{hl}(\beta^l)^{-1}\ln\langle e^{+\beta^l V^{hl}(x_n^h)}\rangle\right.\\ &\left.+\beta_n^{hh}V^{hh}(x_m^h) + \beta_n^{hl}(\beta^l)^{-1}\ln\langle e^{+\beta^l V^{hl}(x_m^h)}\rangle\right\}\\ &-\left\{\beta_m^{hh}V^{hh}(x_m^h) + \beta_m^{hl}(\beta^l)^{-1}\ln\langle e^{+\beta^l V^{hl}(x_m^h)}\rangle\right.\\ &\left.+\beta_n^{hh}V^{hh}(x_n^h) + \beta_n^{hl}(\beta^l)^{-1}\ln\langle e^{+\beta^l V^{hl}(x_n^h)}\rangle\right\},\end{aligned} \tag{4.59}$$

where the ensemble averages over the l type degrees of freedom at T^l and for x_m^h and x_n^h can be obtained using Eq.(4.52) for replicas m and n.

4.3.6 Simulated system and simulation procedure

The adiabatic decoupling technique was tested using liquid water, because the configurational distribution can easily be probed by calculating radial distribution functions and average potential energies and because it is a simple system that yet has not only translational but also rotational degrees of freedom.

A cubic box with an edge length of $3.1057\,\mathrm{nm}$ was filled with 1000 SPC [9] water molecules, resulting in a density of $997\,\mathrm{kg/m^3}$, corresponding to the density of liquid water at $298\,\mathrm{K}$ and $1\,\mathrm{atm}$. [58] Periodic boundary conditions were applied. Molecular dynamics simulations were performed at constant volume and temperature (NVT condition) with the GROMOS05 package [133], modified to incorporate adiabatic decoupling. The geometry of the water molecules was constrained by applying the SHAKE algorithm [23] with a relative geometric tolerance of 10^{-4} on the OH bond lengths and on the intramolecular HH distance. Where not mentioned differently, the temperature was weakly coupled [29] to two baths of $300\,\mathrm{K}$ with a relaxation time of $0.1\,\mathrm{ps}$, one for the h type and one for the l type particles. The nonbonded van der Waals and electrostatic interactions were calculated using triple-range cutoff radii of $0.8/1.4\,\mathrm{nm}$. The short range interactions were calculated every time step by updating the molecular pair list for distances smaller than the first cutoff radius of $0.8\,\mathrm{nm}$. For the intermediate range of distances between 0.8 and $1.4\,\mathrm{nm}$ the pairlist was only updated every fifth timestep and at the same time the interaction was calculated and kept unchanged between these updates. The long range electrostatic interactions beyond the outer cutoff of $1.4\,\mathrm{nm}$ were represented by a reaction field [62,185] with ϵ_{RF}=78.5. For most simulations, the equations of motion were integrated using the leap frog algorithm with a time step of $2\,\mathrm{fs}$, the other cases are explicitly mentioned. The velocities of the atoms at the beginning of the simulation were assigned from a Maxwell distribution at $300\,\mathrm{K}$. During the runs, configurations of the system were saved every $1\,\mathrm{ps}$. The various properties were taken from a $1\,\mathrm{ns}$ simulation that followed a $100\,\mathrm{ps}$ equilibration period.

The simulations performed at constant pressure and temperature (NpT condition), coupled the pressure weakly [29] to a bath of $1\,\mathrm{atm}$ with a relaxation time of $0.5\,\mathrm{ps}$. In the case of temperature scaling, the calculation of the pressure had to be modified to

$$pV = \frac{2}{3}\left[K^{h,corr} + K^{l} - W\right] \tag{4.60}$$

where p is the pressure, K^l the kinetic energy of the l type degrees of freedom, W the virial, and

$$K^{h,corr} = \frac{K^h}{s_T} = \frac{1}{2s_T}(\mathbf{v}_i^h)^2 m_i^h \tag{4.61}$$

is the corrected kinetic energy of the h type degrees of freedom leading to a kinetic energy at the unscaled temperature. In the cases of interaction or force scaling, the calculation of the pressure is to be modified correspondingly by using the virial calculated from the unscaled interactions or forces.

When stochastic dynamics was applied to the oxygen of the h type particles, a friction coefficient γ of $91/s_m\,\mathrm{ps}^{-1}$ was used and no additional thermostat was used on the h type water molecules, while the Berendsen thermostat [29] was used on the l type waters.

Different ratios between h type and l type particles were simulated with N^h/N^l of 10/990, 100/900, 500/500, 900/100, and 990/10. The mass of the h type particles was increased by a factor s_m which was set to 1, 100, or 1000. $s_m = 10$ was left out since increasing the masses of the h type hydrogen atoms with this factor makes their mass similar to that of the l type non-hydrogen atoms, i.e. their degrees of freedom would not be adiabatically decoupled. Either the translational and rotational temperature or only the translational temperature is scaled with s_T, or the force acting on the h particles is scaled with s_V. Two temperature baths for the h type particles, one for the translational and one for the internal and rotational degrees of freedom, were used except in the first case. s_T was set to 1, 2 or 5 and $s_V = s_V^{hh} = s_V^{hl}$ to 1, 0.5 or 0.2.

4.3.7 Analysis

The configurational distributions of liquid water that were generated using the different sampling enhancement methods were characterised by calculating dipolar radial distribution functions. The velocity distribution of liquid water was calculated to investigate its deviation from the Maxwell distribution. The heat flow from and to the thermostats was calculated to monitor heat gain and loss of the various subsets of molecules. The diffusion constant of water was calculated to probe the extent of enhanced sampling.

Dipole-dipole radial distribution function

The dipole-dipole radial distribution function is obtained by a simple histogram summation in radial shells over molecules in the system

$$g^{dip}(r) = \frac{\left\langle \sum_{\alpha}^{N} \sum_{\beta \neq \alpha, r < r_{\alpha\beta} \leq r+dr}^{N} \mathbf{e}_{\alpha}(t) \cdot \mathbf{e}_{\beta}(t) \right\rangle_t}{\left\langle \sum_{\alpha}^{N} \sum_{\beta \neq \alpha, r < r_{\alpha\beta} \leq r+dr}^{N} 1 \right\rangle_t} \tag{4.62}$$

where $\mathbf{e}_{\alpha}$ is the normalised dipole vector of molecule α, N is the number of molecules, $r_{\alpha\beta}$ the distance between the first atoms in molecules α and β, and $\langle \ldots \rangle_t$ denotes averaging over the simulation time.

Heat flow from and to the baths

The temperature scaling factor λ_{sc} for the velocity in the weak coupling [29] scheme was calculated as

$$\lambda_{sc}^2 = 1 + \frac{2c_V^{df}}{k_B} \frac{\Delta t}{\tau_T} \left[\frac{T_{ref}}{T(t - \Delta t/2)} - 1 \right] \tag{4.63}$$

where c_V^{df} is the heat capacity at constant volume per degree of freedom, k_B is the Boltzmann constant, Δt is the time step, τ_T is the coupling time constant, T_{ref} is the reference temperature of the bath and $T(t - \Delta t/2)$ is the temperature at time step $t - \Delta t/2$. This quantity is closely related to the heat flow

$$\Delta Q = \sum_{i=1}^{N} \frac{1}{2} m_i \mathbf{v}_i^2 [\lambda_{sc}^2 - 1] \tag{4.64}$$

into ($\lambda_{sc} > 1$) or out of ($\lambda_{sc} < 1$) the system.

Self-diffusion coefficient D

The diffusion coefficient is obtained from the long-time limit of the mean square displacement

$$MSD = < (\mathbf{r}(\tau + t) - \mathbf{r}(\tau))^2 >_{\tau, molecules} \tag{4.65}$$

according to the Einstein relation, [66]

$$D = \lim_{t \to \infty} \frac{< (\mathbf{r}(\tau + t) - \mathbf{r}(\tau))^2 >_{\tau, molecules}}{6t} \tag{4.66}$$

where $\mathbf{r}(t)$ corresponds to the position vector of the centre of mass of a molecule at time t, and the averaging is performed over both time and water molecules. Only the time window between 0 and 250 ps was used for the calculation of the diffusion constant.

system (N^h/N^l)	s_m	s_T	$\langle E_h^{pot}\rangle$ [kJ/mol]	$\langle E_l^{pot}\rangle$ [kJ/mol]	$\langle T_h^t\rangle$ [K]	$\langle T_h^r\rangle$ [K]	$\langle T_l^t\rangle$ [K]	$\langle T_l^r\rangle$ [K]	D_h [10^{-9}m^2/s]	D_l [10^{-9}m^2/s]
500/500	1	1	-41.5	-41.5	297	308	296	308	4.1	4.3
		2	-34.4	-36.6	511	487	393	417	18	12
		5	-24.0	-27.7	1168	1012	679	737	54	22
	100	1	-41.7	-41.6	302	298	296	306	0.85	1.2
		2	-36.7	-40.0	759	421	323	297	36	30
		5	-34.3	-39.2	2449	507	352	290	380	230
	1000	1	-41.7	-41.6	300	299	294	306	0.34	0.55
		2	-36.2	-40.2	737	456	303	303	6.9	6.0
		5	-32.8	-39.3	2368	616	315	300	50	42

Table 4.1: s_m: scaling factor for the mass of the h type water, s_T: scaling factor for the temperature (translational and rotational) of the h type water, $\langle E_h^{pot}\rangle$: average potential energy per h type water, $\langle E_l^{pot}\rangle$: average potential energy per l type water, $\langle T_h^t\rangle$: average translational temperature of the h type water, $\langle T_h^r\rangle$: average rotational temperature of the h type water, $\langle T_l^t\rangle$: average translational temperature of the l type water, $\langle T_l^r\rangle$: average rotational temperature of the l type water, D_h: diffusion coefficient of the h type water, D_l: diffusion coefficient of the l type water.

4.4 Results

Three sampling enhancement methods that may enhance the sampling of configurational space when used in conjunction with adiabatic decoupling were investigated using the water system: (i) temperature scaling, (ii) translational temperature scaling, and (iii) force scaling. We did not test the method of interaction scaling, because it perturbs the balance between the forces on the h type particles exerted by the h type versus l type particles. For each of the three sampling enhancement methods five different systems, i.e. ratios between N^h and N^l, were investigated: $N^h/N^l = 10/990$, $100/900$, $500/500$, $900/100$, and $990/10$. For each of these systems three mass scaling factors $s_m = 1$, 100, and 1000 were used in conjunction with three temperature scaling factors $s_T = 1$, 2, and 5 or three force scaling factors $s_V^{-1} = 1$, 2, and 5. For one of the methods, temperature scaling, and one system, $N^h/N^l = 500/500$, the variation of the MD time step, of the coupling strength to the heat baths, of the use of a stochastic dynamics thermostat with different friction coefficients, and of constant pressure instead of constant volume boundary conditions, was investigated. For all systems and parameter values the dipolar radial distribution function, the velocity distribution function, the average potential energy, translational and rotational temperature, and the mean-square displacement and diffusion constant for the h type and l type water molecules were calculated. Below we only show and discuss a fraction of this data.

4.4.1 Scaling the translational and rotational temperature with s_T for the system 500/500

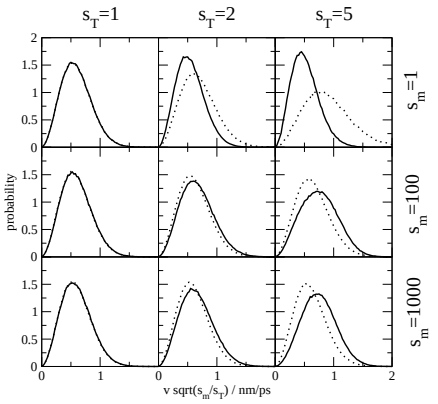

Figure 4.1: Velocity distribution of the water oxygen for 500 *h* type water molecules (solid line) and 500 *l* type water molecules (dotted line) for different mass and temperature scaling factors s_m and s_T, scaling both translational and rotational temperature with one bath but separately for 500 *h* type water molecules and 500 *l* type water molecules. The velocities for the *h* type water molecules were multiplied with $(s_m/s_T)^{1/2}$.

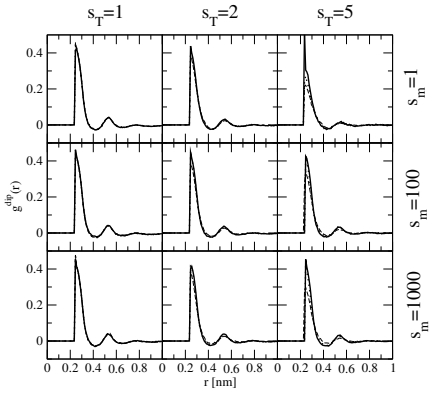

Figure 4.2: Dipole-dipole radial distribution function between 500 *h* type water molecules (solid), 500 *l* type water molecules (dashed), and between 500 *h* and 500 *l* water molecules (dotted) for different mass and temperature scaling factors s_m and s_T, scaling both translational and rotational temperature with one bath but separately for 500 *h* type water molecules and 500 *l* type water molecules.

As a first system an equal mixture of $N^h = 500$ and $N^l = 500$ of *h* and *l* type particles was chosen. The masses of the *h* type particles were varied with s_m and the temperature of their bath T_0^h with s_T. This system is far from the systems of interest

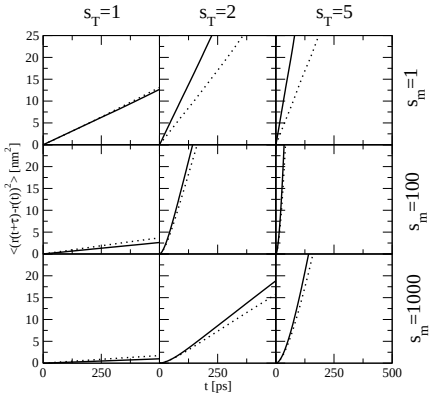

Figure 4.3: Mean square displacment of 500 h type water molecules (solid line) and 500 l type water molecules (dotted line) for different mass and temperature scaling factors s_m and s_T, scaling both translational and rotational temperature with one bath but separately for 500 h type water molecules and 500 l type water molecules.

where $N^h \ll N^l$, but it is more sensitive to the approximations of the methodology and therefore a good test system.

Reweighting was not possible for this system since $e^{-(\beta^l-\beta^h)[V^{hh}(x^h)+(\beta^l)^{-1}\ln\langle e^{+\beta^l V^{hl}(x^h)}\rangle]}$ from Eq. (4.20) becomes infinite in numerical precision.

In Fig. 4.1 the velocity distributions of the oxygens in the h type and the l type waters are shown. The velocities of the h type oxygens were scaled by a factor $(s_m/s_T)^{1/2}$ in order to allow a better comparison of the distributions. They determine the behaviour of the temperature in Table 4.1. The velocities of the h type oxygens are generally shifted to the right for $s_T > 1$ except for $s_m = 1$. As the velocity of the oxygens is almost identical to the center of mass motion, this means that the temperature corresponding to the translational velocity is bigger than $s_T T_0^h$ for $s_T > 1$, $s_m > 1$ and smaller for $s_T > 1$, $s_m = 1$. The increase in temperature of the l type particles leads to a shift in their peak toward the right, which is smaller the better the motions of the heavy and light waters are decoupled.

Table 4.1 shows the results for the average potential energy per water molecule, the average temperatures and the diffusion coefficients upon changing the factors s_m and s_T. For $s_m = 1$ and $s_T = 1$, the standard properties of SPC water should be recovered. The translational and rotational temperature differ slightly from 300 K, because of the heating effect of the cut-off noise induced by the non-bonded cut-off radii 0.8/1.4 nm. Since the translational and rotational degrees of freedom are jointly coupled to the bath of 300 K, the translational temperature is slightly below 300 K, because the rotational degrees of freedom pick up more noise, so need more cooling.

Increasing the temperature with $s_T > 1$ leads to a higher temperature for the h type waters, but also the temperature of the l type waters increases. This increase is larger

the closer s_m is to 1, or, in other words, the less adiabatically decoupled the h and l type water molecules are. The increase in mass not only changes the heat flow from the h water molecules to the l water molecules, it also changes the distribution between the translational and the rotational temperature. This change in distribution influences the behaviour of the average potential energy per particle and of the diffusion coefficients.

Increasing the mass should not influence the average potential energy per particle between the systems with the same s_T. This is observed comparing $s_m = 100$ and $s_m = 1000$. The fact that there is a slight difference between $s_m = 1$ and $s_m = 100$ or 1000 which increases with s_T is most likely due to the change in the distribution of the temperature between the translational and the rotational degrees of freedom with increasing s_T. A moderate increase in the translational temperature together with a corresponding decrease in the rotational one, as they are coupled to the same bath, leads to a more negative average potential energy per water molecule.

Fig. 4.2 shows the dipole-dipole radial distribution function. The structure is fairly well sustained for the l type particles and $s_m > 1$, but also the h type water molecules show the same structure for $s_T < 5$.

The diffusion coefficients (Fig. 4.3) are influenced by s_T as well as s_m. Generally it can be said that a higher mass of the h type particles slows them down, and lowers the diffusion the l type particles along. A higher temperature of the h type particles accelerates these and the l particles. The exceptions are $s_m = 1,\ s_T = 2, 5$ compared to $s_m = 100,\ s_T = 2, 5$, where the larger s_m shows a larger diffusion constant. This is explainable from the distribution of the temperature between the translational and the rotational degrees of freedom. The larger mass leads to less heat flow from the translational degrees of freedom of the h type water molecules to the l type molecules. Therefore their translational temperature is higher in the case of $s_m = 100$. For $s_m = 1000$ both sets of degrees of freedom are decoupled and the rotational temperatures rise.

As an increase in the temperature may require a shortening of the MD time step Δt, the system was also simulated using different time steps, 2 fs for $s_T = 1$, 1.4 fs for $s_T = 2$, and 0.9 fs for $s_T = 5$. For the adiabatically decoupled systems the shortening of Δt did not change the results (not shown). Therefore a time step of 2 fs was further used.

The use of stochastic dynamics (SD) as a thermostat for the h type of degrees of freedom was investigated (Table 4.14, Figs. 4.28, 4.29, 4.30). Using $\gamma = 91\ \mathrm{ps}^{-1}$ it was able to maintain the temperature and velocity distribution better, and for some cases the average potential energy per water molecule was closer to the desired value. With a γ reduced to $9.1\ \mathrm{ps}^{-1}$ the temperature was not well maintained and showed therefore a lower diffusion, while a γ increased to $910\ \mathrm{ps}^{-1}$ led to more random movement and therefore lower diffusion. Since the diffusion, which we take as a measure for sampling speed, is much lower using SD as a thermostat, weak-coupling of the temperature bath was further used.

Reducing the coupling time τ_T of the temperature weak-coupling to 0.01 ps did not significantly change the results (not shown). It only led to an unstable system for $s_m = 1,\ s_T = 5$. In view of these results, the coupling constant of 0.1 ps was kept.

The correction of the pressure, Eq.(4.61), was able to keep the system together when simulated at constant pressure (Table 4.15, Figs. 4.31, 4.32, 4.33) except for $s_m = 1, s_T = 5$ where the system exploded as the temperature of the l type particles became too high.

4.4.2 Scaling the translational and rotational temperature with s_T for the systems 10/990, 100/900, 900/100, 990/10

system (N^h/N^l)	s_m	s_T	$\langle E_h^{pot}\rangle$ [kJ/mol]	$\langle E_l^{pot}\rangle$ [kJ/mol]	$\langle T_h^t\rangle$ [K]	$\langle T_h^r\rangle$ [K]	$\langle T_l^t\rangle$ [K]	$\langle T_l^r\rangle$ [K]	D_h [10^{-9}m^2/s]	D_l [10^{-9}m^2/s]
10/990	1	1	-41.6	-41.5	293	305	297	308	3.0	4.2
		2	-35.5	-41.4	407	379	299	310	5.4	4.3
		5	-33.3	-41.1	850	562	306	314	25	5.2
	100	1	-41.8	-41.5	296	291	297	308	1.6	4.1
		2	-41.7	-41.6	784	325	298	307	27	8.7
		5	-41.5	-42.0	2362	352	300	306	120	36
	1000	1	-41.8	-41.5	292	292	297	308	1.8	4.0
		2	-41.8	-41.6	848	313	297	308	15	4.2
		5	-41.1	-41.5	2566	332	297	308	110	4.7

Table 4.2: s_m: scaling factor for the mass of the h type water, s_T: scaling factor for the temperature (translational and rotational) of the h type water, $\langle E_h^{pot}\rangle$: average potential energy per h type water, $\langle E_l^{pot}\rangle$: average potential energy per l type water, $\langle T_h^t\rangle$: average translational temperature of the h type water, $\langle T_h^r\rangle$: average rotational temperature of the h type water, $\langle T_l^t\rangle$: average translational temperature of the l type water, $\langle T_l^r\rangle$: average rotational temperature of the l type water, D_h: diffusion coefficient of the h type water, D_l: diffusion coefficient of the l type water.

Next the ratio N^h/N^l between h type and l type molecules was varied. On one hand to the ratios 100/900 and 10/990 which are of more practical interest, on the other hand to 900/100 and 990/10 to get a better understanding of systems where h type particles dominate the interactions.

Reweighting was not possible for the systems 100/900, 900/100 and 990/10 where the reweighting factor became infinite. The effect of reweighting is therefore only applied to the 10/990 system.

Tabs. 4.2, 4.3, 4.4 and 4.5 show the average potential energy per water molecule, the average temperatures and the diffusion coefficients upon changing s_m and s_T. Figs.

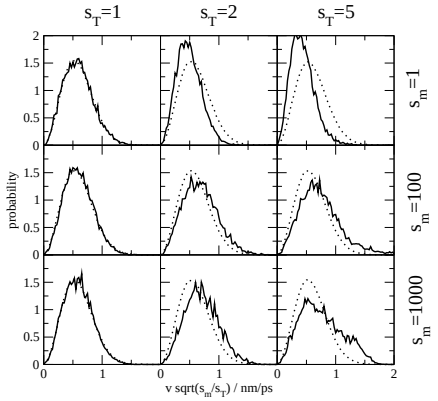

Figure 4.4: Velocity distribution of the water oxygen for 10 *h* type water molecules (solid line) and 990 *l* type water molecules (dotted line) for different mass and temperature scaling factors s_m and s_T, scaling both translational and rotational temperature with one bath but separately for 10 *h* type water molecules and 990 *l* type water molecules. The velocities for the *h* type water molecules were multiplied with $(s_m/s_T)^{1/2}$.

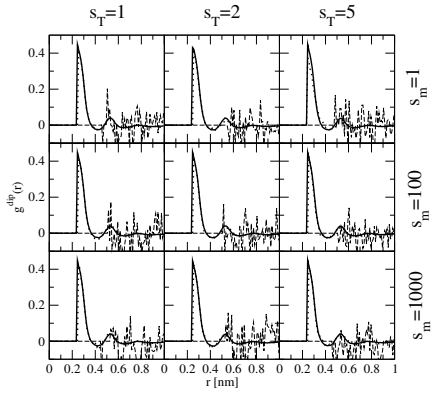

Figure 4.5: Dipole-dipole radial distribution function between 10 *h* type water molecules (solid), 990 *l* type water molecules (dashed), and between 10 *h* and 990 *l* water molecules (dotted) for different mass and temperature scaling factors s_m and s_T, scaling both translational and rotational temperature with one bath but separately for 10 *h* type water molecules and 990 *l* type water molecules.

4.4-4.9 and 4.34-4.39 show the velocity distributions, dipolar radial distribution functions and the mean square displacements of the water molecules. The effects for the different quantities follow similar trends as described for the system 500/500.

The effect of changing the number of *h* type compared to *l* type water molecules can be seen in the average potential energy per water molecule that changes more the more *h* type particles there are in the system. For 10/990 the average potential energy per *l* type water remains virtually unchanged for $s_m > 1$. The same trend is

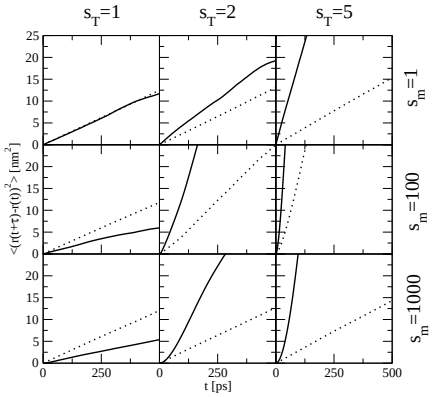

Figure 4.6: Mean square displacment of 10 *h* type water molecules (solid line) and 990 *l* type water molecules (dotted line) for different mass and temperature scaling factors s_m and s_T, scaling both translational and rotational temperature with one bath but separately for 10 *h* type water molecules and 990 *l* type water molecules.

visible for the average potential energy per *h* type water.

system (N^h/N^l)	s_m	s_T	$\langle E_h^{pot}\rangle$ [kJ/mol]	$\langle E_l^{pot}\rangle$ [kJ/mol]	$\langle T_h^t\rangle$ [K]	$\langle T_h^r\rangle$ [K]	$\langle T_l^t\rangle$ [K]	$\langle T_l^r\rangle$ [K]	D_h [10^{-9}m^2/s]	D_l [10^{-9}m^2/s]
100/900	1	1	-41.6	-41.5	295	310	296	309	3.6	4.1
		2	-37.3	-40.4	437	399	318	327	12	6.4
		5	-30.9	-37.9	945	641	379	376	44	10
	100	1	-41.8	-41.6	301	297	297	307	2.0	3.1
		2	-40.1	-41.2	816	341	303	305	30	17
		5	-38.7	-41.0	2505	382	314	300	170	76
	1000	1	-41.8	-41.6	303	295	296	307	1.5	2.8
		2	-40.8	-41.5	858	331	298	306	1.4	6.4
		5	-39.3	-41.2	2579	387	301	306	42	9.7

Table 4.3: s_m: scaling factor for the mass of the *h* type water, s_T: scaling factor for the temperature (translational and rotational) of the *h* type water, $\langle E_h^{pot}\rangle$: average potential energy per *h* type water, $\langle E_l^{pot}\rangle$: average potential energy per *l* type water, $\langle T_h^t\rangle$: average translational temperature of the *h* type water, $\langle T_h^r\rangle$: average rotational temperature of the *h* type water, $\langle T_l^t\rangle$: average translational temperature of the *l* type water, $\langle T_l^r\rangle$: average rotational temperature of the *l* type water, D_h: diffusion coefficient of the *h* type water, D_l: diffusion coefficient of the *l* type water.

For the temperature of the *h* type water molecules we see that the more *h* type particles are present, the less heat flows to the *l* type water molecules and the less heat flows from the translational to the rotational degrees of freedom. The temperature of the translational and rotational degrees of freedom of the *l* type water molecules is

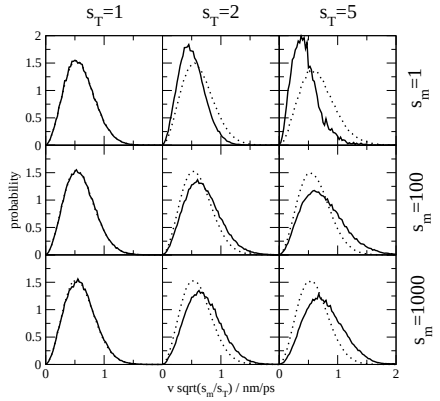

Figure 4.7: Velocity distribution of the water oxygen for 100 *h* type water molecules (solid line) and 900 *l* type water molecules (dotted line) for different mass and temperature scaling factors s_m and s_T, scaling both translational and rotational temperature with one bath but separately for 100 *h* type water molecules and 900 *l* type water molecules. The velocities for the *h* type water molecules were multiplied with $(s_m/s_T)^{1/2}$.

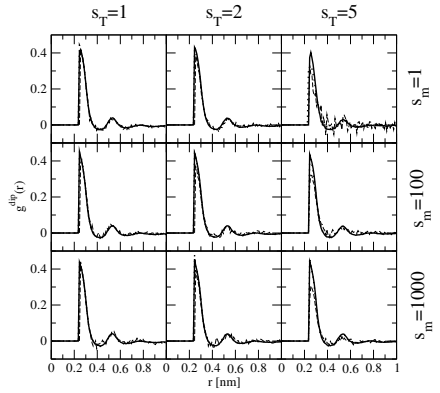

Figure 4.8: Dipole-dipole radial distribution function between 100 *h* type water molecules (solid), 900 *l* type water molecules (dashed), and between 100 *h* and 900 *l* water molecules (dotted) for different mass and temperature scaling factors s_m and s_T, scaling both translational and rotational temperature with one bath but separately for 100 *h* type water molecules and 900 *l* type water molecules.

higher when more *h* type water molecules are in the system.

The dipolar radial distribution functions show the same trends as before, be it that the one between the 10 *h* type water molecules in the 10/990 system becomes noisy due to insufficient sampling (Fig. 4.5).

The diffusion of the *h* and *l* type particles is stronger coupled if more *h* type water molecules are in the system. For the 10/990 system the diffusion of the *l* type water

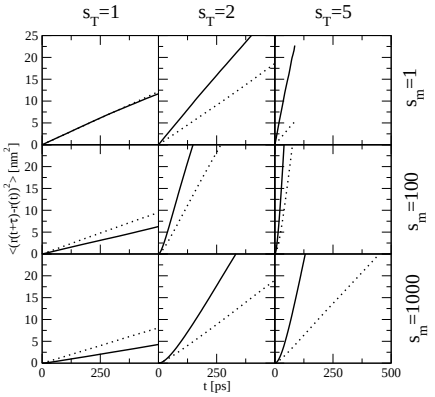

Figure 4.9: Mean square displacment of 100 *h* type water molecules (solid line) and 900 *l* type water molecules (dotted line) for different mass and temperature scaling factors s_m and s_T, scaling both translational and rotational temperature with one bath but separately for 100 *h* type water molecules and 900 *l* type water molecules.

molecules remains almost unchanged except for $s_m = 100$, $s_T = 2, 5$. This is most likely due to some resonance between the *h* and *l* type particles.

system (N^h/N^l)	s_m	s_T	$\langle E_h^{pot}\rangle$ [kJ/mol]	$\langle E_l^{pot}\rangle$ [kJ/mol]	$\langle T_h^t\rangle$ [K]	$\langle T_h^r\rangle$ [K]	$\langle T_l^t\rangle$ [K]	$\langle T_l^r\rangle$ [K]	D_h $[10^{-9}m^2/s]$	D_l $[10^{-9}m^2/s]$
900/100	1	1	-41.5	-41.6	297	308	296	307	4.3	4.2
		2	-31.7	-33.4	581	581	465	512	21	13
		5	-18.2	-20.8	1423	1410	991	1112	47	27
	100	1	-41.7	-41.6	301	299	292	306	0.47	0.53
		2	-31.9	-38.3	611	584	358	283	2.7	2.6
		5	-19.1	-32.9	1593	1380	490	347	6.7	6.3
	1000	1	-41.7	-41.4	299	301	286	312	0.14	0.18
		2	-31.7	-38.6	605	593	304	301	0.78	0.95
		5	-19.0	-34.8	1569	1424	350	298	2.0	2.7

Table 4.4: s_m: scaling factor for the mass of the *h* type water, s_T: scaling factor for the temperature (translational and rotational) of the *h* type water, $\langle E_h^{pot}\rangle$: average potential energy per *h* type water, $\langle E_l^{pot}\rangle$: average potential energy per *l* type water, $\langle T_h^t\rangle$: average translational temperature of the *h* type water, $\langle T_h^r\rangle$: average rotational temperature of the *h* type water, $\langle T_l^t\rangle$: average translational temperature of the *l* type water, $\langle T_l^r\rangle$: average rotational temperature of the *l* type water, D_h: diffusion coefficient of the *h* type water, D_l: diffusion coefficient of the *l* type water.

In Fig. 4.4 the velocity distribution of the *h* type water molecules in the 10/990 system for $s_m = 100$, $s_T = 5$ has a shoulder on the left and is not representing a canonical distribution. This could be due to a lack of energy exchange between the *h* type

particles because they are so few and because of a lack of interaction with the l type particles because they are adiabatically decoupled.

system (N^h/N^l)	s_m	s_T	$\langle E_h^{pot}\rangle$ [kJ/mol]	$\langle E_l^{pot}\rangle$ [kJ/mol]	$\langle T_h^t\rangle$ [K]	$\langle T_h^r\rangle$ [K]	$\langle T_l^t\rangle$ [K]	$\langle T_l^r\rangle$ [K]	D_h $[10^{-9}m^2/s]$	D_l $[10^{-9}m^2/s]$
990/10	1	1	-41.5	-41.7	296	308	292	306	4.2	4.6
		2	-31.2	-32.7	597	604	481	537	21	12
		5	-17.1	-19.6	1486	1497	1056	1204	45	22
	100	1	-41.7	-42.1	300	300	290	299	0.44	0.45
		2	-31.4	-38.1	601	598	357	272	2.3	1.8
		5	-17.4	-32.2	1505	1492	496	345	4.6	6.2
	1000	1	-41.7	-42.2	299	301	283	296	0.14	0.19
		2	-31.3	-38.6	599	600	299	296	0.71	0.41
		5	-17.4	-34.2	1509	1489	343	288	1.5	2.5

Table 4.5: s_m: scaling factor for the mass of the h type water, s_T: scaling factor for the temperature (translational and rotational) of the h type water, $\langle E_h^{pot}\rangle$: average potential energy per h type water, $\langle E_l^{pot}\rangle$: average potential energy per l type water, $\langle T_h^t\rangle$: average translational temperature of the h type water, $\langle T_h^r\rangle$: average rotational temperature of the h type water, $\langle T_l^t\rangle$: average translational temperature of the l type water, $\langle T_l^r\rangle$: average rotational temperature of the l type water, D_h: diffusion coefficient of the h type water, D_l: diffusion coefficient of the l type water.

