Karsten Kreis

Advanced Adaptive Resolution Methods for Molecular Simulation

Advanced Adaptive Resolution Methods for Molecular Simulation

D I S S E R T A T I O N

zur Erlangung des Grades

„Doktor der Naturwissenschaften"

am Fachbereich Physik, Mathematik und Informatik der Johannes Gutenberg-Universität in Mainz

Karsten Kreis

geb. in Bremen

Angefertigt am Max-Planck-Institut für Polymerforschung

Mainz, 30. Mai 2017

Datum der mündlichen Prüfung:

11.01.2018

D77 - Kumulative Dissertation der Johannes Gutenberg-Universität Mainz

Abstract

Computer simulations have become a fundamental tool in molecular soft matter research. For example, they allow the investigation of microscopic details that are not accessible experimentally. However, even the most powerful supercomputers cannot simulate large systems on long time scales with arbitrarily high accuracy. Therefore, a large variety of computational methods have been developed, which range from highly accurate but numerically expensive quantum calculations to efficient but less detailed coarse-grained approaches.

A particular challenge is posed by systems whose macroscopic behavior is sensitively dependent on specific microscopic details. For these multiscale systems, simulation techniques are being developed that concurrently use two or more models with different computational complexity and accuracy in single simulations. In this way, one can use a demanding high-resolution model in a small but relevant subregion, while allowing overall large and long simulations by describing the rest of the system with an efficient, less detailed level. One such method is the adaptive resolution simulation scheme. In this approach, a predefined region of interest, for example a protein in a large box of solvent, is modeled atomistically while molecules far away from this region are described by a coarse-grained force field. A unique feature of the technique is that particles traveling between the two regions change their resolution on the fly, such that the high-resolution subsystem behaves as if embedded in an overall high-resolution environment, although at a significantly lower computational cost.

In this work, we first study the ability of adaptive resolution simulation methods to combine atomistic and coarse-grained models with very different thermodynamic properties. By coupling a highly structured liquid like water with an ideal gas of non-interacting particles, we demonstrate that the details of the coarse-grained region have a surprisingly small effect on the accuracy of the high resolution region. Next, we investigate the theoretical basis behind the described adaptive resolution approaches and show how a single unifying framework can be used to derive different kinds of adaptive resolution schemes. We also demonstrate that the relative entropy—a quantity characterizing differences in the configurational probability distributions of the models in the different regions—can be used as a guide to set up adaptive resolution simulations in an optimal manner. Furthermore, we devise a simulation algorithm that enables atomistic regions with arbitrary geometry and which can adapt during the simulation to follow, for example, the conformational change of a large biomolecule. Finally, we derive a new adaptive resolution scheme that allows a clean coupling of quantum mechanical path integral and classical atomistic models.

The results significantly advance the current state of adaptive resolution simulation methodologies both on the theoretical and the practical front. They shed light on the fundamentals behind such methods and enable more efficient computer simulations of relevant multiscale systems, such as complex biomolecules, membranes, DNA, or polymeric materials.

Zusammenfassung

Computersimulationen spielen heutzutage eine wesentliche Rolle in der Erforschung weicher Materie. Allerdings können selbst die rechenstärksten Supercomputer keine großen Systeme über lange Zeitskalen mit beliebig hoher Genauigkeit berechnen. Dies führte zur Entwicklung unterschiedlichster Simulationstechniken, von extrem genauen, aber sehr rechenaufwendigen quantenmechanischen Modellen bis hin zu effizienten, stark vergröberten Ansätzen.

Eine besondere Herausforderung stellen sogenannte Multiskalensysteme dar, bei welchen Phänomene und Effekte auf unterschiedlichen Zeit- und Längenskalen miteinander gekoppelt sind. Hierfür entwickelte Multiskalensimulationstechniken kombinieren mehrere Modelle mit unterschiedlicher Genauigkeit und Berechnungskomplexität in einzelnen Simulationen. So kann ein rechenaufwendiges, aber genaues Modell in einem kleinen, aber relevanten Teil der Simulationsbox genutzt werden, während eine effiziente, gröbere Beschreibung im Rest des Systems trotzdem eine insgesamt große und lange Simulation ermöglicht. Eine solche Methode ist das Adaptive Resolution Simulation Scheme. Eine festgelegte Region von hohem Interesse, beispielsweise ein gelöstes Protein oder eine Grenzfläche, wird atomistisch modelliert, während weiter entfernte Moleküle vergröbert beschrieben werden. Teilchen, die sich zwischen diesen Regionen bewegen, ändern ihre Auflösung entsprechend, sodass sich der atomistische Bereich verhält, als wäre er in ein komplett hochaufgelöstes System eingebettet, nur zu einem wesentlich geringeren Rechenaufwand.

Als erstes beleuchten wir in dieser Arbeit, inwieweit Adaptive Resolution-Methoden Modelle mit sehr unterschiedlichen thermodynamischen Eigenschaften kombinieren können, und wir zeigen, dass man sogar eine komplexe Flüssigkeit wie Wasser mit einem idealen Gas aus nicht-wechselwirkenden Teilchen koppeln kann. Dann untersuchen wir die theoretischen Grundlagen hinter Adaptive Resolution-Techniken und demonstrieren, wie von einem grundlegenden Ansatz unterschiedliche Adaptive Resolution-Algorithmen hergeleitet werden können. Weiterhin zeigen wir, wie die relative Entropie zwischen den verschiedenen Modellen helfen kann, das adaptive System optimal aufzusetzen. Darüber hinaus entwerfen wir einen Algorithmus, welcher atomistische Regionen in beliebiger Form ermöglicht und welcher außerdem deren automatische Anpassung wahrend der Simulation erlaubt, um zum Beispiel der Konfigurationsänderung eines Proteins zu folgen. Zuletzt leiten wir eine neue Adaptive Resolution-Methode her, welche eine saubere Kopplung von quantenmechanischen Pfadintegral-basierten mit klassischen atomistischen Modellen erlaubt.

Die Ergebnisse der Arbeit entwickeln Adaptive Resolution-Techniken sowohl vom theoretischen Gesichtspunkt wie auch für die praktische Anwendung wesentlich weiter. Sie werfen ein neues Licht auf die Konzepte und Mechanismen hinter Adaptive Resolution-Algorithmen und ermöglichen effizientere Computersimulationen wichtiger Multiskalensysteme, wie komplexe biologische Makromoleküle, Membranen, DNA, oder polymerische Materialien.

Contents

List of Figures

List of Tables

Included publications

This PhD thesis is the result of my doctoral studies completed at the Max Planck Instutite for Polymer Research in Mainz and at New York University in New York City. My studies were performed within the framework of the Graduate School of Excellence Materials Science in Mainz. The results have been published in several peer-reviewed scientific papers and the thesis has been submitted to the Johannes Gutenberg University Mainz as a cumulative dissertation (thesis by publication). The chapters 2-6 consist of the reprinted publications and manuscripts listed below and chapter 7 contains a draft for another research paper that describes extensions of the results from chapter 6. A modified version of this draft has been published in The Journal of Chemical Physics after the submission of this thesis. The contributions by the different authors are explained in appendix A. Furthermore, all algorithms developed in this thesis are implemented in the ESPResSo++ [1] software package and publicly available.

Chapter 2:
Karsten Kreis, Aoife C. Fogarty, Kurt Kremer, and Raffaello Potestio
Advantages and challenges in coupling an ideal gas to atomistic models in adaptive resolution simulations
European Physical Journal Special Topics 224, 2289-2304 (2015)
DOI: 10.1140/epjst/e2015-02412-1

Chapter 3:
Karsten Kreis, Davide Donadio, Kurt Kremer, and Raffaello Potestio
A unified framework for Force-Based and Energy-Based Adaptive Resolution Simulations
EPL (Europhysics Letters) 108, 30007 (2014)
DOI: 10.1209/0295-5075/108/30007

Chapter 4:
Karsten Kreis and Raffaello Potestio
The relative entropy is fundamental to adaptive resolution simulations
Journal of Chemical Physics 145, 044104 (2016)
DOI: 10.1063/1.4959169

Chapter 5:
Karsten Kreis, Raffaello Potestio, Kurt Kremer, and Aoife C. Fogarty
Adaptive Resolution Simulations with Self-Adjusting High-Resolution Regions
Journal of Chemical Theory and Computation 12, 4067-4081 (2016)
DOI: 10.1021/acs.jctc.6b00440
• Featured on the journal cover, Volume 12, Issue 10 (October 11, 2016)

Chapter 6:
Karsten Kreis, Mark E. Tuckerman, Davide Donadio, Kurt Kremer, and Raffaello Potestio
From Classical to Quantum and Back: A Hamiltonian Scheme for Adaptive Multiresolution Classical/Path-Integral Simulations
Journal of Chemical Theory and Computation 12, 3030-3039 (2016)
DOI: 10.1021/acs.jctc.6b00242

Chapter 7:
Draft for another research paper extending the results from chapter 6. A modified version of this draft has been published in The Journal of Chemical Physics after the submission of this thesis:
Karsten Kreis, Kurt Kremer, Raffaello Potestio, and Mark E. Tuckerman
From classical to quantum and back: Hamiltonian adaptive resolution path integral, ring polymer, and centroid molecular dynamics
Journal of Chemical Physics 147, 244104 (2017)
DOI: 10.1063/1.5000701

Chapter 1.

Introduction

The focus of this thesis is the adaptive resolution simulation scheme, a computational method for the efficient multiscale simulation of molecular systems. To set the work in the right context, we first recapitulate in section 1.1 of this introductory chapter the emergence and development of computational molecular modeling in general. In particular, we provide an overview over the most widely used simulation techniques. Next, in section 1.2 we discuss the idea of multiscale modeling and review some of its applications. Section 1.3 is devoted to the adaptive resolution simulation scheme itself, which we present in full detail. In section 1.4, we provide an outline of the following chapters.

1.1. History and overview

The history of computational molecular modeling reaches back to the 50's of the 20$^{\text{th}}$ century. In 1953, Enrico Fermi, John Pasta, Stanislaw Ulam, and Mary Tsingou performed numerical simulations of a vibrating string to study ergodicity and thermalization [2]. Contrary to their expectation, they did not observe an equipartition of energy but a quasi-periodic behavior.[1] This phenomenon is now well-known as the Fermi-Pasta-Ulam problem. Few years later in 1959, the first molecular dynamics (MD) simulation was run by Bernie J. Alder and Thomas E. Wainwright, who studied elastic collisions between hard spheres [4]. Alder was awarded the Boltzmann Medal in 2001 and the National Medal of Science in 2009 for these pioneering works, which mark the invention of the MD simulation method. Independently, MD approaches were also developed by Aneesur Rahman. His simulations of liquid argon using a Lennard-Jones potential are another important landmark in computational molecular modeling [5]. The Aneesur Rahman Prize is now the most prestigious award in computational physics given by the American Physical Society.

Nowadays, computer simulations play an important role in molecular modeling. Molecular dynamics and related techniques are widely applied in chemical physics, material science, as well as in biochemistry and biophysics. They are used to study and characterize new chemical compounds [6–8], for example in the development of materials for organic electronics

[1] Later, in 1965, the puzzle was solved by Norman J. Zabusky and Martin D. Kruskal. They investigated the studied system with the Korteweg-de Vries equation and derived non-linear soliton solutions. This explained the broken ergodicity and the observed quasi-periodicity [3].

[9–11], or to investigate the functionality of biological membranes [12–16] and the folding of proteins [17–20]. However, despite ever growing computer power, the introduction of efficient GPU computation, as well as modern parallelization techniques that enable the coupling of thousands of CPUs, many interesting applications remain challenging or intractable with todays computational capabilities. For example, many proteins fold on timescales far beyond milliseconds, while the longest atomistic protein folding simulations reach only a few milliseconds [20–24]. Similar challenges also occur on much smaller time and length scales: Accurate quantum mechanical simulations are still only possible for systems consisting of up to hundreds or thousands of atoms, but not for larger objects or complex liquids [25–27].

These computational challenges have driven the continuous development of advanced simulation methods for molecular modeling. In general, these techniques can be described in a hierarchy (see Fig. 1.1). At one end, there are approaches that incorporate quantum mechanical

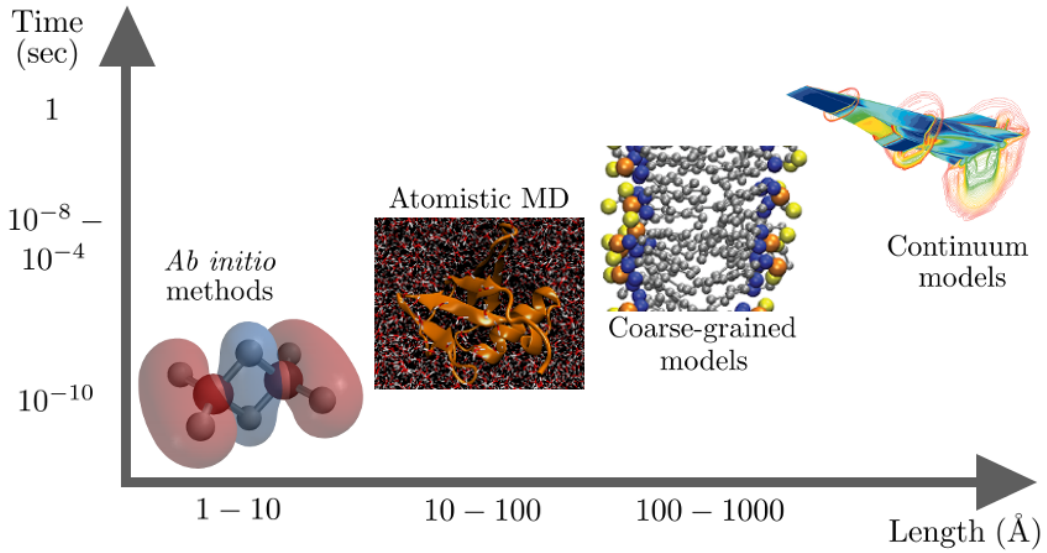

Figure 1.1.: Simulation methods for modeling on different scales. On the one hand, highly accurate techniques, such as *ab initio* schemes, are computationally expensive and hence only applicable on small length and short time scales. On the other hand, numerically efficient but less accurate approaches, like coarse-grained models, enable longer simulations of larger systems.

effects and aim for an accurate atomistic and electronic description of molecules. These schemes are computationally very demanding and hence only suitable for the study of small systems on short time scales.[2] On the other end, there are methods that were designed for the investigation of large, macroscopic systems and for the simulation of much longer physical time scales. In this spectrum, modeling techniques employed for the computational study of soft matter typically range from quantum mechanical so-called *ab initio* schemes that aim to solve the Schrödinger equation from first principles, over empirical particle-based atomistic (AT) and coarse-grained

[2] Whenever we talk about the time or length scales of simulations, we refer to the physical times and lengths that are simulated, not to the real time that a simulation needs to run on a computer.

(CG) simulations, to continuum approaches.[3] In the following, we will provide an overview of the most relevant and widely used simulation approaches, starting with techniques for an accurate modeling of microscopic systems.

1.1.1. Density functional theory and beyond

One of the most popular *ab initio* methods is density functional theory (DFT), providing an approximate solution of the Schrödinger equation [26, 27, 31]. It builds on the Born-Oppenheimer approximation, which assumes that the wavefunctions of nuclei and electrons can be separated. Then, it can be shown that the ground state of the system is described only by the three-dimensional electron density and that a unique energy functional exists which is minimized by the correct ground state density [32]. The latter observations are known as the Hohenberg-Kohn theorems, first proven by Pierre C. Hohenberg and Walter Kohn. Unfortunately, though, the unique energy functional is unknown. The challenge of DFT lies in finding a suitable approximation of the exchange-correlation potential, the part of the unique energy functional which describes the many-body interactions between the electrons. However, various approaches exist that allow reliable calculations of relevant chemical and physical quantities [33]. Furthermore, DFT can be applied in the context of molecular dynamics and used to calculate forces in an atomistic system based on the electronic interactions. The atoms can then be propagated accordingly for a small time step and the electronic problem is solved again. This method is known as Born-Oppenheimer molecular dynamics [34]. Since the computational cost of DFT is still moderate compared to many other quantum chemical simulation techniques, it has become one of the most popular and widely used *ab initio* methods in computational chemistry and physics. In comparison with coarser methods that do not treat the electrons explicitly, DFT is typically applied when the microscopic electronic structure is of special relevance or interest. It is used, for example, to calculate molecular energies [35], to describe chemical reactions [36, 37], to study complex liquids [38], and to accurately model complex materials with unique electronic characteristics such as semiconductors [10, 11, 37, 39]. In 1998, Walter Kohn was awarded the Nobel Prize in Chemistry for his development of DFT, together with John A. Pople, who derived a theory for the approximate calculation of molecular orbitals in π-electron systems [30, 40].

Apart from DFT and its various flavors, there is a whole zoo of further *ab initio* methods, which all use different approaches to approximately solve the many-body Schrödinger equation. Notable techniques include Hartree–Fock [41–43], which assumes that the energy eigenfunctions are products of one-electron wave functions, as well as Post-Hartree–Fock schemes [44–46], which improve on the regular Hartree–Fock approach by better describing the electron correlation. This is particularly relevant for excited systems. Furthermore, quantum Monte Carlo schemes try to solve the many-body problem exactly by using Monte Carlo techniques [47, 48]. A different route is taken by semi-empirical methods, which also incorporate experimental data

[3] The term *ab initio* was coined by Roberto Parr, a theoretical chemist, who developed methods for the approximate calculation of molecular orbitals [28–30].

to obtain computationally more efficient solutions to the Schrödinger equation [49, 50].

1.1.2. Atomistic molecular dynamics

For larger systems, where *ab initio* methods are not possible, quantum mechanical phenomena play often only a negligible role and can hence be treated implicitly. This leads us to the classical molecular dynamics (MD) approach [51, 52]. In atomistic MD, the atoms' nuclei are described as point-like particles that interact classically via empirical force fields. These force fields are usually fitted to approximately reproduce either *ab initio* reference calculations, experimental data, or both [53], and they implicitly incorporate also the electronic interactions. An MD simulation then consists of the iterative step-wise solution of Newton's equations of motion to generate a simulation trajectory. This is done with symplectic integration methods, which offer important properties for the modeling of physical systems, such as time reversibility and energy conservation. The most widely used integrator is the velocity Verlet algorithm, which is exact up to second order in time [52, 54]. In practice, the time step is limited by the highest frequency motion in the system, which is typically related to bonded or angular interactions. To allow for larger time steps, constraints can be introduced which keep the molecules or certain bonds rigid and avoid high frequency motions. For instance, liquid water is often modeled as three rigidly connected atoms [55]. While standard MD provides sampling in the microcanonical ensemble, additional techniques also allow the generation of other statistical ensembles. For example, different thermostating methods can be employed to enforce a constant temperature and simulate a canonical ensemble. This can be achieved by regularly rescaling the particles' velocities [56, 57], by explicitly coupling the particles to a heat bath [58–60], or by modeling the interaction with a heat reservoir in a stochastic fashion via random perturbations and friction acting on the particles [61, 62]. Similar protocols also allow to generate an isobaric ensemble [52].

The MD approach is orders of magnitude more efficient than directly solving Schrödinger's equation and used whenever the detailed electronic structure is of no importance or not of interest, hence making an explicit treatment of electrons unnecessary. Consequently, atomistic MD can be applied on larger-scale problems, such as the calculation of thermodynamic and macroscopic structural properties or the study of diffusion processes. Optimized hardware and software achieve simulations of systems of more than 10^{12} particles [63] or, in the case of protein folding, up to the millisecond time scale [20–24]. In general, the applications of atomistic MD are very diverse. Exemplary systems that are modeled using MD include complex soft matter systems, such as polymeric materials [64–66] and biological objects like proteins and membranes [16, 19, 20, 67]. However, it always has to be kept in mind that classical MD is an approximate method and critically depends on the quality of the force field.

1.1.3. Modeling nuclear quantum delocalization with path integrals

Although *ab initio* simulations approximately solve the Schrödinger equation, the nuclei are usually treated as point-like particles at definite positions and with well-defined velocities. The same is the case in classical MD. A method to account for the quantum mechanical delocalization of light particles is provided by Richard Feynman's path integral (PI) formalism [52, 68]. In this approach, P replicas of the system are considered, each representing one possible configuration of the system (P is usually called the Trotter number, since the system's propagator is split into P terms via Trotter decomposition in the PI methodology). The replicas of single particles are connected with each other in a cyclic fashion via harmonic springs. More specifically, each particle is mapped onto a ring polymer, a closed chain of coupled beads similar to a pearl-necklace (see Fig. 1.2 for a schematic visualization) [69]. The strength of the

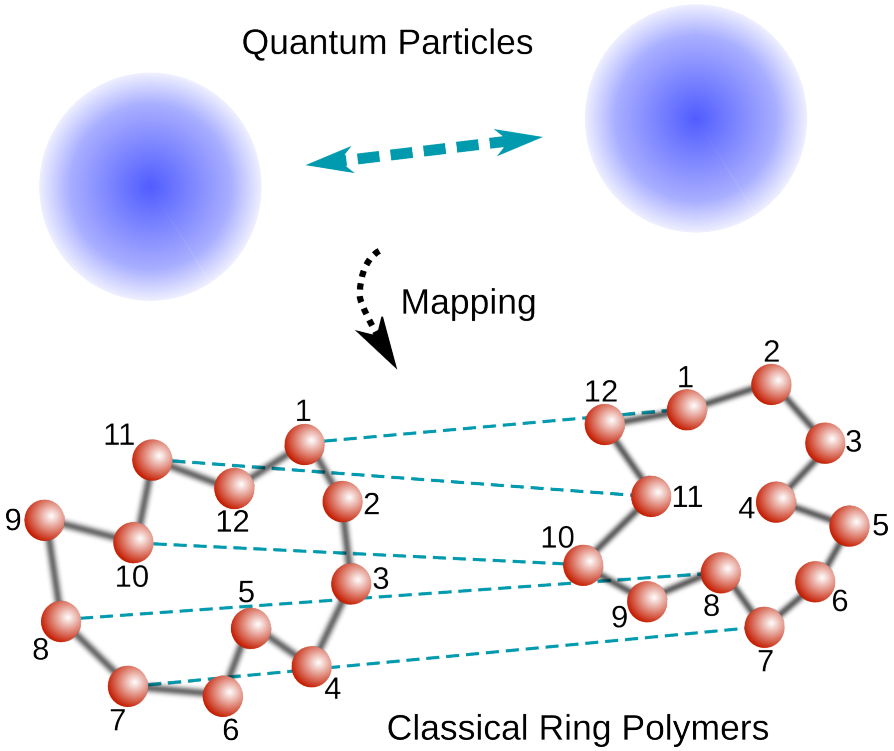

Figure 1.2.: Path integral formalism for modeling nuclear quantum decolization in molecular simulations. The quantum particles are mapped onto classical ring polymers, whose beads are coupled via harmonic springs. Between different ring polymers, only those beads interact with each other that correspond to the same imaginary time slice or replica of the system.

harmonic springs between the beads is proportional to the temperature as well as the mass of the particles. Heavier particles are mapped onto more collapsed rings corresponding to more classical behavior. The same happens when increasing the temperature. Loosely speaking, the extension of the ring polymers can be regarded as the spread of the particles' wavefunctions.

The interaction between different atoms, or rather ring polymers in this picture, proceeds in such a way that beads of different ring polymers only interact with each other if they belong to the same replica of the system. Note that there is another way of looking at the PI formalism, which is based on a connection between the canonical density matrix and the propagator: Evaluating the system's density matrix at an imaginary temperature yields the propagator and, similarly, using an imaginary time in the propagator produces the density matrix. Therefore, the different replicas of the system are often denoted as imaginary time slices and their connection via harmonic springs defines an integration along a path in imaginary time (hence, the name "path integral" formalism). Quantum statistical averages in the PI formalism can eventually be calculated by averaging over all replicas. When applied in simulations, the method is associated with a significant numerical effort: The computational cost compared to classical MD is increased proportionally to the number of replicas employed per atom. Although the approach is formally correct only in the limit of an infinite number of beads per atom ($P \to \infty$), well converged results can be obtained with a computationally feasible number of replicas (a more comprehensive and technical introduction can be found, for example, in Ref. [52]).

The PI formalism is usually applied in an MD or Monte Carlo (MC) setting (path integral molecular dynamics (PIMD) or path integral Monte Carlo (PIMC)) [52, 69–75], where the forces or energies come either from a classical force field [76–78] or from *ab initio* calculations [79–81]. The method can also be extended to allow the calculation of approximate quantum dynamics and quantum time correlation functions [82–84]. It has been shown that for systems with many light atoms, such as hydrogen or helium, and in particular at low temperatures the PI approach can be necessary for a realistic description of the system [76, 77, 85–87]. But even at room temperature, the PI methodology can yield relevant improvements. For example, an accurate description of the structure and dynamics of liquid water needs to take into account nuclear quantum effects, since a classical modeling typically results in overstructured radial distribution functions and increased oscillation frequencies of the O-H bonds as well as the H-O-H angle compared to experimental data [78, 88–91]. Furthermore, the PI formalism is regularly applied in complex biological applications. For example, it is crucial for a proper description of proton transfer reactions [80, 92–96] and for the hydrogen bond network in the active site of proteins [81].

1.1.4. Coarse-graining

Although atomistic MD is a very efficient simulation technique compared to *ab initio* methods, it can still be too computationally demanding for many large-scale problems in molecular modeling. For example, the time scales on which most proteins fold cannot be reached with all-atom MD simulations. Similarly, biological membranes are often too large to allow for an atomistic description. The same challenges occur in many complex polymer systems. A solution is provided by coarse-grained (CG) models [97–101]. In coarse-graining, several atoms are grouped together and described jointly as a single point-like particle (see Fig. 1.3 for a visualization). This significantly reduces the number of particles in the system and hence also

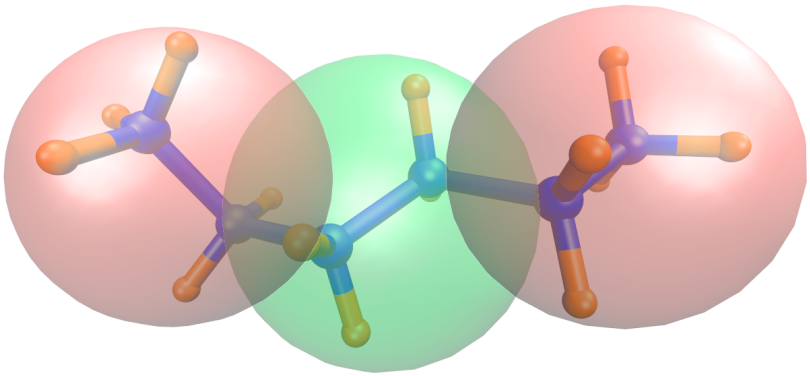

Figure 1.3.: Atomistic and coarse-grained representation of a hexane molecule. The 20 atoms of the original molecule are mapped onto three coarse-grained sites. These are placed at the centers of mass of the three consecutive carbon atom pairs and the attached hydrogen atoms.

the computational cost. Furthermore, the interaction potentials in CG models are typically much softer than atomistic force fields and larger time steps can be chosen when integrating the equations of motion, further alleviating the numerical effort. It is not only their computational efficiency that makes CG models attractive though. In many large-scale applications, we are not interested in the microscopic details of the system anyhow. In a sense, CG approaches automatically average out these details and focus only on the relevant length scales [101].

The main challenge of coarse-graining lies in the derivation of CG models that correctly capture the relevant features of the system, while neglecting the unnecessary details. A large variety of approaches for parametrizing CG representations exists, which can be broadly divided in bottom-up and top-down methods. In bottom-up coarse-graining, high-resolution, usually atomistic, reference data is used to systematically fit a CG interaction potential, while employing a predefined mapping between the atomistic system and a CG description. Noteworthy bottom-up techniques are force matching [102–104], which tries to reproduce the multi-body potential of mean force, inverse Monte Carlo [105] and (iterative) Boltzmann inversion [106, 107], which aim at preserving reference pair correlation functions, as well as relative entropy-based methods, which minimize the information loss between the CG and the reference system [108, 109]. In top-down strategies, the CG potential is parametrized to reproduce experimental measurements or other macroscopic properties [101]. An example for a CG force field derived by such an approach is the celebrated MARTINI force field [110, 111], which was originally developed for the modeling of lipid bilayers and is parametrized by matching experimental partitioning free energies between polar and apolar phases of various chemical compounds.

While CG techniques have been very successful in reproducing relevant structural, thermodynamic and chemical properties of the reference systems, reliably matching the dynamics is still a largely open problem. Since the free energy landscape of a CG system is significantly smoothened out compared to its atomistic reference, diffusion is usually strongly enhanced.

However, if one is only interested in static equilibrium properties of the system, this acceleration can be helpful, as it provides an enhanced sampling of the phase space [101].

Coarse-graining has gained significant popularity in the molecular modeling community and has been applied to many different systems. These include proteins [99, 112–119], biological membranes [15, 120–123], polymer systems [124–135], and complex liquids [103, 104, 136–141].

Beyond CG techniques towards even longer time and length scales, we leave the area of particle-based simulation methodologies. When studying phenomena of macroscopic systems with sizes of μm or larger, the microscopic particle nature of matter becomes less and less relevant. Therefore, in this regime one typically resorts to continuous models based on continuum mechanics. Examples include various fluid dynamics techniques [142] and the finite element method [143]. However, such approaches play only a minor role for molecular modeling and are beyond the scope of this introduction.

1.2. Multiscale modeling

The variety of modeling techniques discussed in the previous section enables computer simulations on many different length and time scales, on various levels of accuracy, and it allows us to study all kinds of molecular systems. It may seem that for all applications a suitable approach may be found along the hierarchy of simulation methods (Fig. 1.1). However, there is one class of systems for which modeling remains particularly challenging, even with all these methods available: multiscale systems [97, 144–146]. In these systems, short and long time and length scale phenomena are inherently coupled. Therefore, their accurate description in computer simulations requires not only large systems and long simulations, but also microscopically accurate models to correctly capture the short length and time scale effects. This is difficult to achieve since one usually has to make a compromise between computational efficiency and modeling accuracy. However, most complex molecular systems, in particular biological ones [116, 147], exhibit some degree of multiscale behavior. Therefore, the development of so-called multiscale modeling techniques has become a very active field of research that tries to address and overcome the described challenges.

1.2.1. Hierarchical approaches

One approach to multiscale modeling is to study the same system with several methods and run both large-scale, albeit coarse, as well as small, but highly accurate, simulations. The results from these calculations on different scales can then be combined to obtain the overall picture. It is crucial that the coarser models approximately capture the most important features of the high-resolution system. Therefore, bottom-up CG methods are most appropriate here, since they are designed to reproduce relevant properties of the atomistic reference system. Indeed, such techniques have been successfully used, for example, for the modeling of complicated polymer systems [66, 97], such as polymer–solid interfaces [148] and polymer nanocomposites [149].

Furthermore, they have been employed in biomolecular simulations of membranes and proteins to study phenomena such as structural transitions, self-assembly, and folding [14, 150–154].

1.2.2. Concurrent coupling

In many multiscale systems, the region that actually requires a high-resolution modeling is limited in size, often to a tiny domain compared to the overall system. Consider, for example, the case of a solvated protein that interacts at its active site with a small ligand. When trying to describe this system accurately in a simulation, we have to fulfill two requirements: On the one hand, we typically need to simulate the complete protein as well as the solvent molecules to properly reproduce the protein's structure as well as the thermodynamic conditions of its environment. This requires a large simulation. On the other hand, we need a highly accurate, often DFT-based, description of the active site to correctly model the chemical reaction with the ligand. Unfortunately, a complete *ab initio* approach is prohibitively expensive for systems of the size of extended macromolecules. However, the high-resolution treatment can be limited *only* to the protein's active site, while further away an atomistic force field will likely suffice to capture the large-scale structure and thermodynamics. This suggests to use a multi-resolution scheme, in which different modeling techniques with different resolutions are used concurrently in a single simulation; in the above example, an *ab initio* model at the active site and a classical force field for the rest of the system.

Such multi-resolution approaches are useful whenever the region that is of particular interest and therefore requires a high-resolution description is limited to a small domain within the simulation box. In fact, there are many more scenarios where this can be the case. For example, when simulating macromolecules, such as proteins or polymers, solvated in complex liquids, one can employ a CG description for solvent molecules that are far enough away from the solute, which itself may be treated on the atomistic level (see Fig. 1.4). Similarly, when studying interface phenomena, such as particle adsorption, one may want to model only the interface with a high-resolution model, while using a coarser approach for molecules further away.

Different methodologies have been developed that establish a concurrent coupling of multiple simulation methods. These techniques share the challenge of smoothly interpolating and connecting the models.

Quantum Mechanics/Molecular Mechanics

One of the most famous concurrent multi-resolution schemes is the quantum mechanics/molecular mechanics (QM/MM) method [156–164].[4] It allows the connection of a small subdomain that is described with an *ab initio* approach to a larger region that is treated by means of a classical atomistic force field. It is applied in scenarios similar to the first one described in this section, i.e., in simulations of proteins in which a certain part, such as the active site, requires a rigorous quantum mechanical modeling. For example, QM/MM techniques have been widely

[4] The nowadays widely used acronym "QM/MM" was coined in 1992 by Jiali Gao and Xinfu Xia, who performed QM/MM Monte Carlo simulations to study polarization effects in aqueous systems [157].

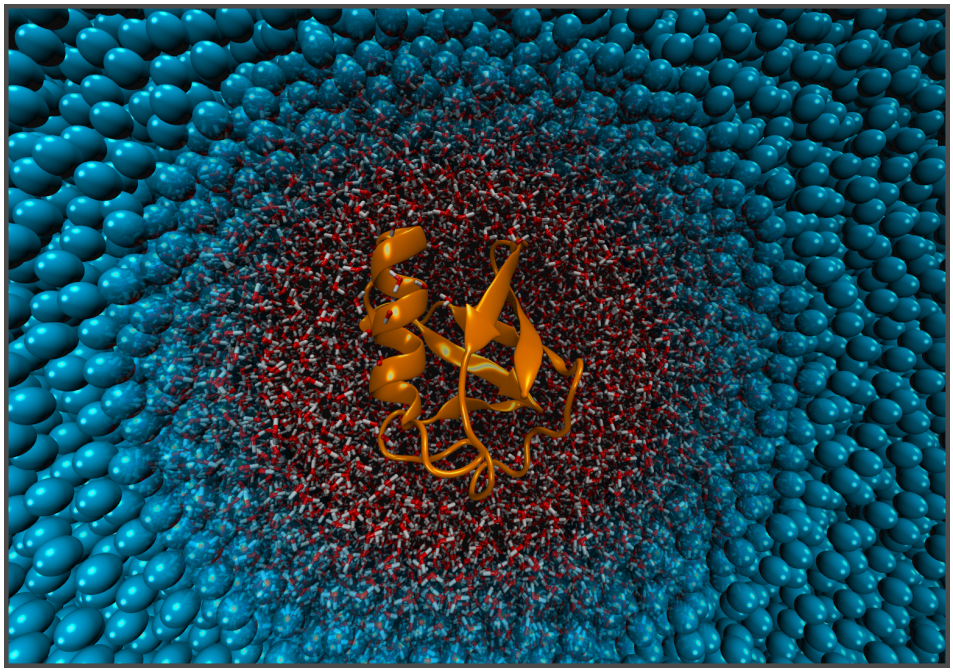

Figure 1.4.: Adaptive resolution simulation of ubiquitin [155]. While the protein and the water in its immediate vicinity is described by an atomistic force field, the solvent molecules further away are treated on a coarse-grained level.

used for the description of enzymes and enzymatic chemical reactions. However, achieving a smooth link between the different regions is difficult. In particular, electrostatic coupling poses a major challenge, since the charges of the classical force field may not be consistent with the charge distribution generated by the *ab initio* calculations [161, 164]. A variety of different schemes have been developed that cope with this problem in various ways: They range from methods that use only classical charges to calculate the electrostatic interactions between the quantum mechanical and the classical region to approaches that use polarizable force fields in the classical region and obtain the electrostatic interactions in a self-consistent manner [161]. Another difficultly associated with standard QM/MM is that the QM region is fixed and particles are not allowed to travel between the different regions. This can be overcome with adaptive QM/MM techniques that allow particles to diffuse freely [165–169], although these methods introduce further difficulties, such as discontinuities in the forces and energies. Nevertheless, QM/MM schemes have become a widespread tool and are now also applied in other areas of materials science, for example, to study crack propagation or chemisorption processes [170–173]. This importance was highlighted in 2013, when Arieh Warshel and Michael Levitt, who first introduced the QM/MM method in 1976 [156], were awarded the Nobel Prize in Chemistry, along with Martin Karplus.

Coupling atomistic and coarse-grained models

Another class of concurrent multi-resolution schemes are those that aim at bridging atomistic and CG models. This can be useful, for example, in simulations of solvated macromolecules that are surrounded by large amounts of solvent molecules, whose only role is to guarantee the correct large-scale thermodynamic conditions [174, 175]. As described above, one may then want to treat solvent particles far away from the central macromolecule with a CG description in order to save computational resources. Only when the solvent molecules approach the solute do they change their resolution on-the-fly and become atomistic. This scenario is visualized in the simulation snapshot in Fig. 1.4.

In practice, this idea boils down to a smooth spatial interpolation of atomistic and CG force fields. Various techniques exist to establish that and they can be categorized into two classes. On the one hand, there are methods that directly interpolate the forces acting between the particles [176, 177]; on the other hand, the interpolation can be performed on the level of potential energies [178–181]. The most popular and widely used methodology for concurrently coupling atomistic and CG models are the adaptive resolution simulation scheme (AdResS) [144, 176, 182] and its relative, the Hamiltonian adaptive resolution simulation scheme (H–AdResS) [181, 183, 184]. In these techniques, the interpolation between the force fields is achieved by associating a spatially dependent resolution property with each solvent molecule (see Fig. 1.5). The particles, which can freely diffuse through the system, then change their resolution on-the-fly according to their positions in space. A hybrid (HY) buffer region between the atomistic (AT) and CG domains ensures a smooth coupling.

The difference between AdResS and H–AdResS is the previously mentioned aspect: In standard AdResS, which was introduced earlier than H–AdResS in 2005 [176], the interpolation takes place on the level of the AT and CG forces. On the contrary, H–AdResS is based on a Hamiltonian in which the potential energy terms are interpolated. Both schemes have certain advantages and disadvantages and depending on the application one may want to use either AdResS or H–AdResS. In fact, both methods have been successfully used for multi-resolution modeling of various soft matter systems, such as solvated polymers, fullerenes, proteins and even DNA [155, 185–194].

Since this thesis focuses on the development of advanced adaptive resolution simulation methods, we will provide a more comprehensive and technical introduction to AdResS and H–AdResS in section 1.3, including a detailed overview over its applications.

From particles to continuum

We conclude our overview of multiscale simulation techniques by briefly reviewing another class of concurrent multi-resolution methods. When dealing with very large-scale systems far beyond the sizes that are typically encountered in particle-based molecular simulation, one usually resorts to much more efficient continuum models. In some situations, though, it is advantageous to use an approach that couples a continuum description with a particle-based scheme in order to retain a more accurate treatment in a locally restricted area. A typical

example where this can be useful is the formation and propagation of cracks in complex materials [195–204]. In this situation, the overall system is typically large and in most parts it is subject to moderate stress. In these regions away from the cracks the material properties can be accurately described by continuum equations. However, the mechanisms of the forming of local cracks and in particular their propagation through the material are highly complex and can be captured reliably only with particle-based models. The tip of a crack is very localized, though, and a high-resolution is required only in a small subregion of the system. Therefore, this application is an ideal candidate for the use of multi-resolution particle-continuum methods.

There exists a large variety of techniques for bridging and connecting particle-based and continuum models. One popular approach is the quasicontinuum method [199], which uses a finite elements scheme for the overall system and an atomistic model in smaller subregions. The coupling between the two models is achieved by picking representative atoms that are employed as quadrature points for the finite elements calculation. Similarly, the strain from the large-scale finite elements method is applied back on the atoms. Other techniques employ Lagrange multipliers to enforce a smooth coupling between atomistic and continuum domains [204], or use molecular dynamics results from a small subregion to provide boundary conditions for a large-scale continuum calculation and vice versa [205]. Some methods even couple models ranging from *ab initio* to continuum level of detail in a single simulation [196–198, 202, 206].

The applications of particle-continuum multi-resolution methodologies are very diverse. Apart from modeling brittle materials, concurrent particle-continuum schemes are, for example, employed to study corner and cavity flow [205, 207, 208]. In such complicated flow scenarios conventional continuum flow equations lead to singularities. Hence, in the singular regions a particle-based treatment is necessary.

1.3. The adaptive resolution simulation scheme

In this section, we provide a detailed, technical introduction to the AdResS methodology. Furthermore, we recapitulate its most important applications so far.

In AdResS, two systems with different classical force fields are coupled with each other, typically an atomistic and a coarse-grained one (see Figs. 1.4 and 1.5) [144, 176]. The high resolution is usually restricted to a small but relevant subregion within an overall large simulation box. In this way, computational resources can be concentrated in the high-resolution domain, resulting in an overall speedup compared to all-atom simulations. An important feature of AdResS is that particles are allowed to freely travel between the different regions without feeling any barriers. In this way, the atomistic subregion behaves as if embedded in a completely atomistic environment [155, 176, 181, 189, 192–194, 209–216].

The interpolation of the different force fields can be implemented either on the level of forces or on the level of potential energies. This has led to two slightly different adaptive resolution schemes, which we will review in the following.

1.3.1. Force interpolation

In the force-based AdResS scheme, proposed in 2005 by Praprotnik et al. [144, 176], the interpolation occurs directly on the intermolecular forces. The method assumes that all forces in the system are short-ranged, which means that also electrostatics need to be treated accordingly using, for example, a reaction field technique [217] (we will address this issue in more detail in section 1.3.5). The non-bonded force $\mathbf{F}_{\alpha\beta}$ between two molecules α and β is then given as

$$\mathbf{F}_{\alpha\beta} = \lambda(\mathbf{R}_\alpha)\lambda(\mathbf{R}_\beta)\mathbf{F}^{AT}_{\alpha\beta} + \left(1 - \lambda(\mathbf{R}_\alpha)\lambda(\mathbf{R}_\beta)\right)\mathbf{F}^{CG}_{\alpha\beta}, \tag{1.1}$$

with

$$\mathbf{F}^{AT}_{\alpha\beta} = \sum_{i\in\alpha}\sum_{j\in\beta}\mathbf{F}^{AT}_{ij}. \tag{1.2}$$

Here, $\mathbf{F}^{AT}_{\alpha\beta}$ is the atomistic (AT) force between the molecules and $\mathbf{F}^{CG}_{\alpha\beta}$ is the CG force. Furthermore, $\lambda(\mathbf{R}_\alpha)$ is a position-dependent function that varies between 0 and 1 and determines the resolution of the molecules. It is typically evaluated based on the molecular center of mass $\mathbf{R}_\alpha$, although other approaches are in principle possible. The term $\mathbf{F}^{AT}_{\alpha\beta}$ is composed of the individual forces $\mathbf{F}^{AT}_{ij}$ between the atoms i and j of the molecules α and β (Eq. 1.2), while the CG term $\mathbf{F}^{CG}_{\alpha\beta}$ is usually a single force applied on the centers of mass of the CG molecules. Note that this force interpolation scheme conserves Newton's third law. The concept is visualized in Fig. 1.5 for an exemplary system of 3-site liquid water coupled to a single-site CG model.

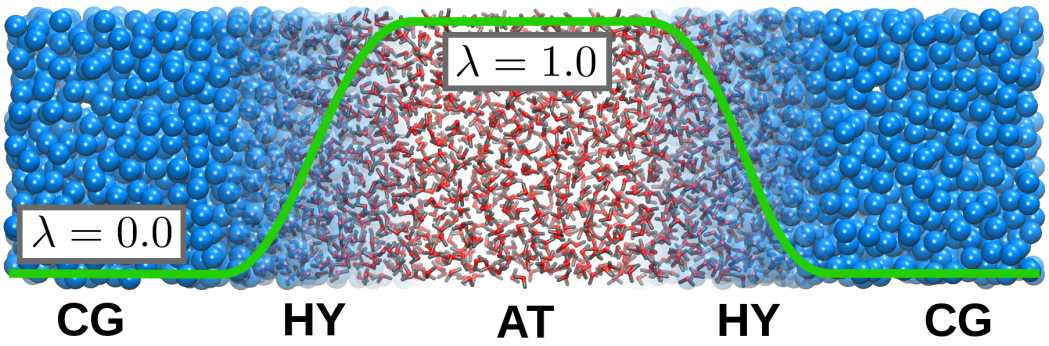

Figure 1.5.: Snapshot from an adaptive resolution simulation coupling 3-site atomistic water to a single-site CG model. The green line represents the resolution function. When molecules diffuse through the system and cross the hybrid region, they change their resolution accordingly.