4.4.3 Scaling the force with s_V for the system 500/500

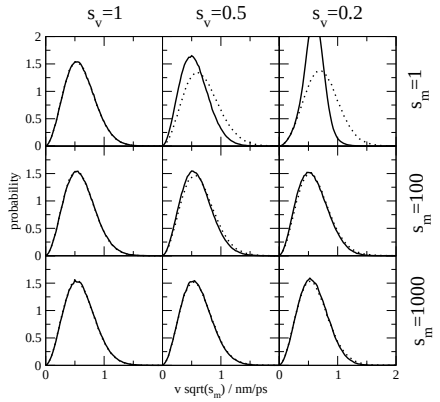

Figure 4.10: Velocity distribution of the water oxygen for 500 h type water molecules (solid line) and 500 l type water molecules (dotted line) for different mass and force scaling factors s_m and s_V. The velocities for the h type water molecules were multiplied with $(s_m)^{1/2}$.

system (N^h/N^l)	s_m	$1/s_V$	$\langle E_h^{pot}\rangle$ [kJ/mol]	$\langle E_l^{pot}\rangle$ [kJ/mol]	$\langle T_h^t\rangle$ [K]	$\langle T_h^r\rangle$ [K]	$\langle T_l^t\rangle$ [K]	$\langle T_l^r\rangle$ [K]	D_h [10^{-9}m^2/s]	D_l [10^{-9}m^2/s]
500/500	1	1	-41.5	-41.5	300	305	299	306	4.3	4.0
		2	-34.6	-37.1	262	245	390	395	44	31
		5	-42.5	-44.2	273	260	442	223	1200	1200
	100	1	-41.7	-41.6	300	300	296	306	0.80	1.2
		2	-34.0	-39.4	298	288	331	297	4.1	3.9
		5	-34.8	-39.6	298	298	312	306	55	45
	1000	1	-41.7	-41.6	300	299	294	307	0.33	0.58
		2	-34.1	-39.9	299	296	305	302	1.2	1.5
		5	-33.1	-39.7	299	299	304	304	6.3	5.3

Table 4.6: s_m: scaling factor for the mass of the h type water, s_V: scaling factor for the force on the h type water, $\langle E_h^{pot}\rangle$: average potential energy per h type water, $\langle E_l^{pot}\rangle$: average potential energy per l type water, $\langle T_h^t\rangle$: average translational temperature of the h type water, $\langle T_h^r\rangle$: average rotational temperature of the h type water, $\langle T_l^t\rangle$: average translational temperature of the l type water, $\langle T_l^r\rangle$: average rotational temperature of the l type water, D_h: diffusion coefficient of the h type water, D_l: diffusion coefficient of the l type water.

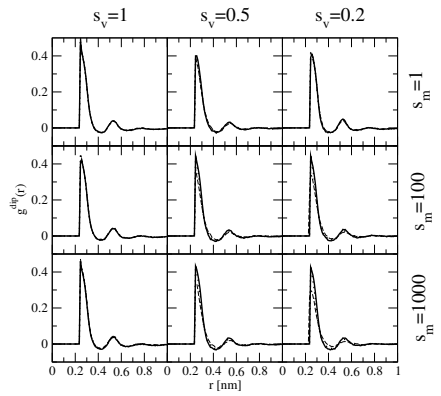

Figure 4.11: Dipole-dipole radial distribution function between 500 h type water molecules (solid), 500 l type water molecules (dashed), and between 500 h and 500 l water molecules (dotted) for different mass and force scaling factors s_m and s_V.

The system of 500 h and 500 l particles was also chosen first for testing the scaling of the force on the h type water molecules. The masses of the h type degrees of freedom were scaled with s_m and the forces acting on the h type water molecules with $s_V = s_V^{hh} = s_V^{hl}$. As coupling the translational and the rotational degrees of freedom to the same bath leads to a heat flow from the rotational degrees of freedom to the translational ones, these sets of degrees of freedom were coupled to separate baths for the h type water molecules.

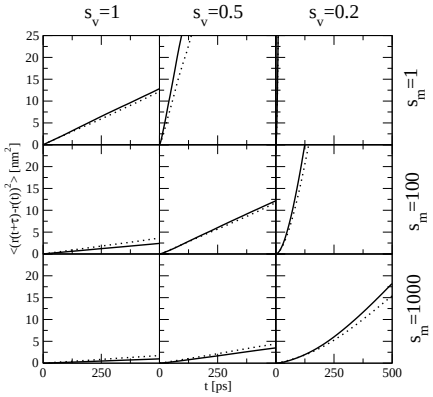

Figure 4.12: Mean square displacment of 500 h type water molecules (solid line) and 500 l type water molecules (dotted line) for different mass and force scaling factors s_m and s_V.

Reweighting was not necessary for this system since $e^{-(\beta^l-\beta^h)[V^{hh}(x^h)+(\beta^l)^{-1}\ln\langle e^{+\beta^l V^{hl}(x^h)}\rangle]}$ is equal to zero in numerical precision.

Table 4.6 shows the result for the average potential energy per water molecule, the temperature and the diffusion coefficient upon changing the factors s_m and s_V.

Lowering s_V without adiabatically decoupling the system, i.e. $s_m = 1$, leads to a heat flow from the h type waters to the translational degrees of freedom of the l type water molecules. This also changes the average potential energy per water as well. The diffusion increases for lower s_V and decreases for higher s_m.

The velocity distributions of the h and l type oxygens in Fig. 4.10 coincide for $s_m > 1$. The dipole-dipole radial distribution functions in Fig. 4.11 vary more for higher s_m. The mean square displacement of the oxygens in Fig. 4.12 is not linear in t anymore for $s_V = 0.2$ but has an amount of quadratic dependence on t indicating a free flying particle which is due to the strongly reduced interaction of the h type particles among themselves and with the l type water molecules.

4.4.4 Scaling the force with s_V for the systems 10/990, 100/900, 900/100, 990/10

To investigate the effect of the relation between the number of h and l type water molecules in the system, the systems with N^h/N^l equal to 10/990, 100/900, 900/100 and 990/10 were simulated too. The data are given in Tables 4.7-4.10 and Figs.4.13-4.18 and 4.40-4.45.

As previously observed with the temperature scaled system, the distortion of the values of the observables of the l type water molecules is the bigger the more h type particles there are in the system. For the system 10/990 the average potential energy

system (N^h/N^l)	s_m	$1/s_V$	$\langle E_h^{pot}\rangle$ [kJ/mol]	$\langle E_l^{pot}\rangle$ [kJ/mol]	$\langle T_h^t\rangle$ [K]	$\langle T_h^r\rangle$ [K]	$\langle T_l^t\rangle$ [K]	$\langle T_l^r\rangle$ [K]	D_h [10^{-9}m^2/s]	D_l [10^{-9}m^2/s]
10/990	1	1	-41.7	-41.5	294	302	297	308	4.0	4.2
		2	-38.3	-41.4	210	201	299	309	7.7	4.6
		5	-32.0	-41.3	209	162	307	308	180	24
	100	1	-41.7	-41.5	289	292	297	308	2.6	4.0
		2	-39.0	-41.5	284	250	298	307	10	4.3
		5	-38.4	-41.5	286	270	298	307	91	4.5
	1000	1	-41.7	-41.5	290	290	297	308	2.9	3.9
		2	-40.0	-41.5	289	275	297	308	7.7	4.4
		5	-39.8	-41.5	290	283	297	308	23	4.8

Table 4.7: s_m: scaling factor for the mass of the h type water, s_V: scaling factor for the force on the h type water, $\langle E_h^{pot}\rangle$: average potential energy per h type water, $\langle E_l^{pot}\rangle$: average potential energy per l type water, $\langle T_h^t\rangle$: average translational temperature of the h type water, $\langle T_h^r\rangle$: average rotational temperature of the h type water, $\langle T_l^t\rangle$: average translational temperature of the l type water, $\langle T_l^r\rangle$: average rotational temperature of the l type water, D_h: diffusion coefficient of the h type water, D_l: diffusion coefficient of the l type water.

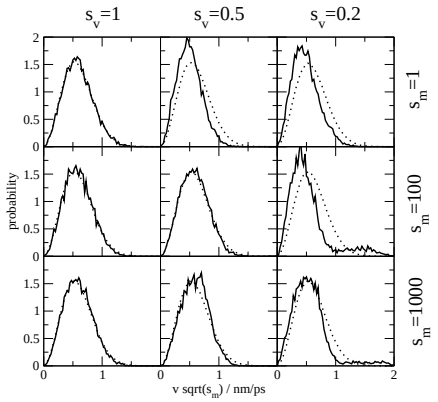

Figure 4.13: Velocity distribution of the water oxygen for 10 h type water molecules (solid line) and 990 l type water molecules (dotted line) for different mass and force scaling factors s_m and s_V. The velocities for the h type water molecules were multiplied with $(s_m)^{1/2}$.

per l type water and the diffusion coefficient of the l type water are not influenced by s_T for $s_m > 1$.

Generally the same trends as for temperature scaling are observed.

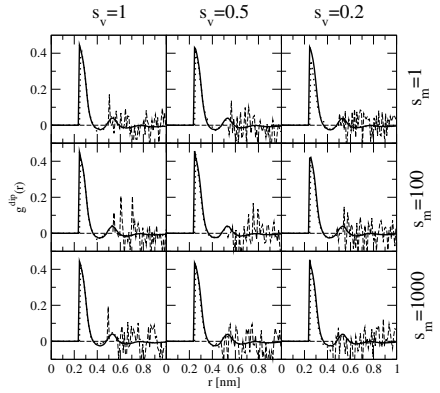

Figure 4.14: Dipole-dipole radial distribution function between 10 *h* type water molecules (solid), 990 *l* type water molecules (dashed), and between 10 *h* and 990 *l* water molecules (dotted) for different mass and force scaling factors s_m and s_V.

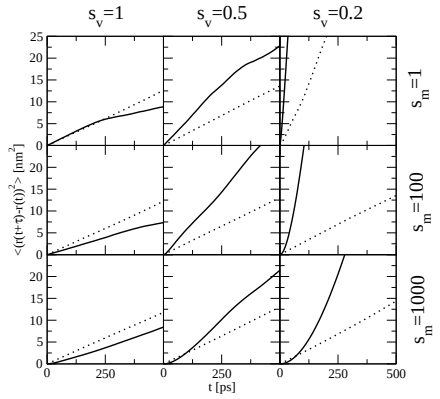

Figure 4.15: Mean square displacment of 10 *h* type water molecules (solid line) and 990 *l* type water molecules (dotted line) for different mass and force scaling factors s_m and s_V.

4.4.5 Scaling the translational temperature with s_T for the system 500/500

Since the temperature scaling of both rotational and translational degrees of freedom of the *h* type water molecules caused a heat flow from one set of degrees of freedom to the other and since the main sampling benefits in the form of fast diffusion seemed to have come from the higher translational temperature, a system with 500 *h* type and 500 *l* type water molecules was set up in which only the translational degrees of freedom of the *h* type water molecules were scaled with s_T.

The velocity distribution in Fig. 4.19 compared to Fig. 4.1 is almost identical for the *h* and *l* type water molecules for $s_m > 1$. The same is true for the dipole-dipole radial distribution function in Fig. 4.20 compared to Fig. 4.2.

system (N^h/N^l)	s_m	$1/s_V$	$\langle E_h^{pot}\rangle$ [kJ/mol]	$\langle E_l^{pot}\rangle$ [kJ/mol]	$\langle T_h^t\rangle$ [K]	$\langle T_h^r\rangle$ [K]	$\langle T_l^t\rangle$ [K]	$\langle T_l^r\rangle$ [K]	D_h [10^{-9}m^2/s]	D_l [10^{-9}m^2/s]
100/900	1	1	-41.5	-41.5	300	306	297	307	4.5	4.0
		2	-37.5	-40.5	226	209	319	322	16	9.4
		5	-43.8	-45.3	253	286	390	226	460	400
	100	1	-41.8	-41.6	299	298	297	307	2.1	3.2
		2	-38.0	-41.0	295	267	307	304	10	6.0
		5	-39.6	-41.2	296	293	302	307	34	15
	1000	1	-41.8	-41.6	299	298	297	307	1.3	2.6
		2	-38.6	-41.3	298	288	300	306	4.7	3.4
		5	-40.0	-41.4	298	297	299	306	9.9	3.5

Table 4.8: s_m: scaling factor for the mass of the h type water, s_V: scaling factor for the force on the h type water, $\langle E_h^{pot}\rangle$: average potential energy per h type water, $\langle E_l^{pot}\rangle$: average potential energy per l type water, $\langle T_h^t\rangle$: average translational temperature of the h type water, $\langle T_h^r\rangle$: average rotational temperature of the h type water, $\langle T_l^t\rangle$: average translational temperature of the l type water, $\langle T_l^r\rangle$: average rotational temperature of the l type water, D_h: diffusion coefficient of the h type water, D_l: diffusion coefficient of the l type water.

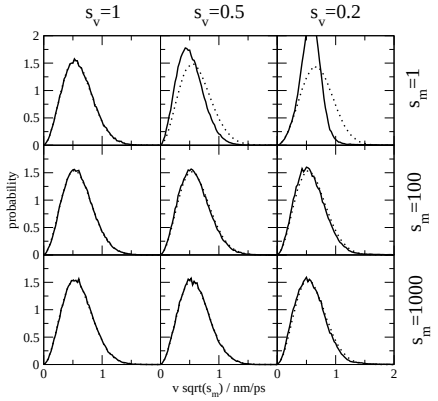

Figure 4.16: Velocity distribution of the water oxygen for 100 h type water molecules (solid line) and 900 l type water molecules (dotted line) for different mass and force scaling factors s_m and s_V. The velocities for the h type water molecules were multiplied with $(s_m)^{1/2}$.

In Table 4.11 the average potential energy per water molecule, the average temperatures, and the diffusion coefficients are given while velocity distributions, dipolar radial distribution functions and mean square displacements of water molecules are shown in Figs. 4.19-4.21. When comparing to the results of Table 4.1 one can see that the average potential energy per water molecule stays closer to its desired value, that the temperatures are closer to their target values and that the diffusion coefficient is larger in most cases.

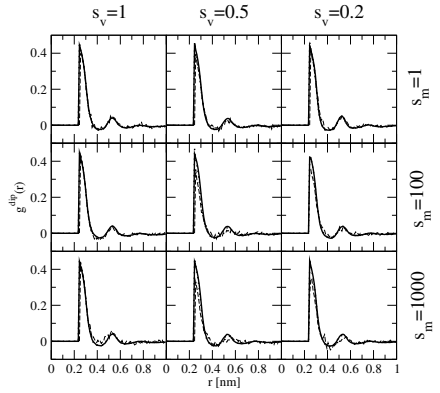

Figure 4.17: Dipole-dipole radial distribution function between 100 *h* type water molecules (solid), 900 *l* type water molecules (dashed), and between 100 *h* and 900 *l* water molecules (dotted) for different mass and force scaling factors s_m and s_V.

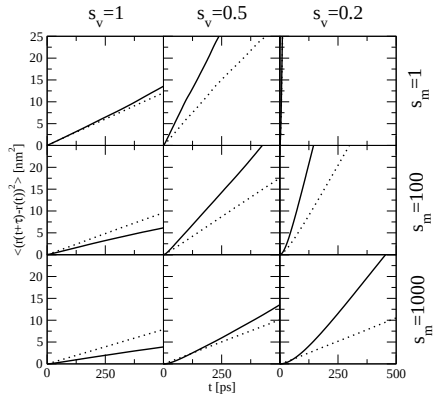

Figure 4.18: Mean square displacment of 100 *h* type water molecules (solid line) and 900 *l* type water molecules (dotted line) for different mass and force scaling factors s_m and s_V.

4.4.6 Scaling the translational temperature with s_T for the systems 10/990, 100/900

Because of the good results obtained when only scaling the translational temperature of the *h* type degrees of freedom, this scaling was also investigated for the systems with N^h/N^l equal to 10/990 and 100/900.

Tables 4.12 and 4.13 show that the fewer *h* type water molecules are present, the less the effect of $s_T > 1$ for $s_m > 1$ on the average potential energy per water molecule, especially for the *l* type water molecules. For 10/990 the average potential energy per *l* type water molecule does not change for $s_T > 1$ and $s_m > 1$, also the diffusion coefficient of the *l* type water molecules is hardly influenced, except for $s_m = 100$, $s_T = 5$. Figs. 4.22-4.27 show this in detail.

system (N^h/N^l)	s_m	$1/s_V$	$\langle E_h^{pot}\rangle$ [kJ/mol]	$\langle E_l^{pot}\rangle$ [kJ/mol]	$\langle T_h^t\rangle$ [K]	$\langle T_h^r\rangle$ [K]	$\langle T_l^t\rangle$ [K]	$\langle T_l^r\rangle$ [K]	D_h [10^{-9}m^2/s]	D_l [10^{-9}m^2/s]
900/100	1	1	-41.5	-41.6	301	304	300	304	4.4	5.1
		2	-31.8	-34.1	292	289	456	486	16	13
		5	-18.5	-24.5	291	284	810	891	26	20
	100	1	-41.7	-41.6	300	300	292	306	0.48	0.63
		2	-31.7	-38.4	300	299	337	289	1.8	2.0
		5	-18.7	-34.3	300	298	382	302	2.5	3.0
	1000	1	-41.7	-41.6	300	300	288	309	0.18	0.24
		2	-31.7	-38.7	300	299	297	307	0.54	0.71
		5	-18.7	-35.3	300	299	316	301	0.73	1.1

Table 4.9: s_m: scaling factor for the mass of the h type water, s_V: scaling factor for the force on the h type water, $\langle E_h^{pot}\rangle$: average potential energy per h type water, $\langle E_l^{pot}\rangle$: average potential energy per l type water, $\langle T_h^t\rangle$: average translational temperature of the h type water, $\langle T_h^r\rangle$: average rotational temperature of the h type water, $\langle T_l^t\rangle$: average translational temperature of the l type water, $\langle T_l^r\rangle$: average rotational temperature of the l type water, D_h: diffusion coefficient of the h type water, D_l: diffusion coefficient of the l type water.

system (N^h/N^l)	s_m	$1/s_V$	$\langle E_h^{pot}\rangle$ [kJ/mol]	$\langle E_l^{pot}\rangle$ [kJ/mol]	$\langle T_h^t\rangle$ [K]	$\langle T_h^r\rangle$ [K]	$\langle T_l^t\rangle$ [K]	$\langle T_l^r\rangle$ [K]	D_h [10^{-9}m^2/s]	D_l [10^{-9}m^2/s]
990/10	1	1	-41.5	-41.5	301	304	298	304	4.3	6.0
		2	-31.3	-33.6	299	300	468	506	15	9.2
		5	-17.3	-23.4	299	299	855	936	20	18
	100	1	-41.7	-41.8	300	300	295	300	0.44	0.52
		2	-31.3	-38.4	300	300	335	286	1.5	1.8
		5	-17.3	-33.8	300	300	374	293	2.0	2.3
	1000	1	-41.7	-41.7	300	300	288	303	0.13	0.24
		2	-31.3	-38.6	300	299	295	301	0.47	0.53
		5	-17.3	-34.8	300	299	313	296	0.63	1.0

Table 4.10: s_m: scaling factor for the mass of the h type water, s_V: scaling factor for the force on the h type water, $\langle E_h^{pot}\rangle$: average potential energy per h type water, $\langle E_l^{pot}\rangle$: average potential energy per l type water, $\langle T_h^t\rangle$: average translational temperature of the h type water, $\langle T_h^r\rangle$: average rotational temperature of the h type water, $\langle T_l^t\rangle$: average translational temperature of the l type water, $\langle T_l^r\rangle$: average rotational temperature of the l type water, D_h: diffusion coefficient of the h type water, D_l: diffusion coefficient of the l type water.

4.5 Discussion

Three methods to enhance configurational sampling of a particular subset of h type degrees of freedom in a molecular system were investigated. They all make use of an

system (N^h/N^l)	s_m	s_T	$\langle E_h^{pot}\rangle$ [kJ/mol]	$\langle E_l^{pot}\rangle$ [kJ/mol]	$\langle T_h^t\rangle$ [K]	$\langle T_h^r\rangle$ [K]	$\langle T_l^t\rangle$ [K]	$\langle T_l^r\rangle$ [K]	D_h [10^{-9}m^2/s]	D_l [10^{-9}m^2/s]
500/500	1	1	-41.5	-41.5	300	305	299	306	4.1	4.0
		2	-39.9	-41.2	485	309	405	291	120	100
		5	-38.4	-41.3	1089	316	687	261	12200	10000
	100	1	-41.7	-41.6	300	300	296	306	0.78	1.3
		2	-39.4	-40.8	588	307	302	303	20	17
		5	-38.4	-40.3	1475	314	312	298	210	140
	1000	1	-41.7	-41.6	300	299	294	307	0.32	0.57
		2	-39.8	-41.0	596	302	296	305	7.5	6.1
		5	-39.1	-40.7	1493	304	298	305	80	48

Table 4.11: s_m: scaling factor for the mass of the h type water, s_T: scaling factor for the temperature (translational) of the h type water, $\langle E_h^{pot}\rangle$: average potential energy per h type water, $\langle E_l^{pot}\rangle$: average potential energy per l type water, $\langle T_h^t\rangle$: average translational temperature of the h type water, $\langle T_h^r\rangle$: average rotational temperature of the h type water, $\langle T_l^t\rangle$: average translational temperature of the l type water, $\langle T_l^r\rangle$: average rotational temperature of the l type water, D_h: diffusion coefficient of the h type water, D_l: diffusion coefficient of the l type water.

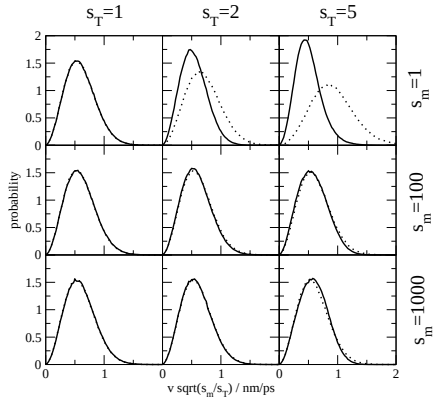

Figure 4.19: Velocity distribution of the water oxygen for 500 h type water molecules (solid line) and 500 l type water molecules (dotted line) for different mass and temperature scaling factors s_m and s_T, scaling the translational temperature. The velocities for the h type water molecules were multiplied with $(s_m/s_T)^{1/2}$.

upscaling of the masses of the h type particles in order to adiabatically decouple their motion from that of the l type particles whose masses are not increased. The statistical-mechanical expressions for ensemble averages of quantities $Q(x^h)$ that depend on the h type particle coordinates for the unscaled system in terms of the corresponding ensemble averages for the systems that were simulated with upscaled temperature or downscaled interactions or forces were given. These expressions show that the meth-

system (N^h/N^l)	s_m	s_T	$\langle E_h^{pot}\rangle$ [kJ/mol]	$\langle E_l^{pot}\rangle$ [kJ/mol]	$\langle T_h^t\rangle$ [K]	$\langle T_h^r\rangle$ [K]	$\langle T_l^t\rangle$ [K]	$\langle T_l^r\rangle$ [K]	D_h [10^{-9}m^2/s]	D_l [10^{-9}m^2/s]
10/990	1	1	-41.7	-41.5	294	302	297	308	1.8	4.2
		2	-40.0	-41.5	392	312	299	308	5.5	4.4
		5	-36.0	-41.3	686	340	304	308	27	5.1
	100	1	-41.7	-41.5	289	292	297	308	2.1	4.1
		2	-41.6	-41.5	553	294	297	308	8.7	4.4
		5	-41.3	-41.5	1341	295	298	307	66	18
	1000	1	-41.7	-41.5	290	290	297	308	3.1	3.8
		2	-41.7	-41.5	577	292	297	308	9.9	4.1
		5	-41.7	-41.5	1437	292	297	308	45	4.4

Table 4.12: s_m: scaling factor for the mass of the h type water, s_T: scaling factor for the temperature (translational) of the h type water, $\langle E_h^{pot}\rangle$: average potential energy per h type water, $\langle E_l^{pot}\rangle$: average potential energy per l type water, $\langle T_h^t\rangle$: average translational temperature of the h type water, $\langle T_h^r\rangle$: average rotational temperature of the h type water, $\langle T_l^t\rangle$: average translational temperature of the l type water, $\langle T_l^r\rangle$: average rotational temperature of the l type water, D_h: diffusion coefficient of the h type water, D_l: diffusion coefficient of the l type water.

system (N^h/N^l)	s_m	s_T	$\langle E_h^{pot}\rangle$ [kJ/mol]	$\langle E_l^{pot}\rangle$ [kJ/mol]	$\langle T_h^t\rangle$ [K]	$\langle T_h^r\rangle$ [K]	$\langle T_l^t\rangle$ [K]	$\langle T_l^r\rangle$ [K]	D_h [10^{-9}m^2/s]	D_l [10^{-9}m^2/s]
100/900	1	1	-41.5	-41.5	300	306	297	307	4.4	4.1
		2	-39.3	-41.0	415	318	316	308	8.6	6.5
		5	-36.2	-40.4	770	345	371	303	620	290
	100	1	-41.8	-41.6	299	298	297	307	1.9	3.2
		2	-41.1	-41.4	579	302	299	306	15	8.5
		5	-40.4	-41.2	1442	308	305	304	100	45
	1000	1	-41.8	-41.6	299	298	297	307	1.4	2.7
		2	-41.5	-41.5	595	299	297	308	7.5	4.6
		5	-41.0	-41.4	1485	302	298	307	32	12

Table 4.13: s_m: scaling factor for the mass of the h type water, s_T: scaling factor for the temperature (translational) of the h type water, $\langle E_h^{pot}\rangle$: average potential energy per h type water, $\langle E_l^{pot}\rangle$: average potential energy per l type water, $\langle T_h^t\rangle$: average translational temperature of the h type water, $\langle T_h^r\rangle$: average rotational temperature of the h type water, $\langle T_l^t\rangle$: average translational temperature of the l type water, $\langle T_l^r\rangle$: average rotational temperature of the l type water, D_h: diffusion coefficient of the h type water, D_l: diffusion coefficient of the l type water.

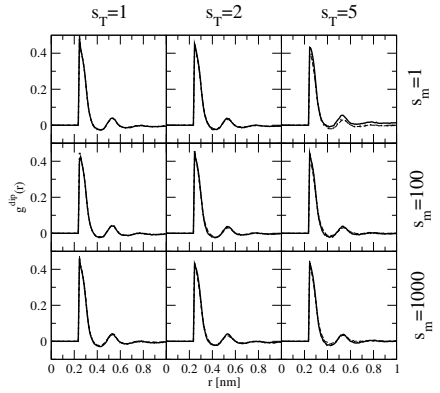

Figure 4.20: Dipole-dipole radial distribution function between 500 h type water molecules (solid), 500 l type water molecules (dashed), and between 500 h and 500 l water molecules (dotted) for different mass and temperature scaling factors s_m and s_T, scaling the translational temperature.

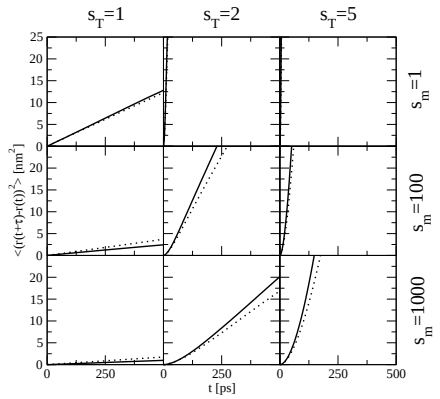

Figure 4.21: Mean square displacment of 500 h type water molecules (solid line) and 500 l type water molecules (dotted line) for different mass and temperature scaling factors s_m and s_T, scaling the translational temperature.

ods will only produce reasonably accurate results if $N^h \ll N^l$. Whether temperature scaling, translational temperature scaling, interaction scaling or force scaling will lead to enhanced sampling of configurational space without distorting the configuration distribution too much will depend on the type of system and h type degrees of freedom, i.e. the particular molecular system, force field, and degrees of freedom of interest.

Here we investigated liquid water at room temperature and a density of 997 kgm^{-3}. It turned out that the average potential energy per h or l type water molecule was a sensitive and thus a simpler probe of the configurational distribution than the dipolar radial distribution function. These quantities were the better reproduced the lower the ratio N^h/N^l was. For $N^h/(N^h + N^l) \leq 0.1$ all three methods tested, temperature scaling,

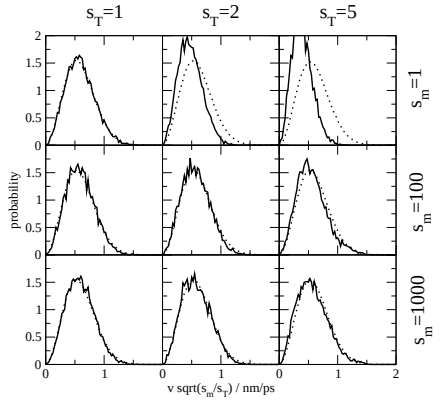

Figure 4.22: Velocity distribution of the water oxygen for 10 *h* type water molecules (solid line) and 990 *l* type water molecules (dotted line) for different mass and temperature scaling factors s_m and s_T, scaling the translational temperature. The velocities for the *h* type water molecules were multiplied with $(s_m/s_T)^{1/2}$.

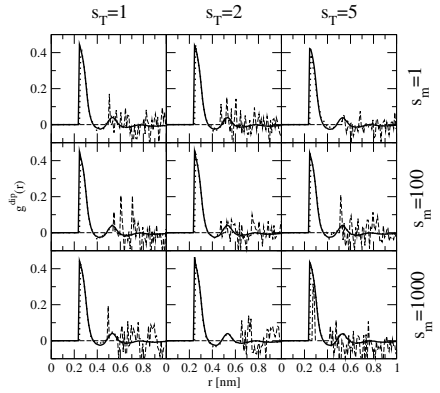

Figure 4.23: Dipole-dipole radial distribution function between 10 *h* type water molecules (solid), 990 *l* type water molecules (dashed), and between 10 *h* and 990 *l* water molecules (dotted) for different mass and temperature scaling factors s_m and s_T, scaling the translationalb temperature.

force scaling and translational temperature scaling, performed well, i.e. a large speed up of sampling could be reached at the cost of a slight perturbation of the configurational distribution. Especially the last method did very well because the configurational distribution was not very sensitive to the translational temperature.

All simulations showed that the mass scaling factor s_m should be sufficiently large, i.e. 10^2 or 10^3, to adiabatically decouple the two types of water molecules. However, the larger s_m, the slower the *h* type molecules will diffuse. Thus for the present system a value of about 10^2 produced most optimal sampling. For the system with $N^h/(N^h + N^l) \approx 0.1$ and $s_m = 10^2$ sampling of the *h* type degrees of freedom was enhanced

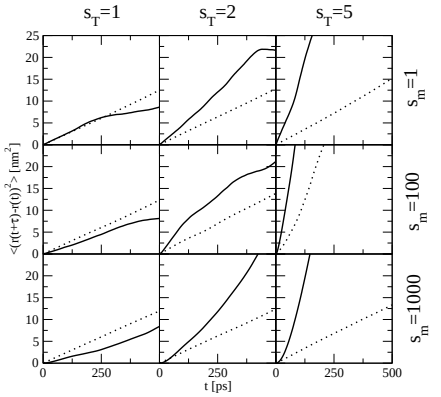

Figure 4.24: Mean square displacment of 10 *h* type water molecules (solid line) and 990 *l* type water molecules (dotted line) for different mass and temperature scaling factors s_m and s_T, scaling the translational temperature.