In the area where the resolution function λ is 1, the CG term in Eq. 1.1 disappears and the liquid is fully atomistic. Similarly, where λ is 0 the AT contribution disappears and the system is fully CG. In between the AT and CG regions, the system switches its resolution as λ smoothly changes from 0 to 1 and vice versa. This transition area is usually called hybrid (HY) region. The AT region can be either fixed in space or attached to a central particle and slowly move with it [176, 190]. Furthermore, it can have different, although regular geometries,

for example a spherical one as in Fig. 1.4 [155, 188, 190, 192, 212], a cuboid one as in Fig. 1.5 [176, 193, 194, 218], or a cylindrical one [191, 215].

An important aspect is that this force interpolation scheme, Eq. 1.1, does not admit a Hamiltonian description [219]. Therefore, it is not energy conserving and excess heat is produced in the HY region. Hence, a thermostat based on the fluctuation-dissipation theorem needs to be applied, removing the excess heat and establishing a thermal equilibrium [218–223]. Since many applications are run in the canonical ensemble, which is usually enforced with a thermostat anyway, this may not be a severe limitation. It has also been demonstrated that it is sufficient to apply a local thermostat only in the CG and HY regions such that the dynamics in the AT domain are not perturbed [155, 210, 216, 224]. However, force-based AdResS does not allow true microcanonical simulations or techniques that are based on a Hamiltonian description such as, for example, most Monte Carlo methods. In the HY region, it is also not consistent with a standard statistical mechanics formulation based on a partition function. Note that the temperature is well defined everywhere in the system, though.

It is important to note that the precise shape of the resolution function λ can vary, as long as it ensures a smooth transition between the force fields. In the literature, squared cosine functions or different polynomials are commonly used [155, 176, 181, 183, 210, 218]. The width of the HY region is another crucial parameter in adaptive resolution simulations. On the one hand, one wants to have a small hybrid region in order to keep the volume in which both AT and CG forces need to be evaluated as small as possible. On the other hand, it needs to be wide enough to guarantee a slow transition of forces between the two regions. An abrupt switch between the force fields would lead to an overgeneration of excess heat. In practice, one often chooses a width slightly larger than the cutoff of the non-bonded interaction potentials to prevent direct interactions between molecules in the AT and CG regions.

In addition to the short-ranged non-bonded intermolecular forces, which are interpolated, there may also be intramolecular forces, like the bond and angular potentials in liquid water. These play a negligible role in the overall computational effort for the force calculations and are therefore typically not subject to any interpolation. In force-based AdResS, molecules can be frozen when being located in the CG region such that the intramolecular forces need not be computed there.

1.3.2. The thermodynamic force

At a given temperature and density, AT and CG systems have typically strongly deviating virial pressures [225–227]. In AdResS, where particles are allowed to diffuse freely between the different regions, this leads to a net force pushing particles from one region to the other until an equilibrium state with an inhomogeneous density profile along the direction of resolution change is reached. This thermodynamic imbalance can be corrected with the so-called thermodynamic force (TF) [182], a compensating force that acts on molecules in the HY region and its vicinity. It is constructed in order to exactly compensate for the pressure difference between the two domains and is derived in an iterative fashion, starting from the

distorted density profile obtained without the application of any correction force:

$$F_{i+1}^{TF}(r) = F_i^{TF}(r) - \frac{M}{\rho_0^2 \kappa_T} \frac{\partial}{\partial r} \rho_i(r), \tag{1.3}$$

where κ_T is the isothermal compressibility, ρ_0 is the reference density and M is the molecular mass (in practice, the prefactor in front of the density gradient can be varied to aid convergence). The force $F_i^{TF}(r)$ is applied in the HY region along the direction of resolution change and r denotes the position along this direction. The protocol converges by construction, when a flat density profile is reached. Hence, the TF also smooths out local density distortions in the HY region. Note that the integral of the TF provides exactly the pressure difference between the two models [182]. The TF scheme has been used extensively in force-based AdResS simulations [155, 187, 188, 211, 214–216, 228], including complex mixtures [174, 175, 189, 191, 213, 229].

1.3.3. Energy interpolation

The previously described deficiencies of the force-based AdResS approach have motivated the development of an adaptive resolution scheme that is based on a Hamiltonian formulation. In 2013, Potestio et al. proposed Hamiltonian–AdResS (H–AdResS), which implements the interpolation between the different force fields on the level of potential energies [181, 183, 184].

In H–AdResS, a global Hamiltonian is defined as

$$H = \sum_\alpha \sum_{i \in \alpha} \frac{\mathbf{p}_{\alpha i}^2}{2m_{\alpha i}} + V^{\text{int}} + \sum_\alpha \left\{ \lambda(\mathbf{R}_\alpha) V_\alpha^{\text{AT}} + (1 - \lambda(\mathbf{R}_\alpha))\, V_\alpha^{\text{CG}} \right\}. \tag{1.4}$$

Here, the sum over α denotes a sum over all molecules in the system, and the sum over i stands for a sum over all atoms within a molecule α. With this notation, $m_{\alpha i}$ is the mass of atom i of molecule α and $\mathbf{p}_{\alpha i}$ is its momentum. Furthermore, V^{int} represents all intramolecular interactions, which are, as in force-based AdResS, not subject to any interpolation. The potentials V_α^{AT} and V_α^{CG} are the sums of all non-bonded AT and CG intermolecular interaction potentials V^{AT} and V^{CG} acting on molecule α, this is,

$$V_\alpha^{\text{AT}} = \frac{1}{2} \sum_{\beta \neq \alpha} \sum_{\substack{i \in \alpha \\ j \in \beta}} V^{\text{AT}}(|\mathbf{r}_{\alpha i} - \mathbf{r}_{\beta j}|), \tag{1.5}$$

$$V_\alpha^{\text{CG}} = \frac{1}{2} \sum_{\beta \neq \alpha} V^{\text{CG}}(|\mathbf{R}_\alpha - \mathbf{R}_\beta|), \tag{1.6}$$

where the sum over β runs over all molecules except α and $\mathbf{r}_{\alpha i}$ denotes the position of atom i of molecule α. Differentiating the Hamiltonian in Eq. 1.4 we obtain the force on atom i of

molecule α as

$$\begin{aligned}\mathbf{F}_{\alpha i} = \sum_{\beta\neq\alpha}\sum_{j\in\beta}\left\{\frac{\lambda(\mathbf{R}_\alpha)+\lambda(\mathbf{R}_\beta)}{2}\mathbf{F}^{\mathrm{AT}}_{\alpha i|\beta j} + \left(1-\frac{\lambda(\mathbf{R}_\alpha)+\lambda(\mathbf{R}_\beta)}{2}\right)\mathbf{F}^{\mathrm{CG}}_{\alpha i|\beta}\right\} \\ + \mathbf{F}^{\mathrm{int}}_{\alpha i} - [V^{\mathrm{AT}}_\alpha - V^{\mathrm{CG}}_\alpha]\,\nabla_{\alpha i}\lambda(\mathbf{R}_\alpha),\end{aligned} \tag{1.7}$$

where $\mathbf{F}^{\mathrm{AT}}_{\alpha i|\beta j}$ is the AT force on atom i of molecule α due to the interaction with atom j of molecule β and, analogously, $\mathbf{F}^{\mathrm{CG}}_{\alpha i|\beta}$ is the CG force due to the interaction with molecule β. Furthermore, $\mathbf{F}^{\mathrm{int}}_{\alpha i}$ stands for all intramolecular forces on atom i of molecule α. The last term, $\mathbf{F}^{\mathrm{drift}}_{\alpha i} = -[V^{\mathrm{AT}}_\alpha - V^{\mathrm{CG}}_\alpha]\nabla_{\alpha i}\lambda(\mathbf{R}_\alpha)$, dubbed *drift force*, is the central difference between force-based AdResS and H–AdResS. It is proportional to the gradient of the resolution function λ and therefore only acts in the HY coupling region. It pushes molecules from one region to the other, just like the pressure difference between the AT and the CG models. In the HY region, it also violates Newton's third law, that is, the force between pairs of molecules cannot be written as a sum of antisymmetric terms due to the additional drift term. Note that the drift force on molecule α does not only depend on its resolution λ_α, but, because of the term $V^{\mathrm{AT}}_\alpha - V^{\mathrm{CG}}_\alpha$, also on its environment.

1.3.4. Free energy corrections

Introducing a static compensation field $\Delta H(\lambda(\mathbf{R}_\alpha))$ in the HY region, the average of the drift force can be removed for all λ_α, effectively compensating for the pushing of molecules from one region to the other [181]: The modified Hamiltonian $\hat{H}$ is

$$\hat{H} = H - \sum_\alpha \Delta H(\lambda(\mathbf{R}_\alpha)), \tag{1.8}$$

and the drift force on a molecule α becomes

$$\hat{\mathbf{F}}^{\mathrm{drift}}_\alpha = -\left[V^{\mathrm{AT}}_\alpha - V^{\mathrm{CG}}_\alpha - \left.\frac{d\Delta H(\lambda)}{d\lambda}\right|_{\lambda=\lambda(\mathbf{R}_\alpha)}\right]\nabla_\alpha\lambda(\mathbf{R}_\alpha). \tag{1.9}$$

Note that we dropped the indices for the individual atoms, since the drift force as well as the correction are usually applied on whole molecules or, for macromolecules such as polymers, on the monomers or similar coarse units. The forces are then simply redistributed equally among the atoms belonging to the molecules or monomers.

As we want to remove the average of the drift force, we obtain the following requirement for $\Delta H(\lambda(\mathbf{R}_\alpha))$:

$$\left.\frac{d\Delta H(\lambda)}{d\lambda}\right|_{\lambda=\lambda(\mathbf{R}_\alpha)} = \left\langle V^{\mathrm{AT}}_\alpha - V^{\mathrm{CG}}_\alpha\right\rangle_{\mathbf{R}_\alpha}. \tag{1.10}$$

The subscript of the term on the right hand side of Eq. 1.10 denotes that the average formally has to be performed in an adaptive simulation constraining the center of mass position of the molecule α at $\mathbf{R}_\alpha$. In principle, Eq. 1.10 is self-consistent, since the average has to be calculated

with the Hamiltonian $\hat{H}$, which already includes the correction $\Delta H(\lambda)$. Thus, the right hand side of Eq. 1.10 is a functional of $\Delta H(\lambda)$.

If the HY region is sufficiently wide, the immediate environment of a molecule with resolution $\lambda_\alpha = \lambda(\mathbf{R}_\alpha)$ consists only of molecules at a similar resolution. Then, we can approximate

$$\left\langle V_\alpha^{\mathrm{AT}} - V_\alpha^{\mathrm{CG}} \right\rangle_{\mathbf{R}_\alpha} \approx \left\langle V_\alpha^{\mathrm{AT}} - V_\alpha^{\mathrm{CG}} \right\rangle_{\lambda=\lambda(\mathbf{R}_\alpha)}, \tag{1.11}$$

where the subscript on the right hand side now indicates an average performed in a simulation box in which all molecules have a uniform resolution $\lambda = \lambda(\mathbf{R}_\alpha)$. In this case, we can also average over all N molecules in the box, such that

$$\left\langle V_\alpha^{\mathrm{AT}} - V_\alpha^{\mathrm{CG}} \right\rangle_{\lambda=\lambda(\mathbf{R}_\alpha)} = \frac{1}{N} \left\langle V^{\mathrm{AT}} - V^{\mathrm{CG}} \right\rangle_{\lambda=\lambda(\mathbf{R}_\alpha)}, \tag{1.12}$$

with

$$V^{\mathrm{AT}} = \sum_{\alpha=1}^{N} V_\alpha^{\mathrm{AT}}, \tag{1.13}$$

$$V^{\mathrm{CG}} = \sum_{\alpha=1}^{N} V_\alpha^{\mathrm{CG}}. \tag{1.14}$$

Hence, if the approximation in Eq. 1.11 is fulfilled, the correction term $\Delta H(\lambda(\mathbf{R}_\alpha))$ can be calculated by standard Kirkwood thermodynamic integration [230]:

$$\begin{aligned} \Delta H(\lambda) &= \int_0^\lambda \frac{d\Delta H(\lambda')}{d\lambda'} d\lambda' \\ &\approx \frac{1}{N} \int_0^\lambda \left\langle V^{\mathrm{AT}} - V^{\mathrm{CG}} \right\rangle_{\lambda'} d\lambda' \\ &= \frac{\Delta F(\lambda)}{N}, \end{aligned} \tag{1.15}$$

where $\Delta F(\lambda)/N$ is the Helmholtz free energy difference per particle between the system at uniform resolution λ and the fully CG system ($\lambda = 0$). We can conclude that the aggregate strength of the drift force, i.e. its integral over the HY region, is approximately proportional to the Helmholtz free energy difference $\Delta F(\lambda)$ between the AT and CG system. Therefore, the derivative of $\Delta F(\lambda)$ with respect to λ can be employed to compensate for it on average.

A similar approach can be taken when the pressure difference between the two subsystems also needs to be corrected for. In this case, one simply applies the Gibbs free energy difference per molecule $\Delta G(\lambda)/N$ as correction:

$$\Delta H(\lambda) = \frac{\Delta G(\lambda)}{N} = \frac{\Delta F(\lambda)}{N} + \frac{p(\lambda)}{\rho^\star}, \tag{1.16}$$

where $p(\lambda)$ is the pressure at resolution λ and $\rho^\star$ is the reference density. In this way, both the explicit drift force as well as the implicit effect by the pressure difference between the AT and

the CG models are corrected and an approximately flat density profile along the direction of resolution change is obtained. Note that the additional pressure term in the correction now leads to an inhomogeneous pressure profile. Eventually, one has to make a compromise and enforce either a flat pressure profile (using only the Helmholtz free energy-based correction) or a flat density profile (using the Gibbs free energy-based correction) at equilibrium [181]. However, in most applications it is only the AT region that we truly care about and a different pressure in the CG domain may not pose a problem.

Because of their interpretations in terms of thermodynamic potentials, the two correction schemes are often called *free energy corrections* (FEC). A particular advantage of the FECs is that all terms can be conveniently calculated during a single thermodynamic integration without any iterative protocol, as was the case for the thermodynamic force. We want to stress that this methodology relies on the approximation made in Eq. 1.11. If the AT and the CG system have very different thermodynamics or if the HY region is not wide enough, the FEC approach may not lead to a perfect correction of the drift force and the pressure gradient between the two systems. In this case, we can still fall back onto the thermodynamic force presented earlier in section 1.3.2 and iteratively refine the density profile if necessary. Recently, there were developed also advanced iterative methods tailored to H-AdResS that calculate the exact Helmholtz or Gibbs FEC on the fly during an adaptive resolution simulation [231]. Finally, we want to mention that the FECs also straightforwardly generalize to mixtures, as was demonstrated by Potestio et al. [183].

1.3.5. Comparison

Having introduced two different routes for concurrently coupling atomistic and coarse-grained systems with each other, we now want to compare them and review their advantages and disadvantages in more detail.

The force-based AdResS scheme is conceptually simple and requires less overhead than H–AdResS, which includes an additional drift term in the HY region. Force-based AdResS also obeys Newton's third law, which is violated in the energy-based approach. However, being incompatible with a Hamiltonian formulation, the AdResS method does not allow for a rigorous partition function-based theory and cannot be used for energy-conserving or Monte Carlo simulations. It also requires thermostating to remove the excess heat produced in the HY region and to guarantee a thermodynamic equilibrium. In comparison, the main advantage of H–AdResS is its solid Hamiltonian foundation and also the FECs, which are usually employed in H–AdResS, have elegant interpretations in terms of thermodynamic potentials.

Since force-based AdResS is not energy-conserving anyhow, the intramolecular interactions of molecules in the CG region can be frozen. In other words, the intramolecular degrees of freedom can be removed in the CG region and reintroduced when a molecule enters the HY region and starts interacting atomistically. This stands in contrast with H–AdResS, in which, in principle, no degrees of freedom are actually removed. The intramolecular motion is retained in the CG region in order to be consistent with the Hamiltonian formulation and

only the intermolecular force field changes. This adds a slight additional overhead compared to force-based AdResS. When applying a thermostat in H–AdResS simulations to generate a canonical ensemble, intramolecular interactions of molecules in the CG region can be frozen like in force-based AdResS, though.

Another more technical difference between the two methodologies is the way in which the resolution of the molecules is used to interpolate the force fields. In the standard AdResS approach, the product is used, $\lambda_\alpha \times \lambda_\beta$, whereas H–AdResS employs the average, $1/2(\lambda_\alpha + \lambda_\beta)$, see Eqs. 1.1 and 1.7. The latter is a result of writing the potential energy as in Eq. 1.4, while the product was chosen in force-based AdResS to prevent any atomistic interaction between molecules in the HY and CG regions whatsoever. In H–AdResS the interaction between HY and CG molecules contains both CG and AT contributions. A similar force-based AdResS scheme could also be constructed by utilizing the average for interpolating the forces.

Concluding, neither of the two methods can be regarded as strictly better than the other. The proper choice of technique depends on the application of interest. For example, when an exact fulfillment of Newton's third law is essential, which may be the case, for example, when studying complicated dynamic or hydrodynamic systems, one probably wants to resort to force-based AdResS. Additionally, when running large-scale simulations in a canonical ensemble with strong thermostating, the force-based AdResS approach may be used because of its lower overhead. For all applications that are based on a solid Hamiltonian formulation, H–AdResS must be employed.

In the previous sections, the interpolated AT and CG potentials were assumed to be two-body non-bonded short-ranged interactions. The adaptive resolution methodology was introduced focusing on such interactions for several reasons. First, it is the non-bonded interactions whose computation requires the most resources. The overhead of bonded potentials is typically very small and other relevant simulation costs, such as inter-processor communication, are rather software and implementation challenges that cannot be addressed by physically inspired simulation strategies such as AdResS. Second, in most large-scale applications only two-body non-bonded interactions are considered, neglecting higher order terms. It is worth pointing out that both force-based and energy-based AdResS can be generalized to higher order potentials. When interpolating three-body potentials, one would just need to calculate the product or average of the resolutions of three interacting molecules instead of two.

The last important point that needs to be addressed are long-range electrostatics. In standard MD simulations, they are usually treated with Ewald summation methods that assume infinite periodicity of the system and calculate the long-range part of the interaction in Fourier space in a non-local fashion [232–234]. Adaptive resolution simulations, however, rely on a local definition of molecular resolution, which can only be provided when the interactions are sufficiently local, that is, short-ranged. Consider, for example, an aqueous salt solution in which water molecules are mapped onto single neutral beads in the CG region. In this case, charges are coarse-grained away in the CG region and water molecules would not screen the electrostatic interactions of the salt ions. This would lead to incorrect results when explicitly calculating long-range

interactions. For this reason, Ewald summation and its relatives cannot be directly applied in adaptive resolution simulations. Instead, electrostatics needs to be treated with methods that are purely short-ranged, such as the reaction field technique [217, 235–237], which assumes that beyond a cutoff distance the long-range electrostatic interaction can be modeled by a uniform and homogeneous dielectric medium, or the damped shifted force [238, 239], which smoothly truncates the electrostatic interaction at a short-ranged cutoff. It was shown by Bevc et al. and by Heidari et al. that these schemes lead to accurate results in AdResS and H–AdResS simulations [213, 231].

1.3.6. Applications

The presented adaptive resolution methodology has been successfully applied on a wide variety of multiscale systems. It was shown, for example, that in adaptive resolution simulations of solvated proteins a thin layer of atomistic water around the solute is sufficient to make it behave, both dynamically and statistically, as if embedded in a fully atomistic environment [155, 188]. Similar tests were also successfully performed for other relevant systems, such as solvated DNA [189, 191, 229] and complex polymers [192–194].

Another interesting application of AdResS are path integral-based simulations, in which classical particles are mapped onto small vibrating ring polymers. In fact, this is an ideal scenario for adaptive resolution methods, since conceptually the mapping gives rise to nothing but a classical, albeit computationally costly, force field. However, the ring polymers can also be easily modeled in a coarse-grained fashion as individual classical particles. Concurrently coupling a path integral model with such a CG force field in adaptive resolution simulations was first suggested by Poma and Delle Site [177]. It was subsequently shown in adaptive simulations of liquid parahydrogen and water that such a connection can be successfully established and that the path integral domain of the simulation behaves as if embedded in a fully quantum simulation box [212, 216, 224, 240]. It is crucial to note, though, that these adaptive quantum–classical simulations rest on a weak foundation: The coupling was achieved by simply interpolating a classical and a path integral force field using the force-based AdResS method that does not allow a Hamiltonian formulation [219]. The path integral formalism itself, however, intrinsically relies on the existence of a Hamiltonian.

The applications mentioned so far aim at using the adaptive resolution scheme for speeding up the simulations compared to fully atomistic ones. AdResS can also be used as a computational tool to answer questions that are otherwise challenging to tackle. For example, how local is the hydrogen bond network and the structure formation of water around a large hydrophobic solute, that is, how does it depend on the bulk [185]? By coupling atomistic water to a simple CG model, AdResS provides a convenient method to switch on and off the hydrogen bonds beyond different distances from a hydrophobic solute while not disturbing the large-scale thermodynamic environment as a whole. This approach has been used by Lambeth et al. as well as by Fritsch et al. to study the locality of the structure of water and toluene around fullerenes [185, 187]. A similar analysis of the structure and dynamics of water in the hydration

shell of proteins was also performed by Fogarty et al. [155]. In all cases, it was found that the investigated structural and dynamical quantities are fairly local, requiring little atomistic support from the bulk.

Another important application of AdResS are open-boundary and semi-grand canonical simulations. Coupling a small AT region to a large CG reservoir that is computationally cheap but guarantees the correct particle fluctuations of the AT domain, it is possible to efficiently calculate accurate thermodynamic quantities, such as Kirkwood-Buff integrals, in a grand canonical fashion in the open AT region [174]. This is particularly relevant when dealing with liquid mixtures. In fact, in this case, adaptive resolution methods have yet another advantage. Consider the situation of a complex macromolecule solvated in such a liquid mixture [175]. As the solute undergoes conformational transitions, it may attract or reject solvent particles, which leads to concentration fluctuations. Usually, these disturbances can be taken care of by inserting or deleting solvent molecules. In dense systems, though, the insertion can become difficult or even impossible. As Mukherji and Kremer have shown, this problem can be elegantly solved by AdResS [175]. One can simply couple the atomistic system to a coarse-grained subsystem with much softer interaction potentials and perform the particle exchange there.

Apart from the discussed applications, AdResS has been used in many more contexts. In particular, various technical enhancements have been proposed and validated. For example, Delgado-Buscalioni et al. have shown that it is even possible to perform triple-scale simulations coupling AT to CG to continuum [241–243]. Zavadlav et al. explored the possibilities of coupling water to more complex CG liquids. They demonstrated that it is possible to use supramolecular CG models in AdResS that combine four water molecules in a single CG entity. The bundled molecules either stay connected in the atomistic region [214] or they are linked together or released in the HY region using clustering algorithms [244]. It is even possible to model the water bundles not by individual CG sites but by polarizable supramolecular CG molecules [215]. This can be useful for systems with strong electrostatic interactions. Finally, we want to mention another interesting technical development that was recently proposed by Fogarty et al.: In adaptive resolution simulations of proteins, it is possible to not only treat the solvent molecules with adaptive resolution but to additionally also use a dual-resolution model for the protein itself, restricting the atomistic description only to the active site [190].

1.4. Outline

Although AdResS and H–AdResS are elegant tools for the coupling of AT and CG models in adaptive resolution simulations, there are many open questions. How different can the AT and CG systems be such that a smooth interpolation is still possible and how far can we go, when trying to connect radically different models? AdResS and H–AdResS, although of course inspired by each other, were derived independently. Hence, what is their deeper relation? Can one be constructed from the other? And how can we interpret the additional drift term present in H–AdResS? As in most concurrent multi-resolution methods, the challenge in AdResS is

to establish a smooth connection between the different force fields. Therefore, can we come up with any recipes in order to optimize the coupling? And how does this coupling actually depend on the size of the hybrid region?

So far, only adaptive resolution simulations with a fixed and well-defined geometry are possible. The high-resolution region is either of cuboid, cylindrical or spherical shape and it is not allowed to change throughout the simulation. However, most interesting systems, in particular biological ones like proteins and membranes, come in complicated forms and undergo conformational transitions. Hence, can we derive a general scheme that allows adaptive simulations with arbitrary geometries? Ideally, in such a fashion that the high-resolution can deform and adapt its shape in order to follow, for example, the conformational changes of a protein?

We have seen that the force-based AdResS methodology was used to couple path integral with classical models. However, it has been shown that the standard force-based AdResS scheme does not allow a Hamiltonian formulation, while the path integral formalism is intrinsically based on a Hamiltonian. Therefore, can we derive a more rigorous adaptive coupling scheme for classical atomistic force fields and path integrals?

These are the main questions that are addressed in this thesis. We build on the current state of the art of adaptive resolution simulations with a twofold goal in mind: We do not only aim at shedding light on the theory and mechanisms behind the two adaptive resolution methods, but we also want to improve the existing schemes and construct more powerful AdResS methods in order to overcome their current limitations. In the following, we provide a brief overview over the chapters 2-7.

In chapter 2, we couple water, a very structured liquid, to an ideal gas of non-interacting particles, a completely unstructured system with dynamics and diffusion on a completely different time scale. In this way, we test whether it is possible to smoothly connect so dramatically different systems like these two exemplary ones with each other without losing accuracy in the AT region.

In chapter 3, we derive a unifying framework for force-based and energy-based AdResS simulations. We show that the drift force in H–AdResS can be used to construct a generalized Langevin equation with memory and that applying it as a colored-noise thermostat on an H–AdResS Hamiltonian, we can obtain the corresponding force-based AdResS scheme.

We continue along this line in chapter 4, and demonstrate that one can use a relative entropy based framework to set up adaptive resolution simulations in an optimal fashion. In particular, we prove that minimizing the relative entropy between the AT and CG potentials results in a smooth coupling. We also derive and validate a quantitative relationship between the width of the hybrid region and the quality of the interpolation.

In chapter 5, we devise a scheme to overcome the previous geometrical constraints in adaptive resolution simulations. We demonstrate how atomistic regions with arbitrary geometries can be constructed and we show that with our technique the shape can even change and adapt during a simulation to follow, for example, the conformational transitions of a solvated macromolecule.

Next, in chapter 6, we derive a path integral-based adaptive resolution method that allows a clean coupling of classical force fields and path integrals. The scheme is derived from the bottom up, is Hamiltonian by construction, and is, opposed to previous works, also consistent with a formal path integral quantization procedure. Most importantly, the technique leads to a significant speedup compared to all-quantum simulations.

In chapter 7, we build on the previous results and show how the derived Hamiltonian path integral-based adaptive resolution protocol can be implemented in an efficient manner in a molecular dynamics framework. For that purpose, we devise a tailored integration scheme and validate it on quantum–classical path-integral simulations of liquid water. We also demonstrate that the method can be used to not only calculate quantum statistical quantities in the high-resolution region but also approximate quantum dynamical properties, such as vibrational spectra.

Chapter 2.

Adaptive Resolution Simulations of Water Coupled to an Ideal Gas

This chapter has been published as a research paper in The European Physical Journal Special Topics. It is reprinted here with kind permission of The European Physical Journal (EPJ).

Karsten Kreis, Aoife C. Fogarty, Kurt Kremer, and Raffaello Potestio
Advantages and challenges in coupling an ideal gas to atomistic models in adaptive resolution simulations
Eur. Phys. J. Special Topics 224, 2289-2304 (2015)
DOI: 10.1140/epjst/e2015-02412-1

Abstract

In adaptive resolution simulations, molecular fluids are modeled employing different levels of resolution in different subregions of the system. When traveling from one region to the other, particles change their resolution on the fly. One of the main advantages of such approaches is the computational efficiency gained in the coarse-grained region. In this respect the best coarse-grained system to employ in the low resolution region would be the ideal gas, making intermolecular force calculations in the coarse-grained subdomain redundant. In this case, however, a smooth coupling is challenging due to the high energetic imbalance between typical liquids and a system of non-interacting particles. In the present work, we investigate this approach, using as a test case the most biologically relevant fluid, water. We demonstrate that a successful coupling of water to the ideal gas can be achieved with current adaptive resolution methods, and discuss the issues that remain to be addressed.

2.1. Introduction

Many soft matter systems, ranging from simple liquids to complex polymer mixtures [127, 245–247] and biomolecules such as proteins [248–259], feature a nontrivial interplay of characteristic

length and time scales. Because of this property, structural or energetic changes occurring at a given scale have repercussions on others. Hence, a realistic modeling of these systems has to take into account all fine-grained details that might affect, or be affected by, larger scale features.

In most cases, however, the smallest meaningful size of the system is too large to allow its simulation with a highly detailed model. At the same time, coarse-grained models, which proved to be extremely successful in understanding the properties of many soft matter systems [66, 101, 260–263], cannot be employed in those cases in which the fine-grained detail plays a major role.

A solution to this problem is sometimes offered by dual resolution simulations [144, 176–178, 181–184, 187, 212, 218, 223, 264–266]. Specifically, we consider here those cases in which the finest level of detail is needed only in a relatively small functionally or physically relevant part of the system. At the same time, the remainder is required to provide the aforementioned relevant part with the necessary thermodynamic support, namely the exchange of energy and matter, albeit not being interesting *per se*. This is, for example, the case of a biological system, such as a protein, immersed in a solution of water and cosolvent, whose solvation properties are correctly reproduced only if the number of solvent/cosolvent molecules is large enough to mimic the Grand Canonical limit [267, 268]. In this situation, the (co)solvent is not the interesting part of the computer experiment; however, it forms its largest part. In such cases one can describe the relevant subsystem with the necessary accuracy, and represent the remainder with a simpler, computationally efficient model. Adaptive resolution simulation schemes have been developed to allow this spatially inhomogeneous treatment of the system. In these approaches, the model used to describe a given component is not fixed for the entire duration of the simulation; rather, molecules freely diffuse across the domain, and their resolution, i.e. the model accuracy, is determined by the specific position they occupy in space at a given point in time.

In most cases, one tries to preserve, in the coarse-grained model, certain fundamental properties of the higher resolution system. An example is given by the structure-based coarse-graining procedures applied to molecular fluids, such as iterative Boltzmann inversion (IBI) [106], which aims to obtain an effective potential energy reproducing the radial distribution function (RDF) of the fluid. Other choices are possible, based on the desired target properties.

However, if the scope of the simulation does not focus on the low resolution part of the system, the choice of the coarse-grained interaction may be driven mainly by system simplicity and computational efficiency, provided that the correct thermodynamical conditions are preserved in the high resolution domain. In this case, it is easy to see that the optimal choice is to have no interactions in the low resolution domain. In practical terms, this means coupling the system in the high resolution region to an ideal gas. Besides computational efficiency, such an approach has several other advantages. One is, for example, the faster diffusion in the low resolution domain, which would accelerate the mixing of solvent molecules in different parts of the system, thereby improving the sampling. Another advantage is the possibility of varying the number of particles in the system at will by simply creating or destroying molecules in a region of

the system where they are all "invisible" to each other. This would enable the simulation of a system with fixed chemical potential rather than number of particles, that is, to simulate a truly Grand Canonical ensemble with minimal computational effort.

The present work is devoted to the investigation of the practical feasibility of the coupling of the most biologically relevant fluid, namely water, to an ideal gas of point-like, non-interacting particles each of which corresponds to a water molecule. Two different but related simulation approaches, the force-based adaptive resolution simulation (AdResS) [144, 176, 218, 264] scheme and the energy-based Hamiltonian AdResS (H-AdResS) [181, 183, 184, 266] scheme are employed to perform this coupling, and various strategies are tested and compared to preserve the correct thermodynamics in the high resolution region, where the system is a fully atomistic fluid.

The most immediate advantage of replacing the vast majority of the solvent molecules in the system with an ideal gas is clearly the lack of any force calculation in the low resolution domain. This benefit naturally comes at some cost, namely the large thermodynamical differences existing between the equations of state of an ideal gas and a generic molecular fluid. These free energy discrepancies nontrivially affect the behavior of the hybrid, dual resolution system, and require particular care in the construction of the interface, in order to allow the atomistic, high resolution region to behave as expected. Several strategies have been developed in the past to modulate the thermodynamical balance between the two coupled models [181, 182], but these were in most cases applied to a "well behaved" coarse-grained system, parametrized for the corresponding atomistic system. The large free energy difference, as well as diffusion dynamics that differ by orders of magnitude, make the construction of a smooth seam between water and ideal gas a challenging problem.

2.2. Adaptive resolution simulations

As briefly sketched in the Introduction, the idea lying at the core of dual resolution simulations is to introduce into the simulation domain of a soft matter system a spatial, geometrical separation between two of its parts. One part, typically the smallest, needs to be described with a computationally expensive, high resolution model, and will be henceforth referred to as the atomistic (AT) region, assuming that the single-atom level is the finest we can and need to reach. (This is not the general case, however, and even finer models can be employed, as e.g. in [177, 212].) The other, larger part of the system is necessary inasmuch as it represents at least the thermal bath and particle reservoir of the small subsystem, but is not interesting in itself and can therefore be modeled in terms of coarse-grained (CG) particles and force fields. This subpart retains simpler, smoother non-bonded interactions, which require a smaller number of force calculations.

These two domains are joined together through a hybrid (HY) region, by which one set of interactions is gradually transformed into the other. This smooth change is parametrized in terms of a resolution, or switching, function $\lambda(x)$, which continuously and monotonically

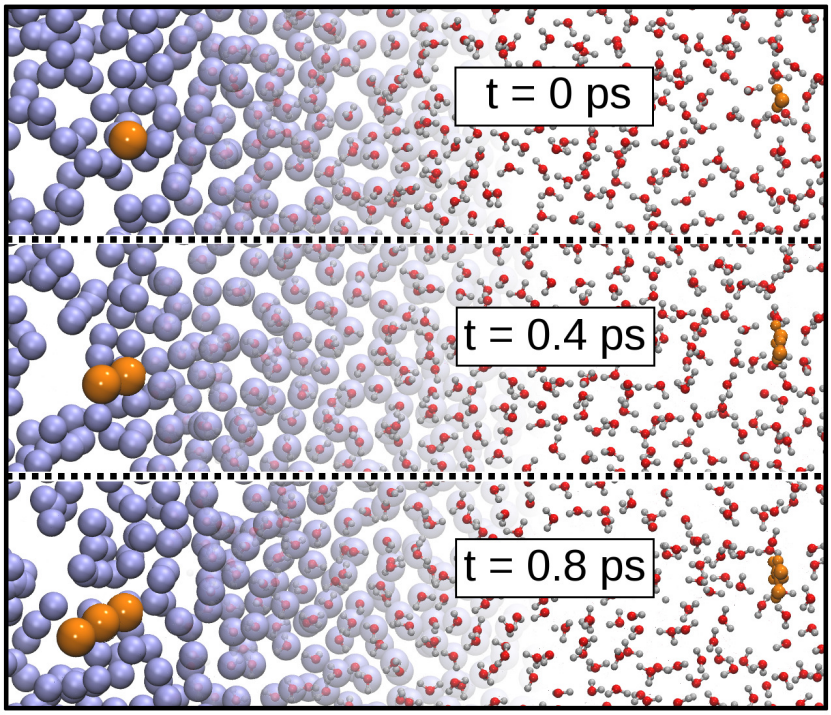

Figure 2.1.: Three consecutive snapshots of the water–ideal gas interface. For clarity we show only molecules in a 0.7 nm thick layer in the direction perpendicular to the sheet. The water, in the right half of the figure, is much more structured than the ideal gas (left), in which several molecules overlap. For the orange highlighted molecules the time evolution is visualized by copying their previous positions into the subsequent snapshots. From this it can be seen that the molecules in the ideal gas diffuse much faster than the water molecules, whose positions almost do not change over 0.8 ps.

goes from 1 (in the AT region) to 0 (in the CG region). In the HY region the interactions are a combination of the atomistic and coarse-grained force fields, the specific form of this interpolation being specific to a given adaptive resolution simulation scheme. Molecules are free to diffuse throughout the whole simulation domain, and their resolution changes dynamically according to their instantaneous position in space.

In general, it is necessary to take into account the fact that the two models follow different equations of state. Therefore the equilibrium state they attain once coupled, each in its pertinent domain, is typically not the state they would have attained if they occupied the entire simulation box [181, 182]. Hence, system-specific modifications have to be enforced to modulate the local equilibrium towards the desired state.

The following subsections 2.2.1 and 2.2.2 are devoted to the description of the two dual resolution simulation schemes employed in the present work. The main difference between them lies in the interpolation of the interactions in the HY interface: the first one, AdResS, is built based on a linear combination of forces; the H-AdResS scheme is formulated in terms

of a Hamiltonian, and the interpolation of the interactions occurs first at the level of potential energies.

2.2.1. Force-based approach

The adaptive resolution simulation (AdResS) [144, 176, 218, 264] scheme is based on the direct interpolation of two force fields. This approach satisfies Newton's Third Law exactly and instantaneously in every part of the system, including the HY region. The atomistic and coarse-grained forces acting between two molecules are linearly interpolated in a symmetric fashion, as follows:

$$\mathbf{F}_{\alpha\beta} = \lambda(\mathbf{R}_\alpha)\lambda(\mathbf{R}_\beta)\mathbf{F}^{AT}_{\alpha\beta} + \left(1 - \lambda(\mathbf{R}_\alpha)\lambda(\mathbf{R}_\beta)\right)\mathbf{F}^{CG}_{\alpha\beta} \tag{2.1}$$

In Eq. 2.1, $\mathbf{R}_\alpha$ (resp. $\mathbf{R}_\beta$) is the centre of mass coordinate of molecule α (resp. β). $\mathbf{F}^{AT}_{\alpha\beta}$ and $\mathbf{F}^{CG}_{\alpha\beta}$ are, respectively, the atomistic and the coarse-grained forces acting on molecule α due to the interaction with molecule β.

The coupling of two different models of the same system naturally leads to a thermodynamical imbalance, e.g. if one of the two has, for a given temperature and density, a higher virial pressure than the other. In this particular case, the system will evolve towards an equilibrium state in which the pressure gradients are flattened out, but the density profile of the system will not be uniform in the direction of resolution change. To enforce a uniform density profile, one can make use of an external field, called Thermodynamic Force [182] (TF), which is obtained iteratively according to the following update scheme:

$$\mathbf{f}^{i+1}_{th} = \mathbf{f}^{i}_{th} - \frac{M}{\rho_0^2 \kappa_T}\nabla\rho^i(r) \tag{2.2}$$

where $\rho^i(r)$ is the density profile as a function of position after the i-th iteration, M is the molecular mass, ρ_0 is the reference density and κ_T is the isothermal compressibility of the fluid. By construction, this iterative protocol reaches a fixed point when the density is uniform, and the update term $\nabla\rho$ is zero. It is worth mentioning that enforcing a flat density profile is not the only option: in fact, one might wish to keep other thermodynamical quantities, such as pressure, or higher-order correlations, e.g. RDFs, constant throughout the system. In principle one can enforce, in the whole system, a uniform profile for two or more of these quantities, but this possibility ultimately depends on the coarse-grained model: in some cases, in fact, having a CG force field that correctly reproduces the reference value of a given thermodynamic quantity implies the impossibility to do the same for one other quantity which is conjugate to the first. A well-known example of this is provided by the pressure and the compressibility of a CG potential obtained *via* IBI, which cannot be simultaneously matched [269]. On the other hand, this limitation of the CG force field does not turn into a limitation of the adaptive approach, since in most cases the only requirement is to have the desired thermodynamics correctly reproduced in the sole AT region.

The force field in Eq. 2.1 is intrinsically non-conservative, as it cannot be obtained as the negative gradient of a potential energy function [219]. This fact determines some limitations to the applicability of this approach, most notably the inability to perform Microcanonical and Monte Carlo simulations. Additionally, a local thermostat is required to enforce a state of dynamical equilibrium, in which the temperature is constant. On the other hand, it can be proven that the system correctly samples the Canonical ensemble in the AT region, and the configurations generated in a molecular dynamics simulation with the AdResS scheme are compatible with thermal equilibrium [228, 265].

2.2.2. Energy-based approach

The second simulation strategy we employ in the present work is the Hamiltonian AdResS (H-AdResS) [181, 183, 184, 266] method, in which two models of a system are coupled directly at the level of potential energies. The H-AdResS scheme is hence formulated in terms of a potential energy function, defined as:

$$V_{H-AdResS} = \mathcal{K} + V^{int} + \sum_{\alpha}^{N} \left\{ \lambda_\alpha V_\alpha^{AT} + (1-\lambda_\alpha) V_\alpha^{CG} \right\} \tag{2.3}$$

where N is the number of molecules, $\mathcal{K}$ is the kinetic energy, V^{int} is the internal potential energy of the molecules, and:

$$\begin{cases} V_\alpha^{AT} = \dfrac{1}{2} \displaystyle\sum_{\beta,\beta\neq\alpha}^{N} \sum_{ij} V^{AT}(|\mathbf{r}_{\alpha i} - \mathbf{r}_{\beta j}|) \\ V_\alpha^{CG} = \dfrac{1}{2} \displaystyle\sum_{\beta,\beta\neq\alpha}^{N} V^{CG}(|\mathbf{R}_\alpha - \mathbf{R}_\beta|) \\ \lambda_\alpha = \lambda(\mathbf{R}_\alpha) \end{cases}$$

Contrary to the force-based approach, H-AdResS also allows NVE and MC simulations. The drawback is the presence in the force field of a term proportional to the gradient of the switching function. In fact, the force acting on molecule α reads:

$$\begin{aligned} \mathbf{F}_\alpha = & \sum_{\beta,\beta\neq\alpha} \left\{ \frac{\lambda_\alpha + \lambda_\beta}{2} \mathbf{F}_{\alpha|\beta}^{AT} + \left(1 - \frac{\lambda_\alpha + \lambda_\beta}{2}\right) \mathbf{F}_{\alpha|\beta}^{CG} \right\} \\ & - \left[V_\alpha^{AT} - V_\alpha^{CG} \right] \nabla_\alpha \lambda_\alpha \end{aligned} \tag{2.4}$$

The first term of Eq. 2.4 bears some similarity with the AdResS force interpolation of Eq. 2.1: both, in fact, are anti-symmetric with respect to molecule label exchange. On the other hand, they differ already at the level of the interpolation weights: in the AdResS case, in fact, they are given by the product of the switching functions of the two molecules, while in the H-AdResS case the average of the λ's naturally appears. This difference mainly results in diverse interactions between molecules in the CG region and those in the hybrid interface. The AdResS scheme treats the interaction with a molecule in the CG domain as purely CG, while in

the H-AdResS scheme the CG molecules interact also at the atomistic level with the ones in the hybrid region, even though only weakly. For all practical purposes, however, this difference bears no consequences.

The last term contains the largest difference with the force-based method, namely a term proportional to the gradient of the switching function. This term, referred to as the drift force, contains the difference between the atomistic and the coarse-grained potential energy of a molecule, and is zero outside the HY region, where $\nabla\lambda \equiv 0$ by construction. In the HY region, though, it locally breaks Newton's Third Law, inasmuch as the force acting between molecule pairs cannot be written as a sum of antisymmetric terms.

The drift force can be compensated on average by including in the definition of the Hamiltonian a new term conceptually similar to the thermodynamic force employed in the AdResS scheme [181, 183]:

$$V_\Delta = V_{H-AdResS} - \sum_{\alpha=1}^{N} \Delta H(\lambda(\mathbf{R}_\alpha)) \tag{2.5}$$

The functional form of this term is defined by the requirement that its corresponding force and the drift force cancel out, i.e.:

$$\left.\frac{d\Delta H(\lambda)}{d\lambda}\right|_{\lambda=\lambda_\alpha} = \left\langle \left[V_\alpha^{AT} - V_\alpha^{CG}\right] \right\rangle_{\mathbf{R}_\alpha} \tag{2.6}$$

where the subscript in the average indicates that the latter has to be performed constraining the coarse-grained site of molecule α in the position $\mathbf{R}_\alpha$. In many cases the term ΔH can be well approximated by the Helmholtz free energy difference between the coarse-grained model and a hybrid model with mixing parameter λ. In practice ΔH can be obtained by means of a Kirkwood Thermodynamic Integration [181, 183, 230] (TI):

$$\Delta H(\lambda) = \int_0^\lambda d\lambda' \frac{d\Delta H(\lambda')}{d\lambda'} \simeq \frac{1}{N}\int_0^\lambda d\lambda' \left\langle \left[V^{AT} - V^{CG}\right] \right\rangle_{\lambda'} = \frac{\Delta F(\lambda)}{N}$$

This choice of ΔH removes the average effect of the drift force, thereby restoring, although only on average, Newton's Third Law also in the HY region. However, it does not guarantee that the density in the AT and the CG domain attains the same value. To this end, it is necessary to add a term proportional to the pressure difference between the two models, which amounts to choosing, for ΔH, the chemical potential difference as a function of the resolution:

$$\Delta H(\lambda) = \frac{\Delta F(\lambda)}{N} + \frac{\Delta p(\lambda)}{\rho_0} \equiv \Delta\mu(\lambda) \tag{2.7}$$

This correction to the Hamiltonian takes the name of free energy compensation (FEC) [181, 183].

2.3. Simulation details

All simulations presented here used a system containing 6526 particles (water molecules and ideal gas particles) in a simulation box of dimensions $\sim 16.1 \times 3.5 \times 3.5$ nm. This corresponds to a density of 33.1 molecules nm^{-3} or 990.7 kg m^{-3}, a value determined via fully atomistic simulations at a pressure of 1 bar and 300 K in the isothermal-isobaric ensemble, and close to the experimental density under ambient conditions. The instantaneous resolution of a given particle was determined by the distance d_x along the X-axis between its centre of mass and the centre of the simulation box. The atomistic region was defined as $d_x \leq 3.0$ nm, flanked by two HY regions at $3.0 < d_x < 5.0$ nm, and the coarse-grained region at $d_x \geq 5.0$ nm. Periodic boundary conditions were used. The water–ideal gas interface region is illustrated in Fig. 2.1. To assign to a molecule its position-dependent resolution λ, its distance from the boundary between the atomistic and the HY region is computed, i.e. $d_x - d_{\text{at}}/2$, where d_{at} is the width of the atomistic region. This quantity is then inserted into the resolution function $\lambda(x)$, which is given as:

$$\lambda(x) = \begin{cases} 1 & : x \leq 0 \\ 1 - \frac{30}{d_{\text{hy}}^5}\left(\frac{1}{5}x^5 - \frac{d_{\text{hy}}}{2}x^4 + \frac{d_{\text{hy}}^2}{3}x^3\right) & : 0 < x < d_{\text{hy}} \\ 0 & : x \geq d_{\text{hy}} \end{cases} \tag{2.8}$$

where d_{hy} denotes the width of the HY region. In our simulations, $d_{at} = 6$ nm and $d_{hy} = 2$ nm.

Simulations were performed using the ESPResSo++ package [1], and a time step of 2 fs. Atomistic water was represented using the SPC/E model [270] and the SETTLE algorithm [55]. Non-bonded interactions used a cutoff of 1.0 nm. Electrostatic interactions were calculated using the reaction field method with a dielectric constant of 67.5998, as previously determined for SPC/E water [174]. Production runs had a length of between 200 ps and 2 ns, depending on the convergence time of the properties studied, and were performed in the canonical ensemble at a temperature of 300 K, using the Langevin thermostat with a friction constant of 0.5 ps^{-1} in H-AdResS and 5.0 ps^{-1} in AdResS. The stronger coupling to the thermostat in the force-based scheme is necessary in order to counteract the excess heat produced in the HY region due to the removal of degrees of freedom and the non-conservative nature of the force interpolation simulations [265]. The use of non-interacting ideal gas particles in the coarse-grained region allows particle positions to overlap, and two overlapping ideal-gas particles entering the HY region, where they begin to interact, may lead to unmanageably large forces. The absolute magnitude of the force between any particle pair was therefore capped at 10^4 kJ mol^{-1} nm^{-1}, in order to allow particle pairs to adapt their inter-particle distance as interaction strength gradually increases across the HY region.

As well as coupling the atomistic water model to an ideal gas, for comparison we also performed simulations coupling atomistic water to a coarse-grained potential developed using Iterative Boltzmann Inversion with pressure correction [106], which reproduces the pressure of the underlying atomistic model and provides an excellent approximation of its RDF. This IBI potential acted on particle centres of mass, and was obtained using the VOTCA coarse-graining

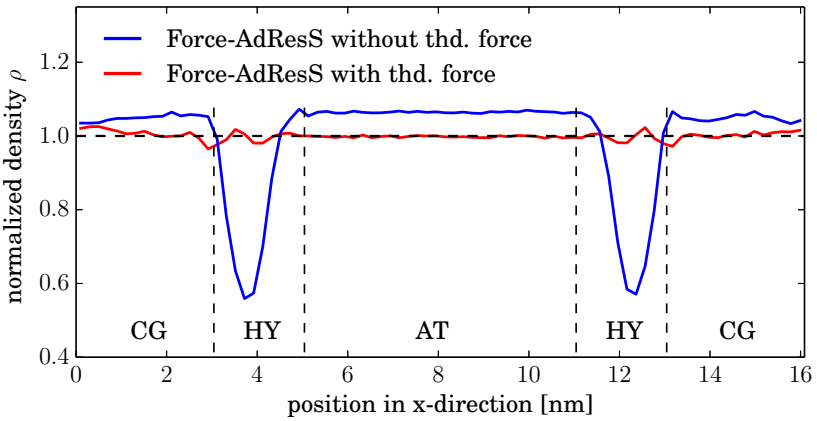

Figure 2.2.: Density profiles for force-AdResS water–ideal gas system, with and without thermodynamic force.

package [271], and 300 IBI steps, each running for 100 ps.

2.4. Results and discussion

2.4.1. Force-based coupling with the AdResS method

First, we couple the atomistic water model to an ideal gas using the force-based AdResS scheme. In this case, i.e. when the coarse-grained interaction is absent, Eq. 2.1 reduces to:

$$\mathbf{F}_{\alpha\beta} = \lambda(\mathbf{R}_\alpha)\lambda(\mathbf{R}_\beta)\mathbf{F}^{AT}_{\alpha\beta} \tag{2.9}$$

A direct coupling of the liquid to the ideal gas produces the density profile reported in Fig. 2.2 (blue line). A considerable depletion can be observed in the HY region, where the density drops to less than 60% of the reference value. Correspondingly, in the AT and CG region the density is higher, but the value attained is almost the same in both domains. The marked dip in the density profile is not symmetric and is located closer to the border with the CG than with the AT region. Its origin can be understood by considering that particles entering the HY region from the CG region, where they were non-interacting, may be located very close together in space. In order to permit stable simulations this is dealt with using force-capping, as outlined above; the high forces nonetheless lead to a peak in the virial pressure profile at the point corresponding to the dip in the density profile, when an AdResS simulation is launched from an initial configuration with a uniform density profile.

In order to remove this depletion we make use of the TF iterative correction, Eq. 2.2, discussed in Sect. 2.2.1. The potential energy function corresponding to the converged TF force field is reported in Fig. 2.3. As shown in Fig. 2.2 (red line), the application of the TF enforces a perfectly flat density profile in the AT region, attaining the reference value. Small fluctuations,

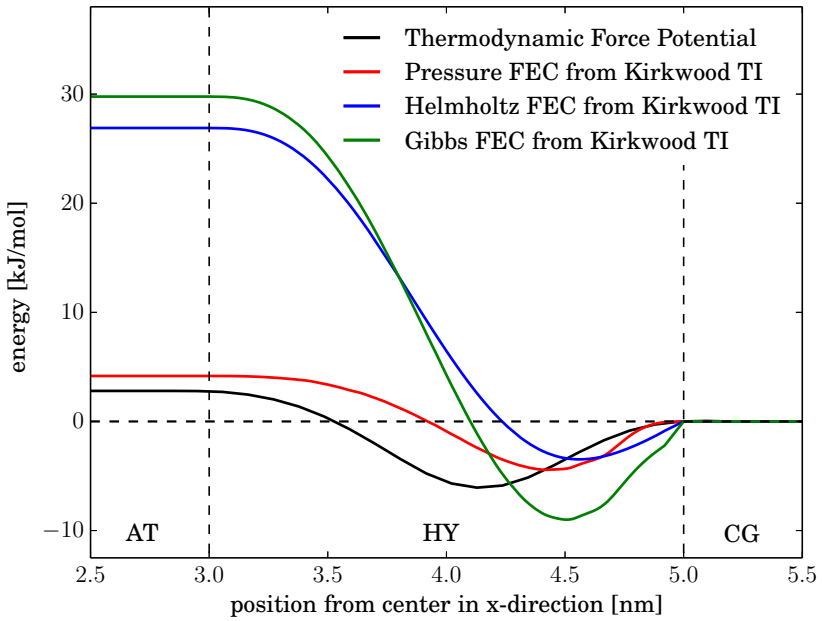

Figure 2.3.: Potential energies corresponding to the Thermodynamic Force and FEC contributions as obtained from Kirkwood TI. The Gibbs FEC is the sum of the Helmholtz and the pressure FECs.

of the order of $1-2\%$, can be observed in the HY region. In the CG region the ideal gas density deviates by almost the same amount from the reference. The slightly noisy profile in the latter region is due to the difficulty to average the position of non-interacting particles.

2.4.2. Energy-based coupling with the H-AdResS method

Next, we move to the case of the H-AdResS method. In the ideal gas case, the H-AdResS Hamiltonian (without FEC) becomes:

$$H = \mathcal{K} + V^{int} + \sum_{\alpha} \lambda_{\alpha} V_{\alpha}^{AT} \tag{2.10}$$

In this case, as in the AdResS case, the simple coupling of atomistic water to the ideal gas results in a depletion of molecules in the HY region, see Fig. 2.4. Additionally, we observe a different density between the AT and the CG regions, due to the presence of the drift force. Applying the FEC, whose plot is reported in Fig. 2.3, we see that the density attains the reference value in the whole AT region and in most of the HY region. A small depletion of approximately 10% is observed at the HY/CG interface, then the density in the CG region flattens out again, to a value $1-2\%$ higher than the reference. Also in this case the higher density in the CG domain is due to the depletion in the HY region. As already mentioned, the origin of this depletion lies in overlapping pairs of molecules diffusing from the CG region

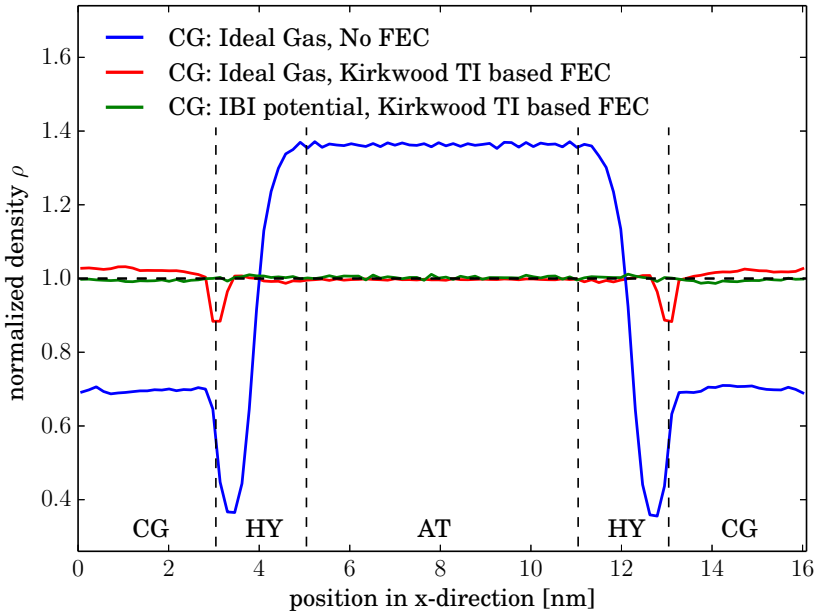

Figure 2.4.: Density profiles for H-AdResS simulations of the water–ideal gas system with and without Kirkwood TI based Gibbs FEC as well as for the water–IBI system, also with Kirkwood TI based Gibbs FEC.

into the HY layer, where their high intermolecular forces lead to high virial pressure. The deviations from the reference density observed in the hybrid simulations making use of the ideal gas as a coarse-grained model are relatively small, do not affect the AT region and, as it will be discussed below, the density dip does not prevent molecules from diffusing across the interfaces separating different resolutions. The application of the thermodynamic force scheme already employed in the AdResS case would in any case remove any minor deviation from the reference density.

For comparison, we performed a simulation of the same atomistic water model coupled to a coarse-grained model obtained *via* iterative pressure-corrected IBI. In this case, the well-parametrized coarse-grained model, together with the FEC field, enforces a very uniform density throughout the system. This demonstrates that when the fully atomistic system is coupled to a coarse-grained model reproducing at least a few thermodynamical properties of the former at the same state point, the FEC term is sufficiently accurate to remove the remaining discrepancies in the free energy. In the ideal gas case, in contrast, the mean-field character of the FEC becomes apparent in the imperfect correction occurring at the HY/CG interface, where local correlations are present that are not compensated for. The trade-off between a non-flat density profile and the advantage of not having to parametrize the coarse-grained model depends on the specific system under consideration.

We now move on to the quantitative assessment of the correctness of the structure of the fluid in the AT region of the water–ideal gas hybrid system. We measured the pair correlation

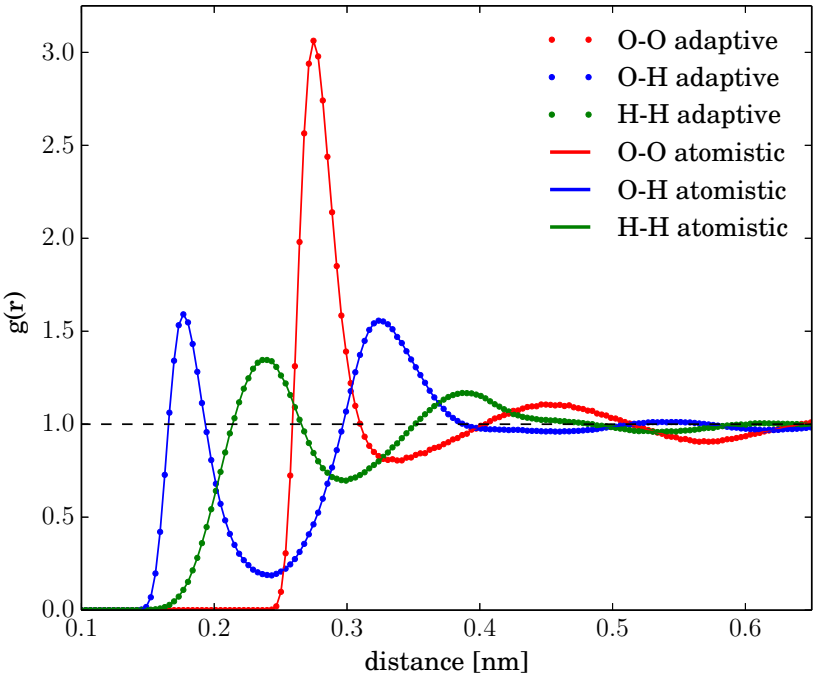

Figure 2.5.: RDFs for pure water and for the atomistic region of the H-AdResS water–ideal gas with Kirkwood TI based Gibbs FEC. Since a rigid water model is employed, all RDFs consider only intermolecular atom pairs.

functions for pairs of atoms in which at least one atom had a distance along the X-axis of less than 1.5 nm from the centre of the atomistic region. Since the sum of this distance and the RDF cutoff is less than the distance used to define the atomistic region, the RDFs calculated are fully atomistic. These RDFs are fully consistent with those measured in completely atomistic reference simulations, as is evident in Fig. 2.5.

A crucial point is to verify that the system in the AT region behaves as if it were a subpart of a completely atomistic system. This means not only measuring the density profile and the RDFs, but also checking that the molecules do not have any impediment in diffusing across the HY region. To this end, we followed the time evolution of a subset of labelled particles at the AT/HY and CG/HY interface, to check that nothing prevents them from moving across the system. Obviously, we can expect a different diffusion rate in the two cases, as the friction of the ideal gas is decidedly smaller than that experienced by the atomistic water molecules.

In Fig. 2.6 (solid lines) we report the diffusion profiles of the molecules initially located in two symmetric slabs of width 1 nm in the AT region, at the interfaces with the HY regions. These molecules, uniformly distributed at $t = 0$, spread out throughout the whole system as time passes. The overall distribution is quasi-Gaussian, but the half moving towards the CG region extends further than the half in the AT region, as the friction in the former allows a faster diffusion.

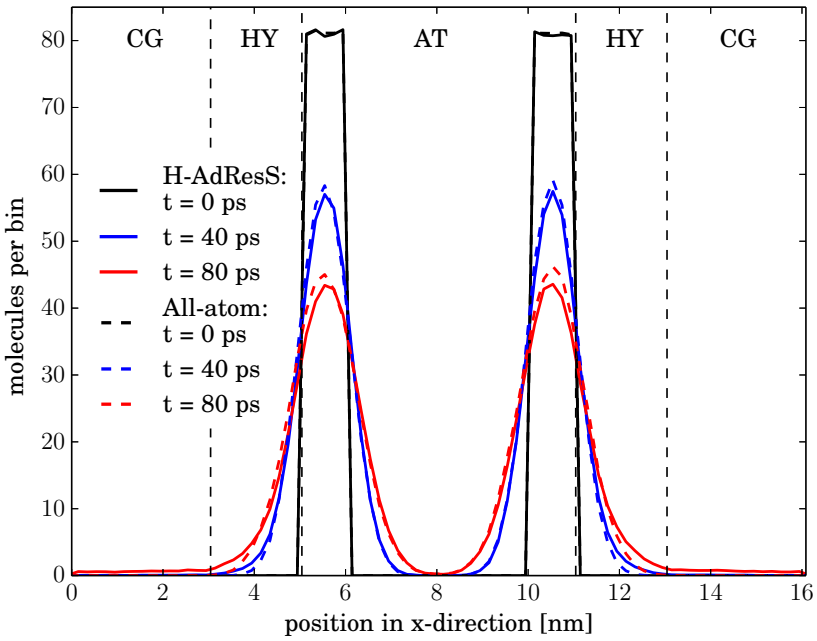

Figure 2.6.: Diffusion profiles in H-AdResS simulations of the water-ideal gas system and in fully atomistic reference simulations of SPC/E water: time evolution of the position of molecules initially located in a 1-nm-wide slab in the atomistic region, immediately adjacent to the HY region. The y-axis is the absolute number of these molecules whose centre of mass X-coordinate is in a given bin at the given time.

The diffusion in the AT region, though, is perfectly compatible with that of a fully atomistic water system. This can be verified by comparing the diffusion profiles of the H-AdResS simulation with those obtained by performing the same analysis on a fully atomistic system. These latter distributions, reported in Fig. 2.6 (dashed lines), overlap very well in the AT region and even in part of the HY region, while, as expected, the distribution of the molecules in the H-AdResS simulation (solid lines) extends deeper in the CG region. This asymmetry is due to the fact that the CG model, the ideal gas, has by construction very different transport properties with respect to the atomistic water model, and differences between the two cases have to be present. Our attention, though, focusses on the atomistic subdomain, where the diffusion occurs exactly as in the reference system, and this is the only aspect we deem crucial.

This difference between the equilibration times of atomistic water and of the ideal gas is most clearly seen when comparing the plots in Fig. 2.6 (solid lines) with those showing the diffusion of molecules at the CG/HY interface, Fig. 2.7. Not only are the distributions for $t > 0$ more skewed than the ones previously shown, they also evolve on a much faster time scale. For the distribution peak to reach the height that the particles at the AT/HY interface attain at $t = 40$ ps, those in the CG/HY interface need only 2 ps.

Finally, we studied the density fluctuations across the system, since these can be expected

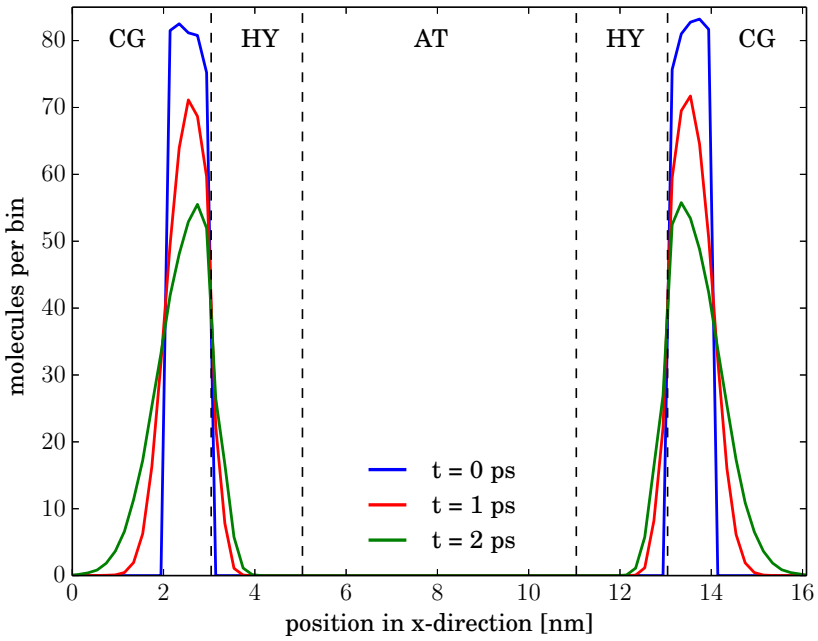

Figure 2.7.: Diffusion profiles in H-AdResS simulations of the SPC/E water-ideal gas system: time evolution of the position of molecules initially located in a 1-nm-wide slab in the coarse-grained region, immediately adjacent to the HY region. The y-axis is the absolute number of these molecules whose centre of mass X-coordinate is in a given bin at the given time.

to differ enormously between a fluid of non-interacting particles and a condensed, strongly interacting fluid. We measured the molecule number fluctuations, a quantity proportional to the compressibility and defined as

$$\Delta N \equiv \frac{\langle N^2 \rangle - \langle N \rangle^2}{\langle N \rangle} \tag{2.11}$$

where N is the number of particles in a 1-nm wide slab of the simulation box. The resulting profiles are shown in Fig. 2.8. Although the local density fluctuations in the ideal gas region are almost an order of magnitude higher than those in the atomistic region, the latter nonetheless correspond perfectly to the local density fluctuations in a fully atomistic system.

2.4.3. Computational gain

When performing adaptive dual resolution simulations, one of the main advantages when coupling to a coarse-grained potential is the computational gain over detailed, fully atomistic simulations. In this respect the ideal gas is the most efficient coarse-grained model, as it corresponds to no interaction at all. In this section we report a comparison of the performance of the adaptive water–ideal gas system and a fully atomistic water setup. Additionally, we also

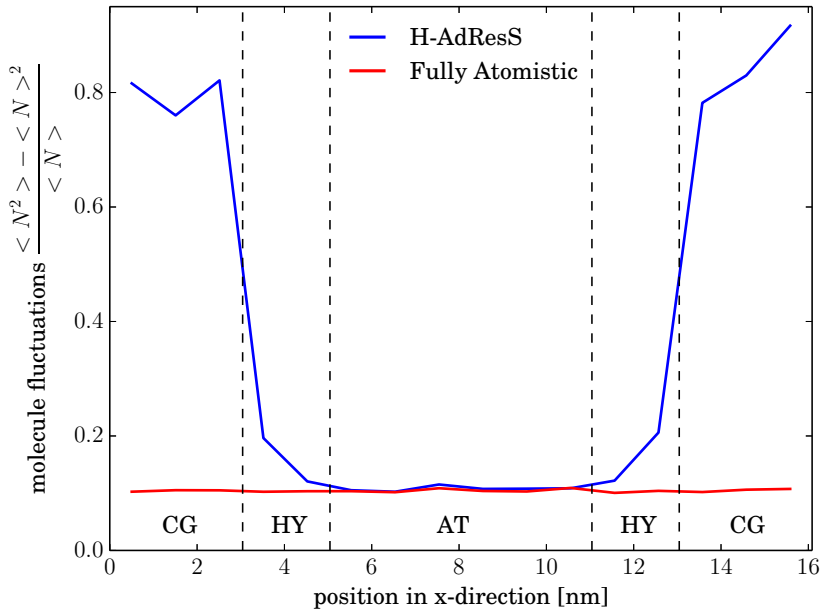

Figure 2.8.: Molecule fluctuations as a function of position in the simulation box for H-AdResS water–ideal gas simulation with Gibbs FEC as well as for fully atomistic reference simulations.

discuss the ideal gas compared to other coarse-grained models, i.e. the coarse-grained potential from IBI.

We performed 4 sets of simulations, with four different box lengths, each consisting of an atomistic simulation as well as a water–IBI and a water–ideal gas H-AdResS simulation. Each of them ran for 10 ps. In the H-AdResS simulations, Kirkwood TI based Gibbs FEC's were employed in order to keep the densities flat (compare Fig. 2.4). In all cases under examination the atomistic region has a total width of 2.0 nm; similarly, the adjacent HY regions always have widths of 2.0 nm each. The size of the CG region, however, was chosen differently for all sets; the total simulation box sizes are presented, together with the corresponding molecule numbers, in Tab. 2.1. In applications of adaptive resolution schemes, the high resolution region is typically, though not necessarily, much smaller than the coarse-grained region. Therefore, in our adaptive test setups the AT and HY regions occupy only a relatively small volume in the simulation boxes compared to the CG part.

To avoid the inclusion of implementation and platform dependent run times we considered, in the measurements reported, only those quantities that differ for all-atom and adaptive H-AdResS simulations. Specifically, we only measured the time that the code spends with the calculation of non-bonded, pairwise forces and with the calculation of the drift term. In order to obtain results independent of parallelization, we ran the simulations on a single CPU. In parallelized adaptive resolution simulations, a load balancing protocol is required to benefit from the computational efficiency of coarse-grained potentials in the low resolution region.

Number of molecules	L_x	L_y	L_z
6526	16.086 nm	3.500 nm	3.500 nm
9803	24.164 nm	3.500 nm	3.500 nm
13064	32.202 nm	3.500 nm	3.500 nm
16349	40.300 nm	3.500 nm	3.500 nm

Table 2.1.: Number of molecules and box geometries for the different sets of simulations for the calculation of the computational gain of water–ideal gas simulations compared to full atomistic and water–IBI simulations.

The results of the simulations are plotted in Fig. 2.9, with the corresponding speedups being presented as an inset in the same figure. The speedups, reported in the inset of in Fig. 2.9, are defined as $T_{\text{atomistic}}/T_{\text{adaptive}}$, with $T_{\text{atomistic}}$ being the time spent for force calculations in the atomistic simulations and T_{adaptive} being the corresponding time in the adaptive simulations.

It can be seen that the force calculations in the adaptive simulations are significantly faster than their corresponding atomistic counterparts. For the largest box, the non-bonded force and drift term calculations in the water–ideal gas simulations are faster by a factor of ≈ 3.5 than the full atomistic simulations. The water–IBI simulations reach a similar, though lower, speed-up. Therefore, the computational efficiency gained by coupling water to an ideal gas is slightly higher than the one gained by coupling to a typical coarse-grained potential. Furthermore, our results show that the time required for the adaptive simulations employing the ideal gas stays nearly constant for the different box sizes. The reason for it is that the simulation time is dictated by the interactions in the AT and HY subdomains, whose size does not change.

2.5. Conclusions

In tackling a wide spectrum of challenging problems in soft matter physics, dual resolution simulation methods can represent an advantageous simulation strategy. In fact, they allow us to provide a relatively small system, described with a highly detailed but computationally intensive model, with an accurate thermodynamical environment at a limited cost in terms of simulation resources. In particular, when the focus is concentrated on the high resolution subsystem, and the realistic modeling of the coarse-grained domain is of no interest, it is natural to push the simplification of the latter to the maximum.

With this goal in mind, we carried out the present work in order to study if an ideal gas can be employed as a highly coarse-grained model for water in an adaptive resolution setup. Our results, obtained from different simulation schemes, show that the basic requirements are indeed satisfied: in the force-based as well as in the energy-based case, the properties of the system in the AT domain are compatible with those that one would measure in a similar subregion cut from a fully atomistic simulation.

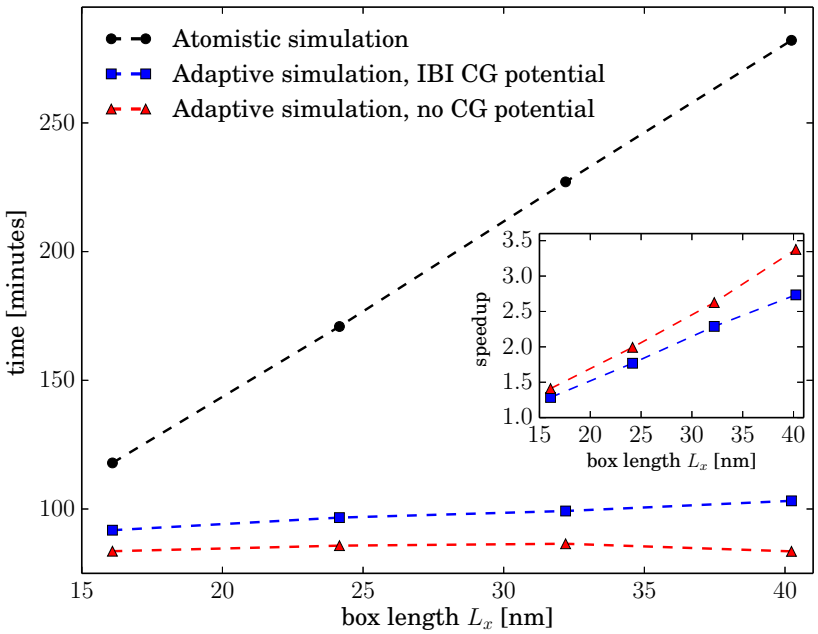

Figure 2.9.: Main figure: times required for the non-bonded force calculations in the atomistic and adaptive simulations for systems of four different box lengths L_x. For the atomistic simulations, these are only non-bonded, pairwise forces, while in the adaptive H-AdResS simulations, also the time required for the drift term calculation is considered. Inset: speedup of the adaptive over the atomistic simulations. The lines are a guide to the eye.

The equation of state of an ideal gas substantially differs from that of a liquid at the same temperature and density. In order to attenuate the deviations from the reference densities observed in both setups we introduce compensating external fields, namely the TF and FEC terms. These fields level out thermodynamical differences between the atomistic water model and the ideal gas, thereby maintaining the density of the fluid at the reference value.

Two-body correlations and relative fluctuations in the number of particles, as measured in the AT domain, perfectly reproduce those measured in a fully atomistic simulation. The diffusion profiles as well indicate that the dynamics in the high resolution region is not affected by the presence of a super-coarse-grained reservoir. These results are obtained at a very small computational cost compared to a fully atomistic simulation of a system having equal size. Additionally, the computational gain does not only come from the shorter time required to perform an integration step compared to a fully atomistic simulation. In fact, a faster diffusion of the solvent accelerates the configurational sampling. Consequently, the simulation duration necessary to equilibrate time independent physical observables is reduced.

Last but not least, the major advancement allowed by the coupling of an ideal gas to a molecular fluid is given by the freedom to "tune" the physics in the low resolution region. The absence of any interaction in the latter enables one to easily insert and delete particles, making

it possible to regulate thermodynamical quantities such as pressure and chemical potential, and employ smaller simulation boxes without introducing finite size effects.

Some issues, however, still remain open. For example, the mean field character of the FEC obtained through Kirkwood TI makes it insufficient, in the H-AdResS simulations, to completely remove a small but noticeable deviation from the reference density at the HY/CG interface. This difficulty might prove to be particularly challenging when the focus moves towards more complex systems, such as solvent/cosolvent mixtures. Specifically for this case, in fact, it might be worth employing more accurate, iterative algorithms to construct an external field which counteracts the average drift force exactly, as suggested in Ref. [184]. A further possibility could be to combine different methods, namely the FEC and the TF.

2.6. Acknowledgments

K. Kreis is recipient of a fellowship funded through the Excellence Initiative (DFG/GSC 266). K. Kremer and A. Fogarty acknowledge research funding through the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement n. 340906-MOLPROCOMP. R. Potestio and K. Kremer acknowledge funding from the SFB–TRR 146 grant. The authors are indebted to T. Bereau and D. Mukherji for a careful reading of the manuscript.

Chapter 3.

A Unified Framework for Force-based and Energy-based Adaptive Resolution Simulations

This chapter has been published as a research paper in letter format in Europhysics Letters. It is reprinted here with permission from the publisher.

Karsten Kreis, Davide Donadio, Kurt Kremer, and Raffaello Potestio
A unified framework for Force-Based and Energy-Based Adaptive Resolution Simulations
EPL (Europhysics Letters) 108, 30007 (2014)
DOI: 10.1209/0295-5075/108/30007

Abstract

Adaptive resolution schemes enable Molecular Dynamics simulations of liquids and soft matter employing two different resolution levels concurrently in the same setup. These methods are based on a position-dependent interpolation of either forces or potential energy functions. While force-based methods generally lead to non-conservative forces, energy-based ones include undesired force terms proportional to the gradient of the interpolation function. In this work we establish a so far missing bridge between these formalisms making use of the Generalized Langevin Equation, thereby providing a unifying framework to traditionally juxtaposed approaches to adaptive simulations.

3.1. Letter

Simulation methods in which two different descriptions of the same system (all atom/coarse grained, quantum mechanical/classical...) are concurrently employed have been developed and applied since the early times of computational science [156, 158–160, 166–168, 170, 175–183, 195–208, 212, 272, 273]. In these schemes only a small region of the simulation domain is treated

with an accurate, computationally expensive model (e.g. an atomistic model), while for the rest a simpler and more efficient representation (e.g. coarse-grained) is employed. In most methods the description of a molecule is fixed for the whole duration of the simulation. Approaches in which the molecules resolution changes 'on the fly' according to its position in space, e.g. due to diffusion, are dubbed adaptive resolution methods. In the present work we focus on the latter, which enable the simulation of soft matter, and specifically of liquids. Furthermore, we concentrate only on classical systems, and do not address Quantum Mechanics-Molecular Mechanics models.

Concerning particle-based, classical simulations, two main classes of adaptive resolution simulation schemes can be identified: force-based and energy-based. The first ones [176, 178, 182] share the common feature that the two models employed are merged *via* a position-dependent interpolation of the force acting between molecule pairs; in the second case [179–181, 183] this interpolation is performed at the level of potentials. The direct interpolation of forces enables the instantaneous preservation of Newton's Third Law [176], often desired in Molecular Dynamics simulations to correctly describe hydrodynamics [241]. This comes at the price of having an intrinsically non-conservative force-field which requires a local thermostat to enforce thermodynamic equilibrium and stability [219, 223].

Energy-based schemes are conservative and Hamiltonian by construction, allowing an explicit, partition function-based theoretical treatment as well as energy-conserving MD simulations or Monte Carlo simulations [181, 183]. On the other hand, the forces obtained differentiating these energy functions contain terms proportional to the gradient of the position-dependent function employed to interpolate between the models. Such terms cannot be written as a sum of pairwise antisymmetric terms; therefore, they break Newton's Third Law and momentum conservation, and are generally seen as an undesired complication. However, they are only present where the gradient of the position-dependent interpolation function does not disappear, i.e. only in the interpolation region, hence leaving Newton's Third Law in the bulk high- and low-resolution regions valid. Removing these terms would effectively transform an energy-based approach into a force-based one.

Exact momentum conservation and energy conservation cannot be simultaneously achieved in an adaptive resolution simulation [219]. This fundamental limitation has hampered the widespread application of adaptive resolution methods in soft matter. Advantages and limitations of force-based and energy-based dual-resolution schemes have been largely investigated [181–183, 219, 223, 228]. Nonetheless, it is still necessary to set a unified framework that would encompass the two classes of methods and contextualize connections and differences between them. In the present work we pursue this goal by establishing a formal relation between an energy-based method, the Hamiltonian adaptive resolution simulation scheme (H-AdResS [181, 183]), and a force-based scheme derived from it.

We begin by considering two Hamiltonians, both having the same number of molecules:

$$H_r = \mathcal{K} + \sum_{\alpha} V_{\alpha}^{r}, \quad r = 0, 1 \tag{3.1}$$

where $\mathcal{K} = \sum_\alpha p_\alpha^2/2m_\alpha$ is the total kinetic energy. Without loss of generality we assume the molecules to be point-like particles; the extension to multi-atomic molecules is straightforward.

The single-molecule potentials V_α^r are the sums of all intermolecular potentials acting on molecule α, properly normalized so that double counting is avoided [181, 183]. In most of the applications [175, 187] a fully atomistic ($r = 1$) and a coarse-grained ($r = 0$) potential are used. In the following, though, no assumption is made on the form of these interactions other than finite range.

These Hamiltonians are interpolated, in the H-AdResS scheme [181], with a resolution function $\lambda_\alpha = \lambda(\mathbf{R}_\alpha)$, varying monotonically between 1 and 0 as a function of the position of the molecule $\mathbf{R}_\alpha$. This Hamiltonian H_{mix} and the force $\mathbf{F}_\alpha$ acting on molecule α are:

$$H_{mix} = \mathcal{K} + \sum_\alpha \left\{ \lambda_\alpha V_\alpha^1 + (1-\lambda_\alpha) V_\alpha^0 \right\} \tag{3.2}$$

$$\mathbf{F}_\alpha = \mathbf{F}_\alpha^N + \mathbf{F}_\alpha^{dr} \tag{3.3}$$

with

$$\mathbf{F}_\alpha^N = \sum_\beta \left\{ \frac{\lambda_\alpha + \lambda_\beta}{2} \mathbf{F}_{\alpha|\beta}^1 + \left(1 - \frac{\lambda_\alpha + \lambda_\beta}{2}\right) \mathbf{F}_{\alpha|\beta}^0 \right\} \tag{3.4}$$

$$\mathbf{F}_\alpha^{dr} = -[V_\alpha^1 - V_\alpha^0] \nabla \lambda_\alpha \tag{3.5}$$

where we introduced the total force $\mathbf{F}_{\alpha|\beta}^r$ acting on molecule α due to the interaction V^r with molecule β. Without loss of generality for our following derivation, we have explicitly written the forces for the case of pairwise interactions [181, 183]. The term $\mathbf{F}_\alpha^N$ contains all the derivatives of the potential energy functions V_α^r with respect to the position of molecule α; the superscript N indicates that if the V^0 and V^1 potentials separately satisfy Newton's Third Law, then so does the force $\mathbf{F}_\alpha^N$ [181]. The second term, the drift force $\mathbf{F}_\alpha^{dr}$, is the contribution coming from differentiating the resolution function. We note here that, in general, the two models have different virial pressure for the same state point, which in turn determines a non-uniform density in the system. This effect can be compensated by adding a conservative external field, acting only in the hybrid region, i.e. in the region where $0 < \lambda < 1$. This methodology, aimed at preserving the appropriate thermodynamic conditions in each subdomain, has been applied in both force-based [182] and energy-based [181, 183] approaches.

A further source of thermodynamic imbalance is given by the last force of Eq. 3.3, dubbed drift force. This term is specific to the energy-based schemes and represents the main difference with respect to force-based approaches to adaptive resolution simulations. Therefore, it is on this quantity that we now focus our attention. Specifically, we define two different models,

named $\mathcal{E}$ and $\mathcal{F}$, with the following equations of motions:

$$\text{model } \mathcal{E}: \; m_\alpha \ddot{\mathbf{x}}_\alpha = \mathbf{F}_\alpha^N + \mathbf{F}_\alpha^{dr} \tag{3.6}$$

$$\text{model } \mathcal{F}: \; m_\alpha \ddot{\mathbf{x}}_\alpha = \mathbf{F}_\alpha^N \tag{3.7}$$

In the first case the equations of motion are integrated including all the contributions obtained differentiating the mixed Hamiltonian H_{mix}; in the second case the drift force is removed, e.g. as it is done in [178], and the resulting model is a representative of the class of force-based approaches to dual resolution simulations. Our goal is now to understand what is the difference between these two models, and specifically under what conditions one can be considered approximately close or equivalent to the other.

As a first step we remove the average of the drift force from the total force of Eq. 3.6 acting on molecule α. This can be done, as demonstrated in refs. [181, 183], by adding to the mixed Hamiltonian an external field $f(\lambda)$ depending only on the resolution of the molecule:

$$\widehat{H}_{mix} = H_{mix} - \sum_{\alpha=1}^{N} f(\lambda_\alpha) \tag{3.8}$$

$$\widehat{\mathbf{F}}_\alpha^{dr} = \mathbf{F}_\alpha^{dr} + f'(\lambda_\alpha)\nabla\lambda_\alpha \tag{3.9}$$

Employing for $f(\lambda)$ the Helmholtz free energy per particle as computed in a Kirkwood thermodynamic integration [230], the drift force is on average removed, i.e. $\langle \widehat{\mathbf{F}}_\alpha^{dr} \rangle = 0$. This Free Energy Compensation (FEC) method demonstrates [181, 183] that, in the H-AdResS framework, a seamless coupling between two arbitrary models of the same system is obtained by removing the difference in chemical potential. This result not only provides a deeper understanding of the fundamental physics of these approaches, but also enables us to operate with them in a computationally efficient manner.

The force $\widehat{\mathbf{F}}_\alpha^{dr}$ averages to zero by construction; therefore, the conservative force field originating from the Hamiltonian in Eq. 3.8 is equal to that of model $\mathcal{F}$, with the exception of a fluctuating term. The absence of the latter from model $\mathcal{F}$ causes its force-field to be non-conservative. It has been empirically observed [181, 223, 228] that if a symplectic, i.e. energy and phase space volume conserving integrator is employed to evolve the equations of motion of a force-based dual resolution scheme, the kinetic energy of the system steadily increases. To remove this "excess heat" produced by molecules crossing the hybrid region, a local thermostat, such as Langevin's, is required. Provided that the coupling is sufficiently strong, this procedure maintains the system in a steady state and avoids energy drift [176, 182].

In order to better understand the role played by the fluctuations of the drift force we performed a microcanonical H-AdResS simulation. We chose as a test system a liquid of 1596 molecules in a box of dimensions $73.69\sigma \times 15\sigma \times 15\sigma$. The molecules are composed of four atoms connected by anharmonic bonds. In the low-resolution region each molecule is described as a single sphere. The interactions between atoms of different molecules is provided by a

purely repulsive Weeks Chandler Andersen (WCA) potential [274]:

$$V^1 = \begin{cases} 4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} + \frac{1}{4}\right] & \text{if } r \leq 2^{1/6}\sigma \\ 0 & \text{if } r > 2^{1/6}\sigma \end{cases} \tag{3.10}$$

where $r = |\boldsymbol{r}_{\alpha,i} - \boldsymbol{r}_{\beta,j}|$. The coarse-grained molecules interact *via* a hard-sphere potential:

$$V^0 = 4\epsilon\left(\frac{1.7\sigma}{r - 0.05\sigma}\right)^{12} \tag{3.11}$$

Given these potential energy functions, we define the characteristic time scale of the system as $\tau = \sigma\sqrt{m/\epsilon}$, where the mass is $m = 1$. In the following, the quantities dimensioned as time will be expressed in units of τ.

Finally, the intra-molecular interaction is given by a quartic bond potential:

$$V^{\text{int}}(r_{\text{int}}) = \frac{1}{4}k(r_{\text{int}}^2 - b^2)^2 \tag{3.12}$$

with $r_{\text{int}} = |\boldsymbol{r}_{\alpha,i} - \boldsymbol{r}_{\alpha,j}|^2$ being the distance between atoms within the same molecule α. Furthermore, we chose $b = \sigma$ and $k = 7500\epsilon/\sigma^2$. In the simulations, the molecules change their resolution along the X direction with the center of the atomistic, high-resolution region being fixed at the center of the box. The total size of the atomistic region is $d_{\text{at}} = 20\sigma$ and the thickness of the hybrid region is $d_{\text{hy}} = 10\sigma$; the rest of the system is treated with the coarse-grained model. To assign to a molecule its position-dependent resolution λ, its distance from the boundary between the atomistic and the hybrid region is computed, i.e. $|\text{R}_x| - d_{\text{at}}/2$, where R_x denotes the X coordinate of the molecule's center of mass in a coordinate system with origin at the center of the simulation box. This quantity is then inserted into the resolution function $\lambda(x)$, which is given as:

$$\lambda(x) = \begin{cases} 1 & : x \leq 0 \\ 1 - \frac{30}{d_{\text{hy}}^5}\left(\frac{1}{5}x^5 - \frac{d_{\text{hy}}}{2}x^4 + \frac{d_{\text{hy}}^2}{3}x^3\right) & : 0 < x < d_{\text{hy}} \\ 0 & : x \geq d_{\text{hy}} \end{cases} \tag{3.13}$$

This system is the same as in Ref. [181], except for the box size along the X direction, which has been doubled; the simulations have been performed with the ESPResSo++ [1] package. To maintain the same density in both regions and remove, on average, the drift force, a FEC has been applied. The equations of motion for a molecule in this setup read:

$$m_\alpha \ddot{\mathbf{x}}_\alpha = \mathbf{F}_\alpha^N + \widehat{\mathbf{F}}_\alpha^{dr} \tag{3.14}$$

The second step is to compute the autocorrelation function of the average-subtracted drift force $\widehat{\mathbf{F}}^{dr}$. This quantity, reported in Fig. 3.1 (top panel, thicker line), provides interesting information about the statistical properties of the drift force. Specifically, we observe that its

normalized autocorrelation function decays with a finite characteristic time, which we estimate $\simeq 0.15\tau$ by a simple exponential fit. This autocorrelation is obtained averaging over various particles in different parts of the hybrid region. To obtain a more detailed description of the dynamics of the system we compute the autocorrelation function of the drift force for a specific value of λ. To this end, we performed 10 simulations of our system employing a mixed-resolution Hamiltonian in which the resolution λ is the same for all molecules:

$$H_\lambda = \mathcal{K} + \lambda U^1 + (1-\lambda)U^0 \tag{3.15}$$

where $U^r = \sum_\alpha V^r_\alpha$ and $\lambda = 0.1n + 0.05,\ n = 0, 1, \cdots 9$. From these simulations we computed, for a given molecule, $\Delta V(t) = V^1(t) - V^0(t) - \langle V^1 - V^0 \rangle$ as a function of time at different λ values; the autocorrelation of these quantities is then measured and related to the force autocorrelation by:

$$\langle \widehat{\mathbf{F}}^{dr}(t)\widehat{\mathbf{F}}^{dr}(0)\rangle = (\nabla\lambda)^2\langle \Delta V(t)\Delta V(0)\rangle \tag{3.16}$$

where we made use of (see also Eq. 3.5 and Eq. 3.9):

$$\widehat{\mathbf{F}}^{dr}(t) = -\nabla\lambda\left(V^1(t) - V^0(t) - \langle V^1 - V^0\rangle\right) \tag{3.17}$$

Also the autocorrelations so measured are reported in Fig. 3.1 (top panel, thinner lines): their decay times increase with increasing value of the resolution. The origin of this behavior lies in the longer time required to molecules interacting *via* the atomistic potential V^1 to rearrange their internal structure; this, in turn, determines slower changes in the drift force acting on a molecule with high resolution. All the autocorrelations, though, decay to negligible values within 1 reduced time unit.