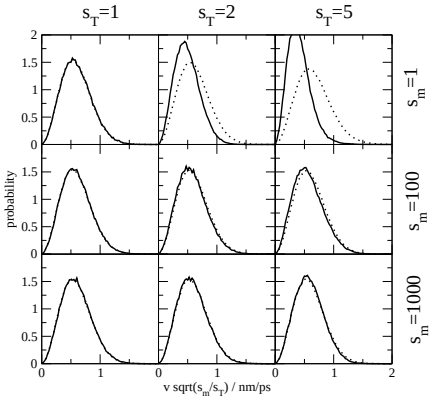

Figure 4.25: Velocity distribution of the water oxygen for 100 *h* type water molecules (solid line) and 900 *l* type water molecules (dotted line) for different mass and temperature scaling factors s_m and s_T, scaling the translational temperature. The velocities for the *h* type water molecules were multiplied with $(s_m/s_T)^{1/2}$.

by roughly a factor of 7 when using temperature or force scaling, while a factor of 25 could be reached using translational temperature scaling. If the system contains less *h* type degrees of freedom these numbers are getting better. For systems with $N^h/(N^h+N^l) \approx 0.01$ and $s_m = 10^2$ a factor of 25 is reached when using temperature of force scaling and a factor of 35 for translational temperature scaling.

Since the sampling enhancement and concomitant perturbation of the configurational distribution depend on the particular energy hypersurface, i.e. molecular model, and number and character of the *h* type degrees of freedom, the parameters s_m and s_T or s_V have to be chosen with care. If the system contains atoms of different mass, s_m should be sufficiently large to scale all masses of the *h* type atoms beyond those of the

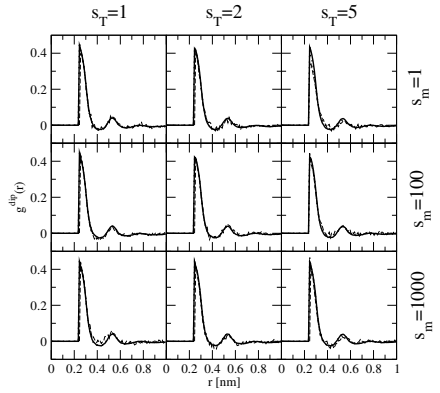

Figure 4.26: Dipole-dipole radial distribution function between 100 *h* type water molecules (solid), 900 *l* type water molecules (dashed), and between 100 *h* and 900 *l* water molecules (dotted) for different mass and temperature scaling factors s_m and s_T, scaling the translational temperature.

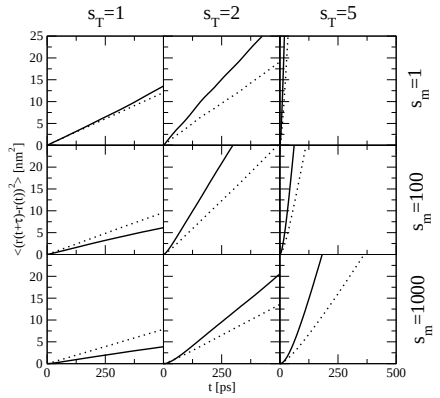

Figure 4.27: Mean square displacment of 100 *h* type water molecules (solid line) and 900 *l* type water molecules (dotted line) for different mass and temperature scaling factors s_m and s_T, scaling the translational temperature.

l type particles. Otherwise the different atom types are insufficiently decoupled, as was the case for $s_m = 10$ for the water system in which the hydrogen and oxygen mass differ by a factor of 16. On the other hand a too large s_m will reduce sampling. The lower N^h versus N^l the better the methods perform.

We are currently testing the methodology for other types of systems, e.g. diffusion of ions in an aqueous solution for which the method of translational temperature scaling seems most suitable, or polypeptide folding for which temperature and force scaling should be compared.

4.6 Supplementary material

system (N^h/N^l)	s_m	s_T	$\langle E_h^{pot}\rangle$ [kJ/mol]	$\langle E_l^{pot}\rangle$ [kJ/mol]	$\langle T_h^t\rangle$ [K]	$\langle T_h^r\rangle$ [K]	$\langle T_l^t\rangle$ [K]	$\langle T_l^r\rangle$ [K]	D_h $[10^{-9}m^2/s]$	D_l $[10^{-9}m^2/s]$
500/500	1	1	-41.7	-41.7	300	302	299	304	0.59	1.3
SD		2	-32.6	-35.3	584	580	412	460	1.7	4.7
91 ps^{-1}		5	-20.2	-24.2	1423	1407	740	936	4.7	10
on oxygen	100	1	-41.7	-41.6	300	300	296	309	0.0091	0.13
		2	-34.3	-40.0	599	598	316	308	0.024	0.48
		5	-26.2	-38.0	1489	1483	363	324	0.059	1.2
	1000	1	-42.1	-41.6	300	303	294	311	0.0011	0.060
		2	-35.1	-40.6	600	602	298	308	0.0028	0.14
		5	-27.4	-39.9	1499	1500	307	307	0.0071	0.36

Table 4.14: s_m: scaling factor for the mass of the h type water, s_T: scaling factor for the temperature (translational and rotational) of the h type water, $\langle E_h^{pot}\rangle$: average potential energy per h type water, $\langle E_l^{pot}\rangle$: average potential energy per l type water, $\langle T_h^t\rangle$: average translational temperature of the h type water, $\langle T_h^r\rangle$: average rotational temperature of the h type water, $\langle T_l^t\rangle$: average translational temperature of the l type water, $\langle T_l^r\rangle$: average rotational temperature of the l type water, D_h: diffusion coefficient of the h type water, D_l: diffusion coefficient of the l type water.

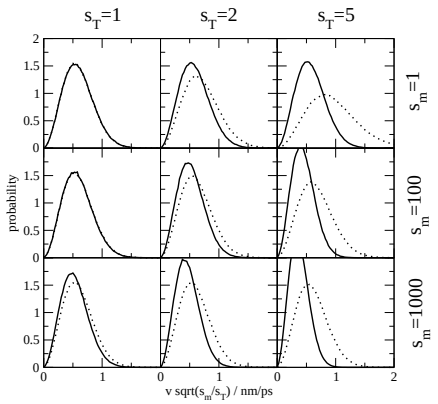

Figure 4.28: Velocity distribution of the water oxygen for 500 h type water molecules (solid line) and 500 l type water molecules (dotted line) for different mass and temperature scaling factors s_m and s_T, scaling both translational and rotational temperature with one bath but separately for 500 h type water molecules and 500 l type water molecules. The velocities for the h type water molecules were multiplied with $(s_m/s_T)^{1/2}$. (SD, 91 ps^{-1}/s_m on the oxygen)

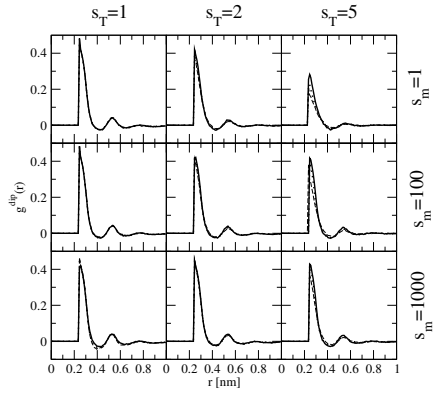

Figure 4.29: Dipole-dipole radial distribution function between 500 *h* type water molecules (solid), 500 *l* type water molecules (dashed), and between 500 *h* and 500 *l* water molecules (dotted) for different mass and temperature scaling factors s_m and s_T, scaling both translational and rotational temperature with one bath but separately for 500 *h* type water molecules and 500 *l* type water molecules.(SD, 91 ps^{-1}/s_m on the oxygen)

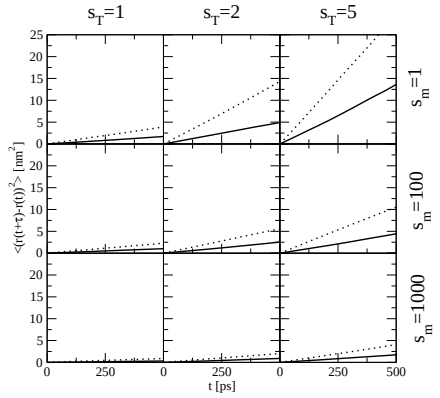

Figure 4.30: Mean square displacment of 500 *h* type water molecules (solid line) and 500 *l* type water molecules (dotted line) for different mass and temperature scaling factors s_m and s_T, scaling both translational and rotational temperature with one bath but separately for 500 *h* type water molecules and 500 *l* type water molecules.(SD, 91 ps^{-1}/s_m on the oxygen)

system (N^h/N^l)	s_m	s_T	$\langle E_h^{pot}\rangle$ [kJ/mol]	$\langle E_l^{pot}\rangle$ [kJ/mol]	$\langle T_h^t\rangle$ [K]	$\langle T_h^r\rangle$ [K]	$\langle T_l^t\rangle$ [K]	$\langle T_l^r\rangle$ [K]	D_h [10^{-9}m^2/s]	D_l [10^{-9}m^2/s]
500/500	1	1	-41.3	-41.3	296	307	297	308	4.6	4.3
NpT		2	-31.9	-34.8	525	486	383	410	30	16
	100	1	-41.5	-41.4	301	298	295	306	0.83	1.2
		2	-35.6	-39.2	776	406	321	296	37	35
		5	-24.5	-33.7	2358	598	360	282	110	81
	1000	1	-41.5	-41.4	300	299	295	306	0.31	0.61
		2	-35.7	-39.5	768	426	302	302	12	11
		5	-24.3	-34.6	2348	638	317	295	22	12

Table 4.15: s_m: scaling factor for the mass of the h type water, s_T: scaling factor for the temperature (translational and rotational) of the h type water, $\langle E_h^{pot}\rangle$: average potential energy per h type water, $\langle E_l^{pot}\rangle$: average potential energy per l type water, $\langle T_h^t\rangle$: average translational temperature of the h type water, $\langle T_h^r\rangle$: average rotational temperature of the h type water, $\langle T_l^t\rangle$: average translational temperature of the l type water, $\langle T_l^r\rangle$: average rotational temperature of the l type water, D_h: diffusion coefficient of the h type water, D_l: diffusion coefficient of the l type water.

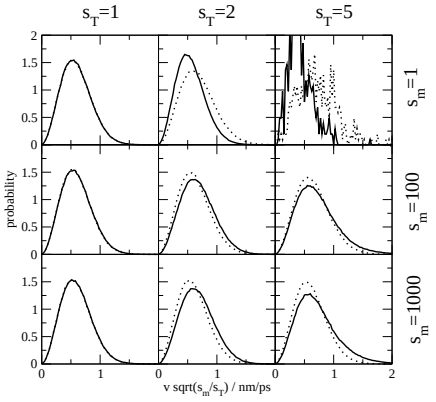

Figure 4.31: Velocity distribution of the water oxygen for 500 h type water molecules (solid line) and 500 l type water molecules (dotted line) for different mass and temperature scaling factors s_m and s_T, scaling both translational and rotational temperature with one bath but separately for 500 h type water molecules and 500 l type water molecules. The velocities for the h type water molecules were multiplied with $(s_m/s_T)^{1/2}$.(NpT)

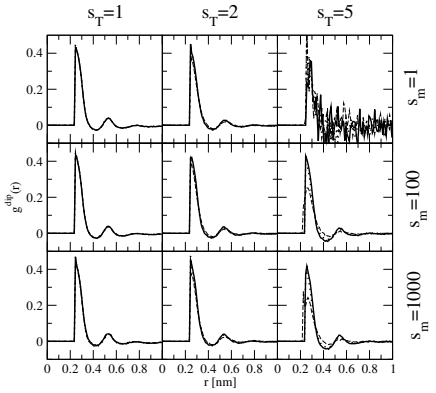

Figure 4.32: Dipole-dipole radial distribution function between 500 *h* type water molecules (solid), 500 *l* type water molecules (dashed), and between 500 *h* and 500 *l* water molecules (dotted) for different mass and temperature scaling factors s_m and s_T, scaling both translational and rotational temperature with one bath but separately for 500 *h* type water molecules and 500 *l* type water molecules.(NpT)

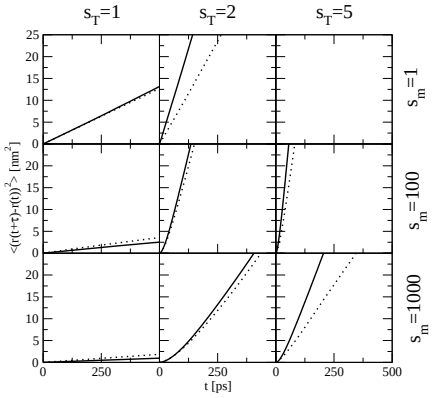

Figure 4.33: Mean square displacment of 500 *h* type water molecules (solid line) and 500 *l* type water molecules (dotted line) for different mass and temperature scaling factors s_m and s_T, scaling both translational and rotational temperature with one bath but separately for 500 *h* type water molecules and 500 *l* type water molecules.(NpT)

system (N^h/N^l)	s_m	s_T	$\langle E_h^{pot}\rangle$ [kJ/mol]	$\langle E_l^{pot}\rangle$ [kJ/mol]	$\langle T_h^t\rangle$ [K]	$\langle T_h^r\rangle$ [K]	$\langle T_l^t\rangle$ [K]	$\langle T_l^r\rangle$ [K]	D_h [10^{-9}m^2/s]	D_l [10^{-9}m^2/s]
900/100	1	1	-41.5	-41.6	301	304	300	304	4.5	4.9
		2	-39.6	-40.8	546	326	474	304	950	850
		5	-38.8	-40.8	1327	344	1007	293	1900	1600
	100	1	-41.7	-41.6	300	300	292	306	0.48	0.52
		2	-38.3	-40.2	590	309	302	301	21	21
		5	-36.7	-39.5	1479	319	314	294	310	190
	1000	1	-41.7	-41.6	300	300	288	309	0.18	0.22
		2	-38.4	-40.2	597	303	289	309	6.7	6.6
		5	-36.8	-39.5	1492	307	292	308	66	53

Table 4.16: s_m: scaling factor for the mass of the h type water, s_T: scaling factor for the temperature (translational) of the h type water, $\langle E_h^{pot}\rangle$: average potential energy per h type water, $\langle E_l^{pot}\rangle$: average potential energy per l type water, $\langle T_h^t\rangle$: average translational temperature of the h type water, $\langle T_h^r\rangle$: average rotational temperature of the h type water, $\langle T_l^t\rangle$: average translational temperature of the l type water, $\langle T_l^r\rangle$: average rotational temperature of the l type water, D_h: diffusion coefficient of the h type water, D_l: diffusion coefficient of the l type water.

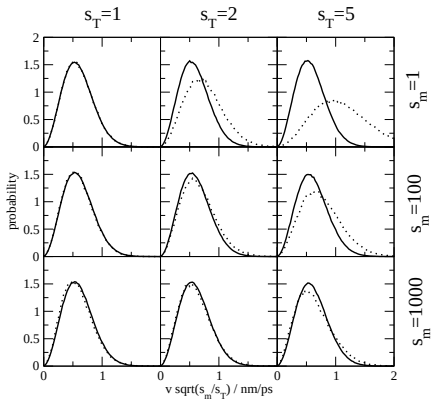

Figure 4.34: Velocity distribution of the water oxygen for 900 h type water molecules (solid line) and 100 l type water molecules (dotted line) for different mass and temperature scaling factors s_m and s_T, scaling both translational and rotational temperature with one bath but separately for 900 h type water molecules and 100 l type water molecules. The velocities for the h type water molecules were multiplied with $(s_m/s_T)^{1/2}$.

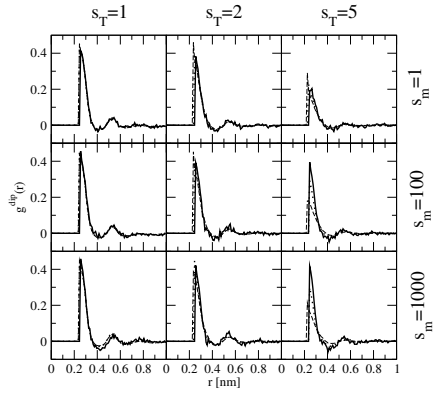

Figure 4.35: Dipole-dipole radial distribution function between 900 *h* type water molecules (solid), 100 *l* type water molecules (dashed), and between 900 *h* and 100 *l* water molecules (dotted) for different mass and temperature scaling factors s_m and s_T, scaling both translational and rotational temperature with one bath but separately for 900 *h* type water molecules and 100 *l* type water molecules.

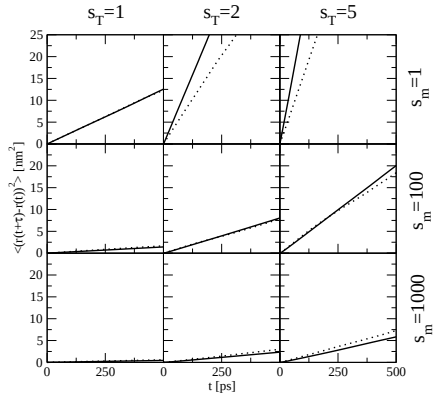

Figure 4.36: Mean square displacment of 900 *h* type water molecules (solid line) and 100 *l* type water molecules (dotted line) for different mass and temperature scaling factors s_m and s_T, scaling both translational and rotational temperature with one bath but separately for 900 *h* type water molecules and 100 *l* type water molecules.

system (N^h/N^l)	s_m	s_T	$\langle E_h^{pot}\rangle$ [kJ/mol]	$\langle E_l^{pot}\rangle$ [kJ/mol]	$\langle T_h^t\rangle$ [K]	$\langle T_h^r\rangle$ [K]	$\langle T_l^t\rangle$ [K]	$\langle T_l^r\rangle$ [K]	D_h [10^{-9}m^2/s]	D_l [10^{-9}m^2/s]
990/10	1	1	-41.5	-41.5	301	304	298	304	4.3	4.3
		2	-39.1	-40.1	553	336	477	320	500	420
		5	-38.2	-40.1	1370	371	1062	324	3900	3100
	100	1	-41.7	-41.8	300	300	295	300	0.44	0.56
		2	-38.1	-40.3	590	310	297	299	27	23
		5	-36.4	-39.3	1479	321	315	290	270	210
	1000	1	-41.7	-41.7	300	300	288	303	0.14	0.21
		2	-38.2	-40.6	597	303	288	303	9.5	11
		5	-36.6	-39.5	1493	306	292	300	64	46

Table 4.17: s_m: scaling factor for the mass of the h type water, s_T: scaling factor for the temperature (translational) of the h type water, $\langle E_h^{pot}\rangle$: average potential energy per h type water, $\langle E_l^{pot}\rangle$: average potential energy per l type water, $\langle T_h^t\rangle$: average translational temperature of the h type water, $\langle T_h^r\rangle$: average rotational temperature of the h type water, $\langle T_l^t\rangle$: average translational temperature of the l type water, $\langle T_l^r\rangle$: average rotational temperature of the l type water, D_h: diffusion coefficient of the h type water, D_l: diffusion coefficient of the l type water.

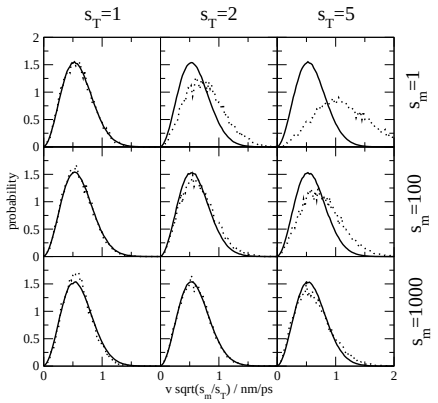

Figure 4.37: Velocity distribution of the water oxygen for 990 h type water molecules (solid line) and 10 l type water molecules (dotted line) for different mass and temperature scaling factors s_m and s_T, scaling both translational and rotational temperature with one bath but separately for 990 h type water molecules and 10 l type water molecules. The velocities for the h type water molecules were multiplied with $(s_m/s_T)^{1/2}$.

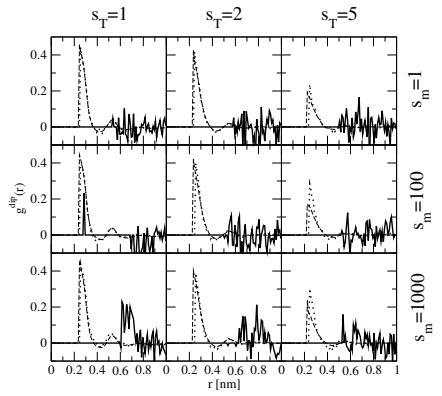

Figure 4.38: Dipole-dipole radial distribution function between 990 *h* type water molecules (solid), 10 *l* type water molecules (dashed), and between 990 *h* and 10 *l* water molecules (dotted) for different mass and temperature scaling factors s_m and s_T, scaling both translational and rotational temperature with one bath but separately for 990 *h* type water molecules and 10 *l* type water molecules.

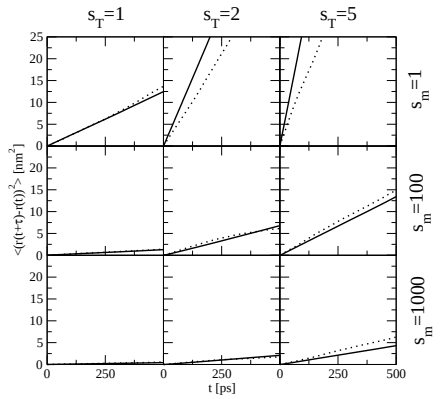

Figure 4.39: Mean square displacment of 990 *h* type water molecules (solid line) and 10 *l* type water molecules (dotted line) for different mass and temperature scaling factors s_m and s_T, scaling both translational and rotational temperature with one bath but separately for 990 *h* type water molecules and 10 *l* type water molecules.

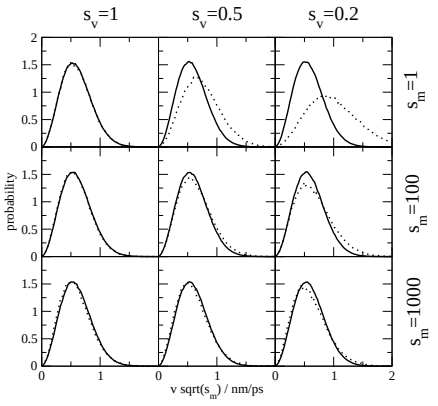

Figure 4.40: Velocity distribution of the water oxygen for 900 *h* type water molecules (solid line) and 100 *l* type water molecules (dotted line) for different mass and force scaling factors s_m and s_V. The velocities for the *h* type water molecules were multiplied with $(s_m)^{1/2}$.

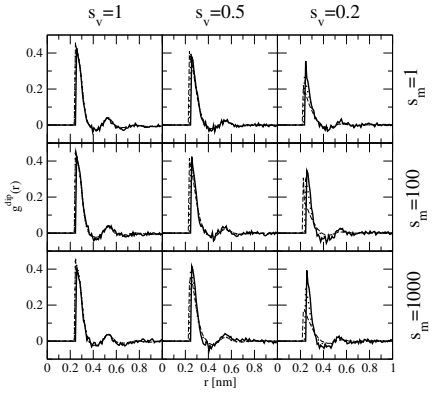

Figure 4.41: Dipole-dipole radial distribution function between 900 *h* type water molecules (solid), 100 *l* type water molecules (dashed), and between 900 *h* and 100 *l* water molecules (dotted) for different mass and force scaling factors s_m and s_V.

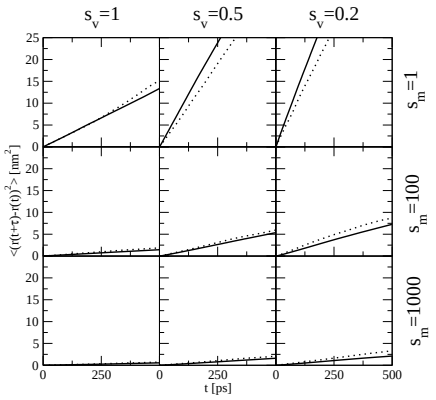

Figure 4.42: Mean square displacment of 900 *h* type water molecules (solid line) and 100 *l* type water molecules (dotted line) for different mass and force scaling factors s_m and s_V.

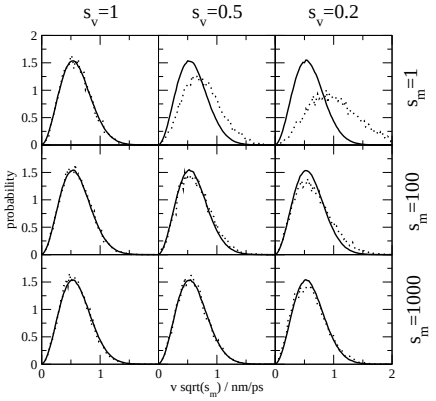

Figure 4.43: Velocity distribution of the water oxygen for 990 *h* type water molecules (solid line) and 10 *l* type water molecules (dotted line) for different mass and force scaling factors s_m and s_V. The velocities for the *h* type water molecules were multiplied with $(s_m)^{1/2}$.

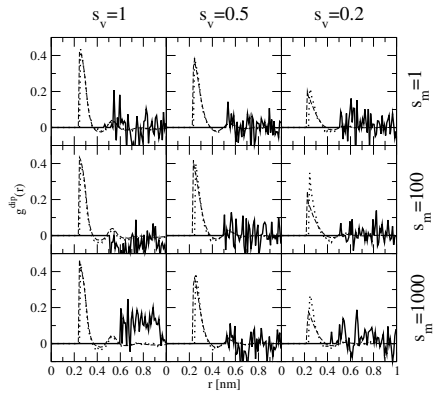

Figure 4.44: Dipole-dipole radial distribution function between 990 *h* type water molecules (solid), 10 *l* type water molecules (dashed), and between 990 *h* and 10 *l* water molecules (dotted) for different mass and force scaling factors s_m and s_V.

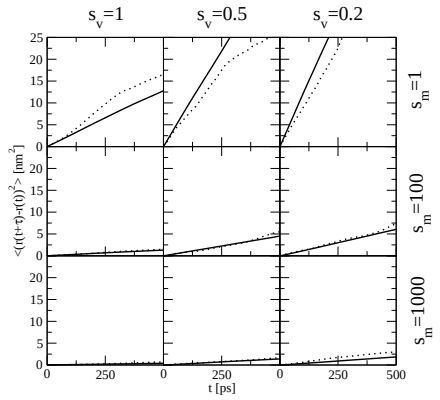

Figure 4.45: Mean square displacment of 990 *h* type water molecules (solid line) and 10 *l* type water molecules (dotted line) for different mass and force scaling factors s_m and s_V.

Hofstadter's Law: It always takes longer than you expect, even when you take into account Hofstadter's Law.

Douglas Hofstadter

Chapter 5

Enhancing the configurational sampling of ions in aqueous solution using adiabatic decoupling with translational temperature scaling

5.1 Summary

Three methods to enhance the configurational sampling of ions in aqueous solution, temperature and Hamiltonian replica exchange and adiabatic decoupling with translational temperature scaling, were compared for a system of $CaSO_4$ in water. It took 11 replicas in the case of temperature replica exchange to make use of a diffusion coefficient that is a factor of 1.5 larger at $350\,\mathrm{K}$ compared to $300\,\mathrm{K}$. 30 replicas were required in the Hamiltonian replica exchange with charge reduction to reach uncharged ions that have a diffusion coefficient that is 2 to 7 times larger than the fully charged ones. The adiabatic decoupling technique with translational temperature scaling yielded a diffusion coefficient that was 15 times larger while keeping the distribution of the water molecules around the ions unaltered with respect to the standard temperature simulation. This result illustrates the efficiency of the adiabatic decoupling technique to enhance configurational sampling.

5.2 Introduction

The role of (bio)molecular simulation has steadily grown over the past few decades and simulation has become a standard technique to describe and study the properties and behaviour of (bio)molecular systems [5, 114–117]. When performing a simulation of a solute, polymer or protein in aqueous solution, four aspects have to be considered [5]: i) the degrees of freedom that one chooses to be explicitly treated, ii) the force field representing the interaction governing the motion along these degrees of freedom, iii)

the method to adequately sample those parts of configurational space that are accessible to the system at the temperature of interest, and iv) the boundary conditions that mimic the interactions of the explicitly treated degrees of freedom with those outside the system. Here we shall focus on how to enhance sampling of configuration space for ions.

While configurational sampling of a liquid is comparably easy, due to the fact that it exists of many identical molecules that may exchange their position in space, and due to the relative simplicity of their interaction function, the conformational sampling of a polymer such as a protein or a carbohydrate is much more difficult due to the connectivity of its covalent topology, which impedes a fast exchange of atom positions, and due to the complexity of its free energy surface. Sampling the configurational distribution of ions around a polymer or protein or DNA is a challenge [186, 187] that lies somewhere in between the two mentioned cases, as their sampling is governed by long range interactions due to their charge, and slow diffusion.

There is a wide range of methods available to enhance sampling [19, 147, 148]. The ones usable for ions are in particular temperature replica exchange [33, 179, 180], Hamiltonian replica exchange [181, 184] with reducing the charge [188] or changing the graining level [182,183], or adiabatic decoupling combined with increased temperature [30,35,158], or with a reduced interaction [36,168,169] or with a reduced force [189] between particular sets of degrees of freedom.

5.3 Methods

5.3.1 Theory

Temperature replica exchange and Hamiltonian replica exchange with reducing the charge were investigated since both an increased temperature and a reduced charge of ions lead to larger diffusion coefficients of the ions. The use of replica exchange with changing grain level was postponed since, to our knowledge, there is not yet a coarse grained model for water or polymers available that is thermodynamically compatible with a fine grained, all atom model for such systems, a necessary condition to obtain meaningful results from a Hamiltonian replica-exchange simulation. The coarse grained simulation would mainly sample configurations that are irrelevant to the fine grained, all atom configurational ensemble.

Third, adiabatic decoupling of ions and water was used together with translational temperature scaling of the ions, since this was shown to improve their diffusion, taken as a measure of extent of sampling, the most.

Replica exchange

In a typical replica exchange molecular dynamic simulation several replicas, i.e. copies of a physical system, are simulated in parallel. These replicas differ either in temperature, for temperature replica exchange T-RE, or in their Hamiltonian, for Hamiltonian replica exchange H-RE.

In Hamiltonian replica exchange we distinguish the different Hamiltonians through a so-called coupling parameter λ, i.e. $H(\lambda)$ or $H_n \equiv H(\lambda_n)$. After a chosen number of time steps, exchanges of the configurations x_m and x_n with adjacent λ values, λ_m and λ_n that define the Hamiltonians H_m and H_n for H-RE, or with adjacent temperatures T_m and T_n for T-RE are attempted using the detailed balance condition,

$$P_{S'} \cdot t(S' \rightarrow S'') = P_{S''} \cdot t(S'' \rightarrow S'), \tag{5.1}$$

where S' denotes the state before the exchange, i.e. the replica with T_m or H_m has configuration x_m and the one with T_n or H_n has configuration x_n, and S'' denotes the state after the exchange, i.e. the replica with T_m or H_m has configuration x_n and the one with T_n or H_n has configuration x_m. P_S denotes the configurational probability of state S and $t(S \rightarrow S')$ the transition probability from state S to state S'. The relative configurational probabilities of states S' and S'' are for H-RE

$$\frac{P_{S''}}{P_{S'}} = \frac{e^{-[H_m(x_n)+H_n(x_m)]/(k_B T)}}{e^{-[H_n(x_n)+H_m(x_m)]/(k_B T)}} \tag{5.2}$$

and for T-RE

$$\frac{P_{S''}}{P_{S'}} = \frac{e^{-H(x_n)/(k_B T_m)} e^{-H(x_m)/(k_B T_n)}}{e^{-H(x_n)/(k_B T_n)} e^{-H(x_m)/(k_B T_m)}}. \tag{5.3}$$

Combining Eq.(5.1) with Eq.(5.2) or Eq.(5.3), one finds for the probability p of exchange of replicas m and n

$$p(m \leftrightarrow n) = t(S' \rightarrow S'') = \begin{cases} 1 & \text{if } \Delta_{mn} \leq 0 \\ P_{S''}/P_{S'} = e^{-\Delta_{mn}} & \text{if } \Delta_{mn} \geq 0 \end{cases} \tag{5.4}$$

$$= min(1, e^{-\Delta_{mn}}) \tag{5.5}$$

with, for H-RE,

$$\Delta_{mn} = \{[H_m(x_n) + H_n(x_m)] - [H_n(x_m) + H_m(x_n)]\}/(k_B T) \tag{5.6}$$

and, for T-RE,

$$\Delta_{mn} = [H(x_n) - H(x_m)](1/(k_B T_n) - 1/(k_B T_m)) \tag{5.7}$$

Adiabatic decoupling with translational temperature scaling

A particular set of N^h degrees of freedom can be adiabatically decoupled from the other N^l degrees of freedom of the system by increasing their mass

$$m^h = s_m m^l \tag{5.8}$$

with $s_m \geq 1$ in such a way that the transfer of kinetic energy between the decoupled degrees of freedom through momentum conserving collisions becomes very small. Here we have denoted the masses of the particles constituting the N^h degrees of freedom collectively by m^h and the ones of the N^l degrees of freedom collectively by m^l. This decoupling is more easily achieved if $N^h \ll N^l$ which applies in the presented case taking ions as h type and water molecules as l type particles. The non-dynamic equilibrium properties of a system remain unaltered by this change of mass of a subset of degrees of freedom.

To enhance the sampling of the N^h h type degrees of freedom their temperature, potential energy or force can be scaled [189]. In this work we focus on the scaling of the translational degrees of freedom of the h type degrees of freedom with a factor s_T such that

$$T^h = s_T T^l \text{ or } \beta^h = \beta^l / s_T. \tag{5.9}$$

The Hamiltonian for this system of N^h h type particles and N^l l type particles is

$$H(p^h, p^l, x^h, x^l) = K^h(p^h) + V^{hh}(x^h) + V^{hl}(x^h, x^l) + K^l(p^l) + V^{ll}(x^l), \tag{5.10}$$

where $K^h(p^h)$ and $K^l(p^l)$ are the kinetic energy of the h type and the l type particles respectively, $V^{hh}(x^h)$ the potential energy between h type particles, $V^{ll}(x^l)$ the potential energy between l type particles and $V^{hl}(x^h, x^l)$ between h and l type particles. In the adiabatically decoupled limit, the partition function of the l type particles is

$$Z^l(x^h; \beta^l) = \int e^{-\beta^l K^l(p^l)} dp^l \int e^{-\beta^l V^{ll}(x^l)} e^{-\beta^l V^{hl}(x^h, x^l)} dx^l. \tag{5.11}$$

and the Hamiltonian of the h type degrees of freedom can therefore be written as

$$H^h(p^h, x^h; \beta^l) = K^h(p^h) + V^{hh}(x^h) - (\beta^l)^{-1} \ln Z^l(x^h; \beta^l). \tag{5.12}$$

The average value of a property $\langle Q(x^h) \rangle_{\beta^l}$ at temperature T^l can be recalculated from $\langle Q \rangle_{\beta^h}$ of $Q(x^h)$ at T^h using the standard unbiasing formula,

$$\langle Q \rangle_{\beta^l} = \frac{\langle Q e^{-(\beta^l - \beta^h)[V^{hh} - (\beta^l)^{-1} \ln Z^l]} \rangle_{\beta^h}}{\langle e^{-(\beta^l - \beta^h)[V^{hh} - (\beta^l)^{-1} \ln Z^l]} \rangle_{\beta^h}}. \tag{5.13}$$

where the ensemble averages are over the h type degrees of freedom, and

$$Z^l(x^h; \beta^l) = \frac{\int e^{-\beta^l K^l(p^l)} dp^l \int e^{-\beta^l V^{ll}(x^l)} dx^l}{\langle e^{+\beta^l V^{hl}(x^h)} \rangle} \tag{5.14}$$

where the ensemble average in the denominator is over the l type degrees of freedom [189]. For the ion-water system considered here, the reweighting factor $e^{-(\beta^l-\beta^h)[V^{hh}-(\beta^l)^{-1}\ln Z^l]}$ for the h type configurations is equal to infinity in numerical precision and therefore no reweighting was applied.

5.3.2 Simulated system and simulation procedure

$CaSO_4$ solvated in water was chosen as a test system for comparing the efficiency of the temperature and Hamiltonian replica exchange and the adiabatic decoupling with translational temperature scaling.

A cubic box with an edge length of $4.082\,\mathrm{nm}$ was filled with 2198 SPC [9] water molecules, resulting in a density of $970\,\mathrm{kg/m^3}$, corresponding to the density of liquid SPC water at $300\,\mathrm{K}$ and $1\,\mathrm{atm}$. [87] To this system 2 Ca^{2+} and 2 SO_4^{2-} were added. The Lennard-Jones parameters used were $\sqrt{C_6} = 0.0317[\mathrm{kJmol^{-1}nm^6}]^{1/2}$ and $\sqrt{C_{12}} = 0.7057\cdot10^{-3}[\mathrm{kJmol^{-1}nm^{12}}]^{1/2}$ for the calcium ion that carries a charge of $+2e$. For the sulfur of the sulfate $\sqrt{C_6} = 0.09992[\mathrm{kJmol^{-1}nm^6}]^{1/2}$ and $\sqrt{C_{12}} = 3.616\cdot10^{-3}[\mathrm{kJmol^{-1}nm^{12}}]^{1/2}$ are the Lennard-Jones parameters and the charge is $+0.54e$. The oxygen has an attractive Lennard-Jones parameter of $\sqrt{C_6} = 0.04756$ $[\mathrm{kJmol^{-1}nm^6}]^{1/2}$ and the repulsive Lennard-Jones parameter is $\sqrt{C_{12}} = 0.8611\cdot 10^{-3}[\mathrm{kJmol^{-1}nm^{12}}]^{1/2}$ for interactions with itself and the sulfur, $\sqrt{C_{12}} = 1.841\cdot 10^{-3}[\mathrm{kJmol^{-1}nm^{12}}]^{1/2}$ for interactions with the oxygen of water, and $\sqrt{C_{12}} = 3.068\cdot 10^{-3}[\mathrm{kJmol^{-1}nm^{12}}]^{1/2}$ for interactions with calcium. The charge the oxygens of the SO_4^{2-} carry is $-0.635e$. The ideal bond length $S-O$ in SO_4^{2-} is $0.15\,\mathrm{nm}$ and the $O-S-O$ angle is 109.5^o with a force constant of 0.1404877 $\mathrm{kJmol^{-1}}$. This is according to the GROMOS force field 53A6 [7].Periodic boundary conditions were applied. Molecular dynamics simulations were performed with the GROMOS05 simulation software package [133], modified to incorporate adiabatic decoupling. The geometry of the water molecules was constrained by applying the SHAKE algorithm [23] with a relative geometric tolerance of 10^{-4} on the OH bond lengths, on the intramolecular HH distance of the water molecules, and on the SO bond length of the SO_4^{2-} ions. The nonbonded van der Waals and electrostatic interactions were calculated using triple-range cutoff radii of $0.8/1.4\,\mathrm{nm}$. The short range interactions were calculated every time step by updating the molecular pair list for distances smaller than the first cutoff radius of $0.8\,\mathrm{nm}$. For the intermediate range of distances between 0.8 and $1.4\,\mathrm{nm}$ the pairlist was only updated every fifth timestep and at the same time the interaction was calculated and kept unchanged between these updates. The long range electrostatic interactions beyond the outer cutoff of $1.4\,\mathrm{nm}$ were represented by a reaction field [62, 185] with ϵ_{RF}=78.5, the dielectric permittivity of water at $300\,\mathrm{K}$ and $1\,\mathrm{atm}$. The equations of motion were integrated using the leap frog algorithm with a time step

of 2 fs. The velocities of the atoms at the beginning of the simulation were assigned from a Maxwell distribution at 300 K.

For the temperature replica exchange 11 replicas with temperatures raised in steps of 5 K from 300 K to 350 K were used. The Hamiltonian replica exchange simulation comprised 30 λ values (0, 0.01, 0.025, 0.05, 0.08, 0.12, 0.16, 0.2, 0.24, 0.28, 0.32, 0.36, 0.4, 0.4.4, 0.48, 0.52, 0.56, 0.6, 0.64, 0.68, 0.72, 0.76, 0.8,0.84, 0.88, 0.915, 0.945, 0.975, 0.99, 1) where $\lambda = 0$ represents a system where the ions are fully charged and $\lambda = 1$ one with ions without charge. The charges of the ions were scaled linearly with $(1 - \lambda)$. No softness was used for this transformation [190]. For both RE simulations exchanges based on a Metropolis Monte Carlo criterion were attempted every 2 ps for 1 ns. These simulations were performed at constant volume. The temperatures were weakly coupled [29] to a bath with a relaxation time of 0.1 ps.

The adiabatic decoupling was investigated with s_m values of 1, 100, 200, 500 and 1000. The translational temperature scaling factors s_T used for the ions were 1, 2, 3 and 5. These simulations were performed at constant pressure, incorporating a correction for s_T in the contribution from the kinetic energy of the h type degrees of freedom to the pressure [189], by weakly coupling [29] it to a bath of 1 atm with a relaxation time of 0.5 ps. The isothermal compressibility was set to $7.513{\cdot}10^{-4}(kJmol^{-1}nm^{-3})^{-1}$. The ions and water molecules were separately weakly coupled [29] to two baths of 300 K with a relaxation time of 0.1 ps. During the runs, configurations of the system were saved every 2 ps. The various properties were taken from a 10 ns simulation that followed a 100 ps equilibration period.

Additional long simulations at 325 K and 350 K (with $\lambda = 0$) and $\lambda = 0.01$, 0.1,, 0.05, 1 (at 300 K) were performed which were used as reference simulations to which the replica exchange or adiabatically decoupled simulations could be compared. These simulations were performed at constant pressure by weakly coupling [29] it to a bath of 1 atm with a relaxation time of 0.5 ps. The isothermal compressibility was set to $7.513{\cdot}10^{-4}(kJmol^{-1}nm^{-3})^{-1}$. They were also weakly coupled [29] to a bath of the mentioned temperature with a relaxation time of 0.1 ps. During the runs, configurations of the system were saved every 2 ps. The various properties were taken from a 10 ns simulation that followed a 100 ps equilibration period.

5.3.3 Analysis

The configurational distribution of the ions in water in the different systems was characterised by calculating their radial distribution functions. The diffusion coefficient was calculated as a measure of the rate of sampling.

Radial distribution function g(r)

The pair distribution function $g(r)$ represents the probability of finding another atom at a distance r from a given atom, relative to the probability expected for a completely uniform distribution at the same density. It can be calculated by a simple histogram summation in radial shells over all molecules in the system using

$$g_{\alpha\beta}(r) = \frac{n(r)}{4\pi r^2 \Delta r \rho} \tag{5.15}$$

where α is the atom type Ca^{2+} or the S of SO_4^{2-}, β is the oxygen of the water, $n(r)$ the number of atoms of type β around an atom of type α at distances between r and $r + \Delta r$, with $\Delta r = 0.01 nm$, and ρ the density of the liquid.

To approximate the difference between two radial distribution functions, the radial distribution difference, was calculated as

$$\Delta g_{\alpha\beta} = \frac{\int_0^{r_{max}} |g_{\alpha\beta}(r) - g_{\alpha\beta}^{ref}(r)| dr}{\int_0^{r_{max}} g_{\alpha\beta}^{ref}(r) dr} \tag{5.16}$$

where r_{max} is the cut off for the radial distribution function, which was set to $1\,\mathrm{nm}$. As reference radial distribution function $g_{\alpha\beta}^{ref}(r)$ the one of the system with mass scaling factor $s_m = 1$ and temperature scaling factor $s_T = 1$ was taken.

Self-diffusion coefficient D

The diffusion coefficient D is obtained from the long-time limit of the mean-square displacement

$$MSD(t) = < (\mathbf{r}(\tau + t) - \mathbf{r}(\tau))^2 >_{\tau, molecules} \tag{5.17}$$

according to the Einstein relation [66],

$$D = \lim_{t \to \infty} \frac{< (\mathbf{r}(\tau + t) - \mathbf{r}(\tau))^2 >_{\tau, molecules}}{6t} \tag{5.18}$$

where $\mathbf{r}(t)$ corresponds to the position vector of the centre of mass of a molecule at time t, and the averaging is performed over both time and molecules. Only the time window between 0 and $250\,\mathrm{ps}$ was used for the calculation of the diffusion constant in order to achieve good statistics.

5.4 Results and discussion

The three approaches to enhance the configurational sampling of ions in aqueous solution– temperature replica exchange, Hamiltonian replica exchange with reducing

the charges and adiabatic decoupling combined with scaling of the translational temperature – were compared.

5.4.1 Temperature replica exchange

T [K]	$D_{Ca^{2+}}$ $[10^{-9}m^2s^{-1}]$	$D_{SO_4^{2-}}$ $[10^{-9}m^2s^{-1}]$	$\Delta g_{Ca^{2+}OW} \cdot 100$	$\Delta g_{SO_4^{2-}OW} \cdot 100$
300	1.06	1.23	0.00	0.00
325	1.17	1.31	1.82	1.39
350	1.63	1.84	4.54	3.34

Table 5.1: Configurational and dynamic properties of the ions in aqueous solution from standard MD simulations at 3 temperatures. T: temperature, D: diffusion coefficient, Δg: radial distribution difference, see Eq.(5.16).

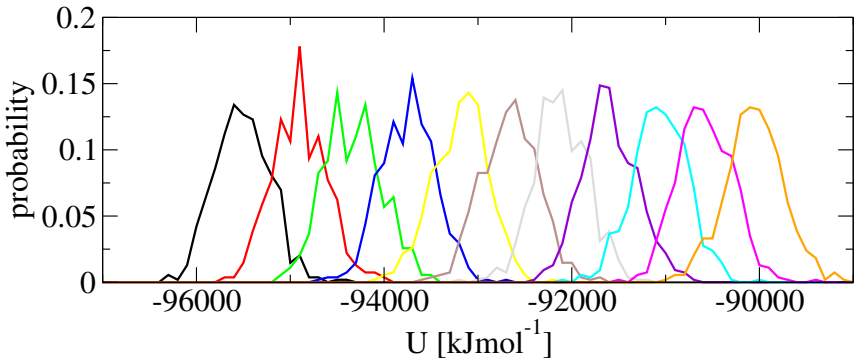

Figure 5.1: Distribution of the potential energy per replica for the 11 replicas of the temperature replica exchange simulation, starting from 300 K (left) and going to 350 K (right) in steps of 5 K.

Temperature replica exchange was considered since an increase in temperature from 300 K to 350 K leads to an increase of the diffusion coefficient by a factor 1.5 (Table 5.1). However, to span this temperature interval, at least 11 replicas are required for a sufficient number of replica exchanges. The replicas have an exchange probability between 0.15 and 0.25. The overlap of their potential energy distributions is shown in Fig. 5.1. For 1 ns of replica exchange simulation with exchange attempts every 2 ps only 3 contributions from the replica at $T = 350$ K are observed at $T = 300$ K (Fig. 5.2). The gain in diffusion by including higher temperature replicas is lost by the necessity to have a sufficient number of replicas. This problem is even more pronounced for bigger systems since the number of replicas necessary is proportional to the square root of

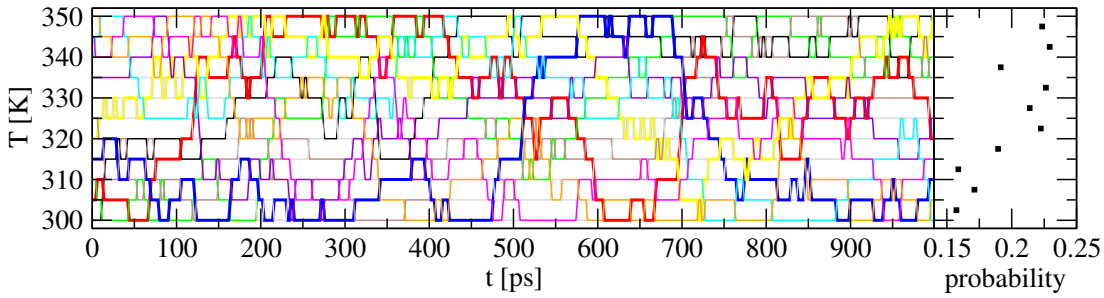

Figure 5.2: Left-hand panel: switches between the 11 replicas in the temperature replica exchange simulation as a function of time. Highlighted are the replicas that managed to go over the whole range of temperatures more than once. Right-hand panel: average acceptance ratio of switches between two adjacent replicas.

the system size [191]. Simply simulating the system at a higher temperature is not an option, since the structure, see the radial distribution difference in Table 5.1, changes significantly if the temperature of the whole system is increased.

5.4.2 Hamiltonian replica exchange using charge reduction

λ	$D_{Ca^{2+}}$ $[10^{-9}m^2s^{-1}]$	$D_{SO_4^{2-}}$ $[10^{-9}m^2s^{-1}]$	$\Delta g_{Ca^{2+}OW} \cdot 100$	$\Delta g_{SO_4^{2-}OW} \cdot 100$
0	1.06	1.23	0.00	0.00
0.01	1.01	1.00	16.44	12.21
0.1	1.55	1.54	41.56	26.06
0.5	2.02	1.88	54.27	31.13
1	7.85	2.64	60.54	31.67

Table 5.2: Configurational and dynamic properties of the ions in aqueous solution from standard MD simulations at 5 different values of the charge reduction factor λ where the charge of the ions is scaled with $(1-\lambda)$. λ: charge scaling factor, D: diffusion coefficient, Δg: radial distribution difference, see Eq.(5.16).

As a reduction of the charge of the ions leads to a considerable increase in the diffusion coefficient by a factor 2 to 7 (Table 5.2), a Hamiltonian replica exchange which reduces the ionic charge was performed. To obtain an exchange probability between 0.07 and 0.4 30 replicas were required. In the simulated time of 1 ns none of the 30 replicas manages to cross the charge range completely (Fig. 5.3). As in the case of the temperature replica exchange the gain in diffusion from replicas with λ values close to 1 is lost by the necessity to have a sufficient number of replicas.

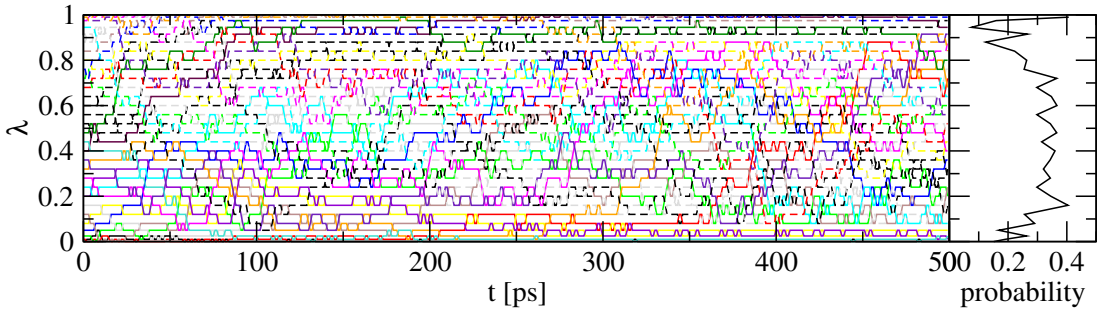

Figure 5.3: Left-hand panel: switches between the 30 replicas in the charge reducing Hamiltonian replica exchange simulation as function of time. Replicas with a starting $\lambda > 0.5$ are dashed. Note that no replica manages to visit both $\lambda = 0$ and $\lambda = 1$. Right-hand panel: average acceptance ratio of switches between two adjacent replicas.

Since the radial distribution differences (Table 5.2) are also significant for a simulation with reduced ionic charge, a simulation with a reduced ionic charge does not offer a viable alternative to replica exchange.

5.4.3 Adiabatic decoupling with translational temperature scaling

To get an impression how much the structure of the water molecules around the adiabatically decoupled ions that have a higher temperature has changed, the radial distribution difference $\Delta g_{\alpha\beta}$ is shown in Table 5.3. Since scaling the mass with s_m without scaling the temperature should not influence the radial distribution function, the results for $s_T = 1$ can be taken as a reference value for contributions of the noise in the radial distribution function. The maximum Δg_{CaOW} is 1.64 observed for $s_m = 1000$ and the maximum for Δg_{SOW} is 1.23 found for $s_m = 100$.

For $s_m = 1$, i.e. without adiabatic decoupling, $\Delta g_{\alpha\beta}$ grows fast with increasing s_T. All systems with $s_m > 1$, except two, $s_m = 1000$ with $s_T = 2$ or 5, display Δg_{CaOW} values within the variation of $s_T = 1$. Δg_{SOW} seems to be a bit more sensitive to raising the temperature: all systems with $s_m > 1$ using $s_T = 3$ or 5 yield Δg_{SOW} values that are somewhat larger than the variation 1.23 found for $s_T = 1$. However, Δg_{SOW} for $s_T = 5$ nicely decreases for increasing s_m, i.e. for better adiabatic decoupling.

The diffusion coefficients in Table 5.3 show a remarkable increase for the adiabatically decoupled systems, $s_m > 1$, with an increased temperature, $s_T > 1$. The diffusion increases with increasing s_T and decreases with increasing s_m. This implies that an optimum is to be found between sufficient decoupling and not too large ionic mass. For $s_m > 1$ and $s_T > 1$ the time dependence of the mean square displacement (Table 5.4) is not linear anymore. This implies that the values for D in Table 5.3 would change if the considered time window for obtaining the D values from the mean-square displacement,

s_m	s_T	$D_{Ca^{2+}}$ $[10^{-9}m^2s^{-1}]$	$D_{SO_4^{2-}}$ $[10^{-9}m^2s^{-1}]$	$\Delta g_{Ca^{2+}OW} \cdot 100$	$\Delta g_{SO_4^{2-}OW} \cdot 100$
1	1	1.06	1.23	0.00	0.00
	2	2.01	2.71	4.20	2.93
	3	4.11	6.03	8.59	6.71
	5	24.39	66.98	10.10	8.23
100	1	1.44	1.80	1.21	1.23
	2	13.22	16.89	1.25	1.20
	3	25.87	35.25	1.55	1.60
	5	51.64	74.56	1.60	2.11
200	1	1.23	1.46	1.34	1.11
	2	7.84	9.66	1.07	1.24
	3	15.27	18.87	1.45	1.40
	5	28.95	38.25	1.41	2.06
500	1	1.13	1.14	1.09	1.14
	2	4.22	4.48	1.18	1.18
	3	7.00	8.26	1.58	1.24
	5	13.19	15.83	1.47	1.97
1000	1	0.83	0.74	1.64	1.05
	2	2.33	2.51	1.75	1.19
	3	4.04	4.29	1.40	1.30
	5	7.43	7.96	1.70	1.54

Table 5.3: Configurational and dynamic properties of the ions in aqueous solution from differently strong (s_m) adiabatically decoupled simulations in which the temperature of the ions is increased by different amounts (s_T). s_m: mass scaling factor, s_T: temperature scaling factor, D: diffusion coefficient, Δg: radial distribution difference, see Eq.(5.16).

Eq.(5.18), would be varied. Thus the D values are mainly given for ease of comparison.

5.5 Conclusions

Adiabatic decoupling with translational temperature scaling offers a method to considerably increase the diffusion of particular degrees of freedom, e.g. ions in aqueous solution, without perturbing the configurational distribution of the surrounding degrees of freedom, i.e. water, significantly. An increase of a factor of 15 in the diffusion coefficient was observed with a mass scaling factor $s_m = 100$ and a temperature scaling factor $s_T = 2$ which represents a system for which almost no difference in the radial distribution function compared to the system at $s_m = 1$, $s_T = 1$ was observed. To achieve this only one simulation had to be performed compared to the 11 replicas required in

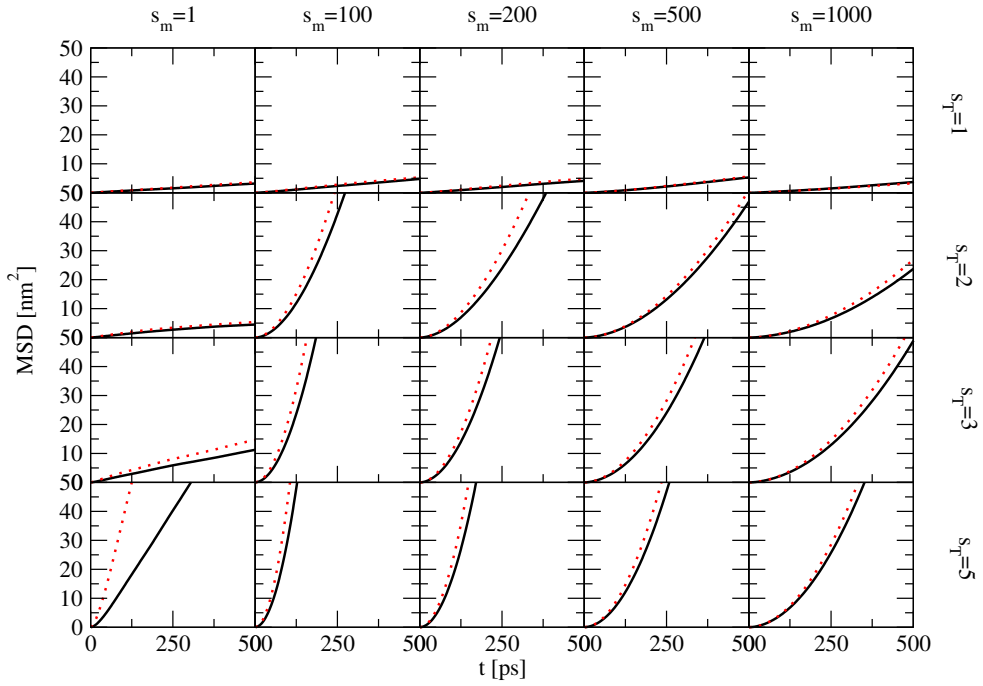

Figure 5.4: Mean-square dispacement of the Ca^{2+} (black, solid) and the SO_4^{2-} (red, dotted) ions in aqueous solution for different mass and temperature scaling factors s_m and s_T in the adiabatically decoupled simulations with different ionic temperatures.

temperature replica exchange and 30 replicas in Hamiltonian replica exchange with charge reduction. Adiabatic decoupling with temperature scaling of the ionic degrees of freedom is, therefore, a very promising method to enhance the configurational sampling of ions around a solute in aqueous solution.

Everything works on something, but
nothing works on everything.

Robert Sheridan

Chapter 6

A method for sampling the internal degrees of freedom of a flexible solute molecule based on adiabatic decoupling and temperature or force scaling

6.1 Summary

Simulation of the folding equilibrium of a polypeptide in solution is a computational challenge. Standard molecular dynamics (MD) simulations of such systems cover hundreds of nanoseconds, which is barely sufficient to obtain converged ensemble averages for properties that depend both on folded and unfolded peptide conformations. If one is not interested in dynamical properties of the solute, techniques to enhance the conformational sampling can be used to obtain the equilibrium properties more efficiently. Here the effect on particular equilibrium properties at 298 K of adiabatically decoupling the motion of a β-hepta-peptide from the motion of the solvent and subsequently up-scaling its temperature or down-scaling the forces acting on it is investigated. The ensemble averages and rate of convergence are compared to those for standard MD simulations at two different temperatures and a simulation in which the temperature of the solute is increased to 340 K while keeping the solvent at 298 K. Adiabatic decoupling with a solute mass scaling factor $s_m = 100$ and a temperature scaling factor of $s_T = 1.1$ seems to slightly increase the convergence of several properties such as enthalpy of folding, NMR NOE atom-atom distances and ^{3}J-couplings compared to a standard MD simulation at 298 K. Convergence is still slower than that observed at 340 K. The system with a temperature of 340 K for the solute and 298 K for the solvent without scaling of the mass converges fastest. Using a force scaling factor $s_V = 0.909$ perturbs the system too much and leads to a destabilisation of the folded structure. The sampling efficiency and possible distortive effects on the configurational distribution of the solute

degrees of freedom due to adiabatic decoupling and temperature or force scaling are also analysed for a simpler model, a dichloroethane molecule in water. It appears that an up-scaling of the mass of the solute reduces the sampling more than the subsequent up-scaling of the temperature or down-scaling of the force enhances it. This means that adiabatic decoupling the solute degrees of freedom from the solvent ones followed by an up-scaling of temperature of down-scaling of the forces does not lead to significantly enhanced sampling of the folding equilibrium.

6.2 Introduction

Proteins play an important role in crucial biological processes, in fact it is their sequence of amino acids that is stored in the DNA of every living creature and passed on from generation to generation. Their functionality highly depends on their three-dimensional structure or fold, or rather on the ensemble of their structures since proteins are not static objects but to a certain extent flexible. The way they fold is mainly determined by their amino-acid residue sequence, even though the thermodynamic conditions such as temperature, pressure, pH or type of solvent play an important role as well. It is therefore one of the holy grails of molecular biology to be able to predict the folding of a protein or peptide from its sequence given the thermodynamic and environmental conditions [192–199]. Finding the folding equilibrium of a protein using molecular dynamics (MD) simulations is still very expensive and can only be done for proteins of short chain length [200]. Since already small β-peptides as short as four β-amino acid residues have shown to adopt well-ordered secondary structures [201–203], their folding equilibria are easier to study using MD than those of proteins or polypeptides composed of α-amino acids where much longer sequences are to be considered. β-peptides can be studied solvated in water or in methanol. The latter solvent allows for fast simulation and experimentally for fast folding. As promising candidates for pharmaceutical application [204–207], a wide range of different β-amino acid sequences have been synthesised and studied experimentally, offering the opportunity to compare their measured properties with the results from simulation.

Since the GROMOS force field is usually able to reproduce in simulation the folding equilibria of a variety of secondary structures of β-peptides such as 3_{14}-helices [208, 209], 2.5_{12}-helices [2], $2.7_{10/12}$-helices [210, 211] or hairpins [212], these β-peptides can be used as test case for sampling enhancement techniques: the result from an enhanced sampling MD simulation can be compared to that of a standard MD simulation and in this way validated. Therefore, a β-hepta-peptide [208, 209, 213] which is known to fold into a 3_{14}-helix was chosen as a test case to study the sampling efficiency of adiabatically decoupling of the peptide degrees of freedom from the solvent ones in combination with temperature or force scaling for the peptide degrees of freedom. For this β-hepta-peptide 41 NOE distance bounds and 21 ^3J-couplings are available from NMR experiments [208, 213]. These data were reproduced previously in

standard MD simulations [208, 209]. The motion along the peptide degrees of freedom is adiabatically decoupled from that of the solvent degrees of freedom by increasing the masses of the peptide atoms by a factor $s_m = 100$ and the sampling of the peptide degrees of freedom is subsequently enhanced by up-scaling their temperature with a factor $s_T = 1.1$ or down-scaling the forces acting on them by a factor $s_V = 0.909$.

To fully understand the effect of adiabatic decoupling with subsequent temperature or force scaling on the internal degrees of freedom of the adiabatically decoupled solute, dichloroethane in water was investigated using different values for the parameters s_m and s_T or s_V of the method of enhanced sampling based on adiabatic decoupling.

6.3 Methods

6.3.1 Theory

For the adiabatic decoupling with subsequent temperature or force scaling the motion of a particular set of N^h h type atoms is adiabatically decoupled from the motion of the other N^l l type atoms in the system by increasing their mass

$$m^h = s_m m^l \tag{6.1}$$

with $s_m \gg 1$ in such a way that the transfer of kinetic energy between the decoupled degrees of freedom through momentum conserving collisions becomes very small. s_m has to be chosen in such a way that the lightest mass of the N^h h type atoms, usually a hydrogen, is significantly larger than the heaviest mass of the N^l l type atoms. It is therefore recommended to choose $s_m \geq 100$. Here we have denoted the masses of the particles constituting the $3N^h$ degrees of freedom collectively by m^h and the ones of the $3N^l$ degrees of freedom collectively by m^l. For $N^h \ll N^l$ the decoupling is easy to achieve [189]. This condition is fulfilled in the present case where the peptide degrees of freedom are h type and the solvent ones are l type. The non-dynamic equilibrium properties of the system remain unaltered by this change of mass of a subset of degrees of freedom.

To enhance the sampling of the $3N^h$ h type degrees of freedom their temperature, potential energy or force can be scaled [189]. Using temperature scaling, the temperature of the h type degrees of freedom is set to

$$T^h = s_T T^l \text{ or } \beta^h = \beta^l / s_T. \tag{6.2}$$

where $\beta = (k_B T)^{-1}$ and k_B is Boltzmann's constant.