It is instructive to measure the change of resolution that a molecule in the hybrid region undergoes by diffusing. In Fig. 3.1 (bottom panel) we show the average resolution change $\Delta\lambda = \langle|\lambda(x - \delta x/2) - \lambda(x + \delta x/2)|\rangle$ of a molecule which moves from the position x by a distance $\delta x = \sqrt{2D_\lambda\tau_\lambda}$; the latter quantity makes use of the diffusion constant D_λ and the drift force autocorrelation decay time τ_λ as measured in the aforementioned simulations performed at uniform resolution. What we observe is that in the time it takes to the autocorrelation to decay to negligible values, the resolution of the molecules changes at most by 4%, a change which occurs mainly in the center of the hybrid region where the gradient of λ is particularly steep.

The results in Fig. 3.1 imply that the drift force decorrelates before a molecule substantially changes its resolution, i.e. the λ parameter, by diffusing. This observation is not expected *a priori* to hold for every system; however, the necessary conditions to obtain this behavior can always be achieved by appropriately choosing the size of the hybrid region as a function of the diffusivity. In general, a system-specific choice of the setup parameters enables the successful simulation of more complicated systems than our test model [186, 187, 209, 210].

We now want to employ this information about the drift force fluctuations to eliminate them without disrupting thermal equilibrium. Specifically, we observe that the drift force $\widehat{\mathbf{F}}^{dr}$ is a

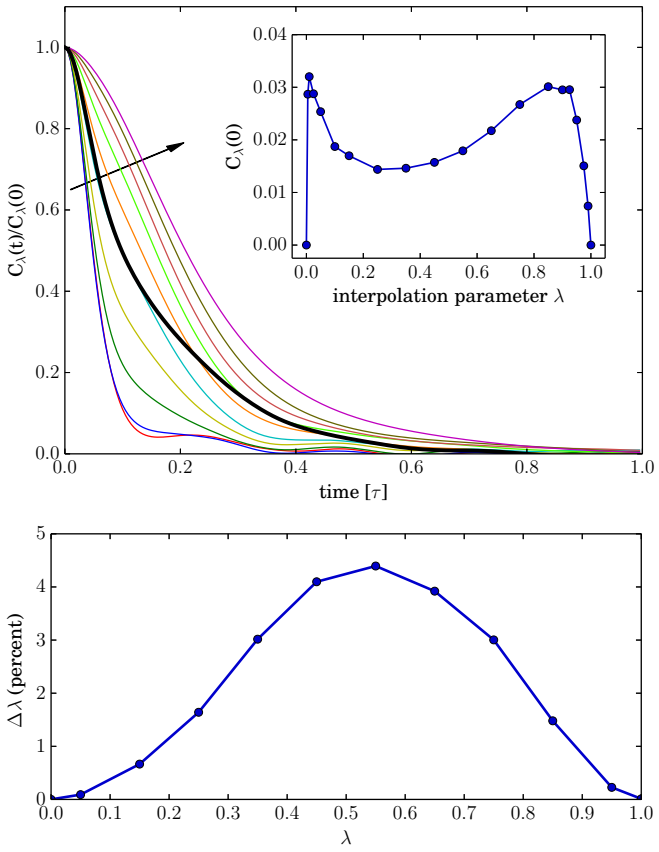

Figure 3.1.: Top: Autocorrelation of the drift force computed for different λ's. The arrow on the plot indicates the direction of increasing λ. The thicker black line is the autocorrelation of the drift force of a particle in the hybrid region. Inset: autocorrelation of the drift force at $t = 0$, as a function of λ. Bottom: average change of resolution due to a molecule's displacement during the characteristic decay time of the corresponding drift force autocorrelation.

zero-average fluctuating force with a finite autocorrelation time: we can therefore apply to the system a colored noise with memory $\mathbf{k}_\alpha^\lambda$ such that it instantaneously and exactly cancels the drift force, i.e. $\mathbf{k}_\alpha^\lambda = -\widehat{\mathbf{F}}_\alpha^{dr}$. This model has the following equations of motion:

$$m_\alpha \ddot{\mathbf{x}}_\alpha = \mathbf{F}_\alpha^N + \widehat{\mathbf{F}}_\alpha^{dr} + \mathbf{k}_\alpha^\lambda = \mathbf{F}_\alpha^N \tag{3.18}$$

In Eq. 3.18 the colored noise is explicitly dependent on the instantaneous resolution of the molecule, as indicated by the superscript λ. The resulting system, rightmost term in Eq. 3.18, is a force-based model $\mathcal{F}$. This interpretation manifestly explains the previously mentioned fact that a force-based dual-resolution simulation scheme undergoes a systematic kinetic energy increase when integrated without a local thermostat [181, 223, 228]: the absence of the drift force can be equated with the introduction of an autocorrelated random noise in a Hamiltonian,

energy-conserving scheme.

To enforce thermal and mechanical equilibrium, the autocorrelated random noise must be balanced by a history-dependent friction term according to the fluctuation-dissipation theorem. The application of the colored noise and the friction with memory amounts to thermostat a model $\mathcal{E}$ by means of a Generalized Langevin Equation (GLE) [275–280]:

$$\begin{aligned} m_\alpha \ddot{\mathbf{x}}_\alpha &= \mathbf{F}_\alpha^N + \widehat{\mathbf{F}}_\alpha^{dr} \\ &\quad - \int_{-\infty}^{t} \mathbf{K}^\lambda(t-t')\mathbf{v}_\alpha(t')dt' + \mathbf{k}_\alpha^\lambda \\ \mathbf{K}^\lambda(t-t') &= \beta\langle \widehat{\mathbf{F}}^{dr}(t)\widehat{\mathbf{F}}^{dr}(t')\rangle_\lambda \end{aligned} \tag{3.19}$$

where the subscript in the average indicates that only those molecules whose resolution is λ are considered. Our definition of the friction kernel relies on the assumption that the molecule's resolution can be considered approximately constant throughout the integration time when $\mathbf{K}^\lambda \neq 0$; such assumption is justified by the previous observation that the force autocorrelation time decays to negligible values before the molecule can substantially change its resolution (see Fig. 3.1). Assuming, as previously done, that the noise cancels the drift force exactly, the equations of motion reduce to:

$$m_\alpha \ddot{\mathbf{x}}_\alpha = \mathbf{F}_\alpha^N - \int_{-\infty}^{t} \mathbf{K}^\lambda(t-t')\mathbf{v}_\alpha(t')dt' \tag{3.20}$$

Eq. 3.20 describes a Hamiltonian scheme thermostatted with a GLE and, at the same time, a force-based model $\mathcal{F}$ with a history-dependent friction term. We thus conclude that the removal of the drift force from the Hamiltonian force-field introduces a non-Markovian behavior whether the system is in thermal equilibrium, as in Eq. 3.20, or not, Eq. 3.18. Incidentally, we observe a relation between the history-dependent friction term and previously developed adaptive resolution simulation methods [178]: in this case, a conceptually similar *ad hoc* term was introduced together with a local thermostat to enforce thermodynamical stability.

To numerically validate Eq. 3.20 we performed a simulation in which the GLE friction kernel $\mathbf{K}^\lambda(t-t')$ has been approximated in terms of the drift force autocorrelation functions $C_\lambda(t) = \langle \widehat{\mathbf{F}}^{dr}(t)\widehat{\mathbf{F}}^{dr}(0)\rangle$ reported in Fig. 3.1. The autocorrelations have been fitted with simple decaying exponentials; the decay time and the initial value $C_\lambda(0)$ for intermediate values of λ have been obtained by interpolating the measured quantities with cubic splines. The numerical implementation of the GLE has been performed following the extended variable method as described in ref. [278]. Here we made use of the property of the system that the force autocorrelation functions decay before the molecules significantly change their position and therefore their resolution λ. In the integration, i.e. throughout the time during which the exponentially decaying kernel is nonzero, λ can in fact be considered constant.

The results of this simulation are reported in Fig. 3.2: specifically, in panel a) the energy of the system is shown as a function of the simulated time for the case of model $\mathcal{F}$ with GLE friction; for comparison, the same quantity has been measured in absence of the friction. In this

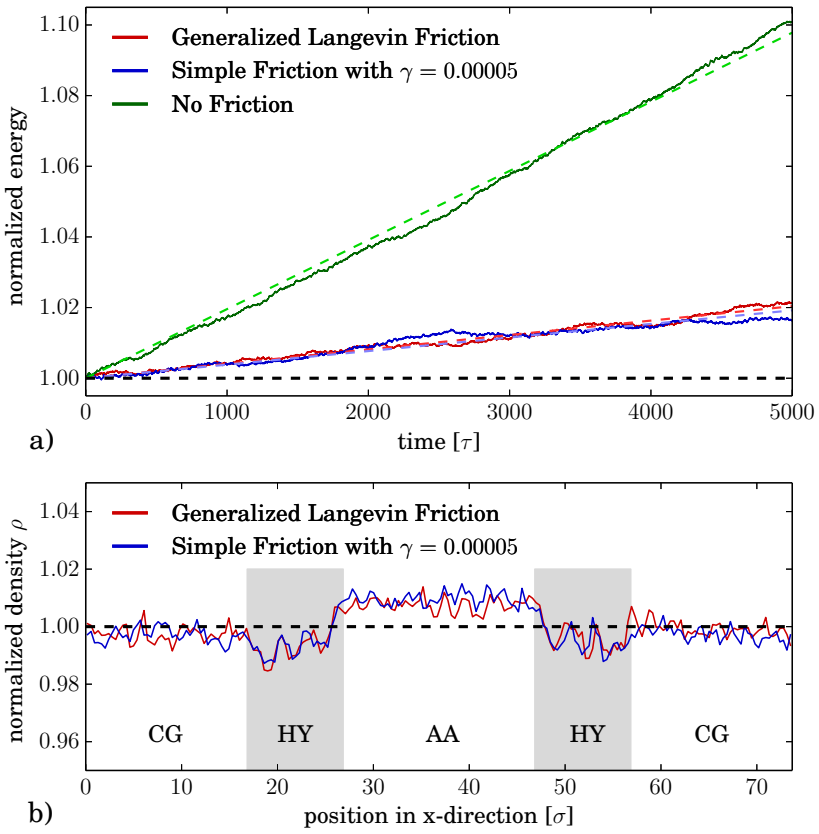

Figure 3.2.: Panel a: energy of the model $\mathcal{F}$ normalized to its initial value as a function of time, in absence of any friction, with GLE friction, and with simple friction proportional to the instantaneous velocity. Panel b: density profiles of the simulations with GLE friction and with simple friction. The average density has been computed over the last 4/5 of the simulations (the noisy profile being a consequence of little statistics). Note that the system is effectively out of equilibrium as the energy is not constant.

second case the energy increases steadily with a rate of $\dot{E} = 1.86 \cdot 10^{-4} \epsilon/\tau$ per particle, which in the first case drops to $\dot{E} = 3.88 \cdot 10^{-5} \epsilon/\tau$ per particle; the non perfect conservation of the energy can be attributed to the approximation that has been employed to fit and model the memory kernel. It is nonetheless worth noting that the energy increase at the end of this fairly long simulation is of the order of 2% for the model with GLE friction, which was obtained by approximating the kernel with no free parameters.

For comparison we introduced in model $\mathcal{F}$ a standard friction term $-\gamma v$ acting only in the hybrid region. We then fine-tuned its value so to obtain an energy increase rate as close as possible to the one measured in the simulation with the GLE. This procedure led us to a friction $\gamma = 5 \cdot 10^{-5}\ m/\tau$, corresponding to a rate $\dot{E} = 3.63 \cdot 10^{-5} \epsilon/\tau$ per particle, not far from what was observed in the GLE friction case. The difference between the two methods lies in the fact that the GLE friction approach is based on a fundamental understanding of the physics of the

system which is intrinsically non-Markovian; the simple friction approach, on the other hand, is just a one-parameter fit obtained by trial and error. It must be observed, though, that by all practical criteria the two lead to identical results: comparing the density profiles of the liquid in the two cases, Fig. 3.2b, we observe no relevant difference.

In summary, we have demonstrated that the energy-based H-AdResS scheme, thermostatted by means of a GLE, is formally equivalent to the force-based approach obtained from it by removal of the drift force and complemented with a history-dependent friction. A numerical verification of the formalism has been carried out within the boundaries of an approximate fit for the memory kernel, which largely reduces the energy drift. Because of the short decay time of the friction kernel, and the weakness of a standard friction term providing equivalent results, the memory effects introduced by eliminating the drift force from a Hamiltonian scheme would be strongly suppressed by a conventional Langevin thermostat. This would be especially the case when the low-resolution model is parametrized to match the thermodynamic properties of the high-resolution model. This result justifies the use of the thermostatted force-based scheme obtained from the H-AdResS method, and possibly also of other, similar approaches.

The interpretation of the drift force as a colored noise enables us to establish a formal, bottom-up connection between energy-based and force-based adaptive resolution simulation approaches. This unified framework provides novel insight in the theoretical foundation of a simulation paradigm whose main actors, force-based and energy-based approaches, are traditionally presented as alternative though complementary strategies. A deeper comprehension of the relation between them can thus provide a solid basis for better understood and more efficient computer simulation strategies.

3.2. Acknowledgments

K. Kreis is recipient of a fellowship funded through the Excellence Initiative (DFG/GSC 266). The authors are indebted with R. Cortes Huerto and R. Everaers for a careful reading of the manuscript.

Chapter 4.

The Role of the Relative Entropy in Adaptive Resolution Simulations

This chapter has been published as a research paper in The Journal of Chemical Physics. It is reprinted here with permission from the publisher.

Karsten Kreis and Raffaello Potestio
The relative entropy is fundamental to adaptive resolution simulations
J. Chem. Phys. 145, 044104 (2016)
DOI: 10.1063/1.4959169

Abstract

Adaptive resolution techniques are powerful methods for the efficient simulation of soft matter systems in which they simultaneously employ atomistic and coarse-grained (CG) force fields. In such simulations, two regions with different resolutions are coupled with each other via a hybrid transition region, and particles change their description on the fly when crossing this boundary. Here we show that the relative entropy, which provides a fundamental basis for many approaches in systematic coarse-graining, is also an effective instrument for the understanding of adaptive resolution simulation methodologies. We demonstrate that the use of coarse-grained potentials which minimize the relative entropy with respect to the atomistic system can help achieve a smoother transition between the different regions within the adaptive setup. Furthermore, we derive a quantitative relation between the width of the hybrid region and the seamlessness of the coupling. Our results do not only shed light on the what and how of adaptive resolution techniques but will also help setting up such simulations in an optimal manner.

4.1. Introduction

Coarse-grained (CG) models are an effective tool for simulations of physical phenomena that occur on time and length scales that would be too demanding to study using high-resolution atomistic descriptions [97–101]. For example, CG methods have been applied to study a variety of complex systems including proteins [99, 112–119], membranes [15, 120–123], polymeric materials [124–135] and liquids [103, 104, 136–141]. The schemes for obtaining CG models are typically divided into two classes, bottom-up and top-down approaches [101]. While in top-down coarse-graining the CG model is usually fitted to reproduce macroscopic target properties or experimental data, bottom-up coarse-graining employs higher-resolution, e.g. atomistic reference data to derive the CG model's potential parameters in a systematic way. Well-known systematic coarse-graining strategies include, for example, iterative Boltzmann inversion [106] (IBI) and inverse Monte Carlo (IMC) [105, 281], which preserve the reference pair correlation functions, and force matching (FM), which aims at reproducing the multi-body potential of mean force (PMF) [102–104]. A fundamental basis for many approaches in systematic coarse-graining is provided by the concept of relative entropy, S_{rel}. It turns out that both IBI and IMC can be interpreted as relative entropy minimization procedures, and also the FM method is closely related to the relative entropy approach to coarse-graining [101, 108, 109, 282–284].

Sometimes, however, CG models alone are not sufficient. In many soft matter problems short and long time scale and length scale phenomena are inherently coupled. Therefore, adaptive resolution approaches have been developed that use low-resolution CG and high-resolution atomistic models concurrently within the same simulation setup [176, 178–181, 285]. In the adaptive resolution simulation scheme (AdResS) [144, 176, 181], a small but relevant part of the system is typically modeled with an atomistic description while another, usually significantly larger part is treated at a computationally more efficient CG level. Particles cross the boundary between the two domains without experiencing any barrier and adapt their resolution on the fly. In this approach, computational resources can be concentrated on the high-resolution region, while nevertheless an overall large simulation box can be used to minimize finite size effects. The atomistic subsystem then behaves as if embedded in a completely high-resolution environment [155, 176, 181, 209–216, 286]. This technique has been used, for example, to study proteins and DNA while treating solvent molecules far away from the solute molecule on a CG level [155, 188, 189]. In a similar fashion, AdResS has been applied to simulate solvated fullerenes [185, 187], and to run efficient semi-grand canonical simulations [174, 175, 211, 228]. It can also be employed to couple path integral based, quantum models with classical models [177, 212, 216, 240], or systems like liquid water with an ideal gas of non-interacting particles [286].

In general, adaptive resolution techniques fall into two categories. In force-based (FB) approaches, different force fields are interpolated directly on the level of forces [144, 176, 178, 182], while energy-based (EB) techniques interpolate potential energies [179–181, 183, 184]. Both methodologies have advantages and disadvantages. The FB AdResS technique preserves Newton's third law, but it does not allow a Hamiltonian description of the system [219]. Hence,

it is not energy conserving and leads to an energy drift; therefore, in practice, it requires thermostatting [218–223]. On the other hand, the EB Hamiltonian adaptive resolution scheme (H-AdResS) [181, 183] is, as its name suggests, based on a Hamiltonian formulation and therefore allows energy-conserving and Monte Carlo simulations. However, this comes at the cost of an additional force term, called drift force, in the hybrid coupling region, which does not satisfy Newton's third law and must be corrected. It has been shown that the two techniques are closely related, since the FB method can be derived from a fundamental Hamiltonian framework [266], which we will review in detail in section 4.2.1.

In this work, we introduce the concept of relative entropy into this unifying formalism, and show that S_{rel} is a powerful tool not only in the theory of coarse-graining, but also to understand adaptive resolution simulation methods. We demonstrate that the behavior of particles in the hybrid region can be described using a relative entropy based approach. In particular, the lower S_{rel} between the models used in the high- and low-resolution regions, the smoother the coupling between the two. In fact, the energy drift in the FB AdResS version can be reduced by minimizing S_{rel}, while in H-AdResS the amplitude of the fluctuations of the undesired drift force can similarly be decreased. Furthermore, we derive a measure that directly relates the energy drift in FB AdResS and the strength of the drift force in H-AdResS to the width of the hybrid interpolation region. The formalism is validated by adaptive resolution simulations of liquid water coupled to various CG models, ranging from IBI- and FM-based descriptions to an ideal gas of non-interacting particles. The results provide insight into adaptive resolution simulation methodologies as well as practical guidance for setting up such simulations in an optimal way.

The paper is structured as follows: In section 4.2.1, we recap the adaptive resolution simulation approach and set the basis for sections 4.2.2, 4.2.3 and 4.2.4, in which we derive a relative entropy based framework for such methods. The details of the simulations used to validate this formalism are presented in section 4.3 and the results are discussed in part 4.4. We conclude in section 4.5.

4.2. Methodology

4.2.1. Adaptive resolution simulations

In this section we review the FB (AdResS) and the EB (H-AdResS) adaptive resolution simulation methodologies and the formalism to bridge the gap between them [266].

In H-AdResS [181, 183, 184], the global Hamiltonian of the system is defined as

$$H = \mathcal{K} + V^{\text{int}} + \sum_{\alpha} \left\{ \lambda_\alpha V_\alpha^{\text{AT}} + (1-\lambda_\alpha) V_\alpha^{\text{CG}} \right\}, \tag{4.1}$$

where

$$\mathcal{K} = \sum_{\alpha} \sum_{i \in \alpha} \frac{\mathbf{p}_{\alpha,i}^2}{2m_{\alpha,i}} \tag{4.2}$$

denotes the kinetic energy and $m_{\alpha,i}$ and $\mathbf{p}_{\alpha,i}$ are, respectively, the mass and the momentum of atom i of molecule α. The sums run over all molecules α and, for each of them, over all atoms i within the molecule. V^{int} represents all intramolecular interactions, such as bond and angle potentials, which are not subject to the interpolation. The single-molecule potentials V_α^{AT} and V_α^{CG} are the sums of all non-bonded intermolecular interaction potentials corresponding to the AT and the CG model acting on molecule α (here two-body interactions are considered for simplicity, but many-body potentials are also possible):

$$V_\alpha^{\text{AT}} = \frac{1}{2}\sum_{\beta\neq\alpha}\sum_{\substack{i\in\alpha\\ j\in\beta}} V^{\text{AT}}(|\mathbf{r}_{\alpha,i}-\mathbf{r}_{\beta,j}|), \tag{4.3}$$

$$V_\alpha^{\text{CG}} = \frac{1}{2}\sum_{\beta\neq\alpha} V^{\text{CG}}(|\mathbf{R}_\alpha-\mathbf{R}_\beta|), \tag{4.4}$$

where β runs over all molecules except α, and i and j correspond to the atoms within the molecules α and β, respectively. $\mathbf{r}_{\alpha,i}$ denotes the position of atom i of molecule α, while $\mathbf{R}_\alpha$ is the center of mass of molecule α. The parameter $\lambda_\alpha = \lambda(\mathbf{R}_\alpha)$ is a position dependent resolution function that takes values between 0 and 1. This resolution parameter is associated with each molecule α and is evaluated at the center of mass of the molecules. It defines the high- and the low-resolution regions of the system: where it is 1—the atomistic (AT) region—the particles interact only according to the potential V_α^{AT}, and where it is 0—the coarse-grained (CG) region—according to V_α^{CG}. In between these two domains λ switches smoothly from 1 to 0. In this transition region, the different force fields are interpolated and the description of the system smoothly changes. This area is called the hybrid (HY) region. In general, the two potentials can be any short-ranged interactions, although usually V^{AT} defines an AT force field, while V^{CG} is a CG model typically evaluated at the molecular center of mass, as is the case in Eq. 4.4.

The λ-function is shown in Fig. 4.1 for a schematic setup in which an AT water model is coupled to a 1-site per molecule CG model. In the presented case, the resolution changes

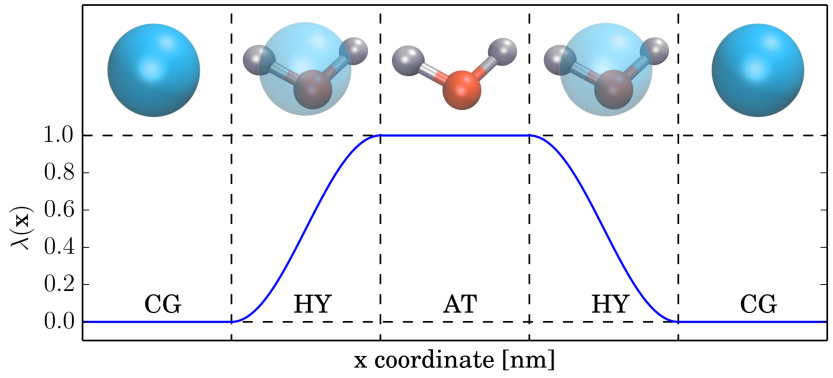

Figure 4.1.: Schematic of the resolution function λ interpolating between a 3-site atomistic and a 1-site coarse-grained water model.

along the x-direction and the AT region is of cuboid form. In general, other geometries for the high-resolution region are possible: for example, it can also take a spherical or cylindrical shape.

The force acting on a molecule α coming from differentiating the non-bonded part of the potential energy in the Hamiltonian in Eq. 4.1 is

$$\mathbf{F}_\alpha = \mathbf{F}_\alpha^N + \mathbf{F}_\alpha^{dr} \tag{4.5}$$

with

$$\mathbf{F}_\alpha^N = \sum_\beta \left\{ \frac{\lambda_\alpha + \lambda_\beta}{2} \mathbf{F}_{\alpha|\beta}^{\mathrm{AT}} + \left(1 - \frac{\lambda_\alpha + \lambda_\beta}{2}\right) \mathbf{F}_{\alpha|\beta}^{\mathrm{CG}} \right\}, \tag{4.6}$$

$$\mathbf{F}_\alpha^{dr} = -[V_\alpha^{\mathrm{AT}} - V_\alpha^{\mathrm{CG}}] \nabla \lambda_\alpha, \tag{4.7}$$

where $\mathbf{F}_\alpha^N$ incorporates all forces resulting from directly differentiating V^{AT} and V^{CG} in the Hamiltonian. The force $\mathbf{F}_{\alpha|\beta}^{\mathrm{AT}}$ belongs to the potential V^{AT} and $\mathbf{F}_{\alpha|\beta}^{\mathrm{CG}}$ to V^{CG}. The term $\mathbf{F}_\alpha^N$ satisfies Newton's third law, provided that the corresponding potentials $V^{\mathrm{AT,CG}}$ do. Furthermore, $\mathbf{F}_{\alpha|\beta}^{\mathrm{AT}}$ is the sum over all atomistic forces between the atoms of the molecules α and β. The term $\mathbf{F}_\alpha^{dr}$ is obtained by applying the derivative on the position-dependent resolution function. This contribution to the forces, dubbed drift force, acts only within the HY region. It does not satisfy Newton's third law [181] and artificially pushes particles from one region to the other, thus inducing a density imbalance in the system. Therefore, it is an undesired quantity which needs to be corrected.

We note in passing that the drift force is not the only force which can lead to an inhomogeneous density in the system. In fact, the different models typically have different equations of state, and thus pressures, for a given state point (T, V). To cancel the pressure gradients the system thus equilibrates to a state where different subdomains feature different densities. Appropriate methods have been devised that effectively correct for this effect [181, 182]. However, their detailed description lies outside of the scope of the present discussion, and, in the following, we assume that such a strategy has been applied to enforce a uniform density across the system.

Instead of interpolating the force fields on the level of potential energies, as in Eq. 4.1, we can also directly interpolate the forces corresponding to the different interaction potentials. In this case, we have only forces according to Eq. 4.6 without the undesired drift term. However, it can be shown that no Hamiltonian exists corresponding to the forces $\mathbf{F}_\alpha^N$ only [219]. In such force interpolation approaches excess heat is produced in the HY region, which leads to an energy drift, and a local thermostat needs to be employed to enforce thermal equilibrium [218–223]. Nevertheless, this force interpolation scheme has been widely and successfully employed [155, 174, 175, 177, 185, 187–189, 211, 212, 216, 228, 240, 286].

The crucial difference between the energy-based potential interpolation method, the EB approach, and the force-based technique, the FB approach, is the existence of the drift term,

Eq. 4.7. Hence, omitting the intramolecular contribution to the forces, we can write for the equations of motion of the molecules[1]

$$\text{EB approach} \ : \ m_\alpha \ddot{\mathbf{R}}_\alpha = \mathbf{F}_\alpha^N + \mathbf{F}_\alpha^{dr}, \tag{4.8}$$

$$\text{FB approach} \ : \ m_\alpha \ddot{\mathbf{R}}_\alpha = \mathbf{F}_\alpha^N. \tag{4.9}$$

Since the drift force in the EB approach leads to an inhomogeneous density, a free energy correction (FEC) is usually applied to counteract it. One approach is to use the approximation that the average behavior of molecules within the HY region at a particular distance, R_α, from the center of the AT region and with resolution $\lambda(R_\alpha)$ can be described by particles within a setup in which the whole system has uniform (hybrid) resolution $\lambda \equiv \lambda(R_\alpha)$, i.e.,

$$\langle \Delta V \rangle_{R_\alpha} \approx \langle \Delta V \rangle_{\lambda \equiv \lambda(R_\alpha)}, \tag{4.10}$$

where $\Delta V = V^{\mathrm{AT}} - V^{\mathrm{CG}}$ and V^{AT} and V^{CG} denote potential energies of single molecules in the system, omitting the molecule index α to lighten the notation. Furthermore, $\langle \cdots \rangle_\lambda$ is the ensemble average in the uniform resolution system described by the Hamiltonian

$$H_\lambda = \mathcal{K} + V^{\mathrm{int}} + \lambda \sum_\alpha V_\alpha^{\mathrm{AT}} + (1-\lambda) \sum_\alpha V_\alpha^{\mathrm{CG}}. \tag{4.11}$$

Note that, depending on the geometry of the high resolution region, the distance R_α is computed from a single point (spherical AT region), a plane (cuboid geometry as in Fig. 4.1), or a line (cylindrical AT region).

Using the approach above, the FEC is given by the Helmholtz free energy difference per particle $f(\lambda)$ between a complete system with fixed resolution λ and the CG system ($\lambda = 0$). It can be derived via thermodynamic integration [181, 183, 230] as

$$f(\lambda) = \int_0^\lambda d\lambda' \langle \Delta V \rangle_{\lambda'}, \tag{4.12}$$

$$\frac{\partial f(\lambda)}{\partial \lambda} = \langle \Delta V \rangle_\lambda. \tag{4.13}$$

The FEC is then implemented as a static field within the HY region. Using the Helmholtz free energy difference per particle $f(\lambda)$, the FEC is, with the assumption made in Eq. 4.10, for each distinct value of $\lambda(R_\alpha)$ as strong as the corresponding average drift force, thereby counteracting

[1] In practice, the way the force fields are interpolated in the force-based model can vary: for example, instead of using the average of the resolution values λ of two molecules, as in Eq. 4.1, one can use the product [144, 176, 182]. These different schemes behave similar though. Therefore, we use the FB approach here as a representative of the class of force-based schemes in general.

it [181, 183]. The Hamiltonian and the drift force including the FEC can be expressed as

$$\widehat{H} = H - \sum_{\alpha=1}^{N} f(\lambda_\alpha), \tag{4.14}$$

$$\widehat{\mathbf{F}}_\alpha^{dr} = \mathbf{F}_\alpha^{dr} + f'(\lambda_\alpha)\nabla\lambda_\alpha. \tag{4.15}$$

The approximation above, Eq. 4.10, holds for sufficiently large hybrid regions, where the values of the resolution λ of neighboring molecules are sufficiently similar to each other. This is the basis of the FEC protocol and has been shown to work well for realistic systems [181, 183], while in more complex cases an on-the-fly strategy can be employed to parametrize the FEC accurately [184]. However, as already pointed out, it only removes the average of the drift force for a given value of λ within the hybrid region. The remaining average-corrected drift force then behaves as a noise term in the hybrid region.

It is essentially this "noise" that makes the EB approach energy-conserving, in contrast to the FB approach that lacks this term [266]. Writing the corrected drift force for an individual molecule as

$$\widehat{\mathbf{F}}^{dr}(t) = -\nabla\lambda\left(V^{\mathrm{AT}}(t) - V^{\mathrm{CG}}(t) - \langle V^{\mathrm{AT}} - V^{\mathrm{CG}}\rangle\right), \tag{4.16}$$

we can calculate its autocorrelation function for different positions in the HY region, i.e. for different λ:

$$\langle\widehat{\mathbf{F}}^{dr}(t)\widehat{\mathbf{F}}^{dr}(0)\rangle_\lambda = \langle\widehat{\mathbf{F}}^{dr}(0)\widehat{\mathbf{F}}^{dr}(0)\rangle_\lambda\tilde{C}_\lambda(t), \tag{4.17}$$

$$\langle\widehat{\mathbf{F}}^{dr}(0)\widehat{\mathbf{F}}^{dr}(0)\rangle_\lambda = \nabla\lambda\nabla\lambda\left[\langle\Delta V^2\rangle_\lambda - \langle\Delta V\rangle_\lambda^2\right], \tag{4.18}$$

where $\tilde{C}_\lambda(t)$ denotes the normalized autocorrelation function and we again used the approximation in Eq. 4.10. Because of the noise-like behavior of the corrected drift force we can now define a friction kernel with memory according to the fluctuation-dissipation theorem as

$$\begin{aligned}\mathbf{K}^\lambda(t) &= \beta\langle\widehat{\mathbf{F}}^{dr}(t)\widehat{\mathbf{F}}^{dr}(0)\rangle_\lambda \\ &= \beta\nabla\lambda\nabla\lambda\left[\langle\Delta V^2\rangle_\lambda - \langle\Delta V\rangle_\lambda^2\right]\tilde{C}_\lambda(t)\end{aligned} \tag{4.19}$$

with $\beta = 1/k_BT$. It was shown that it is precisely this friction that is required to exactly cancel the energy drift in the corresponding FB simulation model [266]. Hence, the amplitude of $\mathbf{K}^\lambda(t)$, that is

$$\beta\langle\widehat{\mathbf{F}}^{dr}(0)\widehat{\mathbf{F}}^{dr}(0)\rangle_\lambda = \beta\nabla\lambda\nabla\lambda\left[\langle\Delta V^2\rangle_\lambda - \langle\Delta V\rangle_\lambda^2\right], \tag{4.20}$$

provides a measure for the energy drift resulting in FB adaptive resolution simulations. At the same time, by definition it also quantifies the magnitude of the fluctuations of the undesired drift force in the corresponding EB approach. Therefore, this expression establishes a quantitative link between the two methodologies.

4.2.2. A relative entropy based framework

The relative entropy S_{rel} has been shown to be a powerful tool in the theory of systematic coarse-graining. It provides a fundamental basis for different techniques, such as IBI and inverse Monte Carlo, and it is also closely related to the FM method [101, 108, 109, 283, 284]. Inspired by its success in systematic coarse-graining, we now introduce S_{rel} in the framework of adaptive resolution simulations.

We consider the relative entropy per molecule between a target (T) system with mixed but uniform resolution λ, described by the Hamiltonian H_λ in Eq. 4.11, and, as reference (R), the fully CG system with H_0:

$$S_{\text{rel}}(\lambda) = \frac{1}{N}\sum_{\mathcal{I}} \mathcal{P}_T(\mathcal{I}) \ln\left(\frac{\mathcal{P}_T(\mathcal{I})}{\mathcal{P}_R(\mathcal{I})}\right), \tag{4.21}$$

where $\mathcal{P}_T$ is the target system's equilibrium probability distribution, $\mathcal{P}_R$ is the one of the reference system, the sum runs over the microstates $\mathcal{I}$, and N is the number of molecules in the system. Since we work in the canonical ensemble, we can replace $\mathcal{P}_T$ and $\mathcal{P}_R$ with the corresponding Boltzmann weights, i.e.

$$\mathcal{P}_T = \frac{e^{-\beta H_\lambda}}{\mathcal{Z}_\lambda},\ \mathcal{P}_R = \frac{e^{-\beta H_0}}{\mathcal{Z}_0}, \tag{4.22}$$

where $\mathcal{Z}_\lambda$ and $\mathcal{Z}_0$ are the systems' partition functions.

We want to point out that both the CG and the AT system have the same degrees of freedom. In the adaptive resolution approach, Eq. 4.1, no degrees of freedom are truly removed, but only an intermolecular coarse-grained interaction is introduced in the CG region that substitutes the atomistic non-bonded interaction and acts at the molecular center of mass. As pointed out earlier, the intramolecular interactions are not subject to any interpolation procedure. Hence, the accessible microstates $\mathcal{I}$ in Eq. 4.21 are the same for both the AT and the CG system and for all values of resolution λ. Furthermore, for $\lambda = 1$ Eq. 4.21 corresponds to the expression for the relative entropy between the fully atomistic and the fully coarse-grained system, which was used in the seminal work by Shell, who has shown the important role of S_{rel} in systematic coarse-graining [108].

The Helmholtz free energy per particle of the system at resolution λ can be expressed either as a thermodynamic integration, Eqs. 4.12 and 4.13, or in terms of the partition function as

$$f(\lambda) = -\frac{1}{N\beta}\ln(\mathcal{Z}_\lambda). \tag{4.23}$$

Furthermore, we have

$$H_\lambda - H_0 = \lambda\left(\sum_\alpha V_\alpha^{\text{AT}} - \sum_\alpha V_\alpha^{\text{CG}}\right). \tag{4.24}$$

We now insert the probability distributions $\mathcal{P}_T$ and $\mathcal{P}_R$, Eq. 4.22, into Eq. 4.21 and make use of

Eqs. 4.13, 4.23 and 4.24 to transform the result into

$$\begin{aligned} S_{\text{rel}}(\lambda) &= \beta f(\lambda) - \frac{\beta}{N}\langle H_\lambda - H_0\rangle_\lambda \\ &= \beta\left(f(\lambda) - \lambda\frac{\partial f}{\partial \lambda}\right), \end{aligned} \tag{4.25}$$

where we have also assumed that the zero of the system's potential energy has been shifted so that $f(0) = 0$. Taking the gradient of S_{rel} with respect to the resolution we obtain

$$\begin{aligned} \frac{\partial S_{\text{rel}}}{\partial \lambda} &= -\beta\lambda\frac{\partial^2 f}{\partial \lambda^2} \\ &= \beta^2\lambda\left[\langle \Delta V^2\rangle_\lambda - \langle \Delta V\rangle_\lambda^2\right] \geq 0. \end{aligned} \tag{4.26}$$

Finally, making use of Eqs. 4.18 and 4.26 we have

$$\beta\langle \widehat{\mathbf{F}}^{dr}(0)\widehat{\mathbf{F}}^{dr}(0)\rangle_\lambda = \nabla\lambda\nabla\lambda\frac{1}{\beta\lambda}\frac{\partial S_{\text{rel}}}{\partial \lambda}\bigg|_\lambda. \tag{4.27}$$

Eq. 4.27 relates the amplitude of the drift force fluctuations, Eq. 4.20, to the gradient of the relative entropy. The connection to the friction kernel is then given by

$$\mathbf{K}^\lambda(t) = \nabla\lambda\nabla\lambda\frac{1}{\beta\lambda}\frac{\partial S_{\text{rel}}}{\partial \lambda}\tilde{C}_\lambda(t). \tag{4.28}$$

This result is particularly interesting because it reveals, within the boundaries of our approximations, under which conditions the FB approach can be considered equivalent to the EB approach. In fact, by definition we have that $S_{\text{rel}}(0) = 0$, and from Eq. 4.26 we learn that $\partial S_{\text{rel}}/\partial\lambda \geq 0$, i.e. the relative entropy cannot decrease as a function of the resolution parameter λ. This suggests that by minimizing the relative entropy between the models used in the different regions of the adaptive resolution setup, the amplitude $\beta\langle \widehat{\mathbf{F}}^{dr}(0)\widehat{\mathbf{F}}^{dr}(0)\rangle_\lambda$ is also minimized. This would reduce the energy drift in FB adaptive simulations, and, as a consequence, it would allow one to decrease the strength of the thermostat required to absorb the excess heat produced. Alternatively, by definition, it would decrease the undesired drift force fluctuations in the EB approach. Furthermore, the lower the relative entropy between the AT and the CG interactions, the more similar are the FB and the EB approaches to each other.

This finding indicates that an appropriate choice for the CG potential could be one obtained via a direct relative entropy minimization procedure as introduced by Chaimovich and coworkers [108, 282–284, 287]. This would then also minimize the amplitude of $\mathbf{K}^\lambda$ with it. Similarly, being equivalent to the aforementioned approach, the IBI or the IMC method could be used [106].

4.2.3. The width of the hybrid region

The previous result, Eqs. 4.27 and 4.28, is dependent on the precise form of the resolution function λ. However, in most applications λ is defined in a similar fashion, trying to establish a smooth coupling between the AT and CG regions. The most common choice is a squared cosine function, defined as

$$\lambda(R_\alpha) = \cos^2\left(\frac{(R_\alpha - r_{\text{at}})\pi}{2d_{\text{hy}}}\right) \tag{4.29}$$

for molecules within the HY region, that is $r_{\text{at}} < R_\alpha < r_{\text{at}} + d_{\text{hy}}$, where r_{at} denotes the radius of the atomistic region, this is the distance from the center of the AT region to the boundary of the HY region measured along the direction of resolution change. The quantity d_{hy} is the width of the HY region and R_α is the distance between the molecule α and the center of the atomistic region. For molecules within the CG region the resolution is 0, while for molecules in the AT region it is 1. The derivative of λ is then

$$\lambda'(R_\alpha) = -\frac{\pi}{d_{\text{hy}}}\cos\left(\frac{(R_\alpha - r_{\text{at}})\pi}{2d_{\text{hy}}}\right)\sin\left(\frac{(R_\alpha - r_{\text{at}})\pi}{2d_{\text{hy}}}\right) \tag{4.30}$$

and it is strictly 0 outside of the HY region. Since the drift force is only present in the HY region (EB approach) and the excess energy is also produced only there (FB approach), it is the only part of the system which is relevant for our discussion. In the HY region λ is a bijective function. Hence, for $r_{\text{at}} < R_\alpha < r_{\text{at}} + d_{\text{hy}}$ we can express λ' as a function of the resolution λ itself:

$$\lambda' = -\frac{\pi}{d_{\text{HY}}}\sqrt{\lambda}\sqrt{1-\lambda}. \tag{4.31}$$

Then, we can write Eq. 4.27 as

$$\beta\langle\widehat{\mathbf{F}}^{dr}(0)\widehat{\mathbf{F}}^{dr}(0)\rangle_\lambda = \frac{\pi^2}{d_{\text{hy}}^2\beta}(1-\lambda)\frac{\partial S_{rel}(\lambda)}{\partial\lambda}. \tag{4.32}$$

This expression shows that the excess heat produced by an individual particle in the hybrid region of a FB adaptive resolution simulation depends on the width of the hybrid region, d_{hy}, as $1/d_{\text{hy}}^2$. This means, for example, that if one wants to reduce the strength of the thermostat used to absorb this heat by a factor of 2, it will be sufficient to increase the width of the HY region by $\sqrt{2}$. Similarly, when doubling the size of the hybrid region, the excess energy produced by an individual particle decreases by 3/4. It is worth pointing out that the above equation, Eq. 4.32, describes the behavior of particles in the HY region independent of the overall geometry chosen in the adaptive setup. It is valid regardless of whether the AT region has a cuboid, spherical or any other shape.

4.2.4. Integrating the hybrid region

As Eq. 4.32 describes the behavior of single molecules in the HY region, we expect to obtain a more precise measure for the overall energy drift in the FB approach when integrating over the whole HY subdomain. To perform this integration we have to assume a certain geometrical arrangement of the different regions. As an example, we choose a cuboid setup, like in Fig. 4.1, in which the resolution changes along the x-direction. Note that in alternative adaptive resolution simulation setups the high-resolution region can have another shape, for example a spherical or cylindrical one. Nevertheless, the HY region locally resembles the form of a cuboid also in such a case, provided that the HY region is sufficiently thin compared to the AT one. This is often the case, e.g. when the HY region is chosen as thin as possible while the AT region needs to be large enough to fit an object of interest inside, such as a macromolecule [155, 185, 187–189].

Integrating Eq. 4.32 over the whole hybrid region in a slab setup as in Fig. 4.1 yields

$$\begin{aligned} \mathcal{K}_{\text{slab}} &= \frac{2\pi^2 \rho_0 L_y L_z}{d_{\text{hy}}^2 \beta} \int_0^{d_{\text{hy}}} dx \, (1 - \lambda(x)) \frac{\partial S_{\text{rel}}(\lambda(x))}{\partial \lambda} \\ &= \frac{2\pi \rho_0 L_y L_z}{d_{\text{hy}} \beta} \int_0^1 d\lambda \sqrt{\frac{1-\lambda}{\lambda}} \frac{\partial S_{rel}(\lambda)}{\partial \lambda}, \end{aligned} \tag{4.33}$$

where L_y and L_z are the box lengths in y- and z-direction and ρ_0 is the system's equilibrium density. The factor 2 is due to the fact that we have two symmetric HY regions in the setup on both sides of the central AT region, see Fig. 4.1. Furthermore, we assumed here that the sum of particles in the HY region can be modeled as a continuum distribution weighted by the density. The second line in Eq. 4.33 has been obtained by substituting the integration in space by an integration in resolution λ using Eqs. 4.30 and 4.31.