The Hamiltonian for a system of N^h h type particles and N^l l type particles is then

$$H(p^h, p^l, x^h, x^l) = K^h(p^h) + V^{hh}(x^h) + V^{hl}(x^h, x^l) + K^l(p^l) + V^{ll}(x^l), \tag{6.3}$$

where $K^h(p^h)$ and $K^l(p^l)$ are the kinetic energy of the h type and the l type particles respectively, $V^{hh}(x^h)$ the potential energy between h type particles, $V^{ll}(x^l)$ the potential energy between l type particles and $V^{hl}(x^h, x^l)$ between h and l type particles. In the adiabatically decoupled limit, the partition function of the l type particles is

$$Z^l(x^h;\beta^l) = \int e^{-\beta^l K^l(p^l)} dp^l \int e^{-\beta^l V^{ll}(x^l)} e^{-\beta^l V^{hl}(x^h,x^l)} dx^l, \tag{6.4}$$

and the Hamiltonian of the h type degrees of freedom can then be written as

$$H^h(p^h, x^h;\beta^l) = K^h(p^h) + V^{hh}(x^h) - (\beta^l)^{-1} \ln Z^l(x^h;\beta^l). \tag{6.5}$$

The average value $\langle Q\rangle_{\beta^l}$ of a property $Q(x^h)$ at temperature T^l can be recalculated from $\langle Q\rangle_{\beta^h}$ of $Q(x^h)$ at T^h using the standard unbiasing formula,

$$\langle Q\rangle_{\beta^l} = \frac{\langle Q e^{-(\beta^l-\beta^h)[V^{hh}-(\beta^l)^{-1}\ln Z^l]}\rangle_{\beta^h}}{\langle e^{-(\beta^l-\beta^h)[V^{hh}-(\beta^l)^{-1}\ln Z^l]}\rangle_{\beta^h}}, \tag{6.6}$$

where the ensemble averages are over the h type degrees of freedom, and

$$Z^l(x^h;\beta^l) = \frac{\int e^{-\beta^l K^l(p^l)} dp^l \int e^{-\beta^l V^{ll}(x^l)} dx^l}{\langle e^{+\beta^l V^{hl}(x^h)}\rangle} \tag{6.7}$$

where the ensemble average in the denominator is over the l type degrees of freedom [189]. For the system considered here, the reweighting factor $e^{-(\beta^l-\beta^h)[V^{hh}-(\beta^l)^{-1}\ln Z^l]}$ for the h type configurations is equal to infinity in numerical precision and therefore no reweighting was applied.

Using force scaling, the force acting on a h type degree of freedom,

$$\mathbf{f}_h = (\mathbf{f}_{hh} + \mathbf{f}_{hl}) s_V \tag{6.8}$$

where $\mathbf{f}_{hh}$ is force from the other h type degrees of freedom and $\mathbf{f}_{hl}$ is the force from the l type degrees of freedom, is scaled by a factor s_V.

The Hamiltonian, in analogy to Eq. 6.5, can be written as

$$H^h(p^h, x^h) = K^h(p^h) + s_V V^{hh}(x^h) - s_V(\beta^l)^{-1} \ln Z^l(x^h;\beta^l). \tag{6.9}$$

with $Z^l(x^h;\beta^l)$ defined as in Eq.(7.7).

The average value $\langle Q\rangle_{\beta^l}$of a property $Q(x^h)$ at β^l can be recalculated from $\langle Q\rangle_{\beta^h}$ of $Q(x^h)$ at $\beta^h = \beta^l s_V$ using the standard unbiasing formula Eq.(6.6). For the system considered here, the reweighting factor $e^{-(\beta^l-\beta^h)[V^{hh}-(\beta^l)^{-1}\ln Z^l]}$ for the h type configurations is equal to infinity in numerical precision and therefore no reweighting was applied.

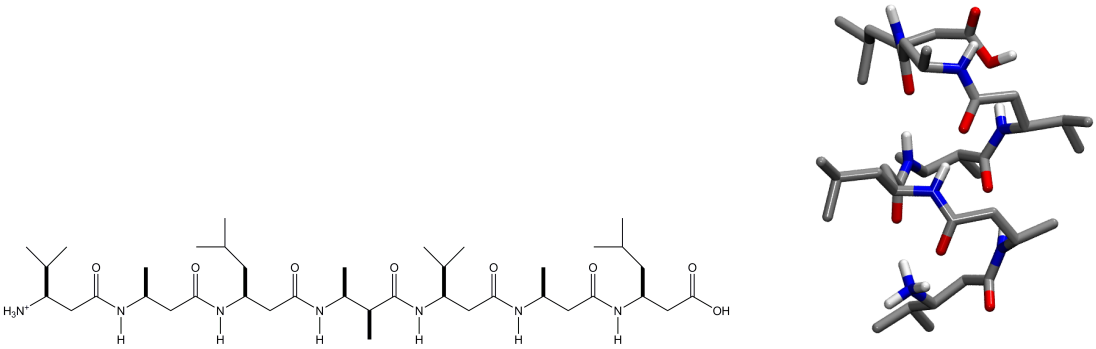

Figure 6.1: Left panel: Structure formula of the β-hepta peptide (H_2^+-β^3-HVal-β^3-HAla-β^3-HLeu-(S,S)-β^3-HAla(α Me)-β^3-HVal-β^3-HAla-β^3-HLeu-OH) studied. Right panel: 3_{14}-helical model structure derived from NMR data at 298 K [213].

6.3.2 Simulation Setup

For the peptide, a cubic box with an edge length of 4.189 nm was filled with 1090 methanol molecules and one H_2^+-β^3-HVal-β^3-HAla-β^3-HLeu-(S,S)-β^3-HAla(α Me)-β^3-HVal-β^3-HAla-β^3-HLeu-OH peptide (6.1). The GROMOS 45A3 [214] force field was used. The geometry of the methanol molecules was constrained by applying the SHAKE algorithm [23] with a relative geometric tolerance of 10^{-4} on the bond lengths and on the intramolecular hydrogen-carbon distance. The bond lengths of the β-hepta-peptide were constrained using the SHAKE algorithm as well.

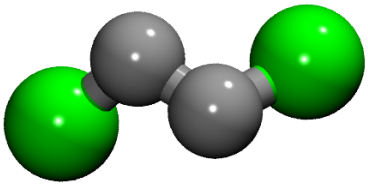

Figure 6.2: Dichloroethane molecule where a green sphere represent a chlorine and the gray sphere a united atom consisting of a carbon atom plus 2 hydrogen atoms.

For the dichloroethane (6.2), a cubic box with edge length of 3.143 nm was filled with 1000 SPC [9] water molecules and one $Cl - CH_2 - CH_2 - Cl$ molecule with the force field parameters described in [215] including a dihedral-angle potential energy function as in [216]. The geometry of the water molecules was constrained by applying the SHAKE algorithm [23] with a relative geometric tolerance of 10^{-4} on the OH bond lengths and on the intramolecular HH distance. The bond lengths of the dichloroethane were constrained using the SHAKE algorithm as well.

All molecular dynamics simulations were performed with the GROMOS05 simulation

software package [133], modified to incorporate adiabatic decoupling. Periodic boundary conditions were applied.

The nonbonded van der Waals and electrostatic interactions were calculated using triple-range cutoff radii of $0.8/1.4\,\mathrm{nm}$. The short range interactions were calculated every time step by updating the molecular pair list for distances smaller than the first cutoff radius of $0.8\,\mathrm{nm}$. For the intermediate range of distances between 0.8 and $1.4\,\mathrm{nm}$ the pairlist was only updated every fifth timestep and at the same time the interaction was calculated and kept unchanged between these updates. The long range electrostatic interactions beyond the outer cutoff of $1.4\,\mathrm{nm}$ were represented by a reaction field [62,185] with ϵ_{RF}=17.7 for methanol and ϵ_{RF}=78.5 for water. The equations of motion were integrated using the leap frog algorithm with a time step of $2\,\mathrm{fs}$. The velocities of the atoms at the beginning of the simulation were assigned from a Maxwell distribution at $60\,\mathrm{K}$ and gradually raised to the simulation temperature.

All simulations were performed at constant pressure by weakly coupling [29] the system to a bath of 1 atm with a relaxation time of 0.5 ps. The isothermal compressibility was set to $4.575{\cdot}10^{-4}(kJmol^{-1}nm^{-3})^{-1}$ The β-hepta-peptide or dichloroethane and the solvent were separately weakly coupled [29] to two baths of the indicated temperature with a relaxation time of $0.1\,\mathrm{ps}$.

Adiabatic decoupling MD was investigated with a s_m value of 100 for the masses of the peptide atoms. The temperature scaling factor s_T used for the peptide was 1.1 and, alternatively, the force scaling factor s_V was 0.909. For comparison two standard MD simulations, at $298\,\mathrm{K}$ and $340\,\mathrm{K}$, without adiabatic decoupling and a simulation with $s_m = 1$ and $s_T = 1.14$ were performed as well.

For the dichloroethane adiabatic decoupling MD was performed varying $s_m = 1,\ 100,\ 1000$ and $s_x = 1,\ 2,\ 5$ where s_x represents s_T or s_v^{-1} depending on the scaling chosen.

During the runs, configurations of the system were saved every $5\,\mathrm{ps}$ for the peptide and every $2\,\mathrm{ps}$ for the dichloroethane respectively. The various properties were taken from $100\,\mathrm{ns}$ simulations that followed a $100\,\mathrm{ps}$ equilibration period for the peptide, and from $50\,\mathrm{ns}$ simulations that followed a $100\,\mathrm{ps}$ equilibration period for the dichloroethane molecule.

6.3.3 Analysis

To investigate whether the simulations sample the folded and unfolded conformations of the β-hepta-peptide, the time series and the distribution of the backbone atom-positional root-mean-square deviation from the initial 3_{14}-helical structure were calculated. The time series of the NOE distance bound violations and the deviation of the ^{3}J-coupling values from the measured ones were calculated to evaluate the convergence of these properties as a function of time, and the average values of these

quantities (r^{-6} averaging for the NOE distances) to evaluate their agreement with the experimental data. Combined conformational clustering analysis of pairs of MD trajectories was performed to see how similar the conformations are that are sampled in the different simulations. For the dichloroethane the distribution of the dihedral angle was calculated to evaluate the effect of adiabatic decoupling with temperature or force scaling on the configurational distribution of this molecule and the transition probability between the different dihedral angle conformations to evaluate the effect on the rate of sampling.

Root-mean-square deviation from the 3_{14}-helical fold

The root-mean-square deviation (RMSD) measures the difference between a reference structure and the structure at time t using

$$RMSD(t) = \left[\frac{1}{N_a} \sum_{i=1}^{N_a} (\mathbf{r}_i(t) - \mathbf{r}_i^{ref})^2 \right]^{1/2} \tag{6.10}$$

where N_a is the number of atoms considered, $\mathbf{r}_i(t)$ is the position of atom i at time t and $\mathbf{r}_i^{ref}$ is the position of atom i in the reference structure. Each structure was first superimposed using the backbone (N, C_β, C_α, C) atoms of the second to the second last residue of the peptide and the RMSD was calculated for these atoms.

Atom-atom distances from NOE intensities

Average inter-proton distances were calculated from the simulation as $< r^{-6} >^{-1/6}$, where r is the inter-proton distance for a given structure. The hydrogen position had to be calculated by defining virtual (for CH_1 or CH_2 united atoms) or pseudo (for CH_3 or CH_2 united atoms) atoms, since the aliphatic hydrogens are not explicitly treated in the GROMOS 45A3 force field. Average NOE distance bound violations were calculated as $< r^{-6} >^{-1/6} - r_{exp}$ where r_{exp} is one of the 41 inter-proton distance bounds derived from the intensities observed in the ROESY NMR spectrum measured at 298 K [213]. Since r_{exp} is only an upper-bound, negative values of $< r^{-6} >^{-1/6} - r_{exp}$ are not considered a violation. The 41 NOE proton-proton pairs are specified in Table S1 of [217].

^{3}J-coupling constants

The ^3J-couplings were calculated using the Karplus relation

$$^3J_{H,H}(\theta) = A cos^2\theta + B cos\theta + C \tag{6.11}$$

where the parameters (A, B, C) were chosen equal to (6.4 Hz, −1.4 Hz, 1.9 Hz) for the calculation of ${}^3J_{H_N,H_C}$ and (9.5 Hz, −1.6 Hz, 1.8 Hz) for the calculation of ${}^3J_{H_C,H_C}$. The 21 ${}^3J$-couplings are specified in Table S2 of [217].

Folding free enthalpy

The folding free enthalpy was calculated using

$$\Delta G_{folding} = -k_B T \ln(P_{C_f}/P_{C_u}) \tag{6.12}$$

where P_{C_f} and P_{C_u} are the relative probabilities of finding the system in folded and unfolded conformations, respectively. A structure is counted as folded if the backbone (residue 2 to 6) atom-positional RMSD from the 3_{14}-helical structure is smaller than 0.1 nm [209, 217].

Conformational clustering

Combined conformational clustering of pairs of ensembles or MD trajectories was performed using different similarity measures.

Two dihedral angle clusterings were performed, one using the root-mean-square deviation for 15 dihedral angles, the ϕ, θ, and ψ torsional angles of the backbone from the second to the sixth residue, and another using the two dihedral angles defined by the C_β atoms from the second to the sixth residue. As a criterion of similarity of two structures a maximum RMSD of 25^o was used.

The clustering algorithm was the same for both types of clustering. The number of neighbours, i.e. structures satisfying the similarity criterion, for each of the structures in the initial pool of structures was determined. As the center of the first cluster the structure with the largest number of neighbours in terms of RMSD was taken and all structures belonging to this cluster were removed from the pool. For each of the remaining structures the number of neighbours was again computed. The structure with the most neighbours became the centre of the second cluster of structures. Structures belonging to this second cluster were then also removed from the pool. This process was iterated until all structures were assigned to a cluster.

Transition probability

The transition probability for transitions of the single dihedral angle of dichloroethane between its three energetic minima was calculated as follows. The occasions at which the dihedral angle conformation was different from the one at the previous, saved time step ($\Delta t = 0.2ps$) were counted and divided by the number (5000) of saved configurations. The conformations were defined as

gauche$^-$	0	$\leq \phi <$	120
trans	120	$\leq \phi <$	240
gauche$^+$	240	$\leq \phi <$	360.

6.4 Results and Discussion

The results for the simulations of the β-hepta-peptide at 298 K using adiabatic decoupling MD with temperature up-scaling using $s_T = 1.1$ or force down-scaling using $s_V = 0.909$ are compared to those of standard MD simulations at two different temperatures, 298 K and 340 K, and a MD simulation without adiabatic decoupling ($s_m = 1$) and temperature scaling for the solute with $s_T = 1.14$.

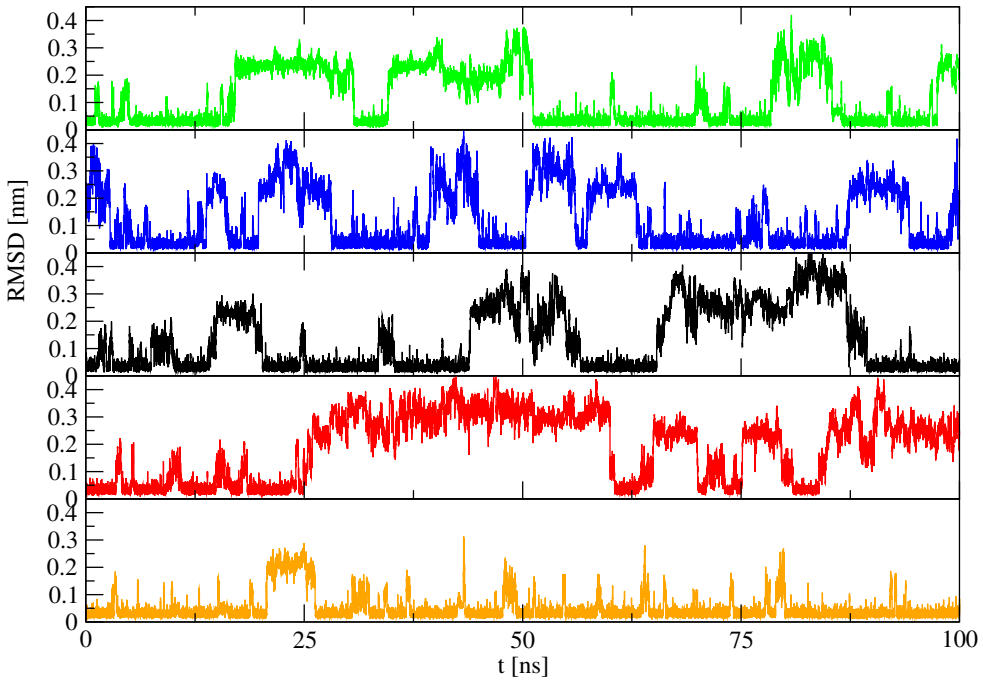

Figure 6.3: Atom-positional root-mean-square deviation from the 3_{14}-helical model structure for the backbone atoms of residues 2-6 for the standard MD simulation at 298 K (green), at 340 K (blue), the adiabatic decoupling MD simulation at 298 K with temperature scaling using $s_T = 1.1$ (black) or force scaling using $s_V = 0.909$ (red), and a standard MD simulation with the solute at 340 K and the solvent at 298 K (yellow).

The backbone atom-positional root-mean-square deviation from the 3_{14}-helical fold is shown in Fig. 6.3. Structures that have an RMSD value larger than 0.1 nm are considered unfolded. All five simulations sample structures that are folded and ones that are unfolded.

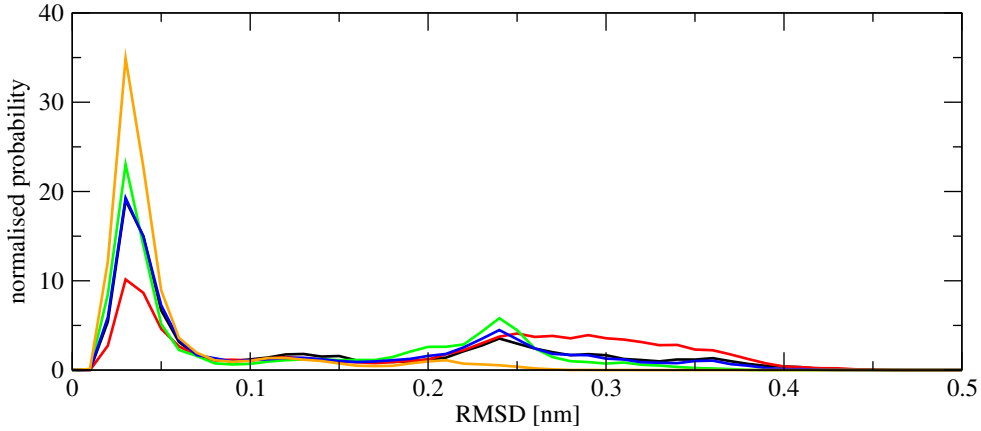

Figure 6.4: Probability distribution of the atom-positional root-mean-square deviation from the 3_{14}-helical model structure for the backbone atoms of residues 2-6 for the standard MD simulation at 298 K (green), at 340 K (blue), the adiabatic decoupling MD simulation with temperature scaling using $s_T = 1.1$ (black) or force scaling using $s_V = 0.909$ (red), and a standard MD simulation with the solute at 340 K and the solvent at 298 K (yellow).

The simulation with adiabatic decoupling with force scaling using $s_V = 0.909$ shows more unfolded conformations. This can also be seen in 6.4 which shows the probability distribution of the RMSD. The first peak, which consists of folded conformations, is much lower for adiabatic decoupling MD at 298 K with force scaling using $s_V = 0.909$, while adiabatic decoupling MD with temperature scaling using $s_T = 1.1$ and the simulation at 340 K have almost identical peaks.The simulation without adiabatic decoupling and $s_T = 1.14$ has the highest folded peak. The unfolded peak around 0.24 nm is highest for the simulation at 298 K followed by the one at 340 K and then the adiabatic decoupled one with temperature scaling. The simulation in which only the solute temperature is scaled has no peak around 0.24 nm. Adiabatic decoupling with force scaling shows unfolded structures with higher RMSD values. The adiabatic decoupling MD simulation at 298 K with temperature scaling using $s_T = 1.1$ generates a folding equilibrium comparable to that of the standard MD simulation at 340 K. Adiabatic decoupling using down-scaling of the force at 298 K using $s_V = 0.909$ favours the unfolded state too much: the forces are too strongly scaled down. Scaling of only the solute temperature in a standard MD simulation favors the folded state. This effect could be due to local heat flow from the solute to the adjacent solvent which is subsequently removed from the set of solvent degrees of freedom by the thermostat. Since all solvent degrees of freedom are jointly coupled to the heat bath, a higher temperature near the solvent will lead to a lower temperature for the solvent far away from the solute, which then forms a relatively cold, stable cage in which unfolding is more difficult. This effect could be avoided by coupling the temperature of individual solvent molecules to a heat bath, e.g. by using Langevin dynamics. However, this way of thermostatting distorts the dynamics of the solvent molecules.

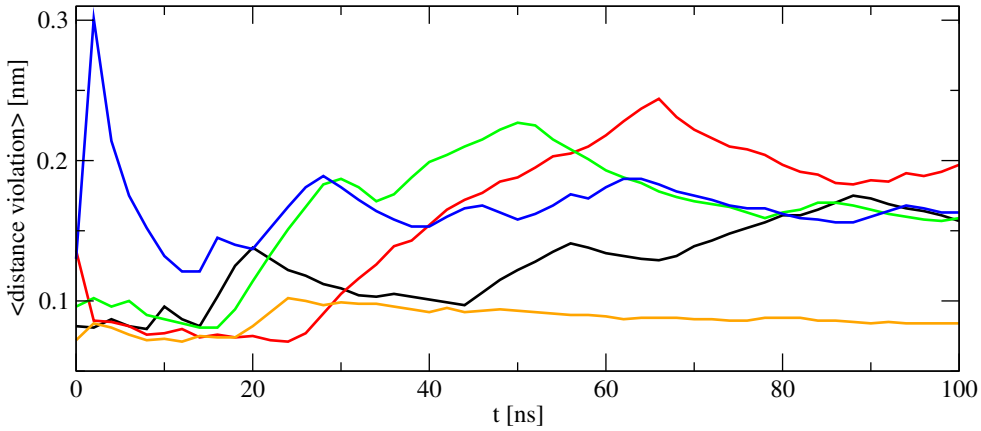

Figure 6.5: Time series of the sum of the (positive) NOE distance bound violations for the standard MD simulation at 298 K (green), at 340 K (blue), the adiabatic decoupling MD at 298 K with temperature scaling using $s_T = 1.1$ (black) or force scaling using $s_V = 0.909$ (red), and a standard MD simulation with the solute at 340 K and the solvent at 298 K (yellow).

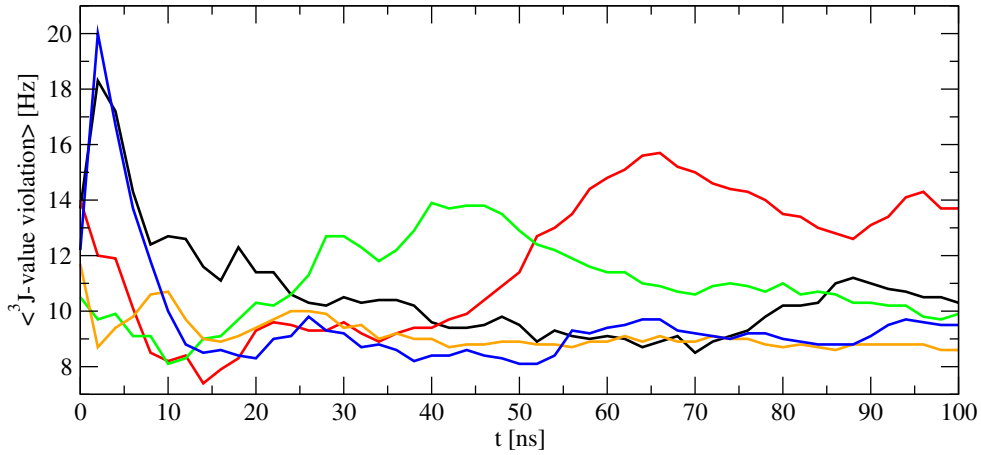

Figure 6.6: Time series of the sum of the absolute value of the deviation of the 21 ^{3}J-values from the experimental ones for the standard MD simulation at 298 K (green), at 340 K (blue), the adiabatic decoupling MD at 298 K with temperature scaling using $s_T = 1.1$ (black) or force scaling using $s_V = 0.909$ (red), the not adiabtically decoupled MD simulation with solvent at 298 K and solute at 340 K (orange).

Both the time series of the sum of the (positive) NOE distance bound violations in Fig. 6.5 and the time series of the sum of the absolute values of the deviation of the 21 ^{3}J-values from the experimental ones in Fig. 6.6 show that the simulations at 298 K, at 340 K and the one using adiabatic decoupling MD at 298 K with temperature scaling with $s_T = 1.1$ give similar results. The convergence is fastest for the standard MD

simulation at 340 K, while the one at 298 K and the one using adiabatic decoupling MD at 298 K with temperature scaling using $s_T = 1.1$ converge at approximately the same rate. The results for adiabatic decoupling MD with force scaling using $s_V = 0.909$ results in a higher deviation from experimental values. The standard MD with the solute temperature scaled converges faster and to a lower value, especially for the time series of the sum of the (positive) NOE distance bound violations. Apparently, the experimental data are compatible with a higher occurrence of the folded conformation.

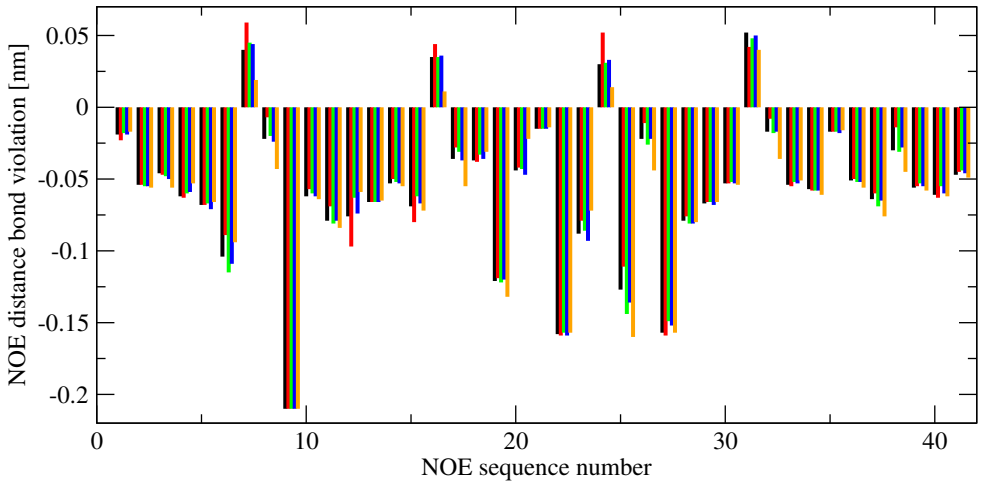

Figure 6.7: NOE distance violations averaged over 100 ns of simulation for the standard MD simulation at 298 K (green), at 340 K (blue), adiabatic decoupling MD at 298 K with temperature scaling using $s_T = 1.1$ (black) or force scaling using $s_V = 0.909$ (red) and a standard MD simulation with the solute at 340 K and the solvent at 298 K (yellow).

This can also be seen in the NOE distance bound violations averaged over the whole simulation time shown in Fig. 6.7 that are largest for adiabatic decoupling MD with force scaling using $s_V = 0.909$ in three out of four cases and lowest for all cases for standard MD simulation with the solute at 340 K and solvent at 298 K. For the deviations of simulated ^{3}J-values from the experimental ones in Fig. 6.8 the findings are similar: all simulations give comparable results, except that adiabatic decoupling MD with force scaling using $s_V = 0.909$ is for most cases further from the experimental values and that the solute temperature scaled standard MD simulation is closer to the experimental values.

The calculated free enthalpy of folding solidifies this finding further, see Fig. 6.9. All simulations, except the one for adiabatic decoupling with force scaling using $s_V = 0.909$ and the solute temperature scaled standard MD simulation, converge to approximately the same value. The standard simulation at 340 K seems to converge the fastest.

The combined clustering for pairs of trajectories for the five simulations using either the 15 backbone ϕ-, θ-, and ψ-dihedral angle root-mean-square differences in 6.10

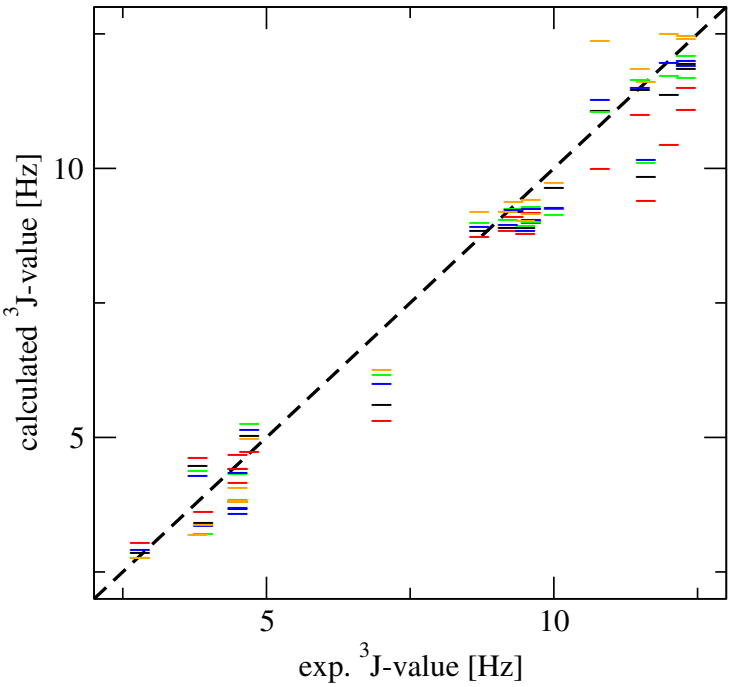

Figure 6.8: ^{3}J-coupling averaged over 100 ns of simulation for the standard MD simulation at 298 K (green), at 340 K (blue), adiabatic decoupling MD at 298 K with temperature scaling using $s_T = 1.1$ (black) or force scaling using $s_V = 0.909$ (red), the not adiabtically decoupled MD simulation with solvent at 298 K and solute at 340 K (orange).

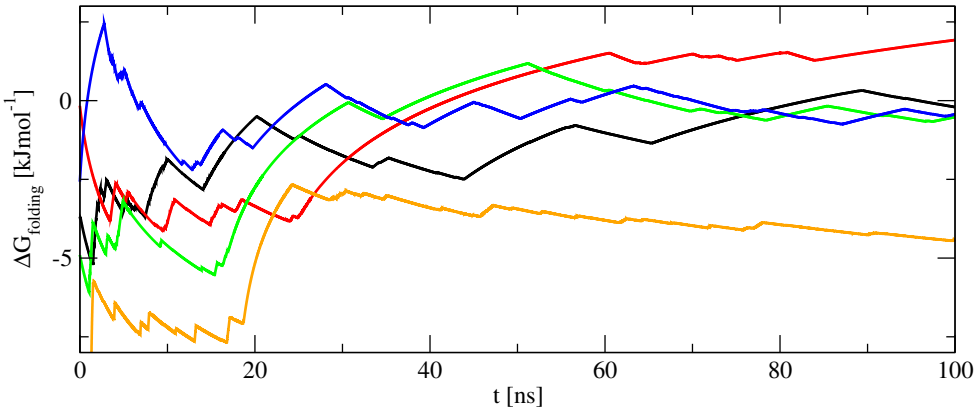

Figure 6.9: Free enthalpy of folding ΔG as function of time for the standard MD simulation at 298 K (green), at 340 K (blue), the adiabatic decoupling MD simulation at 298 K with temperature scaling using $s_T = 1.1$ (black) or force scaling using $s_V = 0.909$ (red), and a standard MD simulation with the solute at 340 K and the solvent at 298 K (yellow).

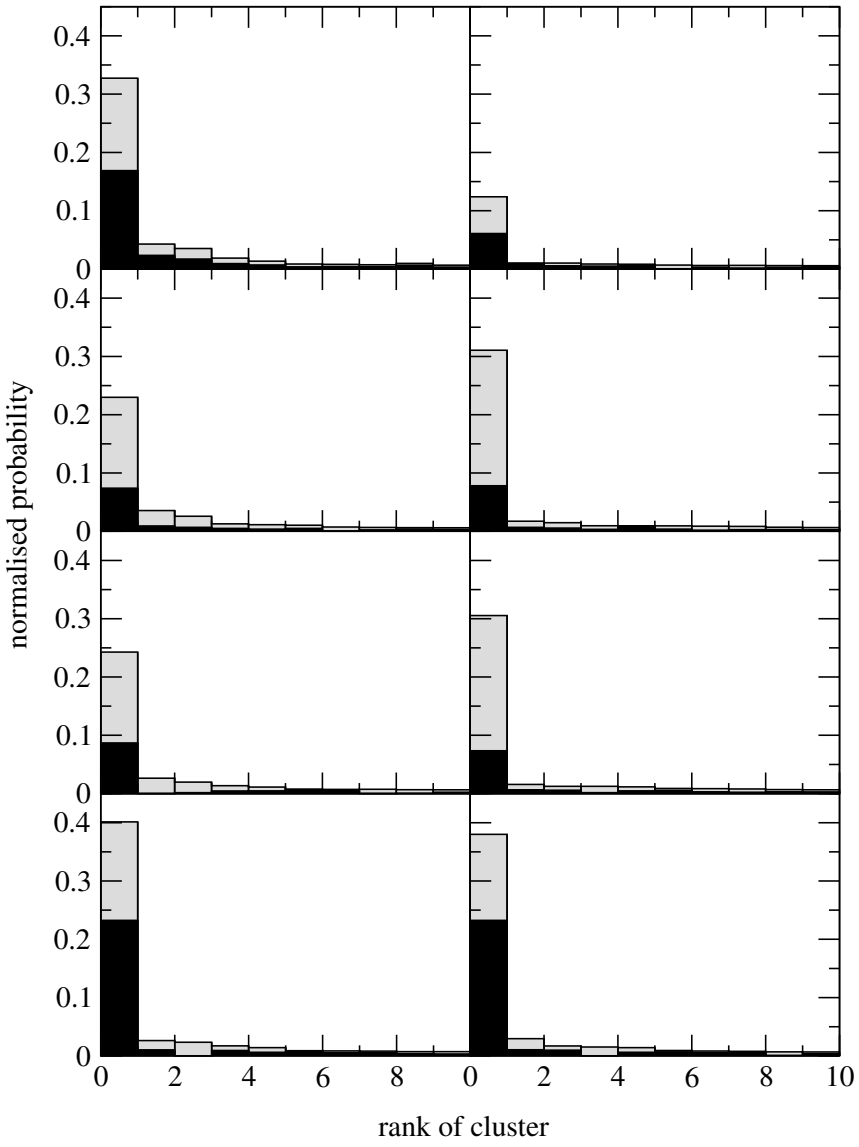

Figure 6.10: Combined conformational clustering of pairs of MD trajectories using the root-mean-square difference between the 15 backbone $\phi-$, $\theta-$, and $\psi-$dihedral angles from the second to the sixth residue. First row, left: standard MD simulation at 298 K (black) and at 340 K (gray), first row, right: adiabatic decoupling MD at 298 K with temperature scaling using $s_T = 1.1$ (black) and force scaling using $s_V = 0.909$ (gray), second row, left: adiabatic decoupling MD at 298 K with temperature scaling using $s_T = 1.1$ (black) and standard MD simulation at 298 K (gray), second row, right: adiabatic decoupling MD at 298 K with temperature scaling using $s_T = 1.1$ (black) and standard MD simulation with solute temperature scaling using $s_T = 1.14$ (gray), third row, left: adiabatic decoupling MD at 298 K using force scaling using $s_V = 0.909$ (black) and standard MD simulation at 298 K (gray), third row, right: adiabatic decoupling MD at 298 K with force scaling using $s_V = 0.909$ (black) and standard MD simulation with solute temperature scaling using $s_T = 1.14$ (gray), forth row, left: standard MD simulation with solute temperature scaling using $s_T = 1.14$ (black) and standard MD simulation at 298 K (gray), forth row, right: standard MD simulation with solute temperature scaling using $s_T = 1.14$ (black) and standard MD simulation at 340 K (gray).

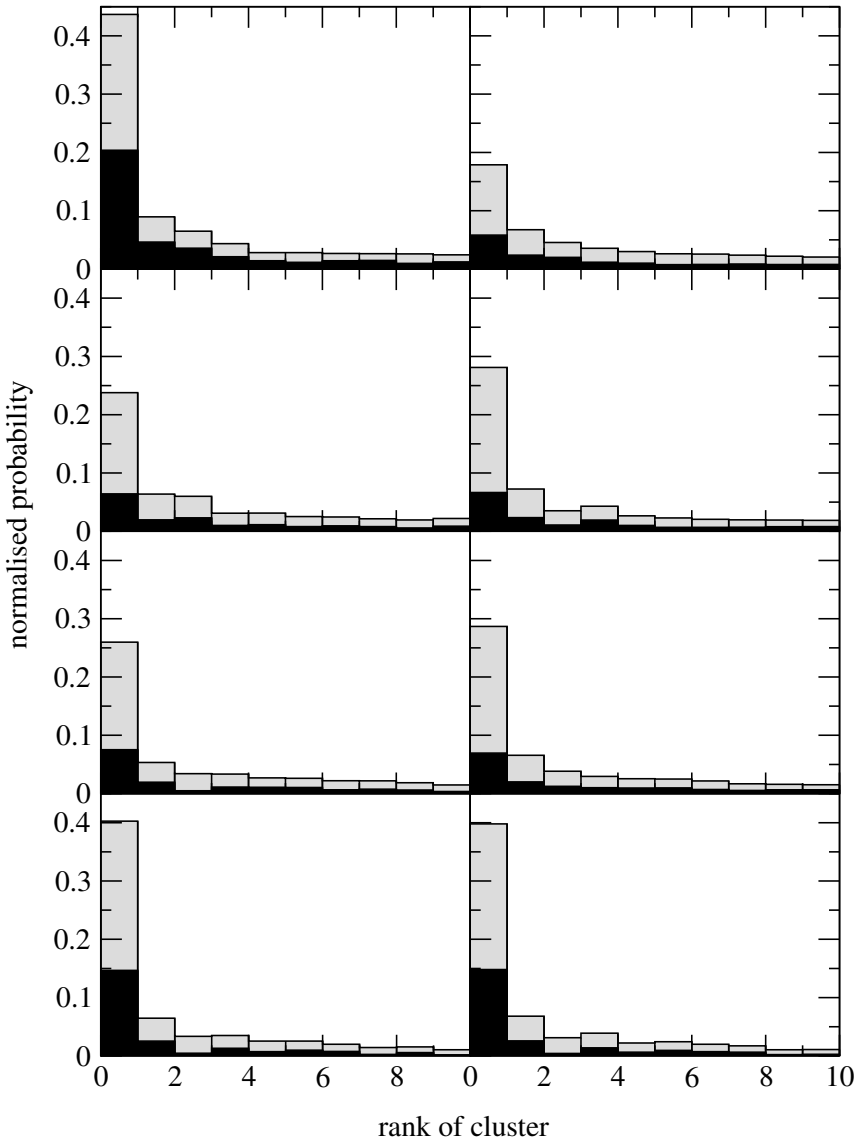

Figure 6.11: Combined conformational clustering of pairs of MD trajectories using the root-mean-square difference between the two dihedral angles defined by the C_β atoms from the second to the sixth residue. First row, left: standard MD simulation at 298 K (black) and at 340 K (gray), rist row, right: adiabatic decoupling MD at 298 K with temperature scaling using $s_T = 1.1$ (black) and force scaling using $s_V = 0.909$ (gray), second row, left: adiabatic decoupling MD at 298 K with temperature scaling using $s_T = 1.1$ (black) and standard MD simulation at 298 K (gray), second row, right: adiabatic decoupling MD at 298 K with temperature scaling using $s_T = 1.1$ (black) and standard MD simulation with solute temperature scaling using $s_T = 1.14$ (gray), third row, left: adiabatic decoupling MD at 298 K using force scaling using $s_V = 0.909$ (black) and standard MD simulation at 298 K (gray), third row, right: adiabatic decoupling MD at 298 K with force scaling using $s_V = 0.909$ (black) and standard MD simulation with solute temperature scaling using $s_T = 1.14$ (gray), forth row, left: standard MD simulation with solute temperature scaling using $s_T = 1.14$ (black) and standard MD simulation at 298 K (gray), forth row, right: standard MD simulation with solute temperature scaling using $s_T = 1.14$ (black) and standard MD simulation at 340 K (gray).

or the root-mean-square difference between the two dihedral angles defined by the C_β atoms of residues 2 to 6 in 6.11 give basically the same result: the first cluster is shared by all simulations, with the highest population for the standard MD simulation with solute temperature scaling using $s_T = 1.14$, followed by the standard MD simulations at 298 K and 340 K, and the least for the adiabatically decoupling MD simulations using temperature or force scaling.

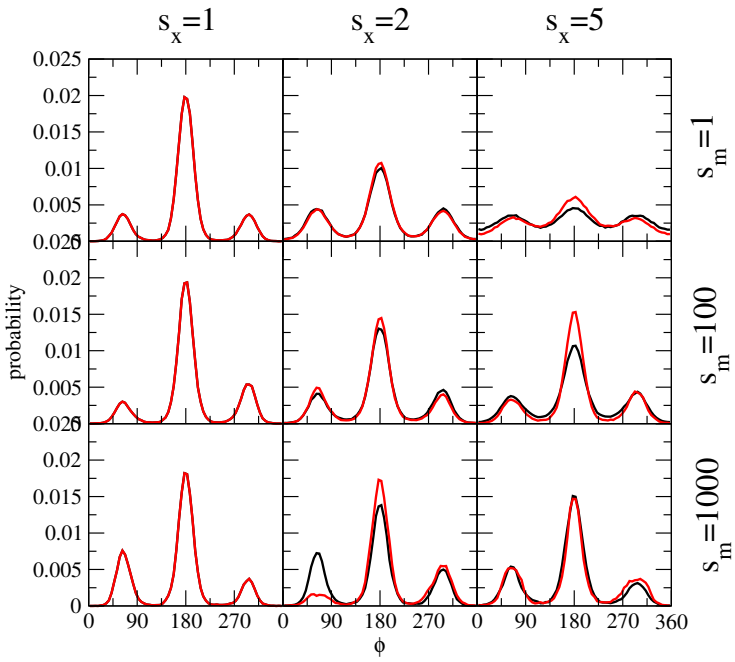

Figure 6.12: Distribution of the dihedral angle of dichloroethane in water for 18 simulations of 30 ns using different mass scaling factors s_m and different temperature scaling factors $s_x = s_T$ (black line) or force scaling factors $s_x = s_V^{-1}$ (red line).

For dichloroethane in water it is found that a scaling of the temperature or force without a scaling of the mass leads to a sizable detoriation of the dihedral-angle distribution as shown in 6.12. An increase of the mass for adiabatic decoupling improves the height of the middle peak. However, equal sampling of the two higher energy minima is not observed even after 30 ns of simulation. This is also visible in 6.13 where the transition probability is significantly lower if the mass is increased. For simulations with adiabatic decoupling parameter values that show a higher transition probability than the standard MD simulation, i.e. $s_m = 100$, $s_T = 5$, the dihedral angle distribution is broader compared to the standard MD simulation. None of the force scaling simulations have a higher transition probability than the standard MD simulation.

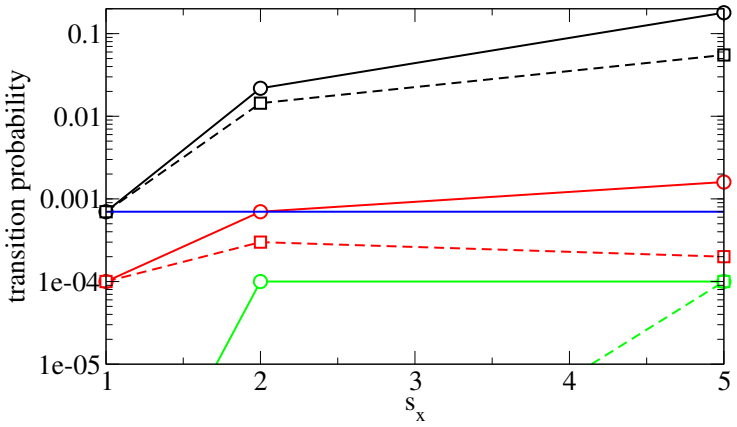

Figure 6.13: Probability of dihedral-angle transitions in dichloroethane for different mass scaling factors $s_m = 1$ (black), $s_m = 100$ (red) and $s_m = 1000$ (green) using different temperature scaling factors $s_x = s_T$ (circles, solid lines) or force scaling $s_x = s_V^{-1}$ (squares, dashed line). As a reference, the value for the standard MD simulation $s_m = 1,\ s_x = 1$ is given as a blue line.

6.5 Conclusions

In the present work it was investigated whether the sampling of the folding equilibrium of a 7-β-peptide in methanol in an MD simulation could be enhanced by adiabatically decoupling the peptide degrees of freedom from the solvent ones combined with subsequent up-scaling of the peptide temperature or down-scaling of the forces felt by the peptide atoms. Adiabatic decoupling MD at 298 K with $s_m = 100$ and a temperature scaling factor $s_T = 1.1$ for all atoms of the β-heptapeptide gives results comparable to those of the standard MD simulations at 298 K or 340 K. A clear speed-up of the convergence of the average values of various properties compared to standard MD at 298 K is, however, not observed. Combining adiabatic decoupling MD at 298 K with $s_m = 100$ and a force scaling factor $s_V = 0.909$ leads to results that significantly differ from the results of standard MD simulation at 298 K. The fastest convergence was found for the rather unphysical non-adiabatically decoupled system with a temperature for the solute of 340 K and for the solvent of 298 K.

For dichloroethane in water, the sampling speed lost by up-scaling the mass is only regained by strongly up-scaling the temperature of the solute at the expense of distorting the dihedral angle distribution, i.e. the correct configurational distribution.

This suggests that adiabatic decoupling of the solute degrees of freedom from the solvent ones does not bear fruit when aiming at enhanced sampling of the folding equilibrium of the solute. This result is in contrast to those of a similar study of ways to enhance the sampling of ionic degrees of freedom in aqueous solution, which showed that adiabatic decoupling of the ionic degrees of freedom from those of water followed

by up-scaling the temperature of the ions leads to much enhanced sampling of the ionic degrees of freedom. The explanation of these seemingly contradictory results lies in the ratio of the interactions within the h type particle degrees of freedom V^{hh} and that between h type and l type particle V^{hl}. In the ionic system V^{hl} is the dominant interaction whereas in the peptidic system the V^{hh}, the intrapeptidic interaction, plays a larger role with respect to the quantities that were analysed.

Chapter 7

Perfect as the wing of a bird may be, it will never enable the bird to fly if unsupported by the air. Facts are the air of science. Without them a man of science can never rise.

Ivan Pavlov

A method for conformational sampling of loops in protein based on adiabatic decoupling and temperature or force scaling

7.1 Summary

A method for conformational Boltzmann sampling of loops in protein in aqueous solution is presented which is based on adiabatic decoupling molecular dynamics (MD) simulation with temperature or force scaling. To illustrate the enhanced sampling the loop from residues 33 to 43 in the bovine protein ribonuclease A is adiabatically decoupled from the rest of the protein and the solvent with a mass scaling factor $s_m = 1000$ and the sampling is enhanced with a scaling of the temperature using $s_T = 2$ or of the force using $s_V = 0.667$. Over 5 ns of simulation the secondary structure of the protein remains unaltered while a combined dihedral-angle conformational cluster analysis shows an increase of conformations outside the first most populated cluster of loop conformations for adiabatic decoupling MD with temperature scaling using $s_T = 2$ or force scaling using $s_V = 0.667$ compared to the standard MD simulation. The atom-positional root-mean-square fluctuations of the C_α atoms of the loop show an increase in the movement of the loop as well, indicating that adiabatic decoupling MD with upscaling of the temperature or downscaling of the force is a promising method for conformational Boltzmann sampling.

7.2 Introduction

Proteins are involved in crucial functions of biological processes. In general their functionality highly depends on their three-dimensional structure. This structure can be constructed using experimental information from X-ray scattering or NMR spectroscopy.

These methods give the best results for regions with clear secondary structure such as α helices and β sheets. However, loops of the polypeptide chain that connect these regions and that are more flexible are often poorly or not resolved due to lack of sufficient experimental data. Since they sometimes play functionally important roles and frequently interact with other biomolecules [218], it is of interest to be able to generate an optimal energy structure, or rather a Boltzmann ensemble of structures, for loops in proteins. For this an abundance of methods is available [219], which can be split into two approaches: (i) generation of a set of structures that subsequently have to be scored with a chosen selection criterion to select the relevant ones, (ii) generation of an ensemble of structures that is already Boltzmann weighted using sampling such as molecular dynamics (MD), Monte Carlo (MC) or stochastic dynamics (SD) simulation. In this work we shall present an approach that belongs to the second class. It is not meant to predict a single minimum energy structure, but a conformational ensemble that is properly Boltzmann weighted. The problem that one faces here is the roughness of the energy hypersurface and the fact that the structure of the rest of the protein should remain intact. Therefore a sampling that is wide for the loop but does not influence the protein structure is required. The goal is to generate rather different loop structures without distorting the protein structure. Whether the conformational ensemble obtained is an accurate, realistic one will critically depend on the quality of the force field used in the simulation. This question is not addressed here.

The technique that is used to generate a Boltzmann ensemble of loop configurations is adiabatic decoupling with temperature scaling [30, 35, 158] or force scaling [189]. For adiabatically decoupling the motion along a small subset of degrees of freedom, in this case the N^h atoms of the loop, their mass is increased by a factor $s_m = 1000$, while the other N^l atoms of the protein and the water in the system keep an unaltered mass. In this way the transfer of kinetic energy between the decoupled degrees of freedom through momentum conserving collisions becomes very small. The sampling can then be enhanced by either increasing the temperature of the N^h atoms of the loop or scaling down the forces acting on them. Here the method of adiabatic decoupling MD with temperature or force scaling to enhance Boltzmann sampling is tested and illustrated on a 11-residue loop, residues 33 to 43, in bovine ribonuclease A, a 124-residue protein [220](entry PDB ID 6RSA in the Protein Data Base) by comparing its sampling properties to those of standard MD simulation. This protein and loop were chosen because they were previously used [60] to test the sampling properties of the Local Elevation (LE) sampling enhancement algorithm [25].

Most methods for conformational search of loops in proteins do not sample or generate a set of Boltzmann-weighted conformations. Only MD or MC based methods do so. Therefore, we only compare the adiabatic decoupling MD simulation results with those of standard MD simulation, which method is known to sample more efficiently than MC simulation when applied to folded, covalently bound macromolecules such as proteins. The method of adiabatic decoupling MD combined with temperature or force scaling is more easily implemented in software and contains less adjustable parameters to be op-

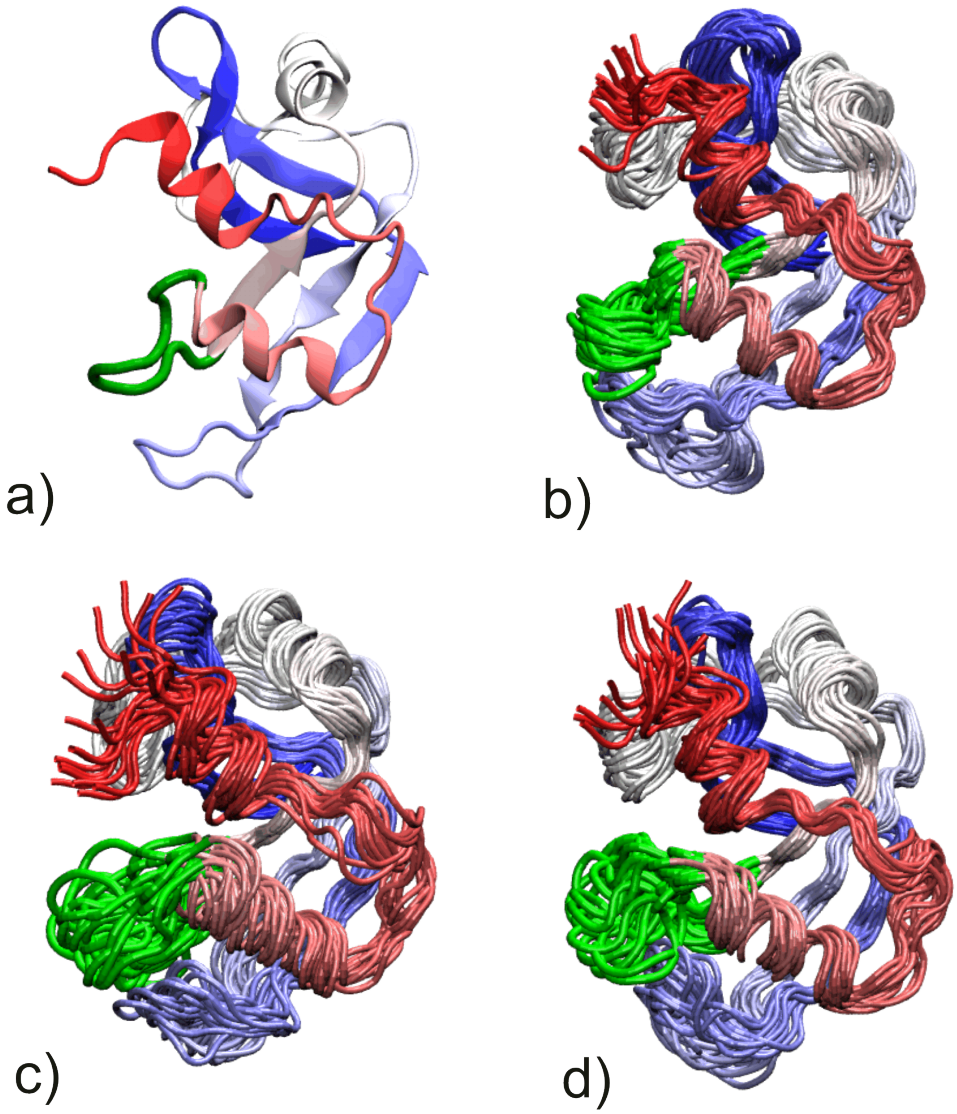

Figure 7.1: Protein structure. a) Starting structure [220] (PDB ID 6RSA) with the loop (residues 33 to 43) to be sampled marked green. The colouring goes from red over white to blue with increasing residue number. b) Structures, 250 ps apart, from a standard MD simulation of 5 ns. c) Structures, 250 ps apart, from an adiabatic decoupling MD simulation of 5 ns with temperature scaling using $s_T = 2$. d) Structures, 250 ps apart, from an adiabatic decoupling MD simulation of 5 ns with force scaling using $s_V = 0.667$.

timised than local-elevation sampling with its separate build-up and sampling phases. Therefore, we refrain from a comparison between these sampling enhancement methods.

7.3 Methods

7.3.1 Theory

A particular set of $3N^h$ degrees of freedom can be adiabatically decoupled from the other $3N^l$ degrees of freedom of the system by increasing their mass

$$m^h = s_m m^l \tag{7.1}$$

with $s_m \gg 1$ in such a way that the transfer of kinetic energy between the decoupled degrees of freedom through momentum conserving collisons becomes very small. s_m has to be chosen in such a way that the mass of the lightest of the N^h atoms, usually a hydrogen, is made significantly larger than the mass of the heaviest of the N^l atoms. It is therefore recommended to choose $s_m \geq 100$. Here we have denoted the masses of the particles constituting the N^h loop atoms collectively by m^h and the ones of the N^l remaining atoms collectively by m^l. This decoupling is more easily achieved [189] if $N^h \ll N^l$ which applies in the presented case taking the atoms of the loop as h type particles and the rest of the protein plus the water as l type particles. The non-dynamic equilibrium properties of the system remain unaltered by this change of mass of a subset of degrees of freedom.

To enhance the sampling of the h type degrees of freedom their temperature, potential energy or force can be scaled [189]. Applying temperature scaling, the temperature T^h of the h type degrees of freedom is set to

$$T^h = s_T T^l \text{ or } \beta^h = \beta^l / s_T \tag{7.2}$$

where $\beta = (k_B T)^{-1}$ and k_B is the Boltzmann constant. The Hamiltonian for a system of N^h h type particles and N^l l type particles is then

$$H(p^h, p^l, x^h, x^l) = K^h(p^h) + V^{hh}(x^h) + V^{hl}(x^h, x^l) + K^l(p^l) + V^{ll}(x^l), \tag{7.3}$$

where $K^h(p^h)$ and $K^l(p^l)$ are the kinetic energy of the h type and the l type particles respectively, $V^{hh}(x^h)$ the potential energy between h type particles, $V^{ll}(x^l)$ the potential energy between l type particles and $V^{hl}(x^h, x^l)$ that between h and l type particles. In the adiabatically decoupled limit, the partition function of the l type particles is

$$Z^l(x^h; \beta^l) = \int e^{-\beta^l K^l(p^l)} dp^l \int e^{-\beta^l V^{ll}(x^l)} e^{-\beta^l V^{hl}(x^h, x^l)} dx^l, \tag{7.4}$$

and the Hamiltonian of the h type degrees of freedom can therefore be written as

$$H^h(p^h, x^h; \beta^l) = K^h(p^h) + V^{hh}(x^h) - (\beta^l)^{-1} \ln Z^l(x^h; \beta^l). \tag{7.5}$$

The average value of a property $Q(x^h)$ at temperature T^l, $\langle Q \rangle_{\beta^l}$, can be calculated from $\langle Q \rangle_{\beta^h}$ of $Q(x^h)$ at T^h using the standard unbiasing formula,

$$\langle Q \rangle_{\beta^l} = \frac{\langle Q e^{-(\beta^l - \beta^h)[V^{hh} - (\beta^l)^{-1} \ln Z^l]} \rangle_{\beta^h}}{\langle e^{-(\beta^l - \beta^h)[V^{hh} - (\beta^l)^{-1} \ln Z^l]} \rangle_{\beta^h}}, \tag{7.6}$$

where the ensemble averages are over the h type degrees of freedom, and

$$Z^l(x^h; \beta^l) = \frac{\int e^{-\beta^l K^l(p^l)} dp^l \int e^{-\beta^l V^{ll}(x^l)} dx^l}{\langle e^{+\beta^l V^{hl}(x^h)} \rangle} \tag{7.7}$$

where the ensemble average in the denominator is over the l type degrees of freedom [189] at T^l. For the system considered here, the reweighting factor $e^{-(\beta^l - \beta^h)[V^{hh} - (\beta^l)^{-1} \ln Z^l]}$ for the h type configurations is equal to infinity in numerical precision and therefore no reweighting was applied.

Applying force scaling [189], the force acting on a h type degree of freedom,

$$\mathbf{f}_h = (\mathbf{f}_{hh} + \mathbf{f}_{hl}) s_V, \tag{7.8}$$

where $\mathbf{f}_{hh}$ is force exerted by the other h type degrees of freedom and $\mathbf{f}_{hl}$ is the force exerted by the l type degrees of freedom, is scaled by a factor s_V.

The Hamiltonian, in analogy to Eq. 7.5, can be written as

$$H^h(p^h, x^h) = K^h(p^h) + s_V V^{hh}(x^h) - s_V (\beta^l)^{-1} \ln Z^l(x^h; \beta^l), \tag{7.9}$$

with $Z^l(x^h; \beta^l)$ given by Eq. 7.7 and $\langle Q \rangle_{\beta^l}$ given by Eq. 7.6 in which β^h is defined as

$$\beta^h = \beta^l s_V \tag{7.10}$$

instead of by Eq. 7.2. For the system considered here, the reweighting factor $e^{-(\beta^l - \beta^h)[V^{hh} - (\beta^l)^{-1} \ln Z^l]}$ for the h type configurations is equal to infinity in numerical precision and therefore no reweighting was applied.

Adiabatic decoupling with either temperature scaling or force scaling of a limited set of degrees of freedom in a protein is a simple technique to enhance sampling of a conformational Boltzmann-weighted distribution.

7.3.2 Simulated system and simulation procedure

A cubic box with an edge length of $7.194\,\mathrm{nm}$ was filled with 11548 SPC [9] water molecules and one ribonuclease A molecule and simulated at $300\,\mathrm{K}$ and $1\,\mathrm{atm}$. The GROMOS 45A4 [214] force field was used to describe the protein. The structure was

taken from the PDB entry 6RSA, hydrogens were added assuming a pH of 7. The disulfide bridge between Cys 40 and Cys 95 was removed [60] because Cys 40 is part of the loop to be conformationally sampled. Periodic boundary conditions were applied. Molecular dynamics simulations were performed with the GROMOS05 simulation software package [133], modified to incorporate adiabatic decoupling. The geometry of the water molecules was constrained by applying the SHAKE algorithm [23] with a relative geometric tolerance of 10^{-4} to the OH bond lengths and to the intramolecular HH distance of the water molecules, and the bond lengths of the protein were constrained using the SHAKE algorithm as well. The nonbonded van der Waals and electrostatic interactions were calculated using triple-range cutoff radii of $0.8/1.4\,\mathrm{nm}$. The short range interactions were calculated every time step by updating the molecular pair list for distances smaller than the first cutoff radius of $0.8\,\mathrm{nm}$. For the intermediate range of distances between 0.8 and $1.4\,\mathrm{nm}$ the pair list was only updated every fifth timestep and at the same time the interaction was calculated and kept unchanged between these updates. The long range electrostatic interactions beyond the outer cutoff of $1.4\,\mathrm{nm}$ were represented by a reaction field [62, 185] with ϵ_{RF}=78.5, the dielectric permittivity of water at $300\,\mathrm{K}$ and $1\,\mathrm{atm}$. The equations of motion were integrated using the leap frog algorithm with a time step of $2\,\mathrm{fs}$. The velocities of the atoms at the beginning of the simulation were assigned from a Maxwell distribution at $300\,\mathrm{K}$.

Adiabatic decoupling was investigated with an s_m value of 1000. The temperature scaling factors s_T used for the loop were 2 and 5, and the force scaling factors s_V were 0.667 and 0.5. For comparison, an MD simulation without adiabatic decoupling was performed as well. All simulations were done at constant pressure by weakly coupling [29] the system to a bath of $1\,\mathrm{atm}$ pressure with a coupling time of $0.5\,\mathrm{ps}$. The isothermal compressibility was set to $7.513{\cdot}10^{-4}\,(kJmol^{-1}nm^{-3})^{-1}$. The loop and the protein and water molecules were separately weakly coupled [29] to three heat baths with a coupling time of $0.1\,\mathrm{ps}$. Since the adiabatically decoupled h type atoms of the loop are covalently bound to the l type atoms, namely the rest of the protein, special attention was paid to possible heat flow through the covalent interaction terms from the two ends of the loop to the rest of the polypeptide chain. This heat flow was avoided by coupling the two residues adjacent to the loop separately to a heat bath with a tight coupling constant of $0.01\,\mathrm{ps}$. During the runs, configurations of the system were saved every $2\,\mathrm{ps}$. The various properties were taken from $5\,\mathrm{ns}$ simulations that followed a $100\,\mathrm{ps}$ equilibration period.

7.3.3 Analysis

To check whether the use of $s_T \neq 1$ and $s_V \neq 1$ to scale the temperature or the force respectively cause the protein to fall apart, the atom-positional root-mean-square deviation from the initial structure and the secondary structure of the protein were calculated as a function of time. The temperature of the C_α atoms serves as a check whether the heat flow from the h type to the l type atoms could be avoided. Combined con-

formational clustering of pairs of trajectories or ensembles using the 8 dihedral angles defined by the 11 C_α atoms of the loop (residues 33-43) was performed and the positional root-mean-square fluctuations of the C_α atoms were calculated to measure the extent of sampling in the adiabatic decoupling MD with temperature or force scaling in comparision to the standard MD simulation.

Atom-positional root-mean-square deviation

The atom-positional root-mean-square deviation (RMSD) measures the difference between a reference structure and another structure at time t using

$$RMSD(t) = \left[\frac{1}{N_a} \sum_{i=1}^{N_a} (\mathbf{r}_i(t) - \mathbf{r}_i^{ref})^2 \right]^{1/2} \tag{7.11}$$

where N_a is the number of atoms considered, $\mathbf{r}_i(t)$ is the position of atom i at time t and $\mathbf{r}_i^{ref}$ is the position of atom i in the reference structure. Each structure was first superimposed onto the reference structure using the backbone (N, C_α, C) atoms and the RMSD was calculated for the C_α atoms.

Temperature per atom

The temperature of atom i is calculated using equipartition

$$T_i = \frac{m_i < \mathbf{v}_i^2 >_t}{3k_B} \tag{7.12}$$

where m_i is the mass of atom i, $\mathbf{v}_i$ its velocity and $< \cdots >_t$ denotes an average over the simulation time t.

Atom-positional root-mean-square fluctuation

The atom-positional root-mean-square fluctuation (RMSF) of atom i measures the fluctuation of its position around its average position

$$RMSF(i) = \left[\frac{1}{N_t} \sum_{n=1}^{N_t} [\mathbf{r}_i(t_n)]^2 - < \mathbf{r}_i >_t^2 \right]^{1/2} \tag{7.13}$$

where N_t is the number of time steps, $\mathbf{r}_i(t)$ is the position of atom i at time t and $< \mathbf{r}_i >_t$ is the time-averaged position of atom i. Each structure was first superimposed onto the first structure using the backbone (N, C_α, C) atoms and the RMSF was calculated for the C_α atoms.

Dihedral-angle clustering

The structure of the loop was characterised by the 8 dihedral angles Φ formed by its C_α atoms, $\Phi(C_{\alpha,i} - C_{\alpha,i+1} - C_{\alpha,i+2} - C_{\alpha,i+3})$ with the residue number i running from 33 to 40. The conformational clustering was performed using the root-mean-square deviation of these 8 dihedral angles of the loop for each pair of structures separated by 2 ps of simulation time, analogously to the conformational clustering using Cartesian coordinates described in Ref. [221].

As a criterion of similarity of two structures an RMSD value of 30 ° for the 8 dihedral angles was used. The number of neighbour structures, i.e. structures satisfying the similarity criterion, for each of the structures in the initial set of structures was determined. As the center structure of the first cluster of structures the structure with the highest number of neighbours was taken and all structures belonging to this cluster were thereafter removed from the set of structures. For each of the remaining structures the number of neighbours was again computed. The structure with the most neighbours became the centre of the second cluster of structures. Structures belonging to this second cluster were then also removed from the set of structures. This process was iterated until all structures were assigned to a conformational cluster.