The quantity $\mathcal{K}_{\text{slab}}$ can be interpreted as a precise measure for the total energy drift obtained in FB AdResS simulations. In other words, the energy drift can be expected to be proportional to $\mathcal{K}_{\text{slab}}$. We can write it in a more compact form as

$$\mathcal{K}_{\text{slab}} = \frac{1}{\beta} \frac{\pi}{d_{\text{hy}}} \mathcal{G}_{\text{slab}} \, \mathcal{M}_{\text{slab}} \tag{4.34}$$

with

$$\mathcal{G}_{\text{slab}} = 2\rho_0 L_y L_z, \tag{4.35}$$

$$\mathcal{M}_{\text{slab}} = \int_0^1 d\lambda \sqrt{\frac{1-\lambda}{\lambda}} \frac{\partial S_{\text{rel}}(\lambda)}{\partial \lambda} \tag{4.36}$$

and see that $\mathcal{K}_{\text{slab}}$ is composed of three parts: The prefactor $\mathcal{G}_{\text{slab}}$ comes from the overall geometric shape of the HY region and is specific for the setup we have chosen to derive the equation; $\mathcal{M}_{\text{slab}}$ is the contribution of the different potentials in the AT and CG regions. Minimizing the latter with a careful choice of the CG potential, a minimal energy-drift for a

given width of the HY region can be obtained in the FB approach. Finally, the term $1/d_{\mathrm{hy}}$ is the effect of the width of the HY region. While on the particle level the excess heat produced is dependent on the width of the HY region with $1/d_{\mathrm{hy}}^2$, on the macroscopic level a larger width of the HY region also increases the number of particles in it linearly, hence the overall $1/d_{\mathrm{hy}}$ behavior. Therefore, if one is interested in the total energy drift in a FB simulation to guide the size of the hybrid region, $1/d_{\mathrm{hy}}$ needs to be considered. On the other hand, when the thermostat that needs to be applied to cancel the drift is analyzed, $1/d_{\mathrm{hy}}^2$ is the quantity to look at, as typical thermostats act on the single-particle level.

At the same time, $\mathcal{K}_{\mathrm{slab}}$ can by definition also be interpreted as a measure for the total, integrated noise in the hybrid region produced by the drift force in an EB adaptive simulation.

When normalizing $\mathcal{K}_{\mathrm{slab}}$ with respect to the total number of particles in the hybrid region,

$$N_{\mathrm{slab}}^{\mathrm{hy}} = 2\rho_0 L_y L_z d_{\mathrm{hy}}, \tag{4.37}$$

we obtain

$$\frac{\mathcal{K}_{\mathrm{slab}}}{N_{\mathrm{slab}}^{\mathrm{hy}}} = \frac{1}{\beta}\frac{\pi}{d_{\mathrm{hy}}^2}\int_0^1 d\lambda \sqrt{\frac{1-\lambda}{\lambda}}\frac{\partial S_{\mathrm{rel}}(\lambda)}{\partial \lambda} \tag{4.38}$$

$$= \frac{1}{\beta}\frac{\pi}{d_{\mathrm{hy}}^2}\mathcal{M}_{\mathrm{slab}}. \tag{4.39}$$

This quantity represents the average energy drift per particle and is, by construction, independent of $\mathcal{G}_{\mathrm{slab}}$; additionally, it is also again proportional to $1/d_{\mathrm{hy}}^2$, as Eq. 4.32.

As briefly explained above, we expect systems with other geometries for the regions in the adaptive setup to behave qualitatively similar to the cuboid scenario discussed here, as long as the HY region is sufficiently thin compared to the high-resolution one. Otherwise, considering for example the case with a spherical geometry, the number of particles in the HY region would increase superlinearly with d_{hy}. For the total energy drift this would lead to a proportionality $\sim d_{\mathrm{hy}}^{\nu}$ with $\nu > -1$, an expression decaying slower than $1/d_{\mathrm{hy}}$.

4.3. Numerical validation

To verify the validity of the equations derived above, we performed adaptive resolution simulations using liquid water as the atomistic system in the high-resolution region. We coupled it to various distinct CG potentials, used different widths of the HY region, and computed both the relative entropies and the derived measures $\mathcal{K}_{\mathrm{slab}}$ and $\mathcal{M}_{\mathrm{slab}}$.

4.3.1. Model systems

Atomistic water was described by the SPC/E model [270] using the reaction field method [217] with a dielectric constant $\epsilon = 67.5998$ [174] to treat electrostatics. All interactions were cut

off at 1.0 nm and all potential energies were shifted to be zero at the cutoff [288]. The bond constraints were enforced using the SETTLE algorithm [55], the temperature was $T = 300\,\mathrm{K}$ and the density $\rho_0 = 33.117\,\mathrm{molecules/nm^3}$.

We parametrized 7 different CG potentials, all of which are short-ranged pair-potentials and approximate water molecules as single beads positioned at the molecular center of mass. The employed CG potentials are:

- A potential obtained via iterative Boltzmann inversion (IBI) [106].
- A potential obtained via the force matching (FM) methodology [102–104].
- A potential minimizing the relative entropy between the CG system and the AT reference system directly (RE potential) using the Newton-Raphson framework [108, 282–284, 287].
- A Weeks-Chandler-Anderson (WCA) potential with $\epsilon = 1\,k_B\mathrm{T}$ and $\sigma = 0.25\,\mathrm{nm}$ [274].
- A WCA potential with $\epsilon = 1\,k_B\mathrm{T}$ and $\sigma = 0.2\,\mathrm{nm}$.
- A WCA potential with $\epsilon = 1\,k_B\mathrm{T}$ and $\sigma = 0.1\,\mathrm{nm}$.
- An ideal gas of non-interacting particles [286]. In this case, there is no pair interaction potential between the beads in the CG region.

For the IBI, FM and RE potential, the coarse-graining procedures were carried out using the VOTCA package together with GROMACS [271, 287, 289]. The cutoffs for all CG potentials were, like for the atomistic water model, 1.0 nm with the exception of the RE potential. The method implemented in VOTCA for deriving the RE potential not only shifts it to be zero at the cutoff, but it also makes the force fade out smoothly. Thereby, the actual desired behavior of the RE potential is lost within the last 0.05 nm. Thus, the potential cutoff was extended to 1.05 nm for the RE method. All potentials are shown in Fig. 4.2.

The short-range interactions in the SPC/E model were capped to avoid diverging forces when overlapping molecules enter the hybrid region in adaptive resolution simulations. Specifically, the oxygen-oxygen Lennard-Jones interaction was truncated at the force corresponding to an interparticle distance of 0.2 nm and the electrostatic reaction field interaction at the force corresponding to a distance of 0.1 nm. Truncations like this have no effect on the properties of the atomistic water itself but only prevent technical problems at the HY/CG boundary.

The chosen methods, IBI, FM, and RE, for deriving the different CG potentials cover the most important and well-established techniques for systematic coarse-graining of molecular liquids [101]. They are carefully parametrized to reproduce certain properties of the atomistic reference system. By construction, the IBI and the RE potentials yield the center of mass RDF's of the atomistic model. This can be particularly important when simulating more complex CG molecular liquids or, for example, polymer systems [126, 132, 133]. The FM potential attempts to optimally approximate the net force on the molecular center of mass from the atomistic reference. An advantage of the FM method is that one can easily include also higher order terms, such as 3-body interactions. This is useful, for example, when creating a CG model of

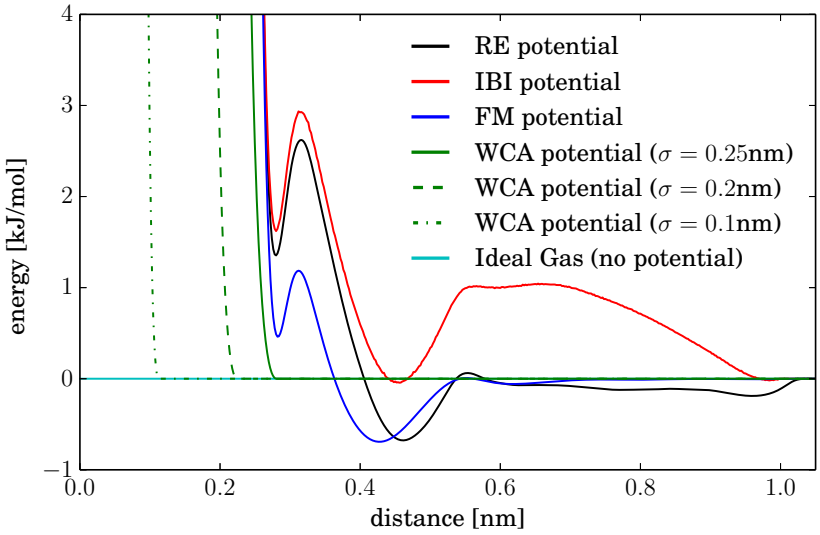

Figure 4.2.: The pair-potentials that were employed to model the water molecules in the coarse-grained region of the adaptive resolution simulations.

water which reproduces its tetrahedral ordering [290]. On the other hand, we consider an ideal gas of non-interacting particles as an extreme case of coarse-graining [286]. Coupling to an ideal gas has particular advantages: for example, it allows the free insertion of particles in the CG region, which can be useful for simulations at constant chemical potential. In between the latter approaches are the different WCA potentials, which are simple hard-sphere models and are not parametrized on any properties of the reference system. However, WCA potentials are frequently employed to model repulsive interactions [65, 274, 291–293] and because of their simplicity they are also well-suited to test our results.

4.3.2. Simulations

Calculation of RDF's. We first ran simulations of the atomistic water system and each of the different coarse-grained systems to calculate the corresponding molecular radial distribution functions (RDF's). The simulations were 1.5 ns long with a 2 fs time step and the simulation box was cubic with length 4.038 nm. We applied here, and also in all other simulations, periodic boundary conditions in each direction. The simulations were carried out in the canonical (NVT) ensemble using ESPResSo++[1] with a Langevin thermostat with friction constant $\gamma = 5.0\,\mathrm{ps}^{-1}$.

Calculation of S_{rel} and $\mathcal{M}_{\mathrm{slab}}$. To calculate the relative entropies between the SPC/E water model and the different CG potentials, we ran 1 ns long simulations using the same box dimensions and the same time step as before. For each CG potential we performed simulations in different ensembles of mixed but uniform resolution λ and measured $\langle \Delta V^2 \rangle_\lambda - \langle \Delta V \rangle_\lambda^2$. Specifically, depending on the CG potential, we sampled between 21 and 28 λ-values within the range from 0 to 1. For each λ and CG potential we ran 5 different, independent simulations

to obtain an error estimate. We then calculated the relative entropies between the atomistic water system and the CG models via Eq. 4.26 and integration. Similarly, we calculated $\mathcal{M}_{\text{slab}}$ according to Eq. 4.36.

Energy drift in force-based adaptive resolution simulations. Next, we ran adaptive resolution simulations in a box of dimensions 16.086 nm × 3.5 nm × 3.5 nm using the FB adaptive resolution methodology of Eq. 4.9. The atomistic high-resolution model was in each simulation the SPC/E water, while in the CG region we used the different CG descriptions. The resolution changed along the x-direction (as in Fig. 4.1) using $r_{\text{at}} = 1.0\,\text{nm}$ and $d_{\text{hy}} = 5.0\,\text{nm}$. To obtain flat density profiles along the direction of resolution change, the force arising from the pressure difference between the models pushing molecules from one region to the other needed to be canceled. For that purpose, we applied a correction derived from the pressure function obtained during Kirkwood thermodynamic integration [230] between the AT and the different CG systems (for further details see Potestio et al. [181]). The thermodynamic integration was performed in the same box and under the same conditions as the previous simulations for the RDF and S_{rel} calculation and carried out using the GROMACS package [187, 289].

Using the derived correction, the adaptive simulations were then performed in the microcanonical (NVE) ensemble without a thermostat. Hence, as we used the force-interpolation scheme, excess heat was produced in the hybrid region, which led to a drift in the total energy. The energy was calculated as the sum of all energies corresponding to the Newtonian forces in Eq. 4.9 plus the kinetic energy, the energy of the intramolecular interactions, and the energy corresponding to the correction force. For each setup, we performed 80 short, independent simulations of length 20 ps each, starting from different equilibrated configurations. In this case, we used a shorter time step of 0.5 fs to avoid any significant artifacts in the energy arising from the finite length of the time steps. Again we employed the ESPResSo++ [1] package for carrying out the simulations. As the energy drift drives the simulation towards a different state point, i.e. it slowly increases the temperature, we performed many short simulations instead of few long ones in order to stay close to the state point for which the compensation was derived.

Various hybrid region widths. Finally, we investigated the effect of the width of the hybrid region on the energy drift using the previously described setup with the FM potential in the CG region and with 5 different values for d_{hy} ranging from 1.0 nm to 5.0 nm. For this part of the study, we opted for the FM potential as it showed both a small energy drift and also only relatively small fluctuations in the energy compared to the other potentials (see section 4.4.2, Fig. 4.4).

4.4. Results

4.4.1. Relative entropies and RDF's

To characterize the different CG potentials, we first measured their molecular RDF's (Fig. 4.3) and calculated the relative entropies with respect to the atomistic water system (Tab. 4.1). Both the IBI and the RE potential match the atomistic reference RDF within line thickness and they

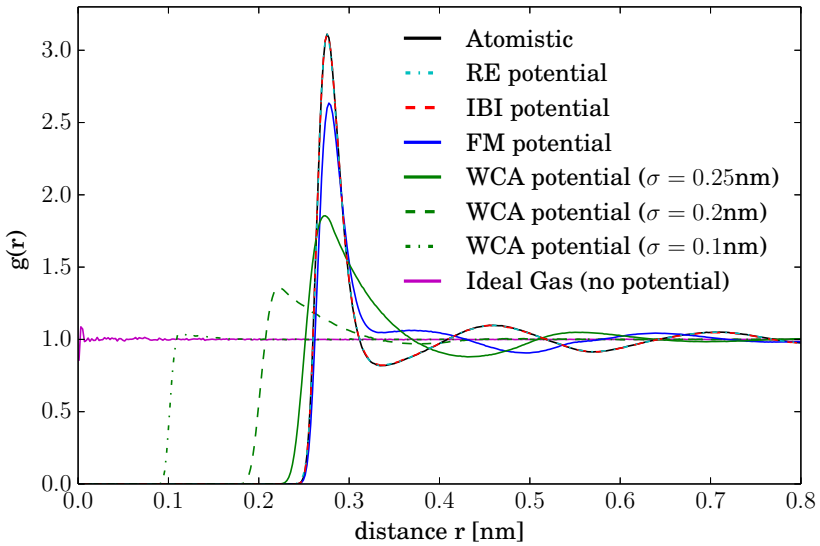

Figure 4.3.: Molecular radial distribution functions generated by the different coarse-grained pair-potentials used to model the water molecules in the CG region.

Potential	$S_{\text{rel}}(\lambda = 1)$	$\mathcal{M}_{\text{slab}}$
RE	5.76 ± 0.04	7.33 ± 0.05
IBI	5.76 ± 0.02	7.32 ± 0.02
FM	5.89 ± 0.06	7.09 ± 0.06
WCA ($\sigma = 0.25\,$nm)	6.23 ± 0.07	7.65 ± 0.04
WCA ($\sigma = 0.2\,$nm)	7.34 ± 0.06	13.72 ± 0.08
WCA ($\sigma = 0.1\,$nm)	8.51 ± 0.05	31.96 ± 0.28
Ideal Gas	8.68 ± 0.09	36.35 ± 0.46

Table 4.1.: Relative entropy per molecule $S_{\text{rel}}(\lambda = 1)$ and the measure $\mathcal{M}_{\text{slab}}$ calculated with respect to the atomistic water model for the different coarse-grained potentials. The errors correspond to the standard deviation.

also both yield the same relative entropy. This can be expected, as both methods are designed to reproduce the RDF, and, therefore, they also yield the same value for S_{rel}. Additionally, the RE technique by construction generates the potential with the lowest relative entropy with respect to the atomistic reference system [108]. Hence, unsurprisingly, the IBI and the RE potentials have the lowest relative entropy among the set of chosen CG potentials. It is interesting to point out that, although the two potentials yield the same S_{rel} and RDF's, they look quite different (see Fig. 4.2). According to Henderson's theorem, to a given RDF uniquely corresponds a specific pair-potential up to a constant [294]. In practice, however, RDF's that are not the same but extremely close can be generated by significantly different potentials [126, 295]. This is in line with our observations.

The FM method aims at approximating the multi-body PMF by projecting it in an optimal way onto a simpler potential, a pair-wise one in this case. Therefore, as we see in Fig 4.3, it still reproduces the first peak of the atomistic center of mass RDF relatively well, but shows a divergent behavior for larger distances. This discrepancy could be fixed by adding a 3-body interaction [290]. Regarding the relative entropy, however, it is still close to the minimum provided by the RE and IBI potentials.

The WCA potentials do not match the RDF from atomistic simulations. Furthermore, larger deviations from the reference RDF yield an increase in relative entropy. This confirms the intuitive expectation that the less two systems match structurally, the larger is the distance between them in terms of relative entropy. However, the WCA potential with $\sigma = 0.25\,\mathrm{nm}$, still roughly reproducing the excluded volume of the reference system, results in an only slightly higher relative entropy than the optimum. This is a similar behavior as observed for the FM potential.

Finally and expectedly, the ideal gas does not yield any structure, and also has the largest S_{rel} value.

We point out that in the case of two-body CG potentials a relative entropy value close to the optimum can be obtained by a simple potential which only reproduces the excluded volume of the reference model. The more detailed structure of the CG model, this is the RDF for larger distances, seems to play a negligible role for the value of the relative entropy. This becomes particularly clear when considering the FM potential: while it perfectly reproduces the position of the first peak of the reference RDF, and hence the excluded volume, it shows significant deviations from the reference system for larger distances. Nevertheless, its relative entropy value is close to the optimum provided by the IBI and the RE potentials.

4.4.2. Energy drift vs. S_{rel}

Next, we performed adaptive resolution simulations with the FB approach, Eq. 4.9, using in the AT region the SPC/E water and in the CG domain the different CG potentials. We did not apply any thermostat to absorb the excess heat produced in the HY region, instead we measured the total energy drifts (Fig. 4.4). Over the short run time the energy of the individual simulations fluctuates strongly. However, as we ran a large set of simulations, we were able to calculate well-defined averages and energy increase rates for the different setups. The ideal gas results in the highest energy drift, while the potentials whose RDF's best match the reference structure result in the lowest drifts.

Plotting the energy drifts against the different relative entropies of the CG potentials (Fig. 4.5), we observe a clear trend. Indeed, as we argued in section 4.2.2, the larger the relative entropy between the models, the higher the energy drift in the corresponding FB adaptive resolution simulation. This confirms that minimizing the relative entropy of the CG model relative to the AT one is a recipe for reducing the energy drift in FB adaptive simulations. As this energy drift can be interpreted as an "error" introduced by the CG model, our result is in line with the findings of Chaimovich and Shell, who have shown that the relative entropy can

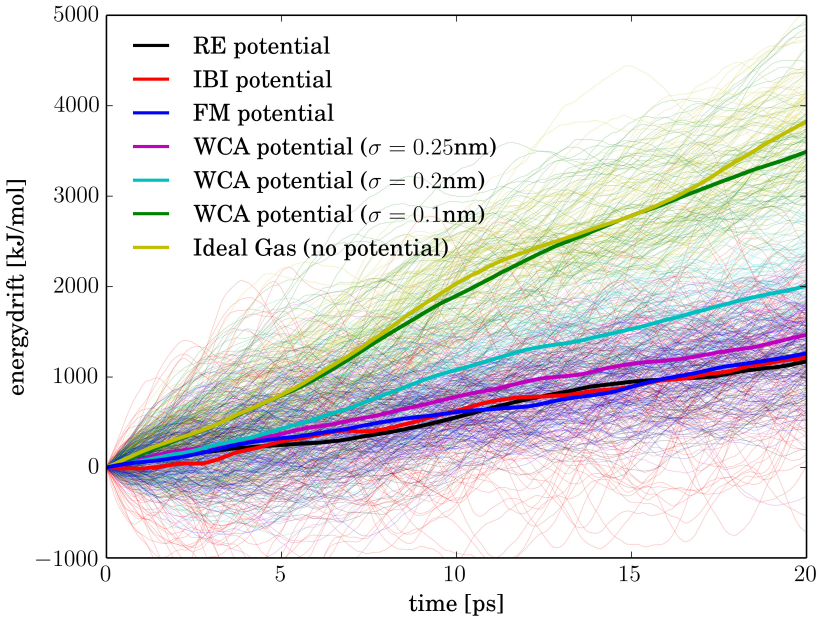

Figure 4.4.: Energy drifts in AdResS simulations with the FB approach using an atomistic water model in the AT region and 7 different CG potentials in the CG region. For each case, 80 simulations were performed (thin lines), which were used to calculate average energy increase rates (thick lines).

serve as a universal metric for multiscale errors introduced by coarse-graining [283, 284].

4.4.3. Energy drift vs. $\mathcal{M}_{\text{slab}}$

Although there seems to be a clear dependence of the energy drift on the relative entropy, a precise measure for the energy drift is only provided by $\mathcal{M}_{\text{slab}}$, Eq. 4.36. We calculated the latter quantity for all CG potentials (Tab. 4.1) and plotted the energy drifts against it (Fig. 4.6). Within the standard deviation, the data fall on a straight line. We conclude that the energy drifts are indeed proportional to $\mathcal{M}_{\text{slab}}$, which validates its derivation.

It is worth to point out that, although the relative entropy S_{rel} itself is minimized by the RE and IBI potentials, the FM potential is the one with minimal $\mathcal{M}_{\text{slab}}$. For practical applications, however, this seems to be irrelevant. The RE, IBI and FM potentials all lead to very similar energy drifts with a wide range of fluctuations within individual simulations. Even the simple WCA potential with $\sigma = 0.25\,\text{nm}$ does not yield a significantly higher energy increase rate. This suggests that, as long as the CG potential reproduces the first peak of the RDF relatively well, the energy drift in the FB approach, or the drift force fluctuations in the EB approach, will be close to the minimum. Hence, in practice it may be sufficient to aim just for roughly matching RDF's and to reproduce the excluded volume of the reference system, as already argued in section 4.4.1.

At this point, we want to make a clarification for the reader to avoid possible confusion and

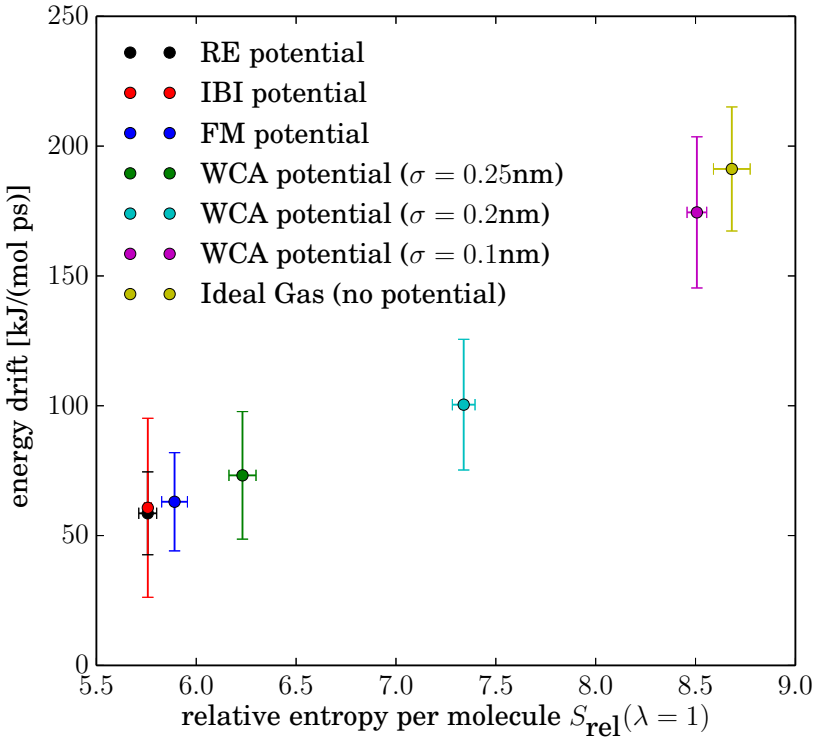

Figure 4.5.: Energy drift in force-based adaptive resolution simulations vs. relative entropy for different CG potentials and using atomistic water as reference system for the $S_{\mathrm{rel}}(\lambda = 1)$ calculation and as system in the AT domain of the adaptive simulations. The error bars correspond to the standard deviation of the measured quantities.

clearly distinguish between the RE potential, which is based on a relative entropy minimization procedure, and the concept of relative entropy as a distance measure between probability distributions in general. The RE potential may be parametrized in an overly accurate way, as also simpler potentials can yield a nearly optimal relative entropy with respect to the AT reference. However, the overall concept of relative entropy, S_{rel}, plays an important role in the understanding of adaptive resolution simulations, as shown in sections 4.2.2–4.2.4.

4.4.4. Energy drift vs. width of the hybrid region d_{hy}

As the last step to validate the results derived in sections 4.2.2–4.2.4, we investigated the energy drift in the FB approach for different widths of the hybrid region using the FM potential in the CG region. The results are presented in Fig. 4.7. The obtained drifts are proportional to $1/d_{\mathrm{hy}}$ and plotting them as a function of the measure $\mathcal{K}_{\mathrm{slab}}(d_{\mathrm{hy}})$, which depends on the HY region width with $1/d_{\mathrm{hy}}$, they fall perfectly on a straight line. This confirms that the total energy drift, given a cuboid geometry for the adaptive setup, is indeed proportional to $1/d_{\mathrm{hy}}$. As we have linearly increased the number of particles in the HY region with growing width, this result also

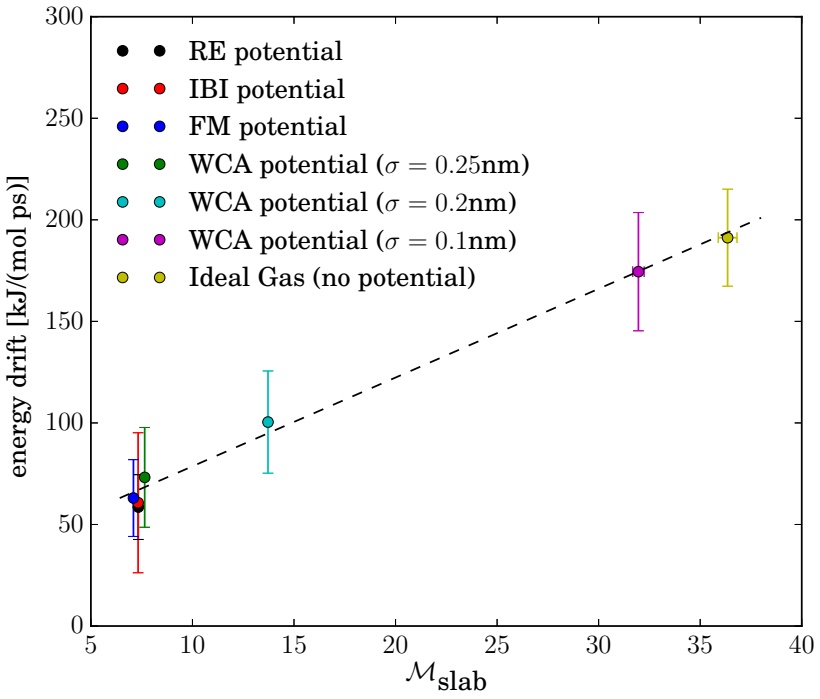

Figure 4.6.: Energy drift in force-based adaptive resolution simulations vs. $\mathcal{M}_{\text{slab}}$ for different CG potentials with atomistic water as reference system for the $S_{\text{rel}}(\lambda)$ calculation and as system in the AT domain of the adaptive simulations. The error bars correspond to the standard deviation. The dashed line is a guide to the eye.

validates that the contribution of the individual particles to the overall energy drift grows with $1/d_{\text{hy}}^2$.

4.5. Conclusions

In summary, we have provided a framework based on the concept of relative entropy that closely connects force-based and energy-based adaptive resolution simulation methods. We have shown that by minimizing S_{rel} of the potential used in the CG region with respect to the atomistic reference system the excess heat produced in the HY region of adaptive simulations with the FB approach can be minimized and, equivalently, the noise generated by the drift term in simulations with the EB approach can be reduced. This also means that the FB and the EB approaches are more similar, the smaller the relative entropy of the CG model with respect to the atomistic one is. This result can help guide the choice of the potential used in the CG region of adaptive simulations in order to achieve a smooth coupling between the two domains. Naturally, systematic coarse-graining methods such as the approaches used in this work (IBI, FM, RE) are therefore well-suited for deriving CG potentials applicable in AdResS and H-AdResS simulations. However, as we have seen, for practical purposes and in particular

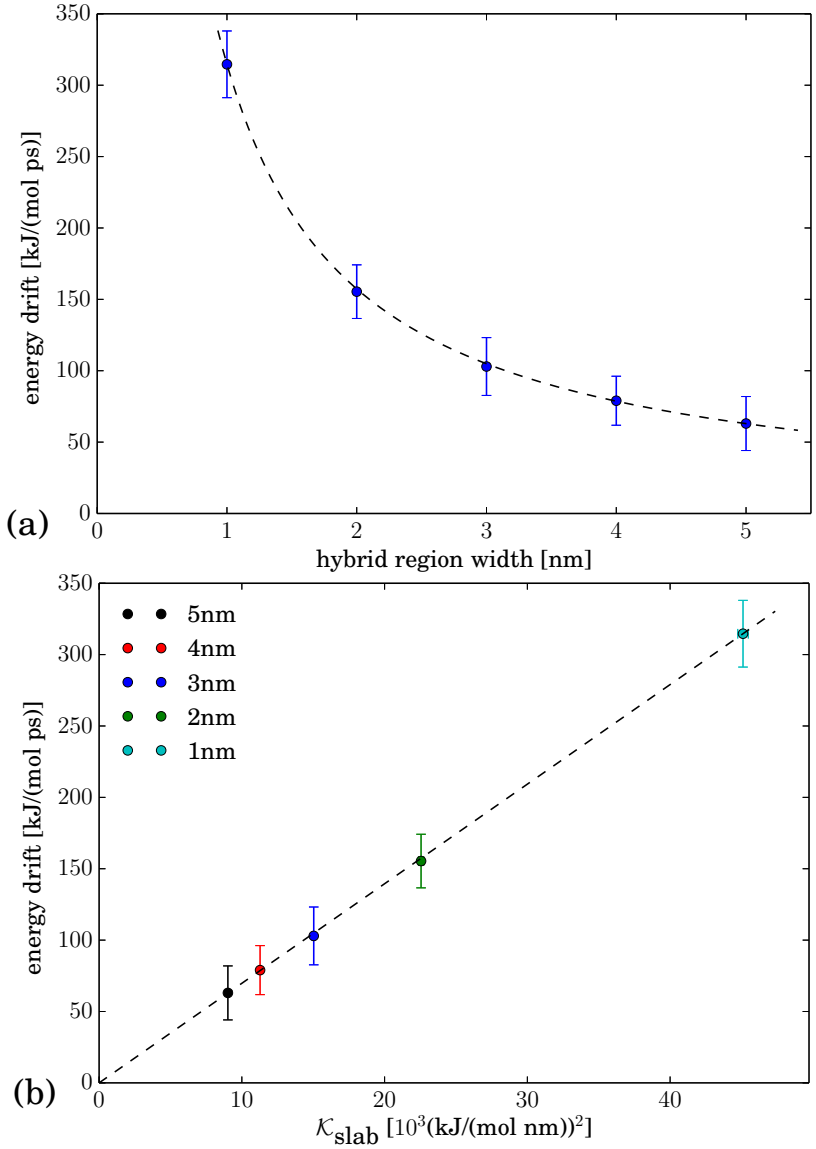

Figure 4.7.: Energy drifts in force-based adaptive simulations with atomistic water in the AT region and the FM potential in the CG region vs. width of the hybrid region d_{hy}. *(a)* Directly plotted against d_{hy}. *(b)* Plotted against the corresponding values for the measure $\mathcal{K}_{\text{slab}}(d_{\text{hy}})$. The dashed line is a fit proportional to $1/d_{\text{hy}}$ and the error bars correspond to the standard deviation.

when treating the CG region just as a particle reservoir, it may be sufficient to employ a much simpler model that only roughly reproduces the excluded volume of the reference system, like an accordingly parametrized WCA potential. In terms of the quality of the coupling in the HY region, this potential would behave in a similar way to the advanced models, while it is at the same time significantly easier to derive, and it also leads, due to its short interaction range, to a

computational advantage.

Additionally, we have introduced a measure $\mathcal{K}_{\text{slab}}$ that is exactly proportional to the excess heat produced in the HY region, FB approach, and the amplitude of the colored noise corresponding to the drift force, EB approach. This measure describes the impact of the width d_{hy} of the HY region on the performance of the adaptive simulations. We have shown that the excess energy (FB approach) and the strength of the drift force (EB approach) are, on the level of single particles in the HY region, proportional to $1/d_{\text{hy}}^2$. This information can help adjusting the thermostat required to absorb the excess heat produced in force-based adaptive simulations. Furthermore, considering the total energy drift in the force-based approach in non-thermostated simulations, we could show that this depends on the width of the HY region with $1/d_{\text{hy}}$, provided the HY region has, locally, a cuboid shape, which is often the case in many practical applications.

In adaptive resolution simulation methods, the hybrid coupling region, where the interpolation between the different force fields takes place, is the most complicated part of the simulation setup, leading to technical and practical challenges. Here we have provided a framework for a deeper understanding of the HY region and of the behavior of particles in it, both for simulations using the EB approach for interpolating the force fields as well as the FB one. We have demonstrated that the relative entropy is not only a powerful tool in the theory of coarse-graining but also for the understanding of adaptive resolution simulation methodologies.

4.6. Acknowledgments

K. Kreis is recipient of a fellowship funded through the Excellence Initiative (DFG/GSC 266). The authors thank Will Noid for fruitful discussions and Aoife Fogarty, Kurt Kremer, Joe Rudzinski and Aviel Chaimovich for a critical reading of the manuscript.

Chapter 5.

Adaptive Resolution Simulations with Self-Adjusting High-Resolution Regions

This chapter has been published as a research paper in The Journal of Chemical Theory and Computation. It is reprinted here with permission from the publisher.

Karsten Kreis, Raffaello Potestio, Kurt Kremer, and Aoife C. Fogarty
Adaptive Resolution Simulations with Self-Adjusting High-Resolution Regions
J. Chem. Theory Comput. 12, 4067-4081 (2016)
DOI: 10.1021/acs.jctc.6b00440

Furthermore, this work is featured on the cover of The Journal of Chemical Theory and Computation, Volume 12, Issue 10 (October 11, 2016).

Abstract

In adaptive resolution simulations, different regions of a simulation box are modeled with different levels of detail. Particles change their resolution on-the-fly when traveling from one subregion to the other. This method is particularly useful for studying multiscale systems in which effects on a broad range of length and time scales play a role. Until now, the geometry of the high-resolution region has been limited to simple geometries of spherical, cuboid, or cylindrical form, whose shape does not change during the simulation. However, many phenomena involve changes in size and shape of system components, for example, protein folding, polymer collapse, nucleation, and crystallization. In this work, we develop a scheme that uses a series of overlapping spheres to allow for an arbitrary division of space into domains of different levels of resolution. Furthermore, the geometry is automatically adjusted on-the-fly during the simulation according to changes in size and shape of, for example, a solvated macromolecule within the high-resolution region. The proposed approach is validated on liquid water. We then simulate the folding of an atomistically detailed polypeptide solvated in a shell of atomistic water that changes shape as the peptide conformation changes. We demonstrate that the peptide folding process is unperturbed by the use of our methodology.

5.1. Introduction

Concurrent multiresolution simulation methods involve the simultaneous use of high- and low-resolution computational models. They are used to study phenomena that are inherently multiscale in nature, meaning that a range of length scales are at play in the same system [144–146]. A typical example is that of macromolecules in solution, where computationally expensive, atomistic-resolution models are needed to capture local physical and chemical phenomena within a macromolecule and at the solute/solvent interface but where more efficient coarse-grained models may be used to capture larger-scale effects of the rest of the solvent [155, 175]. This means defining a region containing the macromolecule and its solvation shell in which interactions are described using atomistic models. To allow free diffusion of solvent particles between this atomistic region and a coarse-grained particle reservoir, one can use the Adaptive Resolution Simulation (AdResS) approach [144, 176], in which particles smoothly change their resolution during the simulation as they cross back and forth between the atomistic and coarse-grained regions via an intermediate transition or hybrid region. This is achieved by an interpolation of atomistic and coarse-grained forces within the transition region [176]. To compensate for the thermodynamic imbalance between atomistic and coarse-grained potentials, an additional compensating force is applied [182].

The adaptive resolution methodology has been widely applied for multiscale problems in soft matter. For example, it has been used to simulate atomistic proteins solvated in a sphere of atomistic water [155, 188], DNA strands in a cylinder [189], and fullerenes in spheres of atomistic solvent [185, 187]. It has also been employed for the efficient implementation of a semi-grand canonical simulation setup [174, 175, 211, 228], and in the context of path integral simulations interpolating between quantum and classical models [177, 212, 216, 224, 240]. The methodology can even couple systems as different as liquid water and an ideal gas of noninteracting particles [286].

All these works have used simple geometries for the atomistic region such as spheres, cuboid slabs, and cylinders. In practice, however, the domain within the simulation box where an atomistic description is needed could have any arbitrary shape, and could change in size and shape during the simulation. In order to study phenomena like protein folding [296] or the collapse and swelling of smart polymers like poly(N-isopropylacrylamide) [297] with maximum computational efficiency, one needs an atomistic region which matches the shape of the atomistic-resolution macromolecular solute, and which adapts its shape as the solute undergoes large-scale conformational change.

Here, we extend the AdResS methodology in order to implement such an atomistic region. Using multiple overlapping spheres to create a region of any arbitrary shape, we show how a smooth resolution change can be achieved and how the necessary compensating thermodynamic force can be dealt with. We then take the example of a polyalanine molecule in aqueous solution and construct an atomistic region which contains the peptide and its hydration shell. This region adapts to the conformational transitions of the peptide during folding (see Fig. 5.1). We show that the folding process is unperturbed in the AdResS simulations.

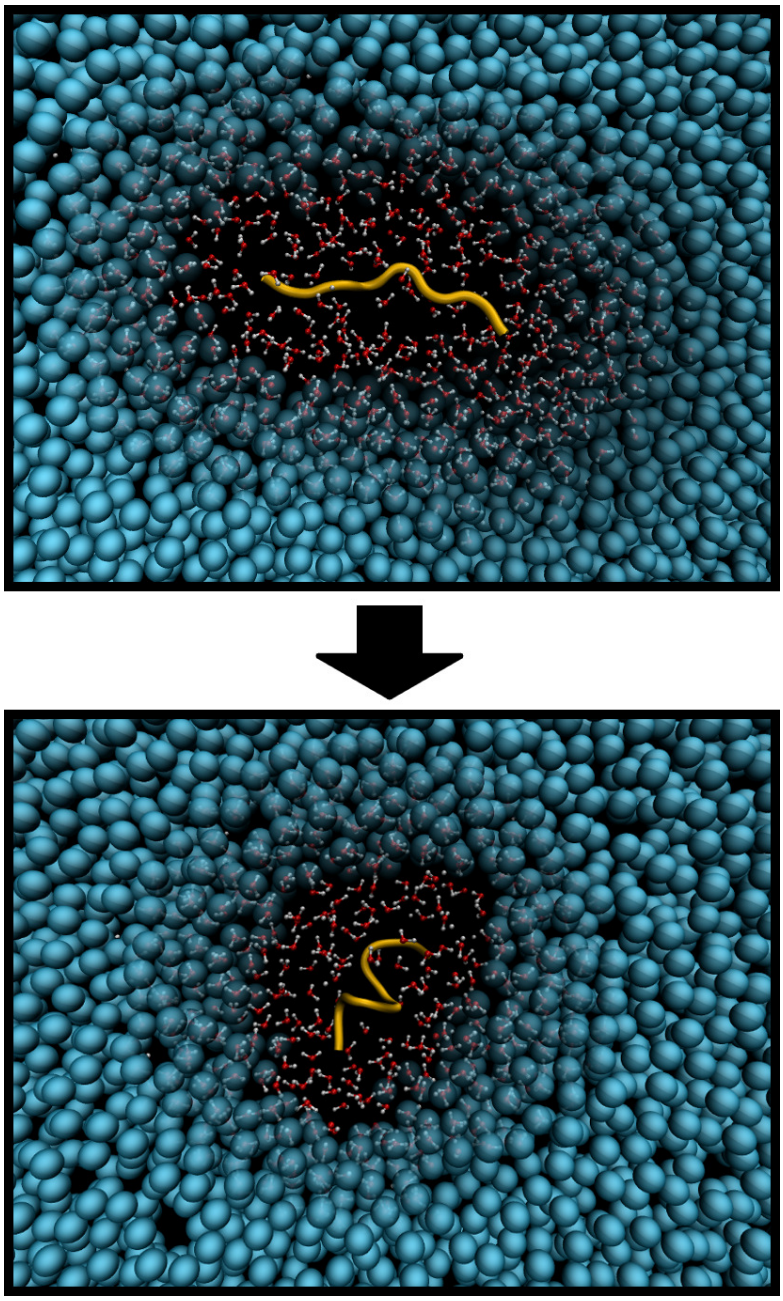

Figure 5.1.: Two snapshots from the self-adjusting peptide folding simulations. At the beginning of the simulation (top), the peptide is extended. When it folds (bottom), the water region around it which is modeled in atomistic resolution adapts accordingly.

The AdResS approach is one of several concurrent multiresolution simulation techniques [179–181, 298, 299]. In particular, Szklarczyk et al. [298] have proposed a methodology for simulating a biomolecular solute with a surface layer of atomistic solvent whose shape and particle number adapt to biomolecular conformational changes; however this involves the use of

restraining potentials, thereby perturbing particle fluctuations and diffusion. The AdResS-based scheme presented here is unique in simultaneously allowing solvent particle diffusion, a fully flexible and self-adjusting atomistic region and an automatic adaptation of the number of atomistic water molecules as the solvent exposed surface area changes.

The paper is structured as follows: In section 5.2.1 we recap the existing AdResS approach and in section 5.2.2 a scheme is derived enabling adaptive resolution simulations with an atomistic region of any given desired geometry. The simulation details are presented in section 5.3 and the results are discussed in part 5.4: The sections 5.4.1 and 5.4.2 present validation simulations of liquid water and a peptide with fixed conformation, while in section 5.4.3 we demonstrate by means of peptide folding simulations that our approach also allows for on-the-fly self-adjusting geometries. In section 5.5 we discuss implementation details and possible extensions to the methodology. We conclude in section 5.6.

5.2. Methodology

5.2.1. The AdResS scheme for a single spherical atomistic region

The AdResS method is a technique to smoothly couple two different models, typically of different resolutions, while still allowing particles to freely diffuse between the regions associated with each model, with particle resolution changing on the fly as they cross the resolution boundary. Usually a predefined atomistic region of spherical, cuboid or cylindrical shape is embedded in a significantly larger system modeled on a numerically more efficient, coarse-grained level. This allows the concentration of the computational resources on the expensive but small atomistic part of the simulation, without compromising on the required overall simulation box size. The structural and dynamic properties of the atomistic region have been shown to be equivalent to those of a much larger fully atomistic system [155, 176, 188, 189, 209–211, 213–216, 286].