7.4 Results and discussion

A loop of 11 residues in ribonuclease A, shown in green in Fig. 7.1a, was chosen to test the applicability of adiabatic decoupling with temperature or force scaling to widen the conformational sampling without distorting the structure of the rest of the protein. The sets of structures in Fig. 7.1b-d show the limited motion of the loop in a standard MD simulation (Fig. 7.1b) compared to the ones generated with adiabatic decoupling plus temperature scaling using a scaling factor of $s_T = 2$ (Fig. 7.1c) or plus force scaling using $s_V = 0.667$ (Fig. 7.1d). The last two display more movement in the loop region while the rest of the protein remains stable.

The backbone atom-positional root-mean-square displacement from the initial structure in Fig. 7.2 shows the same trend for $s_T = 2$ (red) and $s_V = 0.667$ (blue) as that of the standard MD simulation (orange). The protein is not falling apart using those scaling factors. A too strong temperature or force scaling, e.g. $s_T = 5$ (black) or $s_V = 0.5$ (green), results in a much larger deviation from the original structure, i.e. the protein structure is not maintained. Therefore, the simulations based on these scaling factors will not be further considered.

The analysis of the secondary structure in Fig. 7.3 reveals a similar picture: Both the secondary structure of the standard MD simulation and of the adiabatic decoupling MD with temperature ($s_T = 2$) or force scaling ($s_V = 0.667$) remain basically unaltered during the simulation time of 5 ns. The main differences are visible in the loop region

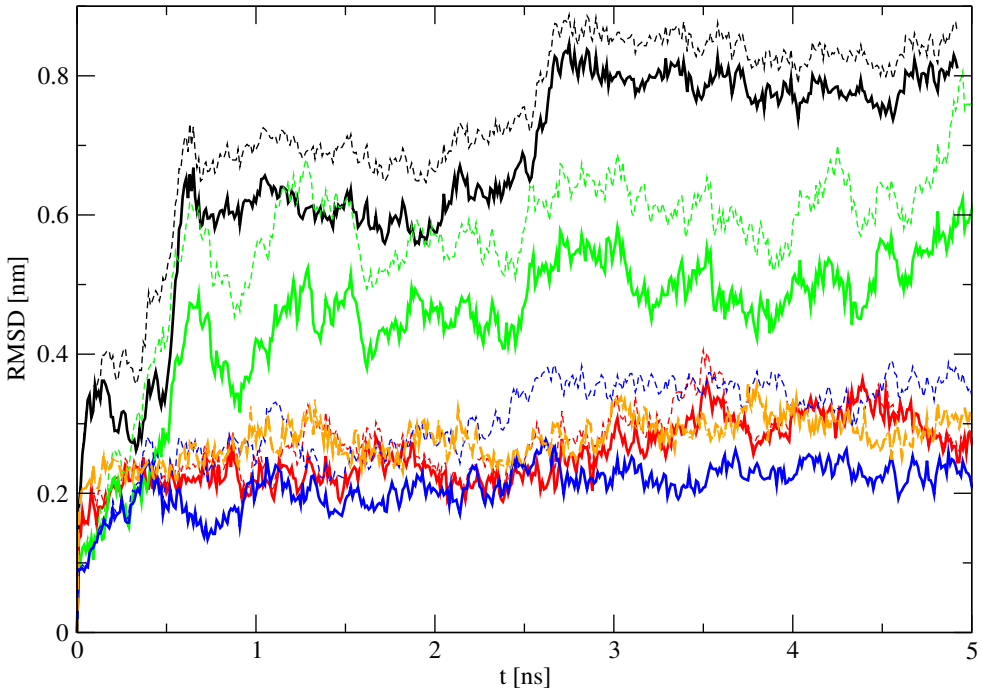

Figure 7.2: Backbone atom-positional root-mean-square deviation from the initial structure excluding (solid line) or including (dashed line) the loop atoms (residues 33 to 43) for adiabatic decoupling MD with temperature scaling using $s_T = 5$ (black) and $s_T = 2$ (red) or with force scaling using $s_V = 0.5$ (green) and $s_V = 0.667$ (blue) compared to standard MD simulation (orange).

between residues 33 and 43, as it is to be expected. Thus the requirement of a stable protein structure is fulfilled.

The heat flow from the loop to the rest of the protein is also under control. Fig. 7.4 shows in the top panel the temperature of the C_α atoms. Only the temperature of the adiabatically decoupled atoms with temperature scaling using $s_T = 2$ is higher as it should be by definition of the method. The bottom panel of Fig. 7.4 shows the root-mean-square fluctuation of the C_α atom positions. In the loop region, the values for adiabatic decoupling MD with temperature or force scaling are larger than the ones from standard MD simulation, indicating once more an enhanced sampling of conformations in this part of the protein.

The distributions of the root-mean-square deviation from the initial structure using the 8 dihedral angles defined by the 11 C_α atoms of the loop in Fig. 7.5 show that both adiabatic decoupling MD simulation using temperature or force scaling sample a much broader range than standard MD.

The combined conformational clustering using the 8 dihedral angles defined by the 11 C_α atoms of the loop comparing adiabatic decoupling MD with temperature scaling

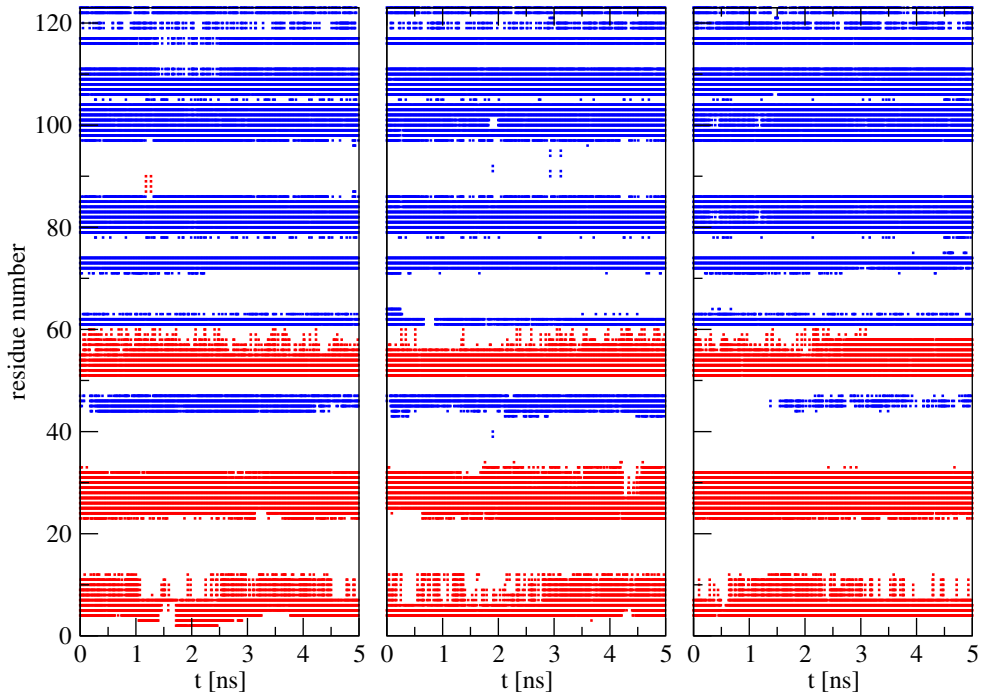

Figure 7.3: Secondary structure as function of time. Right: standard MD simulation, middle: adiabatic decoupling MD with temperature scaling using $s_T = 2$, left: adiabatic decoupling MD with force scaling using $s_V = 0.667$. α-helix (red), β-strand (blue).

or force scaling to standard MD simulation in Fig. 7.6 shows that the first cluster is mainly sampled by the MD simulation. This supports the previous findings that adiabatic decoupling MD for the loop makes it sample a wider set of conformations.

The combined conformational clustering of the loop structures comparing different adiabatic decoupling MD simulations, the one with temperature scaling using $s_T = 2$ and the one with force scaling using $s_V = 0.667$, in Fig. 7.7 (top panel) indicates that the conformations sampled within 5 ns of enhanced sampling simulation are not the same yet. This is no surprise, because convergence must be slower for wider sampling. The other two panels in Fig. 7.7 show the conformational overlap for two trajectories generated using the same sampling enhancement technique, i.e. temperature scaling in the middle panel and force scaling in the bottom panel, but starting with different initial atomic velocities. Again not much conformational overlap is observed, indicating that 5 ns is not sufficient to obtain convergence using these enhanced sampling methods.

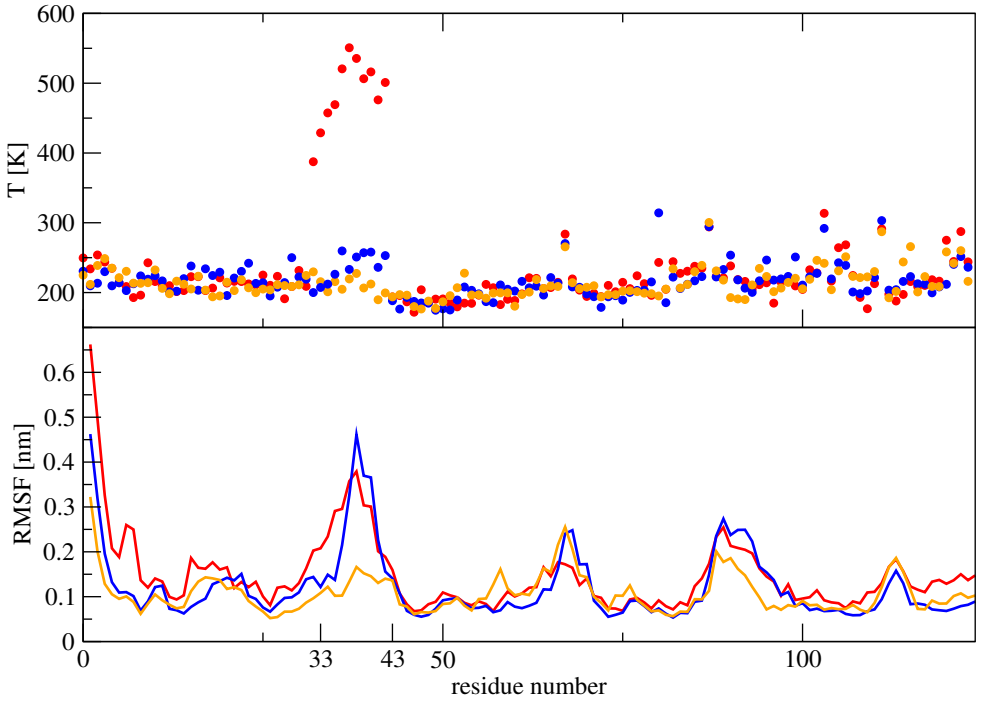

Figure 7.4: Top: Temperature of the C_α atoms, bottom: Root-mean-square fluctuation of the C_α atom positions over 5 ns for adiabatic decoupling MD with temperature scaling using $s_T = 2$ (red) or force scaling using $s_V = 0.667$ (blue) for the atoms of the loop (residues 33 to 43) compared to standard MD simulation (orange).

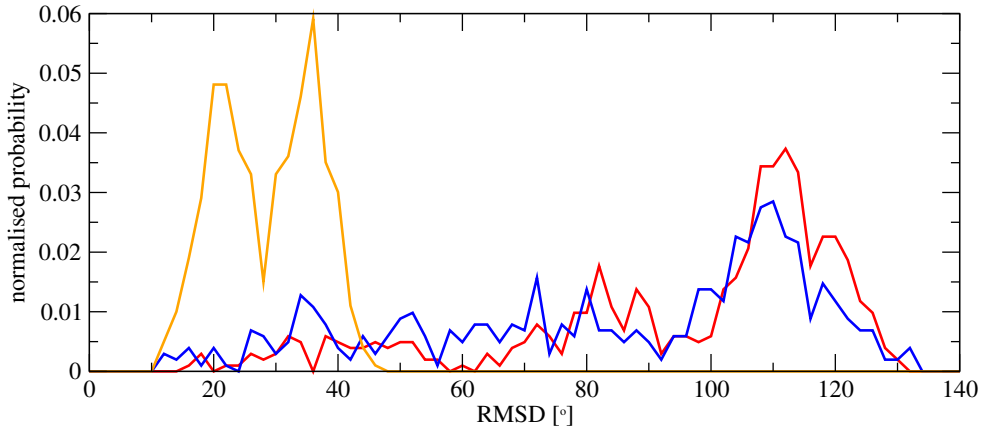

Figure 7.5: Distribution of the root-mean-square deviation from the initial structure using the 8 dihedral angles defined by the 11 C_α atoms of the loop (residues 33 to 43) for adiabatic decoupling MD with temperature scaling using $s_T = 2$ (red) or with force scaling using $s_V = 0.667$ (blue) compared to standard MD simulation (orange).

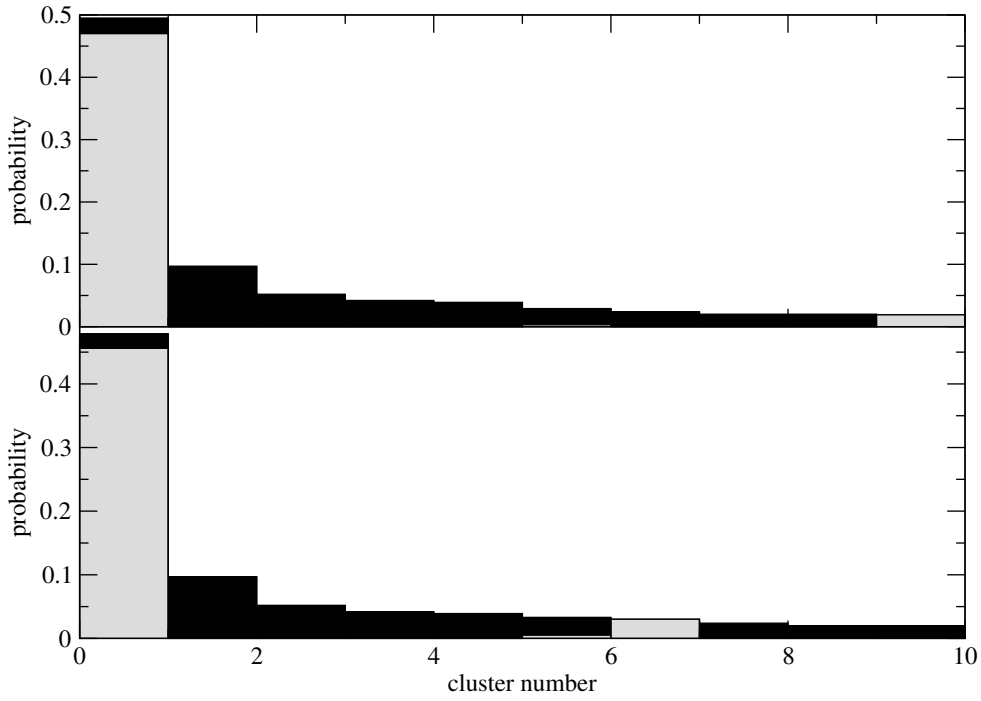

Figure 7.6: Combined conformational clustering of two ensembles using the 8 dihedral angles defined by the 11 C_α atoms of the loop (residues 33 to 43). Top panel: standard MD simulation (gray) combined with adiabatic decoupling MD with temperature scaling using $s_T = 2$ (black). Bottom: standard MD simulation (gray) combined with adiabatic decoupling MD with force scaling using $s_V = 0.667$ (black).

7.5 Conclusions

The loop between residues 33 and 43 in the protein bovine ribonuclease A shows more conformational variability when using adiabatic decoupling MD with temperature scaling using $s_T = 2$ or force scaling using $s_V = 0.667$ than in standard MD simulation. This finding is indicated by larger positional fluctuations of the C_α atoms in the loop and a broader conformational distribution of the dihedral angles characterising the structure of the loop than in standard MD simulation. Yet, adiabatic decoupling with temperature or force scaling leaves the structure of the protein unaltered as indicated by the stable secondary structure and the limited atom-positional deviation from the initial structure for the atoms outside the loop. This shows that adiabatic decoupling MD simulation with temperature scaling or force scaling is a simple and efficient method to sample a Boltzmann weighted conformational ensemble for loop regions of a protein without distorting the remaining protein structure.

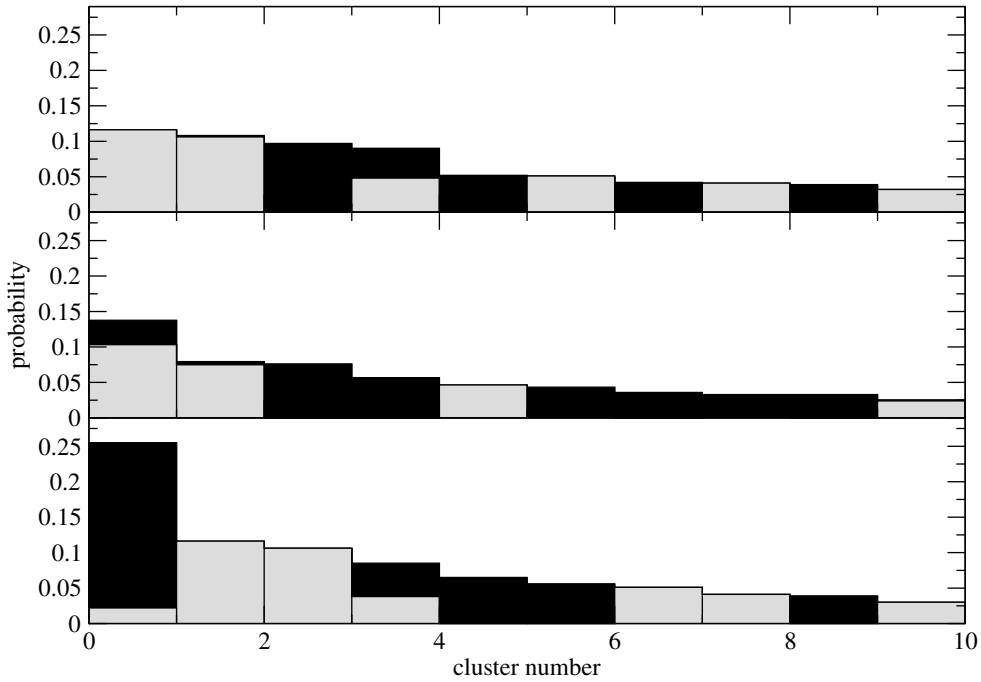

Figure 7.7: Combined conformational clustering of two ensembles using the 8 dihedral angles defined by the 11 C_α atoms of the loop (residues 33 to 43). Top: adiabatic decoupling MD with temperature scaling using $s_T = 2$ (black) and force scaling using $s_V = 0.667$ (gray). Middle: adiabatic decoupling MD with temperature scaling using $s_T = 2$ for two runs starting from different initial velocities. Bottom: adiabatic decoupling MD with force scaling using $s_V = 0.667$ for two runs starting from different initial velocities.

We can only see a short distance ahead, but we can see plenty there that needs to be done.

Alan Turing

Chapter 8

Outlook

The best predictor for the future development of computer simulations has been for a long time Moore's Law. However, it is not only the single computing machine that is getting faster, also the use of parallelised code on compute clusters shortens the time necessary to simulate a given system more and more. With this, simulations of larger systems, over longer time scales and with higher accuracy become feasible.

One way to invest this additional computer power into higher accuracy is to refine force fields by introducing polarisability. Since water is the most relevant solvent in biological systems and since the here presented water model COS/D, Chapter 2, still has some shortcomings like a slightly too low dielectric permittivity, a further optimised parameterisation would be of interest. Using the applied electric field method [222] instead of the fluctuation formula, the dielectric permittivity can be faster and more accurately calculated, so better parametrised. A better polarisable water model can only be the first step toward a general polarisable force field. However, it is an important step as certain properties generally used for the parametrisation of GROMOS force fields like the free energy of solvation highly depend on the chosen water model. Therefore, the quality of a force field rises and falls with the quality of the water model. The next steps involve the development of polarisable solvent models for non-polar solvents such as alkanes, polarisable amino-acid residues, nucleic bases, lipids and carbohydrates. Here the use of non-linear polarisability as proposed in Chapter 2 could be of use in case overpolarisation of particular molecules occurs.

Since scientists are not willing to wait for computers to be fast enough to tackle their problems and want to simulate bigger systems longer with a more detailed description, the further development and testing of sampling enhancement methods such as adiabatic decoupling with temperature of force scaling remains of interest too. Several aspects should still be investigated, such as the effect of adiabatically decoupled ions with a higher temperature on the conformational distribution of a solute that is also adiabatically decoupled from the water and for which the sampling is enhanced by force scaling. The performance of adiabatic decoupling with temperature or force scaling with respect to obtaining a potential of mean force along particular degrees of freedom, e.g. in combination with enveloping distribution sampling (EDS) [223], should also be investigated. Adiabatically decoupled ions with a higher temperature around a solute in

water could also be combined using local elevation umbrella sampling (LEUS) [224] to enhance the sampling of the solute.

Bibliography

[1] A. P. Nanzer, F. M. Poulsen, W. F. van Gunsteren, and A. E. Torda. A reassessment of the structure of chymotrypsin inhibitor-2 (CI-2) using time-averaged NMR restraints. *Biochemistry*, 33(48):14503–14511, 1994.

[2] A. Glättli and W. F. van Gunsteren. Are NMR-derived model structures for β-peptides representative for the ensemble of structures adopted in solution? *Angewandte Chemie-International Edition*, 43(46):6312–6316, 2004.

[3] R. E. Amaro, R. Baron, and J. A. McCammon. An improved relaxed complex scheme for receptor flexibility in computer-aided drug design. *Journal of Computer-Aided Molecular Design*, 22(9):693–705, 2008.

[4] A. Ivetac and J. A. McCammon. Mapping the druggable allosteric space of G-protein coupled receptors: a fragment-based molecular dynamics approach. *Chemical Biology & Drug Design*, 76(3):201–217, 2010.

[5] W. F. van Gunsteren, D. Bakowies, R. Baron, I. Chandrasekhar, M. Christen, X. Daura, P. Gee, D. P. Geerke, A. Glättli, P. H. Hünenberger, M. A. Kastenholz, C. Oostenbrink, M. Schenk, D. Trzesniak, N. F. A. van der Vegt, and H. B. Yu. Biomolecular modeling: Goals, problems, perspectives. *Angewandte Chemie-International Edition*, 45(25):4064–4092, 2006.

[6] J. Hermans, H. J. C. Berendsen, W. F. van Gunsteren, and J. P. M. Postma. A consistent empirical potential for water-protein interactions. *Biopolymers*, 23(8):1513–1518, 1984.

[7] C. Oostenbrink, A. Villa, A. E. Mark, and W. F. van Gunsteren. A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. *Journal of Computational Chemistry*, 25(13):1656–1676, 2004.

[8] H. B. Yu and W. F. van Gunsteren. Accounting for polarization in molecular simulation. *Computer Physics Communications*, 172(2):69–85, 2005.

[9] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans. *Interaction models for water in relation to protein hydration*. Intermolecular forces. Reidel, Dordrecht, 1981.

[10] A. Warshel and M. Levitt. Theoretical studies of enzymic reactions - dielectric, electrostatic and steric stabilization of carbonium-ion in reaction of lysozyme. *Journal of Molecular Biology*, 103(2):227–249, 1976.

[11] F. J. Vesely. N-particle dynamics of polarizable Stockmayer-type molecules. *Journal of Computational Physics*, 24(4):361–371, 1977.

[12] D. van Belle, I. Couplet, M. Prevost, and S. J. Wodak. Calculations of electrostatic properties in proteins - analysis of contributions from induced protein dipoles. *Journal of Molecular Biology*, 198(4):721–735, 1987.

[13] T. P. Straatsma and J. A. McCammon. Molecular dynamics simulations with interation potentials including polarization: Development of a non-iterative method and application to water. *Molecular Simulation*, 5(3-4):181 – 192, 1990.

[14] P. Drude. *The Theory of Optics*. Longmans, Green, and Co., New York, 1902.

[15] Huang K. Born M. *Dynamic Theory of Crystal Lattices*. Oxford University Press, Oxford, UK, 1954.

[16] S. W. Rick, S. J. Stuart, and B. J. Berne. Dynamical fluctuating charge force-fields - application to liquid water. *Journal of Chemical Physics*, 101(7):6141–6156, 1994.

[17] D. P. Geerke and W. F. van Gunsteren. On the calculation of atomic forces in classical simulation using the charge-on-spring method to explicitly treat electronic polarization. *Journal of Chemical Theory and Computation*, 3(6):2128–2137, 2007.

[18] H. B. Yu and W. F. van Gunsteren. Charge-on-spring polarizable water models revisited: From water clusters to liquid water to ice. *Journal of Chemical Physics*, 121(19):9549–9564, 2004.

[19] M. Christen and W. F. van Gunsteren. On searching in, sampling of, and dynamically moving through conformational space of biomolecular systems: A review. *Journal of Computational Chemistry*, 29(2):157–166, 2008.

[20] W. Braun and N. Go. Calculation of protein conformations by proton proton distance constraints - a new efficient algorithm. *Journal of Molecular Biology*, 186(3):611–626, 1985.

[21] T. Huber, A. E. Torda, and W. F. van Gunsteren. Structure optimization combining soft-core interaction functions, the diffusion equation method, and molecular dynamics. *Journal of Physical Chemistry A*, 101(33):5926–5930, 1997.

[22] G. M. Crippen. Conformational-analysis by energy embedding. *Journal of Computational Chemistry*, 3(4):471–476, 1982.

[23] J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen. Numerical-integration of cartesian equations of motion of a system with constraints - molecular-dynamics of n-alkanes. *Journal of Computational Physics*, 23(3):327–341, 1977.

[24] B. Smit, P. A. J. Hilbers, K. Esselink, L. A. M. Rupert, N. M. Vanos, and A. G. Schlijper. Computer-simulations of a water oil interface in the presence of micelles. *Nature*, 348(6302):624–625, 1990.

[25] T. Huber, A. E. Torda, and W. F. van Gunsteren. Local elevation - a method for improving the searching properties of molecular-dynamics simulation. *Journal of Computer-Aided Molecular Design*, 8(6):695–708, 1994.

[26] H. Grubmüller. Predicting slow structural transitions in macromolecular systems - conformational flooding. *Physical Review E*, 52(3):2893–2906, 1995.

[27] H. Tsujishita, I. Moriguchi, and S. Hirono. Potential-scaled molecular-dynamics and potential annealing - effective conformational search techniques for biomolecules. *Journal of Physical Chemistry*, 97(17):4416–4420, 1993.

[28] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983.

[29] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. Molecular-dynamics with coupling to an external bath. *Journal of Chemical Physics*, 81(8):3684–3690, 1984.

[30] J. VandeVondele and U. Röthlisberger. Canonical adiabatic free energy sampling (cafes): A novel method for the exploration of free energy surfaces. *Journal of Physical Chemistry B*, 106(1):203–208, 2002.

[31] R. Elber and M. Karplus. Enhanced sampling in molecular-dynamics - use of the time-dependent hartree approximation for a simulation of carbon-monoxide diffusion through myoglobin. *Journal of the American Chemical Society*, 112(25):9161–9175, 1990.

[32] D. E. Goldberg. Genetics-based machine learning: whence it came, where it's going. *Modelling and Methodology. Knowledge Systems' Paradigms|Modelling and Methodology. Knowledge Systems' Paradigms*, pages 285–300|xii+487, 1989.

[33] Y. Sugita and Y. Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters*, 314(1-2):141–151, 1999.

[34] T. Huber and W. F. van Gunsteren. SWARM-MD: Searching conformational space by cooperative molecular dynamics. *Journal of Physical Chemistry A*, 102(29):5937–5943, 1998.

[35] L. Rosso and M. E. Tuckerman. An adiabatic molecular dynamics method for the calculation of free energy profiles. *Molecular Simulation*, 28(1-2):91–112, 2002.

[36] P. Liu, B. Kim, R. A. Friesner, and B. J. Berne. Replica exchange with solute tempering: A method for sampling biological systems in explicit water. *Proceedings of the National Academy of Sciences of the United States of America*, 102(39):13749–13754, 2005.

[37] D. P. Geerke and W. F. van Gunsteren. The performance of non-polarizable and polarizable force-field parameter sets for ethylene glycol in molecular dynamics simulations of the pure liquid and its aqueous mixtures. *Molecular Physics*, 105:1861–1881, 2007.

[38] A. Grossfield, P. Y. Ren, and J. W. Ponder. Ion solvation thermodynamics from simulation with a polarizable force field. *Journal of the American Chemical Society*, 125(50):15671–15682, 2003.

[39] J. R. Maple, Y. X. Cao, W. G. Damm, T. A. Halgren, G. A. Kaminski, L. Y. Zhang, and R. A. Friesner. A polarizable force field and continuum solvation methodology for modeling of protein-ligand interactions. *Journal of Chemical Theory and Computation*, 1(4):694–715, 2005.

[40] T. M. Chang and L. X. Dang. Liquid-vapor interface of methanol-water mixtures: A molecular dynamics study. *Journal of Physical Chemistry B*, 109(12):5759–5765, 2005.

[41] S. A. Patel and C. L. Brooks. Revisiting the hexane-water interface via molecular dynamics simulations using nonadditive alkane-water potentials. *Journal of Chemical Physics*, 124:204706, 2006.

[42] A. Warshel, M. Kato, and A. V. Pisliakov. Polarizable force fields: History, test cases, and prospects. *Journal of Chemical Theory and Computation*, 3:2034–2045, 2007.

[43] W. S. Xie, J. Z. Pu, A. D. MacKerell, and J. L. Gao. Development of a polarizable intermolecular potential function (PIPF) for liquid amides and alkanes. *Journal of Chemical Theory and Computation*, 3:1878–1889, 2007.

[44] E. Harder, V. M. Anisimov, T. W. Whitfield, A. D. MacKerell, and B. Roux. Understanding the dielectric properties of liquid amides from a polarizable force field. *Journal of Physical Chemistry B*, 112(11):3509–3521, 2008.

[45] S. W. Rick and S. J. Stuart. Potentials and algorithms for incorporating polarizability in computer simulations. In *Reviews in Computational Chemistry, Vol 18*, Reviews in Computational Chemistry, pages 89–146. Wiley-Vch, Inc, New York, 2002.

[46] B. Guillot. A reappraisal of what we have learnt during three decades of computer simulations on water. *Journal of Molecular Liquids*, 101(1-3):219–260, 2002.

[47] H. B. Yu, T. Hansson, and W. F. van Gunsteren. Development of a simple, self-consistent polarizable model for liquid water. *Journal of Chemical Physics*, 118(1):221–234, 2003.

[48] J. Applequist, J. R. Carl, and K. K. Fung. Atom dipole interaction model for molecular polarizability - application to polyatomic-molecules and determination of atom polarizabilities. *Journal of the American Chemical Society*, 94(9):2952–2960, 1972.

[49] B. T. Thole. Molecular polarizabilities calculated with a modified dipole interaction. *Chemical Physics*, 59(3):341–350, 1981.

[50] E. Neufeld, H.B. Yu, T. Hansson, and W.F. van Gunsteren. Polarizable water models and dielectric permittivity. Technical report, ETH, Zürich, 2003.

[51] A. Baranyai and A. Bartok. Classical interaction model for the water molecule. *Journal of Chemical Physics*, 126(18):184508, 2007.

[52] G. Maroulis. Hyperpolarizability of H2O. *Journal of Chemical Physics*, 94(2):1182–1190, 1991.

[53] P. Kaatz, E. A. Donley, and D. P. Shelton. A comparison of molecular hyperpolarizabilities from gas and liquid phase measurements. *Journal of Chemical Physics*, 108(3):849–856, 1998.

[54] G. Maroulis. Hyperpolarizability of H2O revisited: accurate estimate of the basis set limit and the size of electron correlation effects. *Chemical Physics Letters*, 289(3-4):403–411, 1998.

[55] I. M. Svishchev, P. G. Kusalik, J. Wang, and R. J. Boyd. Polarizable point-charge model for water: Results under normal and extreme conditions. *Journal of Chemical Physics*, 105(11):4742–4750, 1996.

[56] Y. S. Badyal, M. L. Saboungi, D. L. Price, S. D. Shastri, D. R. Haeffner, and A. K. Soper. Electron distribution in water. *Journal of Chemical Physics*, 112(21):9206–9208, 2000.

[57] D. P. Geerke and W. F. van Gunsteren. On the calculation of atomic forces in classical simulation using the charge-on-spring method to explicitly treat electronic polarization. *Journal of Chemical Theory and Computation*, 3:2128–2137, 2007.

[58] G. S. Kell. Precise representation of volume properties of water at 1 atmosphere. *Journal of Chemical and Engineering Data*, 12(1):66–69, 1967.

[59] W. F. van Gunsteren, P. Krüger, S. R. Billeter, A. E. Mark, A. A. Eising, W. R. P. Scott, P. H. Hünenberger, and I. G. Tironi. *Biomolecular simulation : the GROMOS96 manual and user guide*. vdf, Zürich, 1996.

[60] W. R. P. Scott, P. H. Hünenberger, I. G. Tironi, A. E. Mark, S. R. Billeter, J. Fennen, A. E. Torda, T. Huber, P. Krüger, and W. F. van Gunsteren. The GROMOS biomolecular simulation program package. *Journal of Physical Chemistry A*, 103(19):3596–3607, 1999.

[61] M. Neumann. The dielectric-constant of water - computer-simulations with the mcy potential. *Journal of Chemical Physics*, 82(12):5663–5672, 1985.

[62] I. G. Tironi, R. Sperb, P. E. Smith, and W. F. van Gunsteren. A generalized reaction field method for molecular-dynamics simulations. *Journal of Chemical Physics*, 102(13):5451–5459, 1995.

[63] J. A. Hayward and J. R. Reimers. Unit cells for the simulation of hexagonal ice. *Journal of Chemical Physics*, 106(4):1518–1529, 1997.

[64] J. P. M. Postma. *MD of H2O*. PhD thesis, Rijksuniversiteit Groningen, The Netherlands, 1985.

[65] A. K. Soper. The radial distribution functions of water and ice from 220 to 673 K and at pressures up to 400 MPa. *Chemical Physics*, 258(2-3):121–137, 2000.

[66] Herman J. C. Berendsen. *Simulating the physical world*. Cambridge University Press, Cambridge, 1st edition, 2007.

[67] Walter G. Rothschild. *Dynamics of molecular liquids*. Wiley, New York, 1984.

[68] B. J. Berne and R. Pecora. *Dynamic light scattering with applications to chemistry, biology, and physics*. Wiley, New York, 1976.

[69] G. R. Fleming. *Chemical applications of ultrafast spectroscopy*. Oxford University Press, New York, 1986.

[70] M. Neumann. Dipole-moment fluctuation formulas in computer-simulations of polar systems. *Molecular Physics*, 50(4):841–858, 1983.

[71] D. van der Spoel, P. J. van Maaren, and H. J. C. Berendsen. A systematic study of water models for molecular simulation: Derivation of water models optimized for use with a reaction field. *Journal of Chemical Physics*, 108(24):10220–10230, 1998.

[72] M. Neumann, O. Steinhauser, and G. S. Pawley. Consistent calculation of the static and frequency-dependent dielectric-constant in computer-simulations. *Molecular Physics*, 52(1):97–113, 1984.

[73] J. T. Kindt and C. A. Schmuttenmaer. Far-infrared dielectric properties of polar liquids probed by femtosecond terahertz pulse spectroscopy. *Journal of Physical Chemistry*, 100(24):10373–10379, 1996.

[74] P. J. W. Debye. *Polar molecules*. Chemical Catalog Company, New York, 1929.

[75] I. G. Tironi and W. F. van Gunsteren. A molecular-dynamics simulation study of chloroform. *Molecular Physics*, 83(2):381–403, 1994.

[76] K. A. Motakabbir and M. Berkowitz. Isothermal compressibility of SPC/E water. *Journal of Physical Chemistry*, 94(21):8359–8362, 1990.

[77] B. Widom. Some topics in theory of fluids. *Journal of Chemical Physics*, 39(11):2808, 1963.

[78] D. Schweke and Y. Haas. The fluorescence of n-phenylpyrrole in an argon/acetonitrile matrix. *Journal of Physical Chemistry A*, 107(45):9554–9560, 2003.

[79] N. F. A. van der Vegt and W. F. van Gunsteren. Entropic contributions in cosolvent binding to hydrophobic solutes in water. *Journal of Physical Chemistry B*, 108(3):1056–1064, 2004.

[80] H. A. Yu and M. Karplus. A thermodynamic analysis of solvation. *Journal of Chemical Physics*, 89(4):2366–2379, 1988.

[81] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma. The missing term in effective pair potentials. *Journal of Physical Chemistry*, 91(24):6269–6271, 1987.

[82] J. Hasted. *Liquid water: Dielectric properties*. Water, A Comprehensive Treatise. Plenum Press, New York, 1972.

[83] K. Ichikawa, Y. Kameda, T. Yamaguchi, H. Wakita, and M. Misawa. Neutron-diffraction investigation of the intramolecular structure of a water molecule in the liquid-phase at high-temperatures. *molecular physics*, 73(1):79–86, 1991.

[84] S. L. Shostak, W. L. Ebenstein, and J. S. Muenter. The dipole-moment of water: Dipole-moments and hyperfine properties of H2O and HDO in the ground and excited vibrational-states. *Journal of Chemical Physics*, 94(9):5875–5882, 1991.

[85] A. J. Russell and M. A. Spackman. Vibrational averaging of electrical-properties development of a routine theoretical method for polyatomic-molecules. *Molecular Physics*, 84(6):1239–1255, 1995.

[86] D. M. Bishop and L. M. Cheung. Vibrational contributions to molecular dipole polarizabilities. *Journal of Physical and Chemical Reference Data*, 11(1):119–133, 1982.

[87] A. Glättli, X. Daura, and W. F. van Gunsteren. Derivation of an improved simple point charge model for liquid water: SPC/A and SPC/L. *Journal of Chemical Physics*, 116(22):9811–9828, 2002.

[88] P. J. van Maaren and D. van der Spoel. Molecular dynamics simulations of water with novel shell-model potentials. *Journal of Physical Chemistry B*, 105(13):2618–2626, 2001.

[89] H. A. Stern, F. Rittner, B. J. Berne, and R. A. Friesner. Combined fluctuating charge and polarizable dipole models: Application to a five-site water potential function. *Journal of Chemical Physics*, 115(5):2237–2251, 2001.

[90] G. Lamoureux, A. D. MacKerell, and B. Roux. A simple polarizable model of water based on classical Drude oscillators. *Journal of Chemical Physics*, 119(10):5185–5197, 2003.

[91] Y. J. Wu, H. L. Tepper, and G. A. Voth. Flexible simple point-charge water model with improved liquid-state properties. *Journal of Chemical Physics*, 124:024503, 2006.

[92] G. Lamoureux, E. Harder, I. V. Vorobyov, B. Roux, and Jr. MacKerell, A. D. A polarizable model of water for molecular dynamics simulations of biomolecules. *Chemical Physics Letters*, 418(1-3):245–9, 2006.

[93] F. Chen and P. E. Smith. Simulated surface tensions of common water models. *The Journal of Chemical Physics*, 126(1):221101, 2007.

[94] D. P. Geerke and W. F. van Gunsteren. Calculation of the free energy of polarization: Quantifying the effect of explicitly treating electronic polarization on the transferability of force-field parameters. *Journal of Physical Chemistry B*, 111(23):6425–6436, 2007.

[95] R. C. Weast. *Handbook of chemistry and physics*. CRC Press, Boca Raton, FL, 61th edition, 1980.

[96] K. Krynicki, C. D. Green, and D. W. Sawyer. Pressure and temperature dependence of self-diffusion in water. *Faraday Discussions of the Chemical Society*, (66):199–208, 1978.

[97] B. Halle and H. Wennerstrom. Interpretation of magnetic-resonance data from water nuclei in heterogeneous systems. *Journal of Chemical Physics*, 75(4):1928–1943, 1981.

[98] R. Ludwig. NMR relaxation studies in water-alcohol mixtures - the water-rich region. *Chemical Physics*, 195(1-3):329–337, 1995.

[99] K. Krynicki. Proton spin-lattice relaxation in pure water between zero degrees C and 100 degrees C. *Physica*, 32(1):167–178, 1966.

[100] D. R. Lide. *Handbook of chemistry and physics*. CRC Press/Taylor and Francis, Boca Raton, FL, 88th (internet version 2008) edition, 2007-2008.

[101] A. D. Buckingham. A theory of the dielectric polarization of polar substances. *Proceedings of the Royal Society of London Series a-Mathematical and Physical Sciences*, 238(1213):235–244, 1956.

[102] M. Neumann. Computer-simulation and the dielectric-constant at finite wavelength. *Molecular Physics*, 57(1):97–121, 1986.

[103] A. Morita and S. Kato. An ab initio analysis of medium perturbation on molecular polarizabilities. *Journal of Chemical Physics*, 110(24):11987–11998, 1999.

[104] L. Delle Site, A. Alavi, and R. M. Lynden-Bell. The electrostatic properties of water molecules in condensed phases: an ab initio study. *Molecular Physics*, 96(11):1683–1693, 1999.

[105] P. L. Silvestrelli and M. Parrinello. Structural, electronic, and bonding properties of liquid water from first principles. *Journal of Chemical Physics*, 111(8):3572–3580, 1999.

[106] P. J. Dyer and P. T. Cummings. Hydrogen bonding and induced dipole moments in water: Predictions from the Gaussian charge polarizable model and Car-Parrinello molecular dynamics. *The Journal of Chemical Physics*, 125(14):144519, 2006.

[107] J. A. Odutola and T. R. Dyke. Partially deuterated water dimers - microwave-spectra and structure. *Journal of Chemical Physics*, 72(9):5062–5070, 1980.

[108] L. A. Curtiss, D. J. Frurip, and M. Blander. Studies of molecular association in H2O and D2O vapors by measurement of thermal-conductivity. *Journal of Chemical Physics*, 71(6):2703–2711, 1979.

[109] K. Rottger, A. Endriss, J. Ihringer, S. Doyle, and W. F. Kuhs. Lattice-constants and thermal-expansion of H2O and D2O ice Ih between 10 and 265 K. *Acta Crystallographica Section B-Structural Science*, 50:644–648, 1994.

[110] C. A. Coulson and D. Eisenberg. Interactions of H2O molecules in ice: dipole moment of an H2O molecule in ice. *Proceedings of the Royal Society of London Series a-Mathematical and Physical Sciences*, 291(1427):445–453, 1966.

[111] E. R. Batista, S. S. Xantheas, and H. Jonsson. Molecular multipole moments of water molecules in ice Ih. *Journal of Chemical Physics*, 109(11):4546–4551, 1998.

[112] E. Whalley. Energies of the phases of ice at zero temperature and pressure. *Journal of Chemical Physics*, 81(9):4087–4092, 1984.

[113] A. Ben-Naim and Y. Marcus. Solvation thermodynamics of nonionic solutes. *Journal of Chemical Physics*, 81(4):2016–2027, 1984.

[114] H. A. Scheraga, M. Khalili, and A. Liwo. Protein-folding dynamics: Overview of molecular simulation techniques. *Annual Review of Physical Chemistry*, 58:57–83, 2007.

[115] E. Lindahl and M. S. P. Sansom. Membrane proteins: molecular dynamics simulations. *Current Opinion in Structural Biology*, 18(4):425–431, 2008.

[116] W. F. van Gunsteren and J. Dolenc. Biomolecular simulation: historical picture and future perspectives. *Biochemical Society Transactions*, 36:11–15, 2008.

[117] F. Khalili-Araghi, J. Gumbart, P. C. Wen, M. Sotomayor, E. Tajkhorshid, and K. Schulten. Molecular dynamics simulations of membrane channels and transporters. *Current Opinion in Structural Biology*, 19(2):128–137, 2009.

[118] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules (vol 117, pg 5179, 1995). *Journal of the American Chemical Society*, 118(9):2309–2309, 1996.

[119] W. L. Jorgensen, D. S. Maxwell, and J. TiradoRives. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society*, 118(45):11225–11236, 1996.

[120] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *Journal of Physical Chemistry B*, 102(18):3586–3616, 1998.

[121] P. Y. Ren and J. W. Ponder. Polarizable atomic multipole water model for molecular mechanics simulation. *Journal of Physical Chemistry B*, 107(24):5933–5947, 2003.

[122] A. P. E. Kunz and W. F. van Gunsteren. Development of a nonlinear classical polarization model for liquid water and aqueous solutions: COS/D. *Journal of Physical Chemistry A*, 113(43):11570–11579, 2009.

[123] T. M. Chang, L. X. Dang, and K. A. Peterson. Computer simulation of chloroform with a polarizable potential model. *Journal of Physical Chemistry B*, 101(17):3413–3419, 1997.

[124] G. Lamoureux, J. D. Faraldo-Gomez, S. Krupin, and S. Y. Noskov. Polarizable model of chloroform based on classical Drude oscillators. *Chemical Physics Letters*, 468(4-6):270–274, 2009.

[125] Z. Lin, A. P. E. Kunz, and W. F. van Gunsteren. A one-site polarisable model for liquid chloroform: COS/C. *Molecular Physics*, 108(13):1749 – 1757, 2010.

[126] S. Patel and C. L. Brooks. A nonadditive methanol force field: Bulk liquid and liquid-vapor interfacial properties via molecular dynamics simulations using a fluctuating charge model. *Journal of Chemical Physics*, 122(2):024508, 2005.

[127] S. Patel and C. L. Brooks. Structure, thermodynamics, and liquid-vapor equilibrium of ethanol from molecular-dynamics simulations using nonadditive interactions. *Journal of Chemical Physics*, 123(16):164502, 2005.

[128] H. B. Yu, D. P. Geerke, H. Y. Liu, and W. F. van Gunsteren. Molecular dynamics simulations of liquid methanol and methanol-water mixtures with polarizable models. *Journal of Computational Chemistry*, 27(13):1494–1504, 2006.

[129] T. M. Chang and L. X. Dang. Molecular dynamics simulations of CCl4-H2O liquid-liquid interface with polarizable potential models. *Journal of Chemical Physics*, 104(17):6772–6783, 1996.

[130] O. Acevedo and W. L. Jorgensen. Exploring solvent effects upon the menshutkin reaction using a polarizable force field. *Journal of Physical Chemistry B*, 114(25):8425–8430, 2010.

[131] I. G. Tironi, P. Fontana, and W. F. van Gunsteren. A molecular dynamics simulation study of liquid carbon tetrachloride. *Molecular Simulation*, 18(1-2):1–11, 1996.

[132] W. A. B. Evans and J. G. Powles. The computer-simulation of the dielectric-properties of polar liquids - the dielectric-constant and relaxation of liquid-hydrogen chloride. *Molecular Physics*, 45(3):695–707, 1982.

[133] M. Christen, P. H. Hünenberger, D. Bakowies, R. Baron, R. Bürgi, D. P. Geerke, T. N. Heinz, M. A. Kastenholz, V. Kräutler, C. Oostenbrink, C. Peter, D. Trzesniak, and W. F. van Gunsteren. The GROMOS software for biomolecular simulation: GROMOS05. *Journal of Computational Chemistry*, 26(16):1719–1751, 2005.

[134] L. van Hove. Correlations in space and time and born approximation scattering in systems of interacting particles. *Phys. Rev.*, 95:249, 1954.

[135] P. Debye. Interference of X rays and heat movement. *Annalen der Physik*, 43(1):49–95, 1913.

[136] C.G. Windsor. *Chemical Applications of Thermal Neutron Scattering*. Oxford University Press, Oxford, 1973.

[137] S. Shibata and L. S. Bartell. Electron-diffraction study of water and heavy water. *Journal of Chemical Physics*, 42(4):1147, 1965.

[138] I. R. McDonald, D. G. Bounds, and M. L. Klein. Molecular-dynamics calculations for the liquid and cubic plastic crystal phases of carbon-tetrachloride. *Molecular Physics*, 45(3):521–542, 1982.

[139] T. M. Chang, K. A. Peterson, and L. X. Dang. Molecular-dynamics simulations of liquid, interface, and ionic solvation of polarizable carbon-tetrachloride. *Journal of Chemical Physics*, 103(17):7502–7513, 1995.

[140] J. C. Soetens, G. Jansen, and C. Millot. Molecular dynamics simulation of liquid CCl4 with a new polarizable potential model. *Molecular Physics*, 96(7):1003–1012, 1999.

[141] A. H. T. Li, S. C. Huang, and S. D. Chao. Molecular dynamics simulation of liquid carbon tetrachloride using ab initio force field. *Journal of Chemical Physics*, 132(2):024506, 2010.

[142] A.W. Adamson. *Physical Chemistry of Surfaces*. Wiley, New York, 1990.

[143] A. Chahid, F. J. Bermejo, E. Enciso, M. G. Hernandez, and J. L. Martinez. Single-particle dynamics of liquid CCl4 - a comparison of molecular-dynamics and neutron quasi-elastic scattering results. *Journal of Physics-Condensed Matter*, 4(5):1213–1231, 1992.

[144] V. Majer, L. Svab, and V. Svoboda. Enthalpies of vaporization and cohesive energies for a group of chlorinated hydrocarbons. *Journal of Chemical Thermodynamics*, 12(9):843–847, 1980.

[145] M. Misawa. Temperature-dependence of structure of liquid carbon-tetrachloride measured by pulsed neutron total scattering. *Journal of Chemical Physics*, 91(9):5648–5654, 1989.

[146] L. Pusztai and R. L. McGreevy. The structure of liquid CCl4. *Molecular Physics*, 90(4):533–539, 1997.

[147] C. D. Christ, A. E. Mark, and W. F. van Gunsteren. Basic ingredients of free energy calculations: A review. *Journal of Computational Chemistry*, 31(8):1569–1582, 2010.

[148] A. Liwo, C. Czaplewski, S. Oldziej, and H. A. Scheraga. Computational techniques for efficient conformational sampling of proteins. *Current Opinion in Structural Biology*, 18(2):134–139, 2008.

[149] G. Jacucci and A. Rahman. *The possibility of using a larger time step in MD studies of water*, pages 32–40. Orsay, France, 1974.

[150] R. Pomes and J. A. McCammon. Mass and step length optimization for the calculation of equilibrium properties by molecular-dynamics simulation. *Chemical Physics Letters*, 166(4):425–428, 1990.

[151] B. Mao and A. R. Friedman. Molecular-dynamics simulation by atomic mass weighting. *Biophysical Journal*, 58(3):803–805, 1990.

[152] B. Mao, G. M. Maggiora, and K. C. Chou. Mass-weighted molecular-dynamics simulation of cyclic polypeptides. *Biopolymers*, 31(9):1077–1086, 1991.

[153] K. A. Feenstra, B. Hess, and H. J. C. Berendsen. Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *Journal of Computational Chemistry*, 20(8):786–798, 1999.

[154] C.H. Bennett. *Mass tensor MD.*, pages 41–63. CECAM, Orsay, France, 1974.

[155] C. H. Bennett. Mass tensor molecular-dynamics. *Journal of Computational Physics*, 19(3):267–279, 1975.

[156] P. J. Gee and W. F. van Gunsteren. Numerical simulation of the effect of solvent viscosity on the motions of a β-peptide heptamer. *Chemistry-a European Journal*, 12(1):72–75, 2006.

[157] P. H. Nguyen. Replica exchange simulation method using temperature and solvent viscosity. *Journal of Chemical Physics*, 132(14):144109, 2010.

[158] L. Rosso, P. Minary, Z. W. Zhu, and M. E. Tuckerman. On the use of the adiabatic molecular dynamics technique in the calculation of free energy profiles. *Journal of Chemical Physics*, 116(11):4389–4402, 2002.

[159] J. B. He, Z. Y. Zhang, Y. Y. Shi, and H. Y. Liu. Efficiently explore the energy landscape of proteins in molecular dynamics simulations by amplifying collective motions. *Journal of Chemical Physics*, 119(7):4005–4017, 2003.

[160] Z. Y. Zhang, Y. Y. Shi, and H. Y. Liu. Molecular dynamics simulations of peptides and proteins with amplified collective motions. *Biophysical Journal*, 84(6):3583–3593, 2003.

[161] L. Rosso, J. B. Abrams, and M. E. Tuckerman. Mapping the backbone dihedral free-energy surfaces in small peptides in solution using adiabatic free-energy dynamics. *Journal of Physical Chemistry B*, 109(9):4162–4167, 2005.

[162] J. B. Abrams and M. E. Tuckerman. Efficient and direct generation of multidimensional free energy surfaces via adiabatic dynamics without coordinate transformations. *Journal of Physical Chemistry B*, 112(49):15742–15757, 2008.

[163] L. Maragliano and E. Vanden-Eijnden. A temperature accelerated method for sampling free energy and determining reaction pathways in rare events simulations. *Chemical Physics Letters*, 426(1-3):168–175, 2006.

[164] J. B. Abrams, L. Rosso, and M. E. Tuckerman. Efficient and precise solvation free energies via alchemical adiabatic molecular dynamics. *Journal of Chemical Physics*, 125(7):074115, 2006.

[165] S. Melchionna. Enhanced sampling of rare events. *Physical Review E*, 62(6):8762–8767, 2000.

[166] J. MacFadyen and I. Andricioaei. A skewed-momenta method to efficiently generate conformational-transition trajectories. *Journal of Chemical Physics*, 123(7):074107, 2005.

[167] S. Melchionna. Numerical integration of projective Hamiltonian dynamics. *Molecular Physics*, 104(19):3045–3049, 2006.

[168] P. Liu, X. H. Huang, R. H. Zhou, and B. J. Berne. Hydrophobic aided replica exchange: an efficient algorithm for protein folding in explicit solvent. *Journal of Physical Chemistry B*, 110(38):19018–19022, 2006.

[169] X. H. Huang, M. Hagen, B. Kim, R. A. Friesner, R. H. Zhou, and B. J. Berne. Replica exchange with solute tempering: Efficiency in large scale systems. *Journal of Physical Chemistry B*, 111(19):5405–5410, 2007.

[170] R. C. van Schaik, W. F. van Gunsteren, and H. J. C. Berendsen. Conformation search by potential energy annealing algorithm and application to cyclosporin a. *Journal of Computer-Aided Molecular Design*, 6(2):97–112, 1992.

[171] X. W. Wu and S. M. Wang. Self-guided molecular dynamics simulation for efficient conformational search. *Journal of Physical Chemistry B*, 102(37):7238–7250, 1998.

[172] X. W. Wu and S. M. Wang. Enhancing systematic motion in molecular dynamics simulation. *Journal of Chemical Physics*, 110(19):9401–9410, 1999.

[173] W. Shinoda and M. Mikami. Self-guided molecular dynamics in the isothermal-isobaric ensemble. *Chemical Physics Letters*, 335(3-4):265–272, 2001.

[174] W. R. P. Scott, A. E. Mark, and W. F. van Gunsteren. On using time-averaging restraints in molecular dynamics simulation. *Journal of Biomolecular NMR*, 12(4):501–508, 1998.

[175] I. Andricioaei, A. R. Dinner, and M. Karplus. Self-guided enhanced sampling methods for thermodynamic averages. *Journal of Chemical Physics*, 118(3):1074–1084, 2003.

[176] X. W. Wu and B. R. Brooks. Self-guided Langevin dynamics simulation method. *Chemical Physics Letters*, 381(3-4):512–518, 2003.

[177] A. Damjanovic, X. W. Wu, B. Garcia-Moreno, and B. R. Brooks. Backbone relaxation coupled to the ionization of internal groups in proteins: A self-guided Langevin dynamics study. *Biophysical Journal*, 95(9):4091–4101, 2008.

[178] A. E. Torda, R. M. Scheek, and W. F. van Gunsteren. Time-dependent distance restraints in molecular-dynamics simulations. *Chemical Physics Letters*, 157(4):289–294, 1989.

[179] R. H. Swendsen and J. S. Wang. Replica Monte-Carlo simulation of spin-glasses. *Physical Review Letters*, 57(21):2607–2609, 1986.

[180] K. Hukushima and K. Nemoto. Exchange Monte Carlo method and application to spin glass simulations. *Journal of the Physical Society of Japan*, 65(6):1604–1608, 1996.

[181] Y. Sugita, A. Kitao, and Y. Okamoto. Multidimensional replica-exchange method for free-energy calculations. *Journal of Chemical Physics*, 113(15):6042–6051, 2000.

[182] T. Z. Lwin and R. Luo. Overcoming entropic barrier with coupled sampling at dual resolutions. *Journal of Chemical Physics*, 123(19):194905, 2005.

[183] M. Christen and W. F. van Gunsteren. Multigraining: An algorithm for simultaneous fine-grained and coarse-grained simulation of molecular systems. *Journal of Chemical Physics*, 124(15):154106, 2006.

[184] C. J. Woods, J. W. Essex, and M. A. King. The development of replica-exchange-based free-energy methods. *Journal of Physical Chemistry B*, 107(49):13703–13710, 2003.

[185] J. A. Barker and R. O. Watts. Monte-Carlo studies of dielectric properties of water-like models. *Molecular Physics*, 26(3):789–792, 1973.

[186] L. Perić, C. S. Pereira, S. Perez, and P. H. Hünenberger. Conformation, dynamics and ion-binding properties of single-chain polyuronates: a molecular dynamics study. *Molecular Simulation*, 34(4):421–446, 2008.

[187] L. Perić-Hassler and P. H. Hünenberger. Interaction of alginate single-chain polyguluronate segments with mono- and divalent metal cations: a comparative molecular dynamics study. *Molecular Simulation*, 36(10):778–795, 2010.

[188] Clara D. Christ and Wilfred F. van Gunsteren. Enveloping distribution sampling: a method to calculate free energy differences from a single simulation. *J Chem Phys*, 126(18):184110, 2007.

[189] A. P. E. Kunz, H. Liu, and W. F. van Gunsteren. Enhanced sampling of particular degrees of freedom in molecular systems based on adiabatic decoupling and temperature of force scaling. *Journal of Chemical Physics*, 2011.

[190] T. C. Beutler, A. E. Mark, R. C. van Schaik, P. R. Gerber, and W. F. van Gunsteren. Avoiding singularities and numerical instabilities in free-energy calculations based on molecular simulations. *Chemical Physics Letters*, 222(6):529–539, 1994.

[191] H. Fukunishi, O. Watanabe, and S. Takada. On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction. *Journal of Chemical Physics*, 116(20):9058–9067, 2002.

[192] C. B. Anfinsen. Principles that govern folding of protein chains. *Science*, 181(4096):223–230, 1973.

[193] M. J. Gething and J. Sambrook. Protein folding in the cell. *Nature*, 355(6355):33–45, 1992.

[194] S. E. Radford and C. M. Dobson. From computer simulations to human disease: Emerging themes in protein folding. *Cell*, 97(3):291–298, 1999.

[195] W. F. van Gunsteren, R. Bürgi, C. Peter, and X. Daura. The key to solving the protein-folding problem lies in an accurate description of the denatured state. *Angewandte Chemie-International Edition*, 40(2):351–355, 2001.

[196] C. M. Dobson. Protein folding and misfolding. *Nature*, 426(6968):884–890, 2003.

[197] D. J. Selkoe. Folding proteins in fatal ways. *Nature*, 426(6968):900–904, 2003.

[198] G. D. Rose, P. J. Fleming, J. R. Banavar, and A. Maritan. A backbone-based theory of protein folding. *Proceedings of the National Academy of Sciences of the United States of America*, 103(45):16623–16633, 2006.

[199] F. U. Hartl and M. Hayer-Hartl. Converging concepts of protein folding in vitro and in vivo. *Nature Structural & Molecular Biology*, 16(6):574–581, 2009.

[200] D. E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. B. Shan, and W. Wriggers. Atomic-level characterization of the structural dynamics of proteins. *Science*, 330(6002):341–346, 2010.

[201] D. Seebach and J. L. Matthews. β-peptides: a surprise at every turn. *Chemical Communications*, (21):2015–2022, 1997.

[202] K. Gademann, T. Hintermann, and J. V. Schreiber. β-peptides: Twisting and turning. *Current Medicinal Chemistry*, 6(10):905–925, 1999.

[203] R. P. Cheng, S. H. Gellman, and W. F. DeGrado. β-peptides: From structure to function. *Chemical Reviews*, 101(10):3219–3232, 2001.

[204] D. Seebach, S. Abele, J. V. Schreiber, B. Martinoni, A. K. Nussbaum, H. Schild, H. Schulz, H. Hennecke, R. Woessner, and F. Bitsch. Biological and pharmacokinetic studies with β-peptides. *Chimia*, 52(12):734–739, 1998.

[205] Y. Hamuro, J. P. Schneider, and W. F. DeGrado. De novo design of antibacterial β-peptides. *Journal of the American Chemical Society*, 121(51):12200–12201, 1999.

[206] M. Werder, H. Hauser, S. Abele, and D. Seebach. β-peptides as inhibitors of small-intestinal cholesterol and fat absorption. *Helvetica Chimica Acta*, 82(10):1774–1783, 1999.

[207] D. F. Hook, F. Gessier, C. Noti, P. Kast, and D. Seebach. Probing the proteolytic stability of β-peptides containing alpha-fluoro- and alpha-hydroxy-β-amino acids. *Chembiochem*, 5(5):691–706, 2004.

[208] X. Daura, W. F. van Gunsteren, D. Rigo, B. Jaun, and D. Seebach. Studying the stability of a helical β-heptapeptide by molecular dynamics simulations. *Chemistry-a European Journal*, 3(9):1410–1417, 1997.

[209] X. Daura, B. Jaun, D. Seebach, W. F. van Gunsteren, and A. E. Mark. Reversible peptide folding in solution by molecular dynamics simulation. *Journal of Molecular Biology*, 280(5):925–932, 1998.

[210] X. Daura, K. Gademann, B. Jaun, D. Seebach, W. F. van Gunsteren, and A. E. Mark. Peptide folding: When simulation meets experiment. *Angewandte Chemie-International Edition*, 38(1-2):236–240, 1999.

[211] D. Trzesniak, A. Glättli, B. Jaun, and W. F. van Gunsteren. Interpreting NMR data for β-peptides using molecular dynamics simulations. *Journal of the American Chemical Society*, 127(41):14320–14329, 2005. Times Cited: 22.

[212] X. Daura, K. Gademann, H. Schafer, B. Jaun, D. Seebach, and W. F. van Gunsteren. The β-peptide hairpin in solution: Conformational study of a β-hexapeptide in methanol by NMR spectroscopy and MD simulation. *Journal of the American Chemical Society*, 123(10):2393–2404, 2001.

[213] D. Seebach, P. E. Ciceri, M. Overhand, B. Jaun, D. Rigo, L. Oberer, U. Hommel, R. Amstutz, and H. Widmer. Probing the helical secondary structure of short-chain β-peptides. *Helvetica Chimica Acta*, 79(8):2043–2066, 1996.

[214] L. D. Schuler, X. Daura, and W. F. van Gunsteren. An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase. *Journal of Computational Chemistry*, 22(11):1205–1218, 2001.

[215] T. C. Beutler and W. F. van Gunsteren. The computation of a potential of mean force: Choice of the biasing potential in the umbrella sampling technique. *Journal of Chemical Physics*, 100(2):1492, 1993.

[216] W. L. Jorgensen. Internal rotation of liquid 1,2-dichloroethane and n-butane. *Journal of the American Chemical Society*, 103:677, 1980.

[217] X. Lin, N. Schmid, and W. F. van Gunsteren. The effect of using a polarizable solvent model upon the folding equilibrium of different β-peptides. *in press, Journal of Physical Chemistry B*, 2010.

[218] B. Alberts, D. Bray, J. Lewis, M. Raff, K. Roberts, and J. D. Watson. Molecular biology of the cell. *Alberts, B., D. Bray, J. Lewis, M. Raff, K. Roberts and J. D. Watson. Molecular Biology of the Cell, VCh Publishers: New York, New York, USA. Illus*, pages XLVIII+1310P, 1986.

[219] A. Fiser, R. K. G. Do, and A. Sali. Modeling of loops in protein structures. *Protein Science*, 9(9):1753–1773, 2000.

[220] B. Borah, C. W. Chen, W. Egan, M. Miller, A. Wlodawer, and J. S. Cohen. Nuclear magnetic-resonance and neutron-diffraction studies of the complex of ribonuclease-A with uridine vanadate, a transition-state analog. *Biochemistry*, 24(8):2058–2067, 1985.

[221] X. Daura, W. F. van Gunsteren, and A. E. Mark. Folding-unfolding thermodynamics of a β-heptapeptide from equilibrium simulations. *Proteins-Structure Function and Genetics*, 34(3):269–280, 1999.

[222] S. Riniker, A. P. E. Kunz, and W. F. van Gunsteren. On the calculation of the dielectric permittivity of molecular models in the liquid phase. *submitted to Journal of Chemical Theory and Computation*, 2010.

[223] C. D. Christ and W. F. van Gunsteren. Comparison of three enveloping distribution sampling Hamiltonians for the estimation of multiple free energy differences from a single simulation. *Journal of Computational Chemistry*, 30(11):1664–1679, 2009.

[224] H. S. Hansen and P. H. Hünenberger. Ball-and-stick local elevation umbrella sampling: Molecular simulations involving enhanced sampling within conformational or alchemical subspaces of low internal dimensionalities, minimal irrelevant volumes, and problem-adapted geometries. *Journal of Chemical Theory and Computation*, 6(9):2622–2646, 2010.