In order to outline the details of the AdResS methodology, we begin with the case where the atomistic (AT) region is a sphere of radius r_{at} centered on a point $\mathbf{r}_{\mathrm{centr}}$, which can be either a fixed point in space or the instantaneous position of a designated particle. The hybrid (HY) or transition region is a spherical shell of thickness d_{hy} across which the resolution change occurs, and the coarse-grained (CG) region occupies the remainder of the simulation box. The smooth resolution change is established by associating a resolution parameter λ with each molecule α, which can be defined as

$$\lambda(\mathbf{r}_\alpha) = \begin{cases} 1 & : r_\alpha \leq r_{\mathrm{at}} \\ \cos^2\left(\frac{(r_\alpha - r_{\mathrm{at}})\pi}{2d_{\mathrm{hy}}}\right) & : r_{\mathrm{at}} < r_\alpha < r_{\mathrm{at}} + d_{\mathrm{hy}} \\ 0 & : r_\alpha \geq r_{\mathrm{at}} + d_{\mathrm{hy}} \end{cases} \tag{5.1}$$

where $r_\alpha = |\mathbf{r}_\alpha - \mathbf{r}_{centr}|$ and $\mathbf{r}_\alpha$ is the center of mass position of molecule α [176]. A value of 1 for λ corresponds to the atomistic force field, a value of 0 to the coarse-grained force field, and the squared cosine function provides a smooth transition between the two. This is illustrated in

Fig. 5.2. Note that other choices for the resolution function are possible.[266, 286]

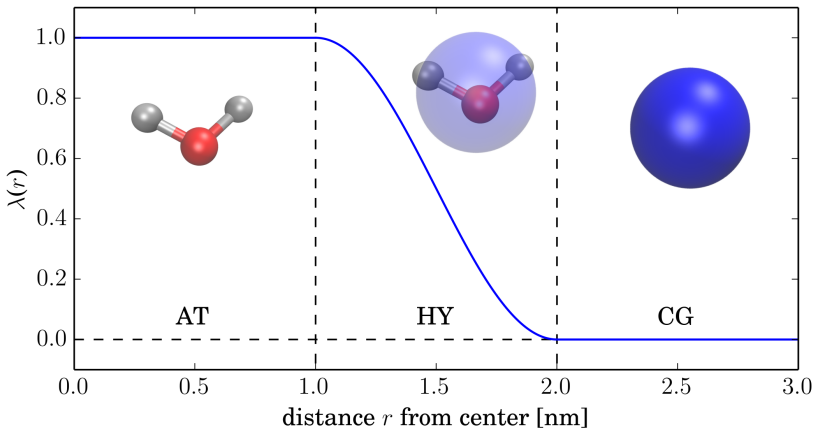

Figure 5.2.: Illustration of the resolution function $\lambda(\mathbf{r}_\alpha)$ used to smoothly interpolate between the AT and CG force fields in the AdResS methodology for the case of water described by a one-particle-per-molecule model in the CG region.

The forces of the atomistic and coarse-grained force fields are then interpolated as

$$\mathbf{F}_{\alpha\beta} = \lambda(\mathbf{r}_\alpha)\lambda(\mathbf{r}_\beta)\mathbf{F}^{AT}_{\alpha\beta} + \left(1 - \lambda(\mathbf{r}_\alpha)\lambda(\mathbf{r}_\beta)\right)\mathbf{F}^{CG}_{\alpha\beta} \tag{5.2}$$

with

$$\mathbf{F}^{AT}_{\alpha\beta} = \sum_{i\in\alpha}\sum_{j\in\beta}\mathbf{F}^{AT}_{ij}, \tag{5.3}$$

where $\mathbf{F}_{\alpha\beta}$ is the total force between two molecules α and β and $\mathbf{F}^{AT}_{\alpha\beta}$ is the contribution of the atomistic force field which is decomposed into forces $\mathbf{F}^{AT}_{ij}$ between the individual atoms i and j of the molecules α and β. Finally, $\mathbf{F}^{CG}_{\alpha\beta}$ is the coarse-grained force between the molecules, typically evaluated between their centers of mass [176]. For a pair of particles both within the AT region, the above scheme simplifies to an interaction via the AT force field, and the equivalent argument holds for a pair of particles both within the CG region. Otherwise, the force interpolation is used.

The scheme of Eq. 5.2 satisfies Newton's third law, but it is not compatible with a Hamiltonian description [219].[1] This leads to the production of excess heat in the HY region, which needs to be removed by a thermostat based on the fluctuation-dissipation theorem to enforce thermodynamic equilibrium [218–223]. It is sufficient, however, to apply the thermostat only in the CG and HY regions, thereby leaving the dynamics in the AT region unperturbed [155, 210, 216].

At a given temperature and reference density, coarse-grained potentials typically have a

[1] We note that a Hamiltonian-based version of AdResS also exists in which the energies are interpolated [181]. It is energy- but not momentum-conserving. The scheme leads to an additional term in the forces, which would in turn result in additional complications in the implementation of a self-adjusting atomistic region scheme as presented here for the force-interpolation version of AdResS.

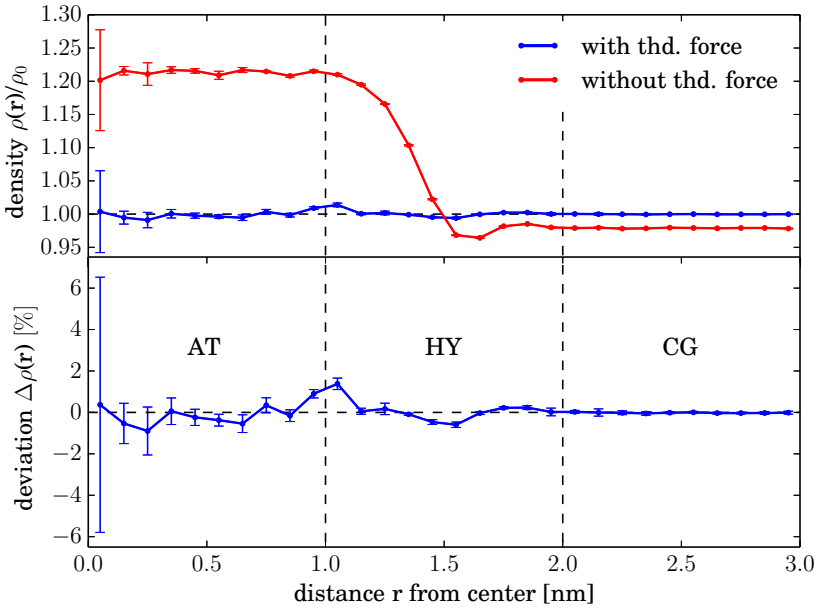

Figure 5.3.: *Top:* Normalized density profiles with and without thermodynamic force correction for an AdResS setup where the atomistic region is a simple sphere. *Bottom:* Percentage deviation from a flat reference density profile for the case with thermodynamic force shown in the top panel.

different virial pressure than the atomistic reference system [225–227]. Therefore, coupling two such different models with each other using Eq. 5.2 will result in an inhomogeneous density profile along the direction in which the resolution changes (see example in Fig. 5.3). A well established method to correct for this thermodynamic imbalance between the two subsystems is provided by a compensating force known as the thermodynamic force (TF) [182], which acts on the centers of mass of solvent molecules, and has a nonzero value only within the HY region or in its immediate vicinity. It is applied along the vector

$$\mathbf{e}_\alpha = \frac{\mathbf{r}_\alpha - \mathbf{r}_{\text{centr}}}{|\mathbf{r}_\alpha - \mathbf{r}_{\text{centr}}|}. \tag{5.4}$$

between the center of mass of a molecule α and the center of the atomistic region, i.e. the vector along which the resolution changes.

To remove the density gradient between the two regions, the TF needs to be parametrized such that its integral equates the pressure difference Δp between the two potentials:

$$\Delta p = \frac{\rho_0}{M} \int_{r_{\text{at}}^{TF}}^{r_{\text{at}}^{TF}+d_{\text{hy}}^{TF}} F^{TF}(r) dr \tag{5.5}$$

where M is the mass of a coarse-grained molecule, ρ_0 is the system's reference density and $F^{TF}(r)$ is the scalar value of the thermodynamic force $\mathbf{F}^{TF}(\mathbf{r}_\alpha) = F^{TF}(r_\alpha)\,\mathbf{e}_\alpha$ of molecule α at

distance r from the center of the atomistic region [182]. Furthermore, it has proven useful to extend the region in which the TF is applied by a short distance $d_{\text{ext}} \leqslant 0.1\,\text{nm}$ on either side of the HY region, such that the width of the region within which the TF is actually applied (referred to as the correction region) is $d_{\text{hy}}^{TF} = d_{\text{hy}} + 2\,d_{\text{ext}}$. We define $r_{\text{at}}^{TF} = r_{\text{at}} - d_{\text{ext}}$, the AT region radius reduced by d_{ext}.

Usually, the TF is derived in an iterative fashion starting from the inhomogeneous density profile obtained when no compensating force is used, as

$$F_{i+1}^{TF}(r) = F_i^{TF}(r) - \frac{M}{\rho_0^2 \kappa_T} \frac{\partial}{\partial r} \rho_i(r) \tag{5.6}$$

where κ_T is the isothermal compressibility at temperature T, $\rho_i(r)$ is the density profile at iteration i. The protocol of Eq. 5.6 converges by construction once a sufficiently flat density is reached and has been used extensively in the context of AdResS simulations [155, 174, 175, 187–189, 211, 213–216, 228, 286]. In practice, the value of the prefactor in front of the density gradient can be varied to aid convergence.

5.2.2. A method allowing for general geometries

We now outline a scheme that allows for AdResS simulations where the high-resolution region can have any arbitrary shape. Such a general geometry can be achieved by constructing the atomistic region as a combination of multiple overlapping spheres with centers $\mathbf{r}_{\text{centr},i}$. The centers of these spheres can be either fixed in space (see sections 5.4.1 and 5.4.2), or associated with particles that move in space (see section 5.4.3), enabling the high-resolution region to change its shape during the simulation as the particles defining the region move. We define all spheres here to have the same radius r_{at} and hybrid region thickness d_{hy}. The resolution parameter $\lambda(\mathbf{r}_\alpha)$ of a molecule α is then defined as in Eq. 5.1 above, but based on the distance r_{min} between the molecule's center of mass $\mathbf{r}_\alpha$ and the closest sphere center $\mathbf{r}_{centr,i}$ out of all the multiple overlapping spheres.

In this picture, we can define surfaces with the same value of r_{min} (see for example the red and blue lines in Fig. 5.4), which we call iso-resolution surfaces in the following. Particles on such a surface all have the same resolution $\lambda(\mathbf{r}_\alpha)$ and the change in resolution within the hybrid region takes place along the perpendicular direction to the iso-resolution surface on which the molecule is located. The additional force arising from the pressure difference between atomistic and coarse-grained models can be supposed to also act perpendicularly to these surfaces, i.e. along the direction of resolution change, as we will indeed later show.

The difficulty now lies in the appropriate application of the TF correction required to compensate for the density gradient arising from this pressure difference. We recall that in the case where the atomistic region is one single sphere, the TF has a value $F^{TF}(r_\alpha)$ derived using the iterative scheme in Eq. 5.6 and is applied along the vector of resolution change, i.e. the vector $\mathbf{e}_\alpha$ between the single sphere center and the molecule on which the TF is applied. Similarly, in the case of multiple overlapping spheres, both the magnitude and direction of

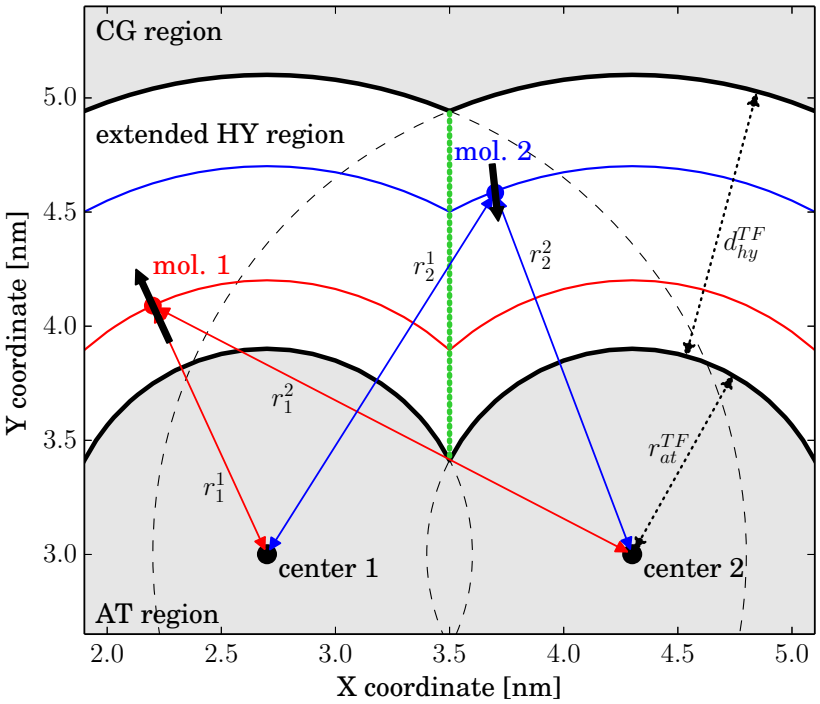

Figure 5.4.: Two-dimensional visualization of the three-dimensional thermodynamic force calculation procedure in the case of an atomistic region composed of two overlapping spheres. The center of mass positions of two molecules are shown as examples and the curved red and blue lines indicate the iso-resolution surfaces on which the molecules reside. The green dotted line shows where the closest center to a molecule changes. The black arrows at the molecular positions denote the thermodynamic force. As molecule 1 is only in the hybrid region with respect to center 1, its thermodynamic force is exactly perpendicular to its iso-resolution surface. Molecule 2 is in the hybrid regions with respect to both centers 1 and 2. Therefore, the direction along which the thermodynamic force on this molecule acts is interpolated between the two vectors toward the centers.

application of the TF correction must be chosen. The value $F^{TF}(r_{\min})$ can be determined based on the distance $r_{\min}$ to the closest sphere center, and can be derived via Eq. 5.6 in a setup with a single spherical region. Thus, Eq. 5.5 still holds along the resolution change direction. We now turn to the more complex question of the direction of the vector. As in the single sphere case, for multiple spheres one could suppose that the TF needs to be applied everywhere in a perpendicular direction to the iso-resolution surface to cancel the imbalance induced by the pressure gradient. However, this would lead to the following two complications: (i) the perpendicular direction to the iso-resolution surface changes abruptly along the line where the closest center changes (green dotted line in Fig. 5.4), leading to undesired abrupt changes in the correction forces, and (ii) on this line, the perpendicular direction is not defined. Therefore, we apply the TF along a direction defined as the linear combination of all normalized unit vectors

$\mathbf{e}_\alpha^i$ from each sphere center $r_{\text{centr},i}$ toward the position of the molecule α, $\mathbf{r}_\alpha$:

$$\mathbf{v}_\alpha^{TF} = \sum_i w_\alpha^i \mathbf{e}_\alpha^i \tag{5.7}$$

where the weights w_α^i are defined as follows.

$$w_\alpha^i = \begin{cases} 0 & : r_\alpha^i \leq r_{\text{at}}^{TF} \\ \left(1 - \frac{r_\alpha^i - r_{\text{at}}^{TF}}{d_{\text{hy}}^{TF}}\right)^\nu & : r_{\text{at}}^{TF} < r_\alpha^i < r_{\text{at}}^{TF} + d_{\text{hy}}^{TF} \\ 0 & : r_\alpha^i \geq r_{\text{at}}^{TF} + d_{\text{hy}}^{TF} \end{cases} \tag{5.8}$$

with ν an integer greater than 0 (discussed in detail later) and $r_\alpha^i = |\mathbf{r}_\alpha - \mathbf{r}_{\text{centr},i}|$. This means that the weight w_α^i is 0 if the molecule is outside the correction region with respect to center i. On the other hand, if the molecule is within this area, the weight changes from 1 closest to center i to 0 furthest from center i. Thereby, stronger weight is given to vectors toward closer sphere centers. Note that these weights are used only for determining the direction and not the magnitude of the TF, therefore the abrupt change of the weight from 0 to 1 at the close edge of the correction region does not lead to any abrupt changes in the TF.

If $\mathbf{v}_\alpha^{TF} = \mathbf{0}$, the molecule α is not in the correction region with respect to any center, and no correction force is applied. Otherwise, if $\mathbf{v}_\alpha^{TF} \neq \mathbf{0}$ we calculate the unit vector of $\mathbf{v}_\alpha^{TF}$,

$$\tilde{\mathbf{e}}_\alpha^{TF} = \frac{\mathbf{v}_\alpha^{TF}}{|\mathbf{v}_\alpha^{TF}|}, \tag{5.9}$$

and apply the thermodynamic force for molecule α along $\tilde{\mathbf{e}}_\alpha^{TF}$. This can be expressed as

$$\mathbf{F}^{TF}(\mathbf{r}_\alpha) = F^{TF}(r_{\min})\tilde{\mathbf{e}}_\alpha^{TF}. \tag{5.10}$$

The outlined protocol generates a TF correction that acts only in the extended hybrid (correction) region. If the molecule is within the correction region with respect to only one center, the TF is applied precisely as in the case of a simulation box containing a single sphere, perpendicularly to the iso-resolution surface. If the molecule is located where the correction regions of two or more spheres overlap (referred to as the overlap region), an interpolation takes place that can be tuned with the parameter ν. We note that the procedure can lead to nonzero $\tilde{\mathbf{e}}_\alpha$ within the combined atomistic region of all individual spheres. However, as the magnitude of the TF correction $F^{TF}(r_{\min})$, which is calculated independently just based on the iso-resolution surfaces, is zero there, no force is applied in this case. The scheme naturally reduces to the regular, well-established approach for single spherical regions [182]. The calculation procedure is visualized with examples in Fig. 5.4.

Finally, we discuss the variable ν, which controls the functional form of the relationship between the weight of a vector $\mathbf{e}_\alpha^i$ in the sum in Eq. 5.7 and the distance of molecule α from center i. If $\nu = 1$, the weight varies linearly with the distance. This leads to a situation in

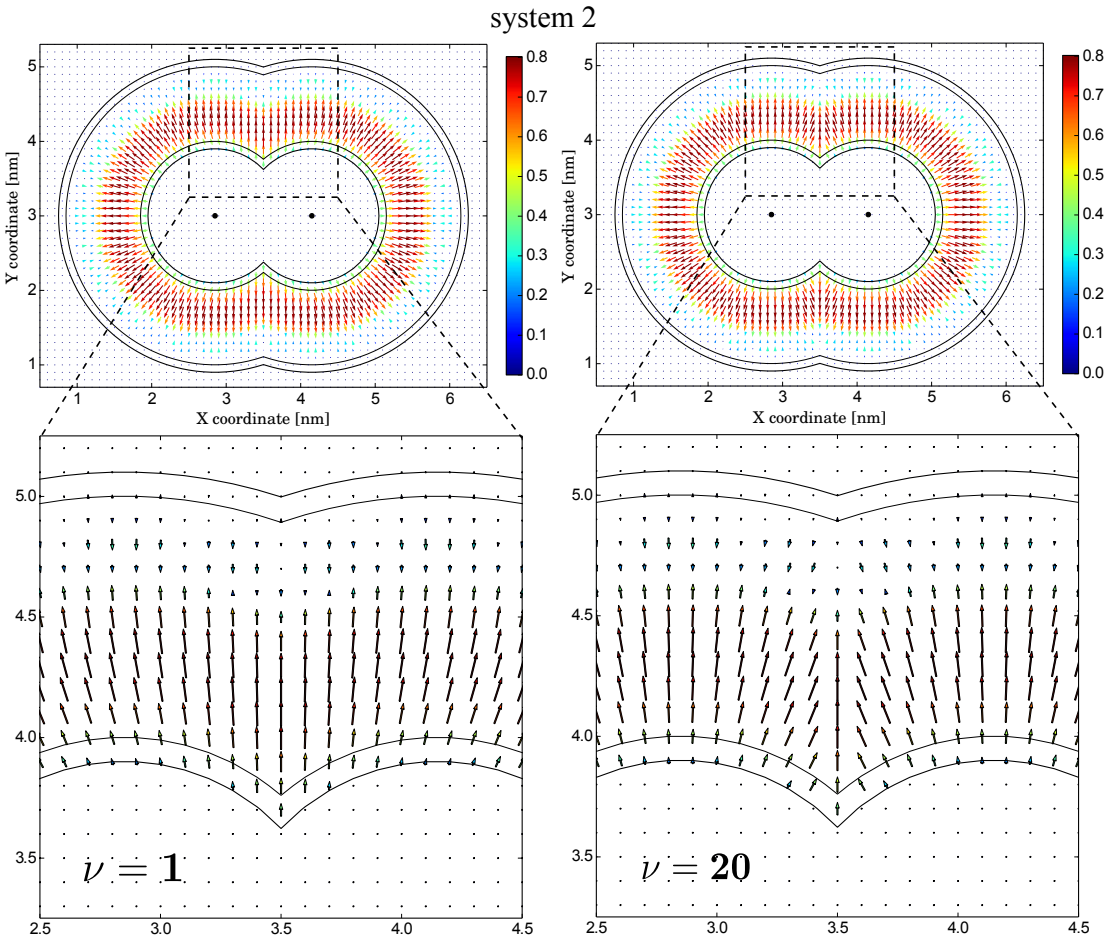

Figure 5.5.: Two-dimensional projections of the thermodynamic force using the scheme proposed in section 5.2.2 with $\nu = 1$ (*left*) and $\nu = 20$ (*right*), in a setup with a high-resolution region composed of two overlapping spheres. The centers of the spheres are separated by 1.3 nm. The thin black lines indicate the hybrid region within which the resolution changes, while the thicker black lines denote the slightly extended area in which the thermodynamic correction is applied. The two heavy black dots represent the centers of the spheres. The length of the arrows is proportional to the absolute values of the forces as follows: The absolute values have been normalized by 1.0 kJ/(mol nm) to make them unitless, shifted by 1.0 and then converted to a logarithmic scale (base 100).

which $\tilde{\mathbf{e}}_\alpha$, the direction of the TF application vector, changes smoothly throughout the correction region (see Fig. 5.5, left). Within most of the overlap region, the direction of $\tilde{\mathbf{e}}_\alpha$ deviates strongly from perpendicularity to the iso-resolution surfaces. Larger values for ν (for example $\nu = 20$ in Fig. 5.5, right) give stronger weight in the sum in Eq. 5.7 to smaller values of $r_\alpha^i - r_{\text{at}}^{TF}$ in Eq. 5.8, i.e. to the vectors toward closer sphere centers. This leads to a situation where the vector $\tilde{\mathbf{e}}_\alpha$ is nearly perpendicular to the iso-resolution surfaces everywhere except very close to the line

along which the closest center to a molecule changes (see green line in Fig. 5.4). The choice of ν has implications for the resulting solvent density throughout the simulation box, as is shown below.

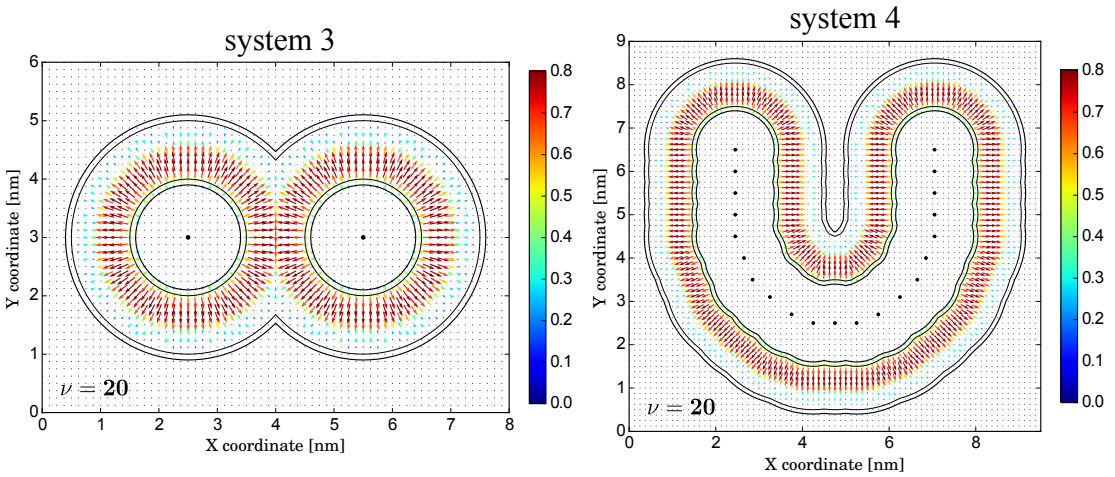

Figure 5.6.: Two-dimensional projections of the thermodynamic force for two different setups. *Left*: Two spherical high-resolution regions whose hybrid regions overlap. The centers are separated by 3.0 nm. *Right*: 21 spheres are placed so as to form a U-shaped high-resolution region. In both cases the interpolation parameter ν is 20. As in Fig. 5.5 the heavy black dots represent the centers of the spheres, the thin black lines indicate the hybrid region within which the resolution changes and the thicker black lines denote the slightly extended area in which the thermodynamic correction is applied. Also the same arrow scaling as in Fig. 5.5 is applied.

5.3. Simulation details

To validate the protocol described above we considered several different simulation setups. First, we studied a series of pure liquid water boxes each containing a differently shaped atomistic region constructed from overlapping spheres, positioned to represent various critical geometrical arrangements. In this case, the centers did not change position during the simulation. This allowed us to explore the effectiveness of the protocol for multiple spheres without taking into account their relative movement. We subsequently studied a peptide in aqueous solution around which we constructed an atomistic region by placing a sphere center on each protein atom. We considered first the frozen case, and finally the case in which the peptide undergoes folding and the atomistic region adapts its shape to correspond to the conformational change.

5.3.1. Liquid water

For the atomistic water we used the SPC/E model [270]. In the CG region, water molecules were modeled by single beads on the center of mass, interacting via a pair potential obtained

via iterative Boltzmann inversion (IBI) without pressure correction [106], which reproduces the water's molecular radial distribution function. Because IBI-based potentials are also simple to calculate they are popular and widely applied. However, coarse-graining typically yields potentials with significantly different thermodynamic properties. In particular, our IBI potential leads to a pressure ≈ 6000 times higher than the water model's reference pressure. To compensate for this pressure imbalance between the IBI potential and the atomistic water, we derived a TF using the iterative procedure described in Eq. 5.6, using a single spherical atomistic region with dimensions $r_{\mathrm{at}} = 1\,\mathrm{nm}$ and $d_{\mathrm{hy}} = 1\,\mathrm{nm}$. To assess the quality of the resulting TF, we ran simulations of a cubic box (dimension 6.00216 nm) of water containing a spherical atomistic region with dimensions as above (**system 1**). The total simulation time for the production runs was 7.5 ns, with 20 ps equilibration.

To test the proposed method for adaptive resolution simulations with arbitrary geometries, we considered three particularly critical test cases, using again $r_{\mathrm{at}} = 1\,\mathrm{nm}$ and $d_{\mathrm{hy}} = 1\,\mathrm{nm}$:

- **System 2:** a system with a high-resolution region composed of two spheres whose centers are separated by 1.3 nm along the x-axis, at $z = 3.0\,\mathrm{nm}$, in a box of dimensions $6.93\,\mathrm{nm} \times 6.0\,\mathrm{nm} \times 6.0\,\mathrm{nm}$. We ran three sets of simulations for this setup, one without TF correction and two with TF and interpolation parameters $\nu = 1$ and $\nu = 20$. The two-dimensional projection of the TF onto the plane at $z = 3.0\,\mathrm{nm}$ of this arrangement is visualized in Fig. 5.5

- **System 3:** a system where only the hybrid but not the atomistic regions of two neighboring high-resolution regions overlap. Two sphere centers were placed separated by 3.0 nm along the x-axis, at $z = 3.0\,\mathrm{nm}$, in a box of dimensions $7.95\,\mathrm{nm} \times 6.0\,\mathrm{nm} \times 6.0\,\mathrm{nm}$ (see Fig. 5.6, left). Simulations were run with TF using an interpolation with $\nu = 20$.

- **System 4:** a larger, strongly deformed high-resolution region. We placed 21 sphere centers, at $z = 3.0\,\mathrm{nm}$, in a box of dimensions $9.40\,\mathrm{nm} \times 9.0\,\mathrm{nm} \times 6.0\,\mathrm{nm}$ to form a U-shaped high-resolution region (see Fig. 5.6, right). Also in this case, we used the TF with $\nu = 20$.

In each case, the total simulation time for production runs was 6 ns, with an initial equilibration of 20 ps.

All simulations were run at $300\,K$ at a density of $\rho_0 = 33.117\,\mathrm{molecules/nm^3}$. The electrostatics were treated with the reaction field [217] method using a dielectric constant $\epsilon = 67.5998$ [174]. The cutoff for all interactions was 1 nm. Bond constraints were enforced using the SETTLE algorithm [55]. The time step was 2 fs, production simulations were performed in the NVT ensemble with a Langevin thermostat with a friction constant $\gamma = 5.0\,\mathrm{ps^{-1}}$. For system 4, the U-shaped high-resolution region, we additionally performed a set of short simulations during which we apply a Langevin thermostat with friction constant $\gamma = 30\,\mathrm{ps^{-1}}$ only in the CG and HY regions. This leaves the dynamics in the AT region unperturbed [155]. Here, the total production run time was 100 ps.

The coarse-grained potential and the TF were obtained using the VOTCA package [271] together with GROMACS [187, 289]. All other simulations were carried out using the open source ESPResSo++ package [1] in which we implemented the scheme derived in section 5.2.2.

For all systems where fully atomistic reference simulations are referred to, these were performed with identical force field and simulation parameters to the corresponding AdResS simulations, at the reference density of $\rho_0 = 33.117\,\text{molecules}/\text{nm}^3$.

5.3.2. Polyalanine peptide

The peptide system studied was a polyalanine-9 molecule in aqueous solution. In order to achieve peptide folding on a relatively short time scale, we used the Amber99 force field [300], which has the well-known property of overfavorizing α-helix formation [301]. We note that in this work we are not interested in elucidating the precise details of polyalanine folding, something which has already been explored in detail elsewhere [301–303]. Rather, we are interested in studying the performance of our methodology in the case of an example macromolecular system which rapidly and consistently undergoes realistic large-scale conformational change, hence the force field choice.

For the AdResS peptide simulations, an atomistic region was constructed from multiple overlapping spheres by placing one sphere center on each protein atom. As before, each of these regions had a radius r_{at} of 1 nm and a hybrid region width d_{hy} also of 1 nm. Thereby, the peptide was modeled on the fully atomistic level at all times and was enveloped in a 1 nm thick layer of atomistic water. Previous work showed this to be the minimum necessary distance for protein and hydration shell properties to be preserved [155]. Surrounding this layer of atomistic water was the hybrid region in which the solvent smoothly changes resolution, and this was embedded in the CG region which was the remainder of the simulation box. The interpolation parameter for the TF was $\nu = 20$ in all AdResS peptide simulations. For all peptide simulations, the time step was 1 fs, consistent with the highest frequency motion in the system, the vibrations of bonds containing hydrogen. Where a barostat was used in fully atomistic equilibration simulations, this was the Parrinello–Rahman barostat with coupling time 2.0 ps. All other simulation parameters and interaction potentials were as given for the pure water simulations in section 5.3.1.

We generated a fully extended polyalanine conformation using the Avogadro package [304], and solvated it in 10178 water molecules. The cubic simulation box had sides of length 6.73 nm. The peptide was in zwitterionic form with charged termini. When the polypeptide is fully stretched, the distance between the end C_α atoms is 2.8 nm. During 100 ps NPT and 2 ns NVT equilibration with the fully atomistic approach, we applied a harmonic constraint with its minimum at 2.6 nm to this distance to produce equilibrated but extended structures. Fourteen such configurations were generated, sampled at intervals of 100 ps. The constraint was then removed to allow the peptide to fold, and 20 folding trajectories of length 8 ns each were produced with the fully atomistic and AdResS approaches, as well as eight additional trajectories of length 5 ns to increase sampling of the folding pathways.

We also performed simulations in which the conformation of the peptide was frozen, using an equilibrated extended conformation produced as described above. Here, a further 20 ps equilibration were performed followed by 5 ns total production. As in the pure water case above, we also ran 2.5 ns with the thermostat ($\gamma = 30\,\mathrm{ps}^{-1}$ in this case) acting only in the CG and HY regions in order to leave the dynamics of the AT region completely unperturbed.

5.4. Results and Discussion

5.4.1. Liquid water

System 1: Single spherical atomistic region. While the TF has been shown to work very well in counteracting the density gradient caused by the pressure difference between atomistic and coarse-grained force fields [155, 174, 175, 187–189, 211, 213–216, 228, 286], fluctuations on the order of 1% still tend to occur in the density profile even with the compensating force. We therefore first tested the quality of the thermodynamic force used in this work in order to be able to distinguish between two effects: (i) inherent density fluctuation effects which occur even in the case of a simple spherical atomistic region and the well-established existing TF protocol and (ii) potential additional effects arising from the use of multiple overlapping spheres and the TF application procedure outlined in section 5.2.2. We therefore calculated density profiles along the direction of resolution change in AdResS simulations using a single spherical atomistic region. These have already been presented in Fig. 5.3. Without TF the pressure difference between the AT and CG regions leads to an increased density in the AT regions by about 20% compared to a flat profile. Using the TF compensation, however, we obtain a virtually flat profile. Only small deviations up to 1% in the AT and HY regions remain.

System 2: Atomistic region composed of two overlapping spheres. We now turn to the setup with two overlapping spheres (see Fig. 5.5). The chosen geometry results in a pronounced kink in the iso-resolution surfaces, along the circumference of the circle where the closest sphere center changes for a particle on that surface (intersection of the spheres forming the iso-resolution surface, on the green line in Fig. 5.4). Therefore, this scenario provides a good test of the performance of the correction protocol outlined in section 5.2.2.

We ran simulations with TF (interpolation parameters $\nu = 1$ and $\nu = 20$) and without TF and calculated the density on a three-dimensional grid in cubic boxes of dimension 0.3 nm. We visualize this as projections of the density onto two-dimensional planes at a selection of values of z along the z-axis of the simulation box. The density maps obtained without TF are shown in Fig. 5.7. As we have already seen in the previous case of a single spherical high-resolution region, the pressure gradient between the subregions leads to an approximately 20% higher density in the AT region.

Using the TF correction and applying it according to Eqs. 5.7–5.10, we obtained the density maps shown in Fig. 5.8. An interpolation parameter $\nu = 1$ results in strong density distortions in the region where the closest center for a particle on an iso-resolution surface changes. In most of this region, the TF is applied in a direction which is not perpendicular to the iso-resolution

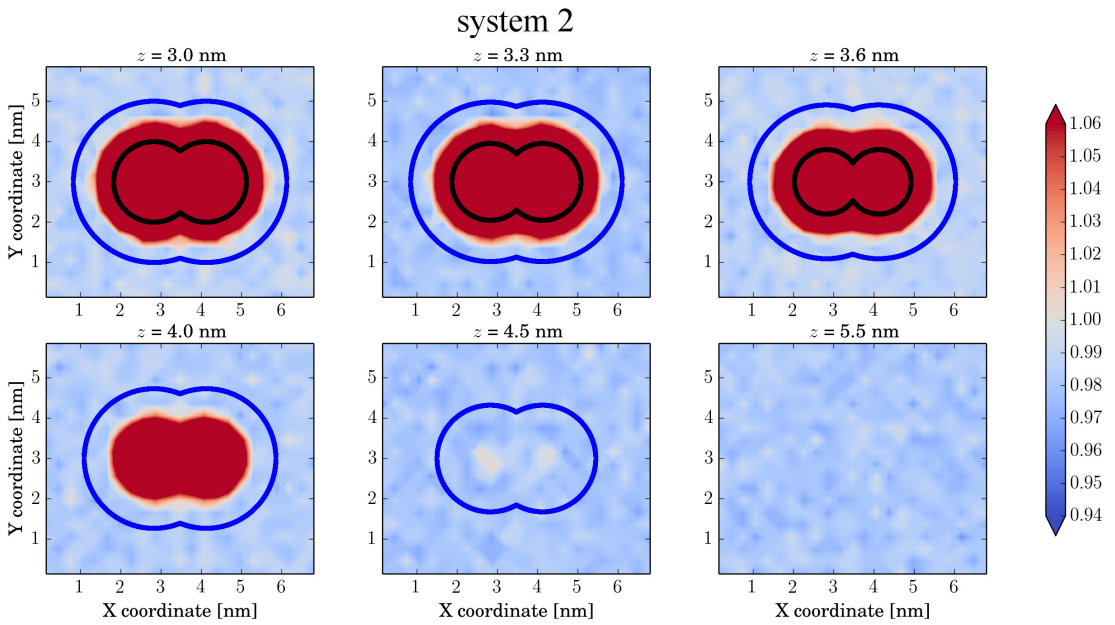

Figure 5.7.: Two-dimensional density maps calculated at a selection of z-planes for an adaptive resolution simulation of liquid water (system 2), with no TF correction applied. They are normalized to the reference density such that the ideal value is 1. The thick black lines indicate the intersections of the z-planes with the boundary between AT and HY regions. Similarly, the thick blue lines represent the intersections with the HY/CG boundary. As the total box length in the z-dimension is $L_z = 6.0\,\text{nm}$, the setup is symmetric with respect to the $z = 3.0\,\text{nm}$ plane. Therefore, we only show results for $z \geq 3.0$ nm.

surfaces (Fig. 5.5). This perpendicular direction, however, is the direction along which we expect the pressure gradient to push particles from the CG to the AT region. Hence, the TF does not appropriately cancel this.

This problem can be solved by using a larger value for ν. This significantly increases the domain where the TF is applied perpendicularly to the iso-resolution surfaces, i.e., interpolation is performed only directly adjacent to the point where the closest center changes. We explored different values of ν in the range between 1 and 50 and found that with $\nu = 20$ the approach for applying the TF works well (see Fig. 5.8 (b)). The previous distortions are resolved and a nearly flat density map is obtained. The remaining small deviations are of the order 1%–2%. In Fig. 5.9 we show an equivalent density map for a fully atomistic reference system. The fluctuations in the density are also ∼1%–2%, due to the finite length (6 ns) of the simulations. We note that the average distance traveled by a water molecule in this time is about 4 nm, based on the diffusion constant of SPC/E water [305]. Comparison of Figs. 5.9 and 5.8 (b) makes it clear that most of the observed density fluctuations in the latter figure are due to finite averaging. Nevertheless, there seems to be a small (not larger than ∼1%) systematic deviation toward lower densities within the center of the AT region (bluish color in the center for $z = 3.0$, $z = 3.3$,

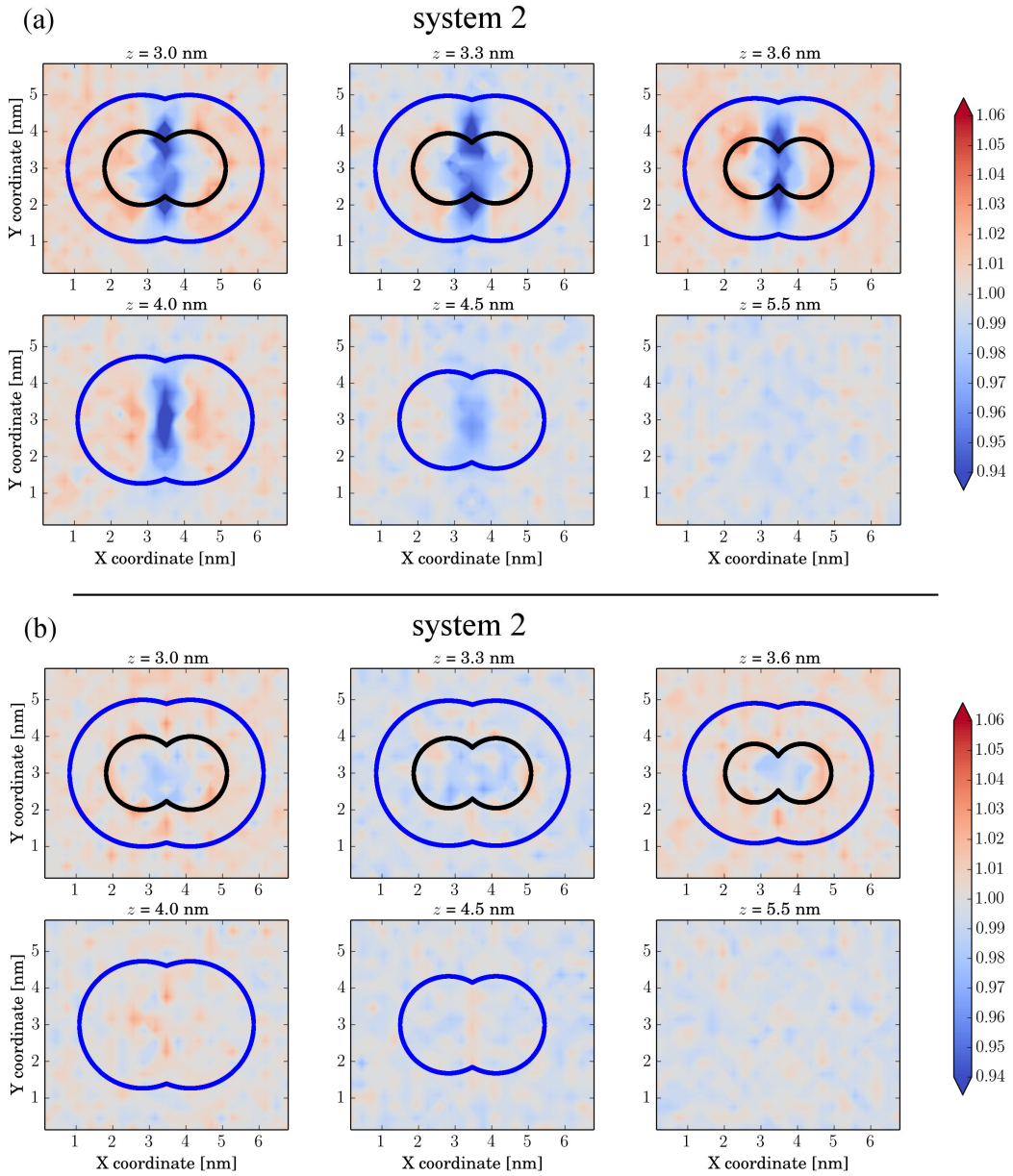

Figure 5.8.: Two-dimensional density maps calculated at a selection of z-planes for an adaptive resolution simulation of liquid water (system 2). They are normalized to the reference density such that the ideal value is 1. The protocol derived in section 5.2.2 is applied using *(a)* $\nu = 1$ and *(b)* $\nu = 20$. The thick black lines indicate the intersections of the z-planes with the boundary of the AT region. Similarly, the thick blue lines represent the intersections with the outer boundary of the HY region.

and $z = 3.6\,\text{nm}$). This is in line with the observations for the spherical case (see Fig. 5.3), where the density is also about 0.5%–1.0% lower than the reference. Hence, we suppose that the small

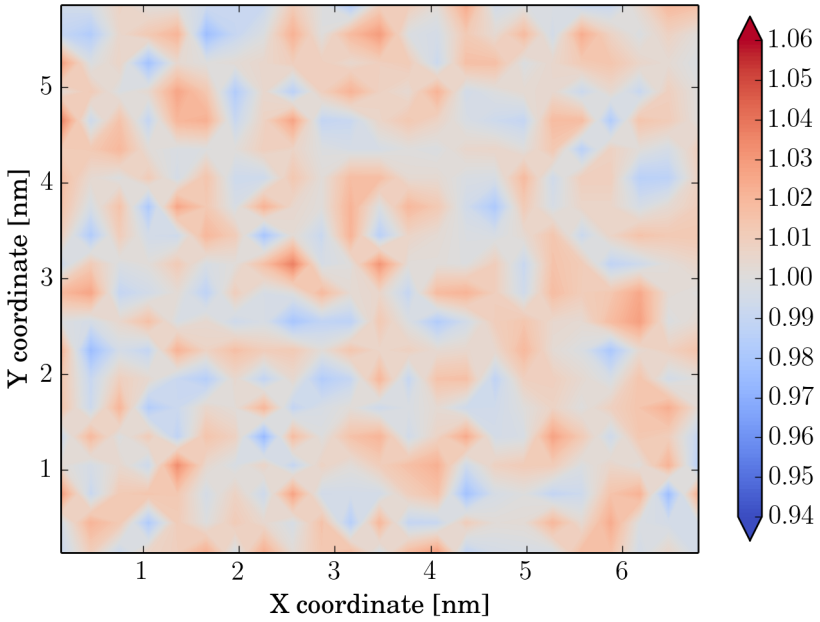

Figure 5.9.: Two-dimensional density map calculated at the $z = 3.0\,\mathrm{nm}$ plane for a fully atomistic reference simulation of liquid water using the same box dimensions as previously for system 2. They are normalized to the reference density such that the ideal value is 1.

systematic deviation in this case is rather a feature of the parametrized TF itself than an effect from the protocol outlined in Sec. 5.2.2.

Larger values for ν did not lead to any further improvements; hence, we chose $\nu = 20$ for all subsequent work.

System 3: Two spherical atomistic regions, overlapping hybrid regions. Next, we consider the case in which only the hybrid regions of two neighboring high-resolution regions overlap. The setup we used is shown in Fig. 5.6 (left).

We again performed simulations with $\nu = 20$ and measured two dimensional density maps for different z-planes. They are shown in Fig. 5.10 (a). No significant distortions are visible.

System 4: U-shaped atomistic region. Finally, we consider a U-shaped AT region which is assembled from 21 overlapping spheres arranged as visualized in Fig. 5.6 (right). This resembles a scenario that could occur, for example, when a long polymer or peptide folds over into a loop. Figure 5.10 (b) shows the density maps obtained from simulations with an interpolation parameter $\nu = 20$. We also measured the density as a function of distance with respect to the closest AT region center (Fig. 5.11 (a)).

The density in the atomistic region is systematically 1%–2% lower than the reference. Considering the previous (Fig. 5.8 (b)) or the simple spherical case (Fig. 5.3), density fluctuations of around 1% are to be expected. However, in this case there is an additional minor effect arising from the use of multiple overlapping spheres. We therefore quantified to what extent

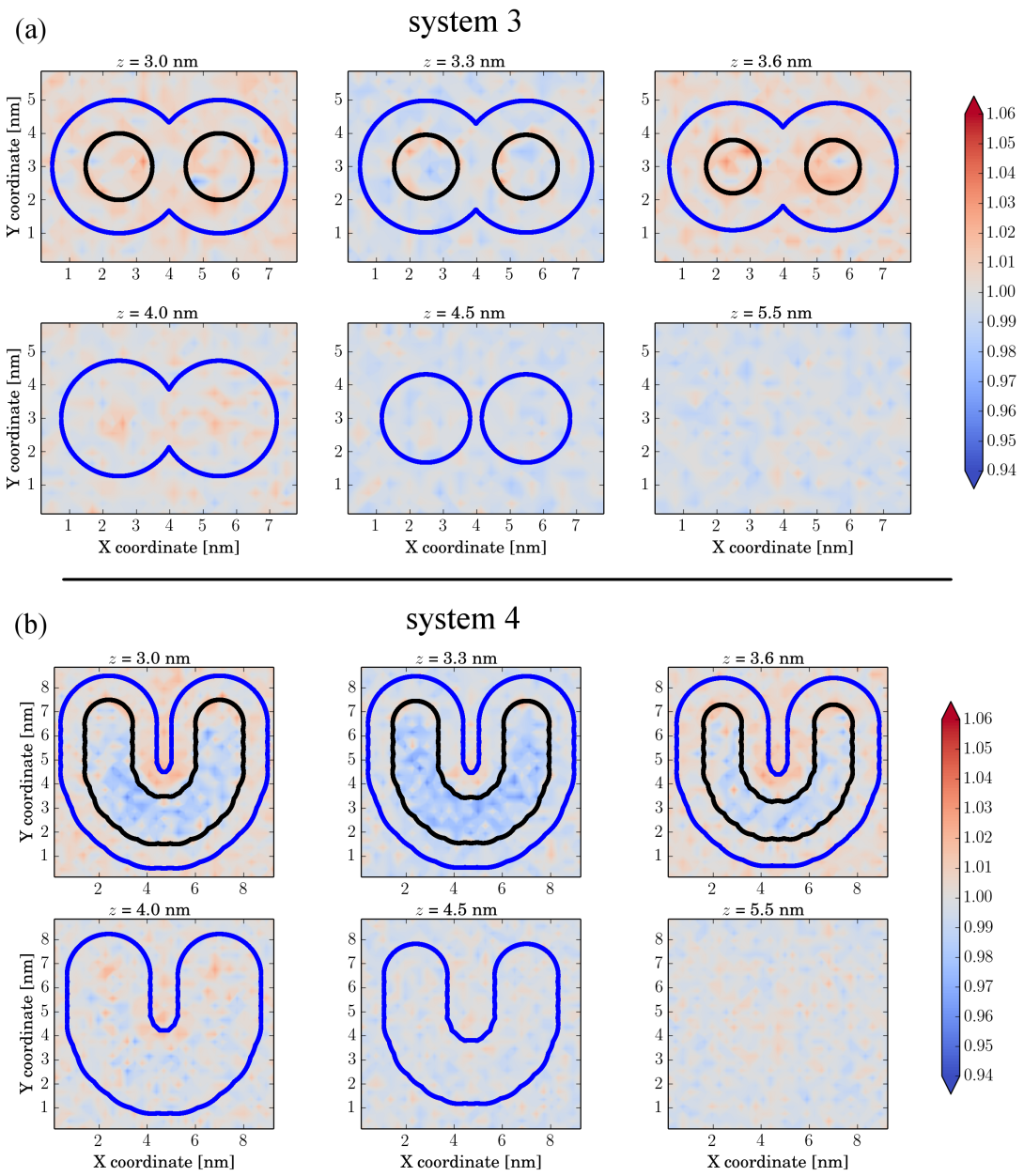

Figure 5.10.: Two-dimensional density maps calculated at different z-planes for two different adaptive resolution simulations of liquid water. They are normalized to the reference density such that the ideal value is 1. *(a)* Two spherical high-resolution regions overlap only with their hybrid regions (system 3). *(b)* U-shaped high-resolution region assembled from 21 overlapping small spheres (system 4). In both cases, we used $\nu = 20$. The thick black lines indicate the intersections of the z-planes with the boundary of the AT region, while the thick blue lines represent the intersections with the outer boundary of the HY region.

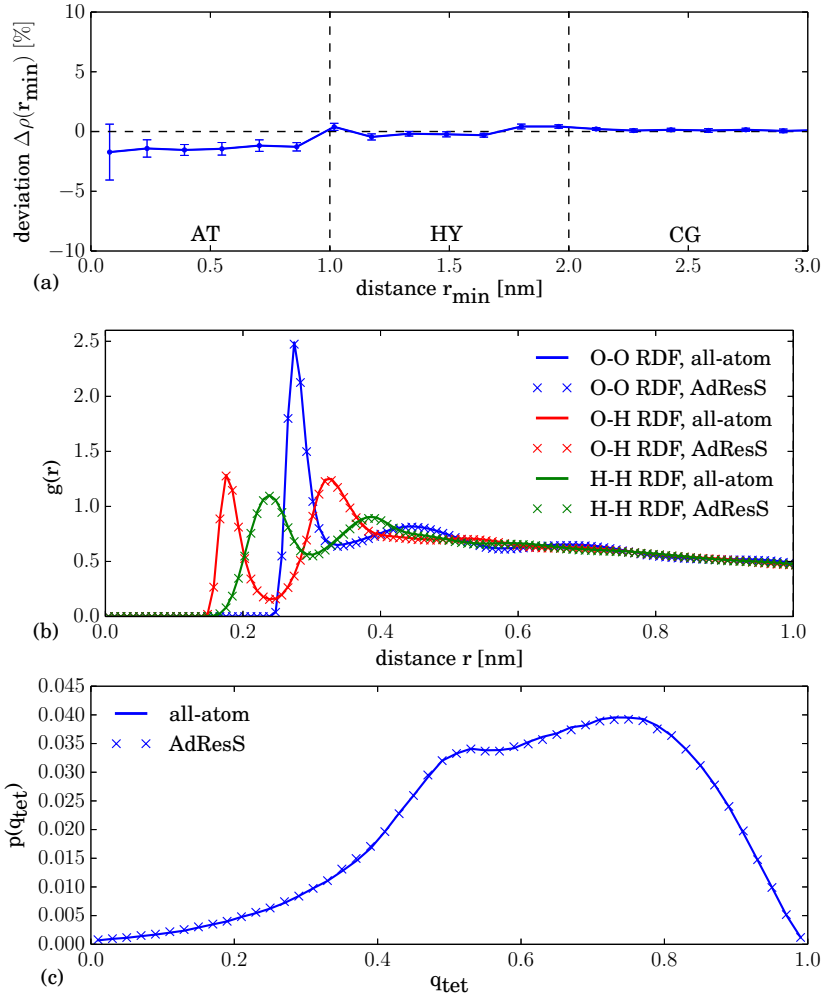

Figure 5.11.: *(a)* Density as function of distance with respect to the closest sphere center for the U-shaped high-resolution region setup (system 4, Figs. 5.6 (right) and 5.10 (b)). *(b)* and *(c)* Radial distribution functions (oxygen–oxygen, oxygen–hydrogen, hydrogen–hydrogen) and tetrahedral order parameter q_{tet} measured within the AT region of the U-shaped setup and for all-atom reference simulations.

the structural and dynamical properties of the AT region may be perturbed with respect to a fully atomistic reference system. For that purpose, we calculated the water radial distribution functions (RDFs) as well as the tetrahedral order parameter q_{tet} within the AT region. For a molecule i, the parameter q_{tet} is defined as

$$q_{\text{tet}} = 1 - \frac{3}{8} \sum_{j=1}^{3} \sum_{k=j+1}^{4} \left(\cos(\theta_{j,k}) + \frac{1}{3} \right)^2 \tag{5.11}$$

where j and k run over i's four nearest neighbor molecules. The angle $\theta_{j,k}$ is formed by the

oxygen atoms of molecules i, j and k with i in the center. This results in $q_{tet} = 1$ when the molecule forms a perfect tetrahedron with its four nearest neighbors and it is on average 0 in an ideal gas. The results are shown in Fig. 5.11 (b) and (c). Both the RDFs and q_{tet} match the all-atom reference quantitatively. Considering that the density within in the atomistic region is slightly lower than the reference we would expect small deviations in the RDFs. However, the density mismatch is small enough that the differences in the RDFs are insignificant at this level of resolution. Hence, the structural properties are well preserved in the AT region of the adaptive simulation.

To check for any possible perturbation of the water dynamics in the AT region we analyzed simulations where the thermostat was only applied within the CG and HY regions. We measured the water's oxygen and hydrogen velocity autocorrelation functions $\langle \mathbf{v}(t_0)\mathbf{v}(t_0 + t)\rangle / \langle \mathbf{v}(t_0)\mathbf{v}(t_0)\rangle$ (VACF) and again compared to an all-atom reference. The measurements were performed within the inner part of the AT region considering only atoms further than 0.5 nm away from the AT/HY interface. As is visible in Fig. 5.12, the obtained VACFs from

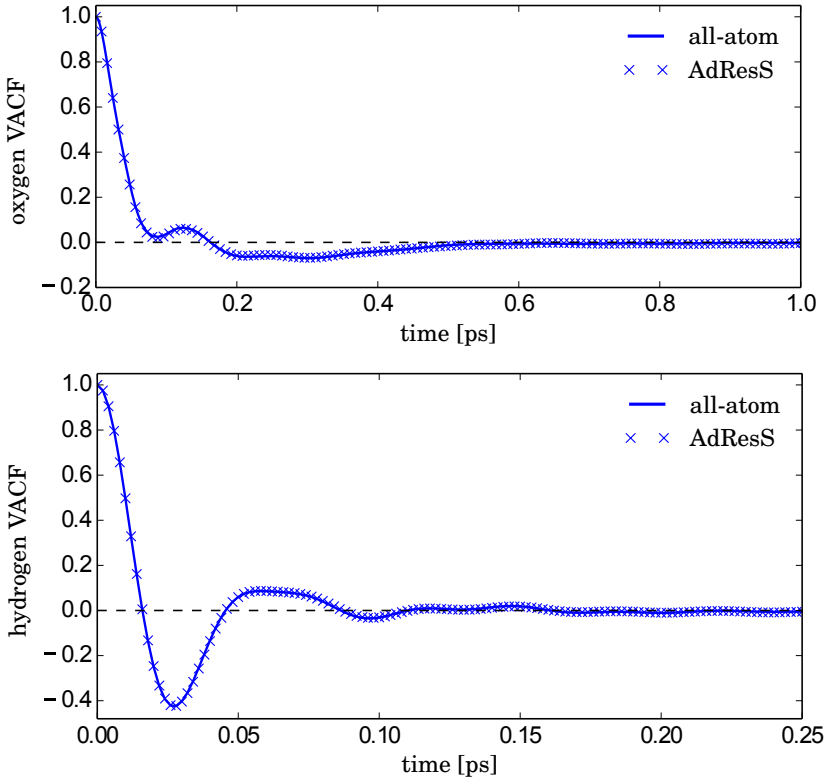

Figure 5.12.: Oxygen and hydrogen velocity autocorrelation functions measured within the AT region of the U-shaped adaptive setup (system 2) and for all-atom reference simulations.

the adaptive and the reference simulations match within line thickness. Therefore, also the dynamical properties within the AT region of the adaptive setup are unperturbed. The observed

small density deviations are too small to have a significant measurable effect on the observed structural and dynamical quantities in the AT region.

We can conclude that when choosing a sufficiently high value for the interpolation parameter ν the proposed approach for applying the TF in cases of irregular geometries for the AT region works well. The precise value of ν needed will depend on the AT and CG potentials and, in particular, on the difference in thermodynamic properties between the two models. In our case, choosing $\nu = 20$ achieves a smooth density across the system and also leaves the behavior of the water in the AT region unperturbed. This confirms that the additional force arising from the pressure difference between the models is indeed acting perpendicular to the iso-resolution surfaces, and this is the direction along which the thermodynamic force must be applied. When ν is too small, however, the area where the TF is applied in a inappropriate direction is too large. Nevertheless, for many other systems, a smaller value of ν than that used here could be sufficient, for example, when the different models have a smaller pressure difference. On the other hand, the computational cost of increasing ν is negligible; therefore, a value of $\nu = 20$ can be seen as a "safe choice" for most realistic applications, where the difference in thermodynamic properties will be similar to or less than that of the two models used here. Further testing would be recommended in systems coupling atomistic and coarse-grained models that are thermodynamically even more different than the case studied here.

The three test cases explored here are representative of situations occurring in practical simulation applications. Whenever an atomistic region is composed of spheres associated with atoms or molecules arranged in an irregular geometrical pattern, such overlaps can occur. One can think, for example, of the case of a fully atomistic macromolecule inside a high-resolution region composed of spheres associated with each of its atoms. Moreover, side chain rotations, loop motions, and other large-scale conformational fluctuations will lead to a variety of different overlaps occurring during the simulations. Other examples include rough crystallization fronts and aggregation of molecules that each have their own associated high-resolution region.

5.4.2. Static peptide

Next, we turn to one such more complex system, a polyalanine-9 peptide, fixed in space; i.e., all of its atoms have fixed positions (for the case of a freely moving and folding peptide, see section 5.4.3). We associate a sphere with each of the peptide's 93 atoms, thereby forming a shell of atomistic water around the molecule. The water molecules then gradually change from their atomtistic description to the coarse-grained model as the distance from the peptide increases (Fig. 5.1, top). As the peptide atoms are fixed, the high-resolution region does not change its geometry during the simulation. Again, we employ the previously proposed scheme for applying the TF with $\nu = 20$.

To assess the performance of the TF in this case, we measured the water density as a function of distance with respect to the closest peptide atom for each water molecule. The resulting density profile is shown in Fig. 5.13 (a). Overall, we obtain a smooth profile but with a density in the AT region lower by about 1%–2% than the reference. This is consistent with the previous

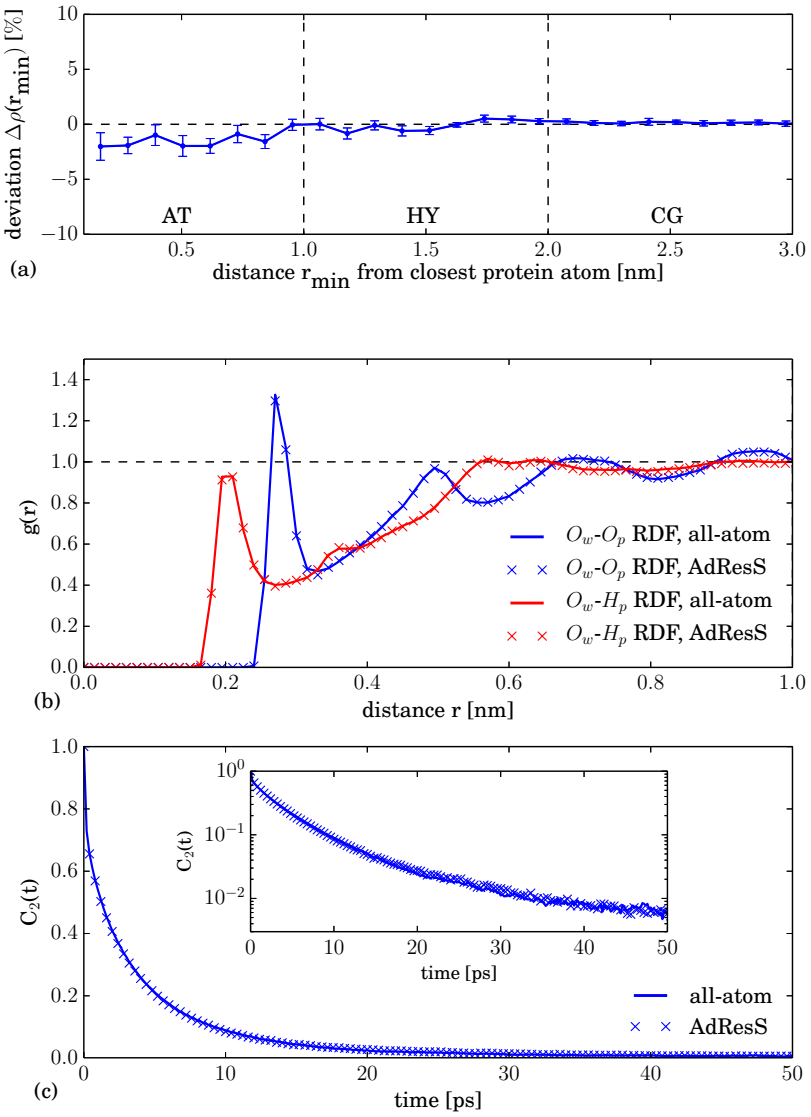

Figure 5.13.: Results for the simulations of the static extended peptide. *(a)* Density as a function of distance with respect to the closest protein atom. *(b)* Radial distribution functions between the water's oxygen atoms O_w and the oxygen O_p and hydrogen H_p atoms in the peptide's backbone amide bonds. *(c)* Second order reorientational time correlation function of the O–H bonds in water molecules in the first hydration shell of the peptide. *Inset:* Same as main plot but with y-axis in logarithmic scale.

observations for the density maps for the U-shaped atomistic region (Fig. 5.11 (a)).

To test whether the small density deviations affect the dynamics and structure of the water around the protein, we measured the radial distribution functions between the peptide's amide hydrogen and oxygen atoms and the water as well as the water dynamics in the peptide's hydration shell. For the latter, we analyzed the second-order reorientational time correlation

function of the *O–H* bonds in water molecules in the first hydration shell. It is defined as

$$C_2(t) = \langle P_2[\mathbf{u}(0) \cdot \mathbf{u}(t)] \rangle \,, \tag{5.12}$$

where $\mathbf{u}$ is the vector along the water *O–H* bond and P_2 is the second Legendre polynomial. It is related to experimental quantities accessible via ultrafast infrared spectroscopy and NMR spectroscopy [306]. For this measurement we applied the thermostat only in the CG and the HY regions. The results are presented in Fig. 5.13 (b) and (c). Both the radial distribution functions and the second order reorientational time correlation function match their all-atom references closely. Hence, we infer that the slightly too low density has no significant effect on the water's structural and dynamical properties within the peptide's hydration shell. We conclude that the method for applying the TF also works well in realistic scenarios in which the AT region is constructed by a large number of overlapping spheres.

Finally, we want to stress that even though this does not seem to be necessary in this and the previous setup with the U-shaped atomistic region in pure water, in principle, an even more accurate TF could be derived with the protocol in Eq. 5.6 that achieves a yet smoother density profile [182]. The geometry of the high-resolution regions in system 4 and the frozen peptide setup resemble, locally, a cylinder more closely than a sphere. For those specific cases, therefore, one option to improve the quality of the TF could be to derive it also in a cylindrical setup, as opposed to the spherical setup we used here.

We also note that, in fact, what we learn from the study of arbitrarily shaped but static atomistic regions is also directly applicable to the case of an arbitrarily shaped dynamically changing atomistic region, if the changes in shape of the region during the simulation are significantly slower than the time scale for solvent rearrangement.

5.4.3. Peptide folding in a self-adjusting AdResS setup

Finally, we turn to the case in which the spheres forming the atomistic region can move relative to each other. Using the polyalanine system investigated in the previous section, we now study its folding process using the same simulation setup as before, i.e., an atomistic peptide with an atomistic hydration shell, but without fixing the peptide in space. As the peptide folds, the high-resolution region folds with it since the overlapping spheres forming the AT region are associated with the peptide's moving atoms. Therefore, the high-resolution region automatically adapts itself when the peptide changes its conformation.

Alanine has a strong propensity to form helices [303], and the folding of short polyalanine peptides has already been extensively studied [301, 303]. Here, we use it as a paradigm system undergoing a realistic spontaneous large-scale conformational change.

We follow the process of peptide folding (shown in Fig. 5.1) via the radius of gyration, R_g,

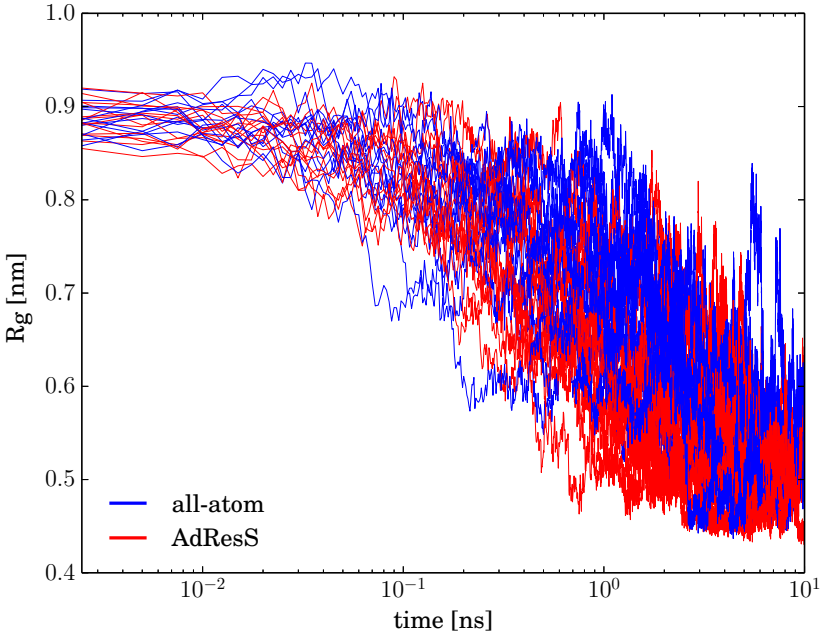

Figure 5.14.: Radius of gyration during peptide folding for the AdResS and the fully atomistic simulations (14 trajectories for each case). The average folding time in the atomistic reference simulations is 1.6 ± 0.5 ns, while the average folding time corresponding to the AdResS trajectories is 1.2 ± 0.7 ns.

and a measure of helicity, Q_{hel}, defined as [307]

$$Q_{hel} = \frac{1}{N} \sum_{i-j=3} \exp\left[-\frac{(R_{ij} - R_0)^2}{2\sigma^2}\right] \tag{5.13}$$

where the sum is over all pairs of C_α atoms separated by a shift of three residues along the backbone, N is the number of such 1–4 pairs, R_{ij} is the distance between C_α atoms i and j, R_0 =0.5 nm and σ^2 =0.02 nm^2 [307]. These measures were calculated as a function of time in multiple independent trajectories with AdResS and fully atomistic methodologies, using for both approaches the same set of equilibrated initial configurations in which the peptide is extended. The radius of gyration over time is shown in Fig. 5.14. The peptide rapidly collapses to the native state in a downhill folding process. We defined a folding time scale as the first passage time from the initial extended conformation to the state defined by $R_g < 0.52$ nm and $Q_{hel} > 0.5$, these being the average folded values. This gave a folding time of 1.6 ± 0.5 ns averaged over the fully atomistic trajectories and 1.2 ± 0.7 ns over the AdResS trajectories. The standard deviations are high, pointing to a large ensemble of folding pathways, and the difference between the two methodologies is less than the spread of times for a given methodology. The average time scale in the AdResS simulations is somewhat faster, which is most likely an indirect effect of the strongly increased diffusivity of the solvent model used

in the CG region [269]. The slight decrease in the folding time scale is not unexpected, since the inaccurate dynamics of CG models is a well-known phenomenon [129]. In order to better reproduce the folding time scale, the diffusion constant of the CG model could be tuned.

We note that the time scale for peptide folding is highly force field dependent, and the values given here should not be taken as a measure of the true folding time scale for polyalanine-9, but only as a quantitative comparison between atomistic and AdResS simulations.

We visualized the ensemble of folding pathways via the probability density function as a function of radius of gyration and helicity. This was calculated for each trajectory until the first folding time, and Fig 5.15 presents its negative logarithm averaged over all trajectories, showing that peptide molecules in AdResS and fully atomistic trajectories sample the same energy landscape while collapsing to the equilibrium state.

We now turn to the properties of the folded peptide. Typically, small peptides are not fully stable as helices in solution and contain a mixture of random coil and helical regions [303, 308]. In order to demonstrate that the peptide has folded to the same region of conformational space in the AdResS as in the fully atomistic simulations, we present in Fig. 5.16 logarithmic-scale Ramachandran plots calculated for the folded peptide (time since simulation start > 3 ns) in the fully atomistic and AdResS setups. Ramachandran plots show the two-dimensional probability density function (PDF) of the backbone dihedral angles ϕ and ψ (about the bonds N–C_α and C_α–C respectively). Here, we calculate the PDF averaged in time and over all angles and trajectories for each simulation methodology. With both AdResS and fully atomistic approaches, the peptide clearly forms predominantly α-helix-like structures, which occur in the broadly defined region $-180^\circ < \phi < 0^\circ$, $-100^\circ < \psi < 45^\circ$ [309]. Some torsional angles also occupy the region (here, around $\phi = \psi = 50^\circ$) sometimes referred to as the left-handed α helix region but which mostly contains structures more correctly labeled turns [309]. Finally, we also observe angles with values typical of a random coil with β conformation, (here $\psi > 45^\circ$) although of course there is no formation of a β-sheet here. We note that atomistic force fields, including that used here, are known to have Ramachandran plots which do not perfectly correspond with experimental observations from crystal structures [310]. In Fig. 5.16, the folded peptide in both cases clearly occupies the same region of conformational space.

Finally, we probe the local water structure in the hydration shell of the folded peptide. For this we calculate the distribution of tetrahedral order parameters for water molecules in the hydration shell. The tetrahedral order parameter is as defined in Eq. 5.11, including both water oxygen atoms and protein heavy atoms as potential nearest neighbors. The hydration shell is defined as all water molecules H-bonded to the peptide or whose oxygen atom is within 0.45 nm of a peptide carbon atom. Fig. 5.17 shows that the distributions of q_{tet} values in the AdResS and the fully atomistic system perfectly match. The structure of the hydration shell is fully preserved in the AdResS system.

The folding process and the conformational properties of the folded peptide depend critically on the aqueous solvent and its electrostatic and hydrogen-bonding properties [311], for which an atomistic-resolution model is necessary. Our methodology allows the simulation of both the

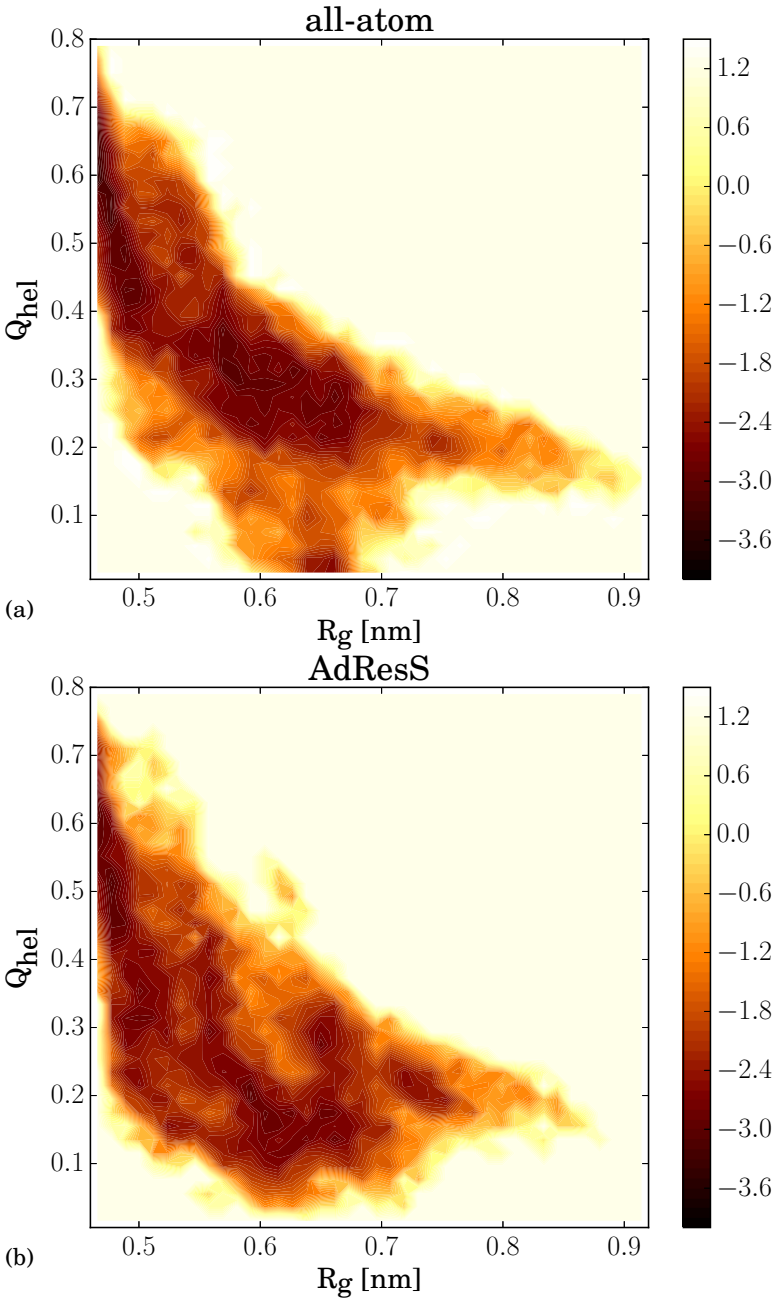

Figure 5.15.: Energy landscape sampled during folding as a function of the helicity Q_{hel} and the radius of gyration R_g. The color map is the negative of the logarithm of the probability density function. *(a)* Fully atomistic reference simulations. *(b)* AdResS simulations.

peptide and its hydration shell at the atomistic level, without in itself perturbing the folding process and while concentrating computational resources only on those parts of the system

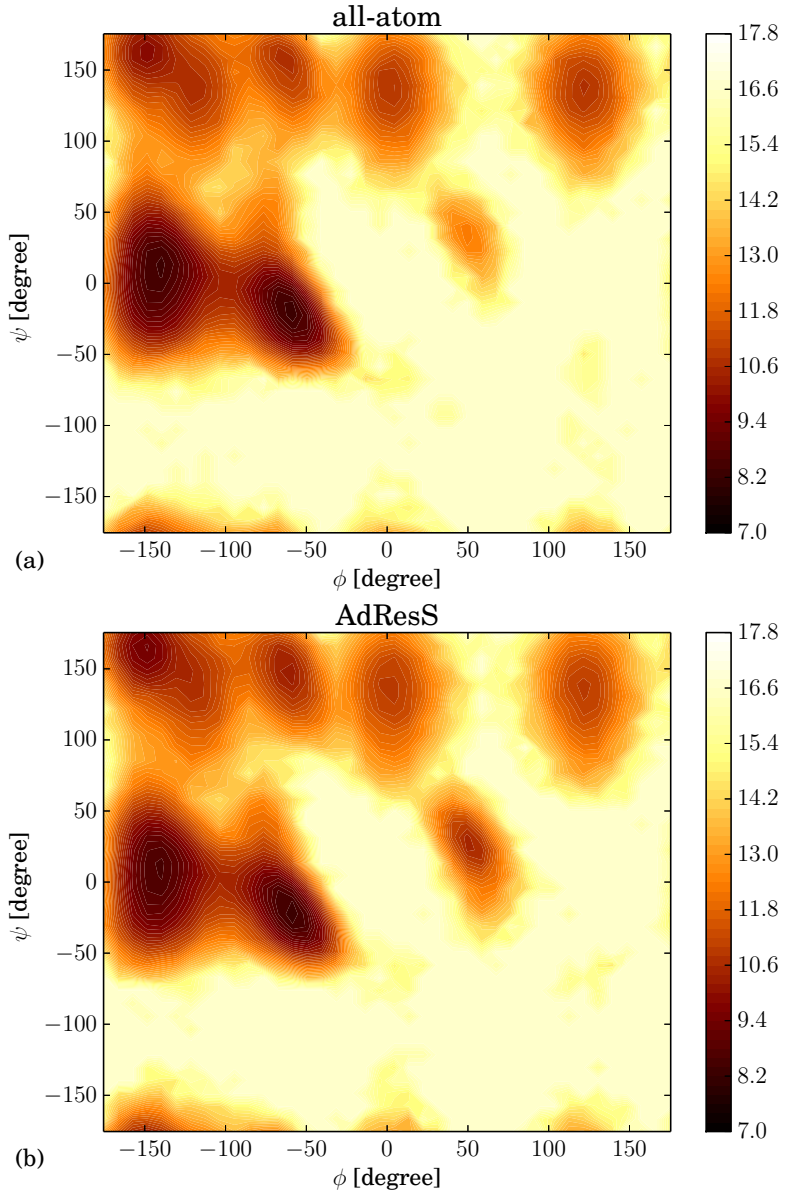

Figure 5.16.: Ramachandran plots for the folded peptide. The color map is the negative of the logarithm of the probability density function. *(a)* Fully atomistic reference simulations. *(b)* AdResS simulations.

where atomistic accuracy is necessary. Additionally, the structural and dynamical properties of the hydration shell are unperturbed despite the presence of the dynamic boundary at which the solvent changes resolution. This is thanks to the properties of the AdResS methodology, which allows realistic molecular fluctuations and free exchange of particles across the resolution boundary. By extension, we expect the results presented here to be applicable to other cases of solvated macromolecules undergoing large-scale conformational change.

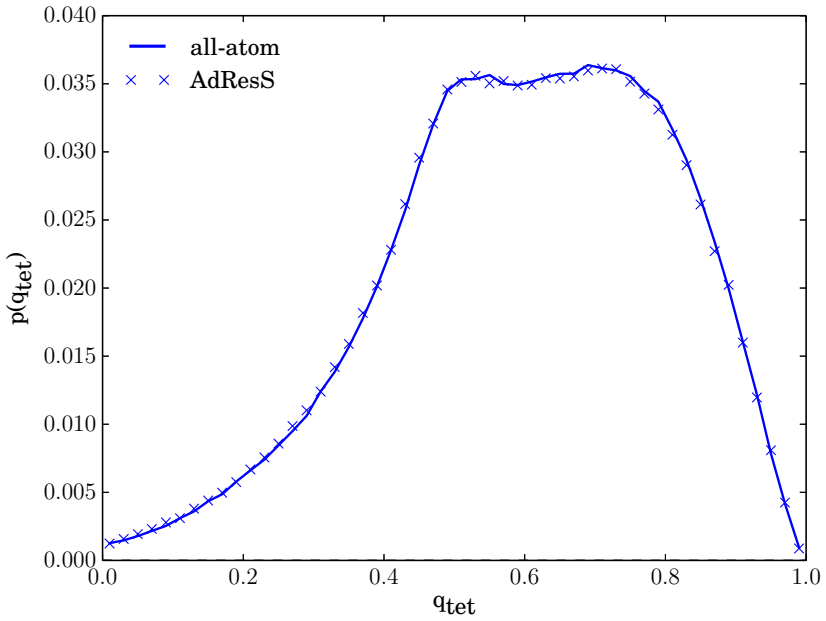

Figure 5.17.: Tetrahedral order parameter q_{tet} in the hydration shell of the folded peptide for AdResS and fully atomistic simulations.

5.5. Implementation and methodological extensions

We first want to point out that the scheme proposed here may not be the only technique to facilitate adaptive resolution simulations with several overlapping spherical atomistic regions. However, using a linear combination of the vectors with respect to the different sphere centers (Eq. 5.7) to derive the direction in which the TF acts is a simple, general, and intuitive solution to the problem. Although one possible variation of the scheme would be to change the way the weights within the linear combination are calculated (Eq. 5.8), we anticipate that simply changing the value of ν should be sufficient to accommodate almost all possible systems of interest.

The implementation of the method that we proposed in this paper is straightforward and did not lead to any significant additional computational overhead in our simulations. For the specific peptide system studied, the adaptive resolution simulations were twice as fast as the atomistic ones. This value was obtained on a single CPU to exclude additional effects due to parallelization. For parallel simulations, the computational speedup depends on the employed parallelization protocol and the load balancing scheme. Furthermore, the actual computational speedup achievable in practice is also strongly dependent on system size and the volume ratio of the atomistic and coarse-grained regions [286]. In future applications where the ratio of coarse-grained to atomistic region size is even larger, the speedup will be correspondingly greater than here. Such a situation could, for example, occur when choosing an overall box size so large that the AT subregion behaves as in the grand canonical ensemble. Then, contrary to

the test scenarios considered throughout this work, the CG region would be significantly larger than the AT region. Other cases, where a significantly larger speedup could be expected are those in which the CG model is even coarser. In our case, only three water atoms are coarse grained into one CG interaction site, but for liquids composed of larger molecules, such as toluene [187], there would be more atoms per CG bead. Likewise, a significant reduction of the number of interactions also occurs when interpolating classical and path integral based force fields [177, 212, 216, 224, 240]. Finally, the AdResS scheme is more advantageous the less computationally expensive CG interaction itself is. One could think, for example, of a basic, very short-ranged interaction like a Weeks–Chandler–Anderson potential on the coarse-grained level [274]. An even more drastic approach would be to use no potential at all in the CG region and let the molecules evolve as in an ideal gas [286]. In particular in solvent mixtures, coupling to an ideal gas provides a powerful method for controlling the cosolvent concentration, since particles can more easily be inserted into the gas of noninteracting particles than into a dense liquid. Thereby, combined with the methodology presented in this work, one could, for example, study protein folding and unfolding or smart polymer swelling and collapse as a function of changes in the concentration of cosolvent/denaturants.

We want to point out several possible variations and improvements of the proposed methodology. In the previously described simulations of the folding peptide system, a spherical high-resolution region was associated with each peptide atom. The geometry of the overall atomistic region was updated during each simulation step as the peptide atoms moved. However, it may be sufficient to associate AT regions only with some of the peptide atoms, for example, only the heavy atoms or only the alpha carbons. Furthermore, the values of $r_{centr,i}$ do not necessarily need to be updated during every integration time step. For example, it could be sufficient to update them only when also the neighbor list is rebuilt. These variations in the implementation would significantly reduce the interprocess communication in parallel simulations required to inform all threads about the current shape of the AT region.

Finally, in future applications, it could be useful to construct atomistic regions from differently sized spheres. The methodology presented here can trivially be generalized to the case where different spheres have different values of r_{at}.

5.6. Conclusions

We have proposed and validated a scheme which allows the simulation of an arbitrarily shaped atomistic high-resolution region with a reservoir of coarse-grained particles. The particles can freely travel between the two regions, changing their resolution on-the-fly (adaptive resolution). As we have shown for the example of a folding peptide, the geometry of the high-resolution region can also adapt during the simulation to follow, for example, conformational changes of a solvated macromolecule, without perturbing structural and dynamical properties or the conformational transition itself. The method includes an appropriate application of the compensating force that is needed to cancel the thermodynamic imbalance that occurs when

the models used in the two subregions have different pressures. In conclusion, the approach generalizes the AdResS methodology and allows fully flexible and self-adjusting adaptive resolution simulations.

The use of the scheme will be advantageous in any system in which the region where a high-resolution model is required has a nonstandard changing geometry. This includes, for example, DNA strands, rugged surfaces during crystallization, fluctuating interfaces and membranes as well as aggregation processes.

5.7. Acknowledgments

K. Kreis is recipient of a fellowship funded through the Excellence Initiative (DFG/GSC 266). K. Kremer and A.C.F. acknowledge research funding through the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC Grant Agreement No. 340906-MOLPROCOMP. The authors thank Debashish Mukherji and Joe Rudzinski for a careful reading of the manuscript and their helpful suggestions.

Chapter 6.

Towards Quantum/Classical Adaptive Resolution Simulations

This chapter has been published as a research paper in The Journal of Chemical Theory and Computation. It is reprinted here with permission from the publisher.

Karsten Kreis, Mark E. Tuckerman, Davide Donadio, Kurt Kremer, and Raffaello Potestio
From Classical to Quantum and Back: A Hamiltonian Scheme for Adaptive Multiresolution Classical/Path-Integral Simulations
J. Chem. Theory Comput. 12, 3030-3039 (2016)
DOI: 10.1021/acs.jctc.6b00242

Abstract

Quantum delocalization of atomic nuclei affects the physical properties of many hydrogen-rich liquids and biological systems even at room temperature. In computer simulations, quantum nuclei can be modeled via the path-integral formulation of quantum statistical mechanics, which implies a substantial increase in computational overhead. By restricting the quantum description to a small spatial region, this cost can be significantly reduced. Herein, we derive a bottom-up, rigorous, Hamiltonian-based scheme that allows molecules to change from quantum to classical and vice versa on the fly as they diffuse through the system, both reducing overhead and making quantum grand-canonical simulations possible. The method is validated via simulations of low-temperature parahydrogen. Our adaptive resolution approach paves the way to efficient quantum simulations of biomolecules, membranes, and interfaces.

6.1. Introduction

Nuclear quantum delocalization plays a crucial role in low temperature systems, e.g. helium or hydrogen [76, 77, 85–87], which can undergo a superfluid transition, and it affects in nontrivial ways a large variety of processes at more standard thermodynamic conditions. This is the

case for proton transfer in biomolecules and membranes and in DNA oxidation [312–318], the thermodynamics of ice [319], the structure of water adlayers on catalysts [320, 321], the structure and dynamics of bulk water at room temperature [78, 88–90], and aqueous proton and hydroxide transport [80, 92–96].

In order to account for these effects in computer simulations, Feynman's path integral (PI) formulation of quantum statistical mechanics [52, 68, 85] can be employed, which enables an accurate description of quantum delocalization via Monte Carlo (MC) or Molecular Dynamics (MD) simulations [52, 85] albeit at an increased computational cost. A strategy to overcome this limitation is to restrict the PI description of the atoms to a (small) spatial region where quantum effects are important, and to model the remaining atoms as classical particles interacting *via* an appropriately chosen effective potential. Molecules diffusing across the boundary separating these two regions change their representation "on the fly" from classical to quantum and *vice versa*. This approach, which is obviously viable only for sufficiently short De Broglie wavelength, is beneficial particularly when the region that needs to be modeled quantum mechanically is small compared to the size of the full system. Examples of such systems include liquid-solid or liquid-liquid interfaces [322–324], aqueous solutions [80, 92–96], and proteins [81, 325, 326]. The simplified model in the classical region also allows the number of molecules in the system to change during the simulation [175, 228], thereby generating a grand-canonical distribution. More generally, an approach in which a classical and a PI model are concurrently used would allow a substantial computational gain and enable the simulation of larger systems for significantly longer sampling times.

A first step in this direction was taken in the framework of the Adaptive Resolution Simulation (AdResS) scheme [144, 176, 182] by merging quantum and classical effective forces [177, 212, 216, 240]. These works and subsequent advancements [224, 327] demonstrated the possibility of investigating the properties of a system of light particles by explicitly considering their quantum nature locally without affecting the overall thermodynamic balance between the quantum and classical regions. The AdResS scheme, however, is based on an interpolation of forces that does not admit a Hamiltonian structure [219]; the PI formulation, on the contrary, relies on a configurational potential energy. Hence, a theoretically rigorous formulation of a hybrid quantum-classical system should be based on a Hamiltonian function defined everywhere in the simulation domain.

In this paper, we provide a quantum-classical coupling protocol via a global Hamiltonian, allowing us to treat a system of particles quantum mechanically only in a restricted region of space and classically everywhere else. Additionally, the particles can freely diffuse across the simulation domain and switch the nature of their interactions according to their position in space. The validation of the proposed approach is carried out by means of Monte Carlo simulations.

We have organized the paper as follows: In Sec. 6.2, we review the classical Hamiltonian version of AdResS, denoted H–AdResS, and show how to extend it to a quantum system with variable mass. The quantization of such a system leads to a condition on the minimal size of

the hybrid region. In Sec. 6.3, we provide details of our validation studies, including system setup, free-energy compensation terms, a Monte Carlo algorithm for sampling the path-integral H–AdResS distribution, and specific simulations that demonstrate the performance of the method. In Sec. 6.4, we discuss the speedup of the new approach over a full path-integral calculation. In Sec. 6.5, we place the path-integral H–AdResS scheme in the context of, and compare it to, other methods designed to reduce the cost of path-integral calculations. Finally, in Sec. 6.6, we present a brief summary and conclude.

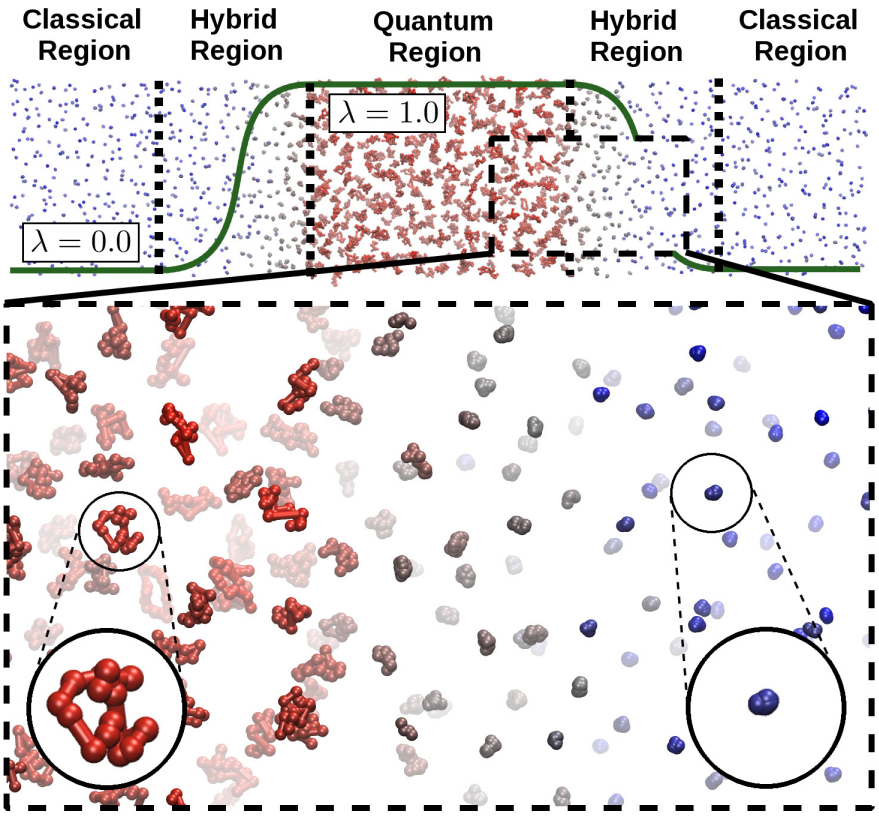

Figure 6.1.: Illustration of the simulation setup for the quantum–classical simulations. Red (resp. blue) color corresponds to a larger (resp. smaller) radius of gyration. The green line shows the interpolation function λ, smoothly changing from 0 in the classical region to 1 in the quantum region. The smooth transition from extended to collapsed molecules demonstrates the transition from quantum mechanical to classical behavior. The particles freely move between the regions and change their descriptions accordingly.

6.2. Methodology

Here we make use of the Hamiltonian AdResS (H–AdResS) method [181, 183, 266], which was developed to perform adaptive resolution MD/MC simulations based on a global Hamiltonian

and is thus the most suitable framework to correctly model each subsystem with the appropriate interactions. A typical H–AdResS system is partitioned in two regions connected *via* a hybrid buffer region. The resolution of a particle depends on the value of a representative coordinate $\mathbf{R}$ and is parametrized by a continuous function $\lambda(\mathbf{R})$ smoothly switching from 0 to 1 in the hybrid region (see Fig. 1). The total potential energy of each molecule is obtained by interpolating between the two resolutions. The H-AdResS Hamiltonian H of a system of point-like particles reads

$$H = \mathcal{K} + \sum_{\alpha=1}^{N} \left[\lambda_\alpha V_\alpha^1 + (1-\lambda_\alpha) V_\alpha^0 - \Delta H(\mathbf{R}_\alpha) \right] \tag{6.1}$$

where $\mathcal{K}$ is the kinetic energy, α indexes the N particles, and $\lambda_\alpha = \lambda(\mathbf{R}_\alpha)$. The single-particle potentials V_α^{Res} (with $Res = 0, 1$) are the sums of all intermolecular potentials acting on particle α, properly normalized so that double counting is avoided [181, 183]. The term ΔH, referred to as the Free Energy Compensation (FEC) [181, 183], is an external field acting in the hybrid transition region to eliminate the density imbalance that naturally occurs in such dual-resolution systems. In fact, different models of the same physical system exhibit a free energy difference that needs to be neutralized in order to enforce identical thermodynamical and/or structural properties (e.g. density) everywhere in the simulation domain. The FEC, whose calculation is described in section 6.3.2, aptly compensates for these free energy imbalances.

The incorporation of the H–AdResS Hamiltonian in the PI formalism is straightforward. The effective potential energy obtained from the PI quantization of the Hamiltonian in Eq. 6.1, assuming Boltzmann statistics, describes a set of N ring polymers, each containing P points or "beads" connected by harmonic springs of frequency $\omega_P = \sqrt{P}/\beta\hbar$, where $\beta = 1/k_B T$. The exact quantum behavior is recovered only for $P \to \infty$; however, sufficiently accurate results are recovered with finite values of P, which typically range from 16 to 48 beads for standard PI simulations [78, 317, 328–332]. Due to the H–AdResS setup, the interaction between the ring polymers changes adaptively in space according to the potential energy interpolation in Eq. 6.1. The quantum behavior, dictated by the strength of the springs connecting the beads of each ring, is nevertheless the same everywhere.

A strategy for switching between quantum and classical descriptions is to modify the masses of the atoms, as larger masses correspond to stiffer springs with constant $m\omega_P^2$; a large mass causes the ring polymers to collapse, and the particles approach their classical limit. We thus define $m \to \mu(\lambda) = \lambda m + (1-\lambda)M$, where $\mu(\lambda)$ smoothly switches from a mass $\mu(0) = M$ to a mass $\mu(1) = m \ll M$ and $\lambda = \lambda(\mathbf{R}_\alpha)$ with $\mathbf{R}_\alpha$ being the particle positions. When $\mu = m$, the physical mass, the particles are light, and the quantum zero-point motion becomes important. The mass M should be chosen large enough that the particles become essentially classical.

Assigning the masses in this way, which causes them to become position dependent, introduces the problem of quantizing a system with coordinate-dependent masses, and we turn to this problem now (as a reference, for the regular approach for constant-mass particles see, for example, Tuckerman [52]). Consider first the Hamiltonian operator for a particle of mass $\mu(x)$ in one dimension subject to a potential $V(x)$. The approach easily generalizes to N Boltzmann

particles in any number of dimensions. When the mass depends on x, we have the problem of representing the Hamiltonian as a Hermitian operator, which we can obtain by writing it in the following form:

$$\hat{\mathcal{H}} = \frac{1}{2}\hat{p}\mu^{-1}(\hat{x})\hat{p} + V(\hat{x}) \tag{6.2}$$

where $\mu^{-1}(\hat{x})$ is the inverse mass operator, and $\hat{p}$ is the momentum operator. Using this Hamiltonian, we seek to formulate the partition function $Q = \mathrm{Tr}[\exp\{-\beta\hat{\mathcal{H}}\}]$ as a path integral. Introducing the usual set of $P-1$ resolutions of the identity operator and the Trotter factorization of the Boltzmann operator, we can write the trace in the coordinate basis as

$$\begin{aligned} Q &= \lim_{P\to\infty} \int dx_1 \cdots dx_P \prod_{k=1}^{P} \langle x_k | \exp\left(-\frac{\beta}{P}V(\hat{x})\right) \exp\left(-\frac{\beta}{2P}\hat{p}\mu^{-1}(\hat{x})\hat{p}\right) |x_{k+1}\rangle \Big|_{x_{P+1}=x_1} \\ &= \lim_{P\to\infty} \int dx_1 \cdots dx_P \prod_{k=1}^{P} \exp\left(-\frac{\beta}{P}V(x_k)\right) \langle x_k | \exp\left(-\frac{\beta}{2P}\hat{p}\mu^{-1}(\hat{x})\hat{p}\right) |x_{k+1}\rangle \Big|_{x_{P+1}=x_1} \end{aligned} \tag{6.3}$$

To derive the matrix elements in Eq. 6.3, we introduce the momentum identity resolution:

$$\langle x_k | \exp\left(-\frac{\beta}{2P}\hat{p}\mu^{-1}(\hat{x})\hat{p}\right) |x_{k+1}\rangle = \int_{-\infty}^{\infty} dp \, \langle x_k|p\rangle \, \langle p| \exp\left(-\frac{\beta}{2P}\hat{p}\mu^{-1}(\hat{x})\hat{p}\right) |x_{k+1}\rangle \tag{6.4}$$

Given that the limit $P \to \infty$ is ultimately taken, we can work with an infinitesimal version of the exponential operators by expanding the exponential to first order. Thus, we obtain

$$\langle p| \exp\left(-\frac{\beta}{2P}\hat{p}\mu^{-1}(\hat{x})\hat{p}\right) |x\rangle \approx \langle p| \left(1 - \frac{\beta}{2P}\hat{p}\mu^{-1}(\hat{x})\hat{p}\right) |x\rangle \tag{6.5}$$

Now, we introduce the commutator $[\mu^{-1}(\hat{x}), \hat{p}]$ and write

$$\begin{aligned} \mu^{-1}(\hat{x})\hat{p} &= \hat{p}\mu^{-1}(\hat{x}) + [\mu^{-1}(\hat{x}), \hat{p}] \\ &= \hat{p}\mu^{-1}(\hat{x}) + i\hbar\frac{d\mu^{-1}}{d\hat{x}} \end{aligned} \tag{6.6}$$

Substituting Eq. 6.6 into Eq. 6.5 yields

$$\begin{aligned} \langle p| \left(1 - \frac{\beta}{2P}\hat{p}^2\mu^{-1}(\hat{x}) - \frac{i\hbar\beta}{2P}\hat{p}\frac{d\mu^{-1}}{d\hat{x}}\right) |x\rangle &= \langle p|x\rangle \left(1 - \frac{\beta p^2}{2P}\mu^{-1}(x) - \frac{i\hbar\beta}{2P}p\frac{d\mu^{-1}}{dx}\right) \\ &\approx \langle p|x\rangle \exp\left[-\frac{\beta}{2P}\left(p^2\mu^{-1}(x) + i\hbar p\frac{d\mu^{-1}}{dx}\right)\right] \end{aligned} \tag{6.7}$$

where the operators are now replaced by the corresponding eigenvalues. Substituting Eq. 6.7

into Eq. 6.4 gives

$$\begin{aligned}
&\langle x_k| \exp\left(-\frac{\beta}{2P}\hat{p}\mu^{-1}(\hat{x})\hat{p}\right)|x_{k+1}\rangle \\
&\quad = \int_{-\infty}^{\infty} dp\, \langle x_k|p\rangle \langle p|x_{k+1}\rangle \exp\left[-\frac{\beta}{2P}\left(p^2\mu^{-1}(x_{k+1}) + i\hbar p \frac{d\mu^{-1}}{dx}\Big|_{x_{k+1}}\right)\right] \\
&\quad = \left(\frac{\mu(x_{k+1})P}{2\pi\beta\hbar^2}\right)^{\frac{1}{2}} \exp\left\{-\frac{\beta\mu(x_{k+1})P}{2(\beta\hbar)^2}\left[(x_k - x_{k+1}) - \frac{\beta\hbar^2}{2P}\frac{d\mu^{-1}}{dx}\Big|_{x_{k+1}}\right]^2\right\}
\end{aligned} \tag{6.8}$$

for the matrix elements in Eq. 6.3, where the last equality has been obtained by introducing the matrix elements $\langle x|p\rangle = \exp(ipx/\hbar)/\sqrt{2\pi\hbar}$ and performing the momentum integration by completing the square.

From Eq. 6.8, we see that the inverse mass derivative term can be neglected if the following condition holds:

$$\left|\left(\frac{d\mu^{-1}}{dx}\right)_{x_{k+1}}\right| \ll \frac{2P\Delta x_{k,k+1}}{\beta\hbar^2} \tag{6.9}$$

where we defined $\Delta x_{k,k+1} = |x_k - x_{k+1}|$. Since,

$$\left|\left(\frac{d\mu^{-1}}{dx}\right)_{x_{k+1}}\right| = \left|\frac{1}{\mu^2(x_{k+1})}\left(\frac{d\mu}{dx}\right)_{x_{k+1}}\right| \tag{6.10}$$

the condition becomes

$$\left|\left(\frac{d\mu}{dx}\right)_{x_{k+1}}\right| \ll \frac{2\Delta x_{k,k+1}}{\Lambda_\mu^2(x_{k+1})}\mu(x_{k+1}) \tag{6.11}$$

using the definition $\Lambda_\mu(x) \equiv \sqrt{\beta\hbar^2/(P\mu(x))}$. Since

$$\langle \Delta x\rangle \equiv \sqrt{\langle \frac{1}{P}\sum_{l=1}^{P}\Delta x_{l,l+1}^2\rangle} = \Lambda_\mu\sqrt{(P-1)/P} \approx \Lambda_\mu \tag{6.12}$$

for a free ring of constant mass μ and typical values of P, we can approximate $\Delta x_{k,k+1} \approx \Lambda_\mu(x_{k+1})$ and write

$$\left|\frac{d\mu(x)}{dx}\right| \ll \frac{2\mu(x)}{\Lambda_\mu(x)} \tag{6.13}$$

for an arbitrary position x.

The inequality in Eq. 6.13 places a lower bound on the width of the hybrid region (where $\lambda(x) \neq$ const): the interpolation must be sufficiently smooth so that the mass derivative can be safely neglected.

The derivation and the criterion generalize to an arbitrary number N of interacting particles in three dimensions. In fact, typical potentials do not significantly change the radius of gyration

of the ring polymers; additionally, it is necessary to consider only the components of the bead-bead distances parallel to the direction in which the mass changes. The size of the hybrid region compatible with Eq. 6.13 is still generally small (1-2 nm in our simulations for parahydrogen at 20 K, even smaller at room temperature), thus having no significant impact on the computational overhead.

Concluding, introducing the H–AdResS potential energy we obtain the following partition function for N interacting Boltzmann particles in three dimensions:

$$Q = \lim_{P\to\infty} \left[\prod_{k=1}^{P} \prod_{\alpha=1}^{N} \left(\frac{mP}{2\pi\beta\hbar^2} \right)^{\frac{3}{2}} \int \mathrm{d}\mathbf{r}_{\alpha,k} \right] e^{-\beta V_P^{\mu}} \tag{6.14}$$

with

$$\begin{aligned} V_P^{\mu} &= \sum_{k=1}^{P} \sum_{\alpha=1}^{N} \left\{ \frac{\mu_{\alpha,k}\,\omega_P^2}{2} |\mathbf{r}_{\alpha,k} - \mathbf{r}_{\alpha,k+1}|^2 - \frac{3}{2\beta} \log \frac{\mu_{\alpha,k}}{m} \right. \\ &+ \left. \frac{1}{P} \left[\lambda_{\alpha,k} V_{\alpha,k}^1 + (1-\lambda_{\alpha,k}) V_{\alpha,k}^0 - \Delta H(\mathbf{r}_{\alpha,k}) \right] \right\} \end{aligned} \tag{6.15}$$

and $\mu_{\alpha,l} = \mu(\mathbf{r}_{\alpha,l})$, where the representative coordinates $\mathbf{R}_{\alpha,k}$ in the functions $\lambda_{\alpha,k}$ are the beads positions $\mathbf{r}_{\alpha,k}$. In Eq. 6.15, the position-dependent prefactor has been explicitly introduced in the potential V_P^{μ} as a logarithmic function of the bead masses, so that it can be treated as a conventional energy term and fully removed from the Hamiltonian by means of the FEC function ΔH in Eq. 6.1 [333]. The light mass m has been used as the reference mass scale. The ring polymers described by the energy function V_P^{μ} (Eq. 6.15) are expanded in the region where the mass is small and collapse to nearly classical point-like particles in the large-mass region.

6.3. Validation

To validate the proposed quantum-to-classical coupling scheme, adaptive path integral MC simulations of liquid parahydrogen at 20 K with $P = 16$ are performed. Liquid hydrogen between 14 K and 25 K exhibits pronounced quantum behavior [76, 77, 334], thus it constitutes a probing test case. A Trotter number of $P = 16$ is likely not large enough for full convergence of physical observables in this system under the aforementioned conditions, however, our focus is less on the exact properties of parahydrogen and more on the validation of the proposed path-integral H–AdResS approach.

6.3.1. System setup

We consider a system composed of 1458 hydrogen molecules in a slab of dimensions 11.0 nm × 2.5 nm × 2.5 nm (molecular density 28.4 $\mathrm{cm}^3/\mathrm{mol}$) with periodic boundary conditions in all directions. The width of the low-mass quantum region is set to $d_{\mathrm{QM}} = 3.0$ nm and the thickness

of each hybrid transition region is $d_{\mathrm{HY}} = 2.0\,\mathrm{nm}$. In order to assign to a bead its position-dependent resolution λ, its distance from the boundary between the quantum and the hybrid region is computed, *i.e.*, $\delta x_{\alpha,k} = |x_{\alpha,k}| - d_{\mathrm{QM}}/2$, where $x_{\alpha,k}$ denotes the X coordinate of the bead in a coordinate system with its origin at the center of the simulation box. This quantity is then employed in the resolution function $\lambda(\delta x_{\alpha,k})$. This function is given generically by

$$\lambda(x) = \begin{cases} 1 & : x \leq 0 \\ \cos^2\left(\frac{\pi}{2}\,\frac{x}{d_{\mathrm{HY}}}\right) & : 0 < x < d_{\mathrm{HY}} \\ 0 & : x \geq d_{\mathrm{HY}} \end{cases} \tag{6.16}$$

The mass m is set to the molecular hydrogen mass $m_{\mathrm{H}_2} = 2.001\,au$. In the classical region the increased mass is chosen as $M = 100\,m_{\mathrm{H}_2}$. In the quantum (QM) region we employ the Silvera-Goldman potential [335, 336] with a cutoff at 0.9 nm for the intermolecular interaction potential V^1, while in the classical (CL) region we make use of a shifted, purely repulsive Weeks-Chandler-Andersen (WCA) potential [274]:

$$V^0(r) = \begin{cases} 4\epsilon\left[\left(\frac{\sigma}{r-r_0}\right)^{12} - \left(\frac{\sigma}{r-r_0}\right)^{6} + \frac{1}{4}\right] & : r \leq R_{\mathrm{c}} \\ 0 & : r > R_{\mathrm{c}} \end{cases} \tag{6.17}$$

where $r = |\boldsymbol{r}_{\alpha,k} - \boldsymbol{r}_{\beta,k}|$, $(\alpha \neq \beta)$ denotes the distance between beads of the same imaginary time slice k in different molecules α and β. Furthermore, we choose $\epsilon = 1.0\,\mathrm{kJ/mol}$, $\sigma = 0.14\,\mathrm{nm}$, and $r_0 = 0.15\,\mathrm{nm}$. The cutoff is given by $R_{\mathrm{c}} = 2^{\frac{1}{6}}\sigma + r_0$. The two potentials are graphically presented in Fig. 6.2. The WCA potential in the classical region is not interpreted as a classical model for low-temperature parahydrogen. Rather, it is parametrized only to reproduce approximately the hard-core radius of the reference quantum particles and to serve as a crude model of particles occupying a reservoir with which the quantum region can exchange matter. We purposely avoid fitting the classical potential to the structure of the reference to demonstrate the generality of the protocol: indeed, the potential could be chosen arbitrarily, thereby allowing various options and applications. One could, for example, focus on reproducing specific properties in the classical region, and/or employ a computationally advantageous model such as an ideal gas of non-interacting particles [286].

The mass of the particles in the CL region is 100 times larger than that of the molecules in the QM region and the chosen set of parameters also satisfies Eq. 6.13. Finally, we stress that in the CL region the WCA interaction between ring polymers is computed only using the center of mass of the ring, thus gaining an effective reduction of the computational cost. This simplification is allowed by the essentially point-like structure of the rings in the CL regions, as can be seen from the radius of gyration profile (Fig. 6.3). The number of computations per pair of molecule is reduced from $P = 16$ to one. A snapshot of the simulations is shown in Fig. 6.1 which shows the gradual change in size of the ring polymers indicating the transition between classical and quantum mechanical behavior.

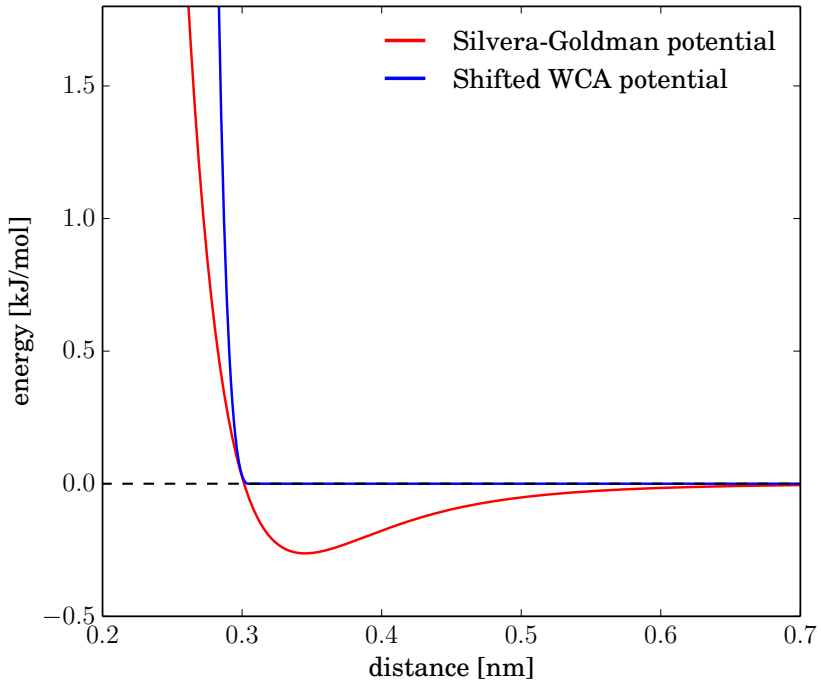

Figure 6.2.: Non-bonded intermolecular interaction potentials used in the adaptive quantum–classical simulations. The red curve is the Silvera Goldman potential, which is employed in the low-mass quantum region. The blue curve shows the shifted WCA potential, which is used in the high-mass classical region.

6.3.2. Free energy compensation

To modulate the thermodynamic imbalance between the classical high-mass and the low-mass quantum subsystems, a FEC is applied [181, 183]. First, we compute the compensation term $\Delta H_{\mathrm{KTI}}(\lambda)$ by Kirkwood thermodynamic integration (KTI) [230] of a smaller system of 360 molecules in a box with dimensions 2.570 nm $\times$ 2.570 nm $\times$ 2.570 nm. However, using only the correction term derived by KTI is insufficient to reach a flat density profile (see Fig. 6.4). The interpolation between subsystems with very different masses and intermolecular interaction potentials leads to large thermodynamic imbalances and strong forces pushing molecules from one region to the other. The KTI-based FEC alone is incapable of sufficiently correcting these imbalances between the classical and the quantum domains because it only provides a mean field estimate of the necessary compensation. This becomes especially clear when comparing the density profile obtained when applying only the KTI-based FEC with the radius of gyration profiles (see Fig. 6.3). Indeed, the density profile is mostly distorted in the area where the radius of gyration of the ring polymers changes more steeply, that is, where the local environment of molecules changes sharply. Therefore, in order to remove also the remaining fluctuations in the obtained density profile after applying the KTI-based FEC, an iterative approach similar to the

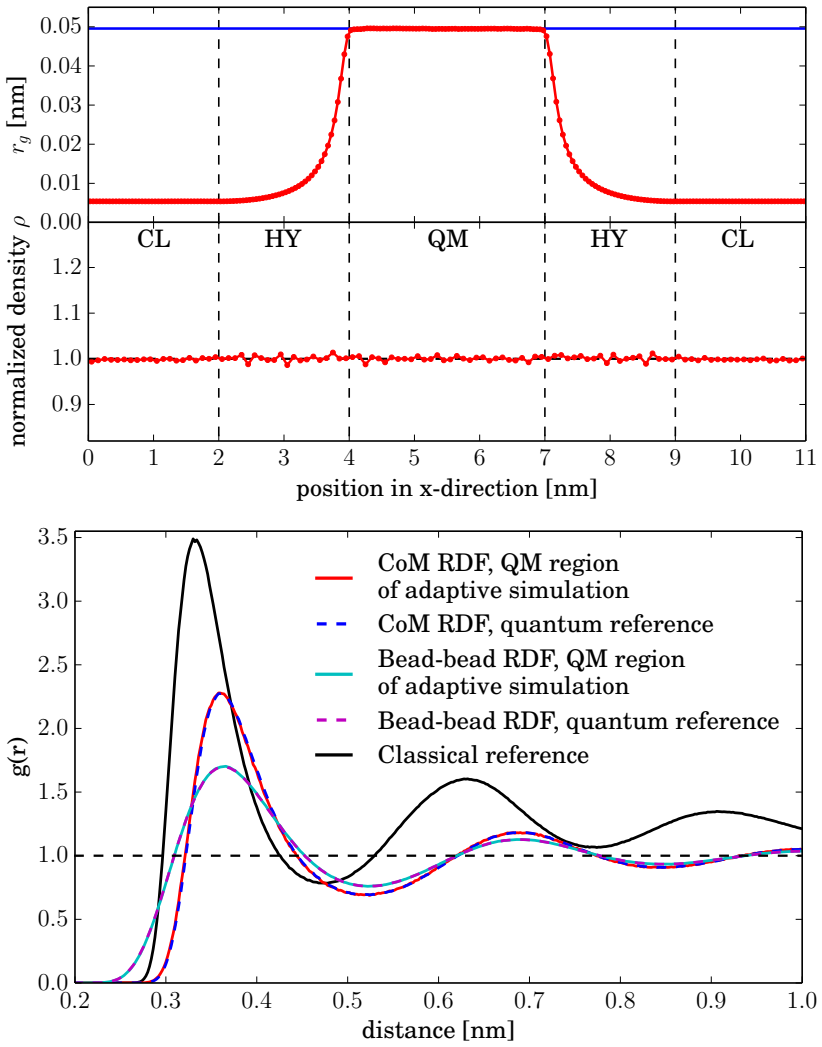

Figure 6.3.: Top: Radius of gyration r_g and the normalized density ρ profiles. Blue corresponds to the radius of gyration of molecules in a fully quantum reference simulation. Bottom: Ring polymers' center of mass (CoM) and bead-bead RDF's of the quantum-to-classical simulation calculated in the quantum region and of a fully quantum reference simulation. The black curve is the RDF of a fully classical ($P = 1$) system of particles interacting *via* the Silvera-Goldman potential.

one devised by Fritsch et al. is employed [182]. The scheme reads

$$\Delta H^{i+1}(x) = \Delta H^{i}(x) - \Delta h^{i}(x) \tag{6.18}$$

with $\Delta H^{0}(x) = \Delta H_{\mathrm{KTI}}(x)$, the initial correction term obtained by Kirkwood thermodynamic integration converted from a function of resolution into a function of position.

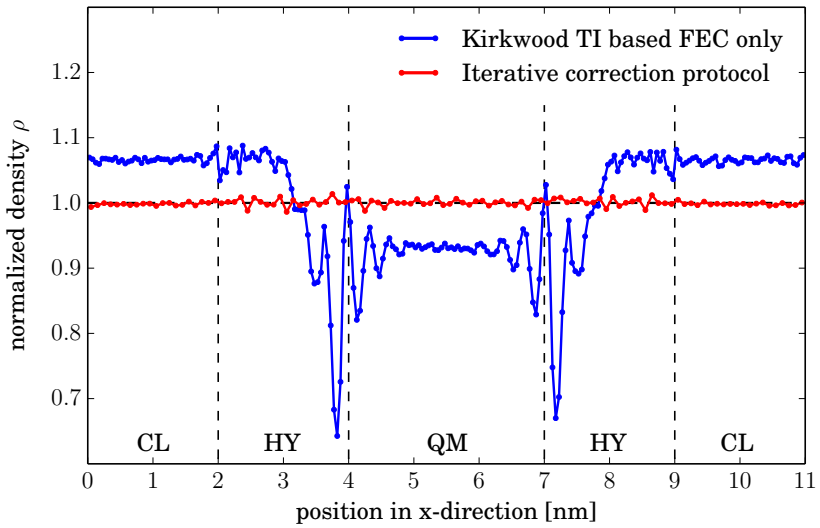

Figure 6.4.: Normalized density profiles obtained when applying a correction field in the hybrid region based on Kirkwood thermodynamic integration only (blue) and when applying the iterative protocol described in the text (red).

Furthermore,

$$\Delta h^i(x) = \begin{cases} -\frac{1}{\beta}\ln\{\rho^i(x_0)\} & : x \leq x_0 \\ -\frac{1}{\beta}\ln\{\rho^i(x)\} & : x_0 < x < x_1 \\ -\frac{1}{\beta}\ln\{\rho^i(x_1)\} & : x \geq x_1 \end{cases} \tag{6.19}$$

where

$$x_0 = \frac{d_{\mathrm{QM}}}{2} - 0.5\,\mathrm{nm} \tag{6.20}$$

$$x_1 = \frac{d_{\mathrm{QM}}}{2} + d_{\mathrm{HY}} + 0.5\,\mathrm{nm} \tag{6.21}$$

and x is measured as the absolute distance from the center of the simulation box. $\rho^i(x)$ denotes the density profile obtained when applying the FEC $\Delta H^i(x)$ of iteration i (therefore, $\rho^0(x) = \rho^{\mathrm{KTI}}(x)$, where $\rho^{\mathrm{KTI}}(x)$ corresponds to the initial density profile obtained from simulations in which only the KTI-based FEC term is applied). The term $\Delta h^i(x)$ acts in an extended hybrid region, defined as the regular hybrid region plus additional 0.5 nm both within the quantum and the classical region. As we have two hybrid transition regions in the presented simulation setup above, averages over both regions are used to derive one consistent and symmetric correction, which is applied in the same way in both regions.

The protocol converges by construction when a flat density profile is achieved. However, to speed up the convergence, on top of the iterative density based correction, we also add a linear

ramp term

$$\Delta h^i_{\text{ramp}}(x) = \begin{cases} 0 & : x \leq x_0 \\ -(a_i/\beta) \times (x - x_0)/(x_1 - x_0) & : x_0 < x < x_1 \\ -(a_i/\beta) & : x \geq x_1 \end{cases} \tag{6.22}$$

for some of the later iterations when the overall density difference between the two region has become small. Note that this additional term is not strictly required but can help to speed up the iterative procedure. The values of the coefficients a_i that have been employed here to favor convergence are reported in the following section.

The protocol is executed in an iterative fashion until a sufficiently flat density profile is obtained (see Fig. 6.4). Note that the converged profile is also significantly smoother and more uniform than the ones obtained with a related approach by Agarwal and Delle Site [216].

6.3.3. Monte Carlo sampling

To sample the system's configurational space we employ a standard Metropolis Monte Carlo algorithm [52]. For the Kirkwood TI of the small system we run 16 simulations with 10^5 sweeps each. The λ parameter increases linearly every sweep by 10^{-5}. The results are averaged after the simulations. Employing the Kirkwood TI FEC term thus obtained we then run 14 iterations of simulations applying the protocol set out above to refine the density profile. The parameters a_i for the ramp term in the correction protocol are chosen as $a_7 = 0.06$, $a_8 = 0.04$, $a_9 = 0.05$, $a_{10} = 0.05$, $a_{11} = 0.02$, $a_{12} = 0.03$ and $a_{14} = 0.02$. For all other iterations no ramp term is added, *i.e.*, $a_i = 0$. Each iteration consists of 32 parallel simulations, each of these running $1.5 \cdot 10^4$ equilibration sweeps and another $7.5 \cdot 10^3$ sweeps during which the density profile is measured. Also here, after each iteration the results are averaged. Having reached a sufficiently smooth density profile, we then utilize the FEC from the Kirkwood TI and the iterative protocol to perform the main production simulations. For these we perform 32 simulations in parallel, each running $1.5 \cdot 10^4$ sweeps after another $1.5 \cdot 10^4$ equilibration sweeps. Afterwards, the results (*i.e.*, the radial distribution functions (RDF's), the density profiles, the radius of gyration profiles as well as the particle number histograms) are once again averaged over all simulations.

Each sweep is constituted by $N \times P$ attempted Monte Carlo moves on randomly chosen molecules, with N being the total number of molecules and P the Trotter number ($N = 1458$ and $P = 16$ in the production run simulations). Three different kinds of moves are randomly performed:

Whole molecule displacements: The chosen molecule is displaced as a whole by moving its center of mass. The direction is chosen randomly from a uniform spherical distribution and the distance is drawn from a Gaussian distribution with zero mean and width σ_{CoM}.

Molecule rotations: The chosen molecule is rotated as a whole around a randomly oriented axis passing through its center of mass. The angle is chosen randomly from a Gaussian distribution with zero mean and width σ_{rot}.

Individual Trotter-bead moves: An individual bead of the molecule is randomly chosen and displaced. The direction is chosen randomly from a uniform spherical distribution and the distance is drawn from a Gaussian distribution with zero mean and width σ_{bead}.

The different values for the σ_i's of all simulations are presented in Tab. 6.1. In the adaptive resolution simulation, they are chosen such that they result in adequate acceptance ratios for the moves both in the classical high-mass as well as in the quantum low-mass region. When picking a molecule for a Monte Carlo move the probabilities for performing whole molecule displacements or molecule rotations are $\frac{1}{13}$ each while the probability for Trotter-bead moves is $\frac{11}{13}$. This choice leads to a convenient balance between whole molecule motions and Trotter-bead fluctuations.

Simulation	σ_{CoM}	σ_{rot}	σ_{bead}
Kirkwood TI	0.1 nm	0.5 rad	0.03 nm
Adaptive Simulation	0.1 nm	0.5 rad	0.03 nm
QM Reference	0.1 nm	0.5 rad	0.07 nm
Classical Reference	0.1 nm	-	-

Table 6.1.: Widths of the Gaussian distributions employed to draw the random displacements and rotations from for the Kirkwood Thermodynamic Integration, for the reference simulations as well as for the adaptive quantum-classical simulations.

6.3.4. Reference simulations

In order to evaluate the results of the adaptive quantum-classical simulations, we perform full-quantum as well as full-classical reference simulations of liquid parahydrogen for comparison. The same setup is used for the adaptive simulations, and likewise the temperature is set to $T = 20\,\text{K}$ and the Silvera-Goldman potential is employed. For the full-quantum simulations we choose $P = 16$ as in the adaptive simulations while the classical simulations are performed with $P = 1$. In both cases, 32 simulations are run in parallel, each one for $1.5 \cdot 10^4$ equilibration sweeps and another $1.5 \cdot 10^4$ production sweeps during which measurements are performed. Afterwards the results are averaged.

The values used for the σ_i's in the reference simulations are presented in Tab. 6.1. In the classical simulations, all moves are as described above with the obvious exception of bead and rotating moves, which do not exist for individual classical particles.

The calculations of the particle number fluctuations in the QM simulations are performed in exactly the same fashion and using the same area as for the adaptive simulations. The RDF's of the classical and QM reference simulations are calculated in the complete simulation domains.

6.3.5. Results

The radius of gyration of the ring polymers in the quantum region, shown in Fig. 6.3, perfectly reproduces that of a corresponding full quantum simulation, while in the CL region it drops by $\approx 90\%$, indicating nearly classical behavior. At the same time, the FEC enforces a uniform density profile everywhere in the system. A quantitative measure of the fluid structure is provided by the ring center of mass and bead-bead (intermolecular bead pairs with the same Trotter index) RDF's. In spite of the remarkable differences between the quantum fluid and the classical model, the RDF's measured in the QM region perfectly match those from the full quantum reference simulation.

Furthermore, the particle number fluctuations in the inner quantum region (the subdomain of the QM region where particles are 1.0 nm far from the hybrid region, that is, a distance larger than the 0.9 nm cutoff) are calculated and compared to the same quantity in the full quantum simulations (see Fig. 6.5). The fluctuations in the adaptive simulation match the reference

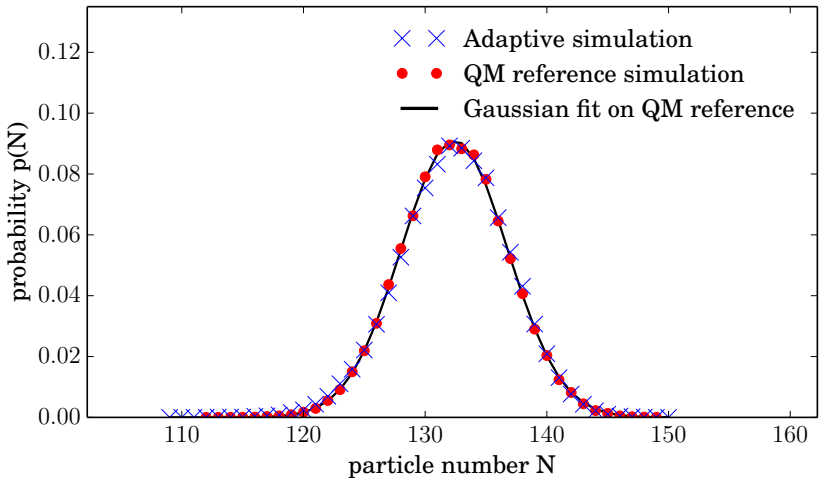

Figure 6.5.: Particle number probability distribution of the inner part of the QM region in the adaptive simulation (blue crosses) and of the same volume in a reference full quantum simulation (red points) including a Gaussian fit on the latter (black curve, $\mu = 132.4$, $\sigma = 4.383$).

closely. This means that the quantum subdomain exchanges particles with its surrounding in the same fashion as it would if embedded in a full quantum system. Therefore, by coupling to a sufficiently large particle reservoir, the scheme can be employed to reliably and rigorously simulate the quantum grand-canonical ensemble with only little additional computational cost due to the numerically efficient classical particle reservoir.

6.4. Speedup over full-quantum simulations

As mentioned earlier, in the proposed quantum-to-classical coupling scheme, interactions in the classical region do not need to be calculated P times (with P being the Trotter number). Rather, because of the collapse of the polymer rings they are computed only once between the centers of mass of the (quasi point-like) rings. Additionally, a numerically simpler potential with a shorter cutoff can be used in the classical region. In general, the overall speedup of the simulations strongly depends on the details of the implementation of the algorithm and is therefore platform dependent. For example, if there is a high overhead in the code, the overall computational gain by more efficient potential energy calculations will be small. If the program spends most of its time with these calculations, a significant improvement is possible.

The expected speedup S for the energy calculations of short range intermolecular potentials, compared to full quantum simulations, can be estimated as

$$S \sim \frac{PN}{P\langle N_{\mathrm{QM}}\rangle + c_{\mathrm{CLpot}}\langle N_{\mathrm{CL}}\rangle} \tag{6.23}$$

where N is the total number of particles in the system, $\langle N_{\mathrm{QM}}\rangle$ is the average number of particles in the QM and hybrid regions, $\langle N_{\mathrm{CL}}\rangle$ is the average number of particles in the CL and hybrid regions and c_{CLpot} is a constant depending on the molecular interaction potential in the CL region. If the same potentials are used in the QM and CL subsystems, $c_{\mathrm{CLpot}} = 1$, while a computationally cheaper potential in the CL region would result in a smaller c_{CLpot} with the extreme being a non-interacting ideal gas, in which case $c_{\mathrm{CLpot}} = 0$. Hence, the speedup would approach $S_{\mathrm{max}} = N/\langle N_{\mathrm{QM}}\rangle$, its upper bound, in a good implementation of the scheme with a simple interaction in the classical region. Therefore, in simulation setups featuring a small quantum-to-total volume ratio, e.g. employing a small spherical quantum region in a large, three-dimensional box, the computational gain could be close to S_{max}. Such a situation could, for example, occur in studies of proton transfer reactions in membranes, where only a small part of the membrane needs to be described quantum mechanically, while the remainder of the system could be treated via a simpler classical model. Similarly, in simulations of enzymes, it is likely sufficient to treat a large region of the molecule classically in a computationally efficient manner, while quantum nuclear effects can be taken into account at the active site where they play a significant role [81]. The method also enables efficient sampling of a quantum grand-canonical ensemble by embedding a QM subsystem in a particle reservoir, which could be significantly larger than the QM domain in order to allow for realistic particle fluctuations in the latter. At the same time, the structure and dynamics of the reservoir, being unimportant, could be described by a much coarser and numerically more efficient classical model. In all such cases, an adaptive quantum-classical coupling scheme, as proposed in this work, would provide a significant advantage.

To demonstrate the improved computational efficiency we perform 4 sets of simulations for different box sizes with each set consisting of a full-quantum and an adaptive simulation in which the quantum region as well as both hybrid regions have a width of 2.0 nm each. The total

box sizes as well as the corresponding molecule numbers for the simulations are presented in Tab. 6.2. The temperature and density are the same as before. In all cases but the first the classical regions are significantly larger than the quantum ones. For each setup, we perform 8 simulations and average the results. Furthermore, the simulations are run for 500 sweeps, and in the case of the adaptive simulations the previously derived FEC is applied. The set of Monte Carlo moves is chosen for both the quantum and the adaptive simulations to be the one previously used for the adaptive simulations.

Number of molecules	L_x	L_y	L_z
1325	10.0 nm	2.5 nm	2.5 nm
3974	30.0 nm	2.5 nm	2.5 nm
6623	50.0 nm	2.5 nm	2.5 nm
9272	70.0 nm	2.5 nm	2.5 nm

Table 6.2.: Number of molecules and box geometries for the different sets of simulations for the calculation of the computational gain of adaptive quantum-classical over full-quantum simulations.

In order to obtain platform-independent results, we only measure the time our code spends with potential energy calculations. These times are plotted in Fig. 6.6. Additionally, the corresponding speedups, defined as $T_{\text{quantum}}/T_{\text{adaptive}}$ with T_{adaptive} (T_{quantum}) being the time spent for the energy calculations in the adaptive (quantum) simulations, are presented. Furthermore, we plot $V_{\text{total}}/V_{\text{QM}}$, the ratio of the total box volume and the volume of the quantum plus the hybrid regions. In the latter part, all quantum mechanical energy calculations associated with the ring polymers are performed explicitly. This quantity is essentially the speedup derived above since $V_{\text{total}}/V_{\text{QM}} \approx N/\langle N_{\text{QM}}\rangle$.

It can be seen that the adaptive simulations are significantly faster than their corresponding fully quantum counterparts. For the largest box, the energy calculations in the adaptive quantum-classical simulations are faster by a factor of ≈ 9 than the full-quantum simulations. Furthermore, it can be seen that the time required for the adaptive simulations stays nearly constant for the different box sizes. The reason for this is that the computational cost of the interactions between classical molecules is negligible compared to the time required for the computation of the potential energies in the quantum region. Therefore, as shown in the graph, the speedup in our simulations is indeed close to its upper bound.

6.5. Relation to other path integral methods

In this section, we describe the present methodology in the context of, and in relation to, other established techniques that also reduce the computational effort of traditional path integral simulations.

As a first observation we recall from the Introduction that the first attempts to couple a

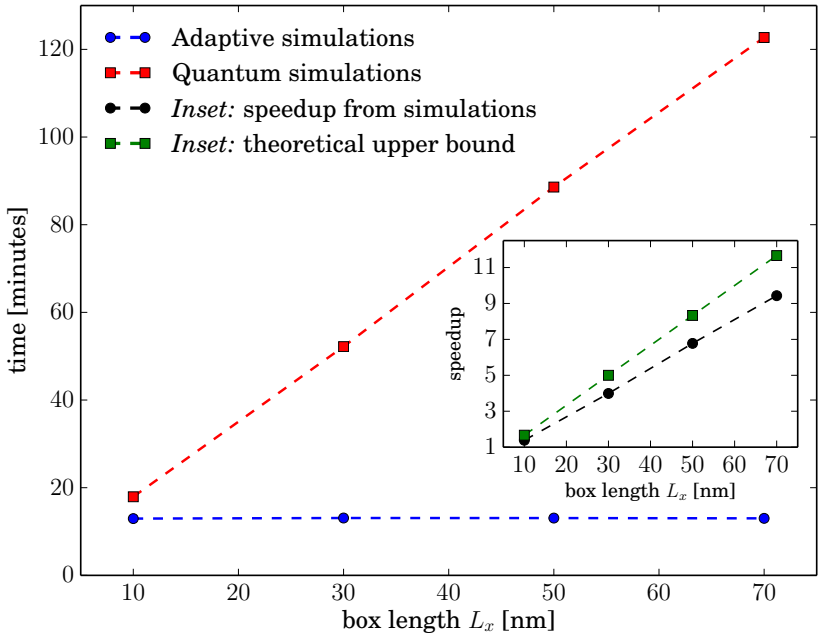

Figure 6.6.: Main figure: Simulation times required for the potential energy calculations in quantum and adaptive simulations for slab systems of four different box lengths L_x. Inset: Speedup of the adaptive quantum-classical simulations as well as the upper bound on the speedup. The linear increase in the speedup follows the increase in size of the full box (at constant QM region size). If all box sides are varied in a setup where the QM region is spherical, one would expect a cubic curve. The lines are a guide to the eye.

PI and a classical model of a liquid were performed in the framework of the force-based, non-Hamiltonian AdResS method [177, 212, 216, 240]. This scheme, which is introduced *after* the system has already been quantized, provides a pragmatic coupling of the two descriptions. The methodology presented here represents an advance in the classical/path integral hybrid methodology as it stems from a unique Hamiltonian function that governs the entire system and couples the different models *before* quantization is performed. This enables a more fundamental description of the setup and a more consistent treatment of the various parts of the system. Furthermore, the scheme presented can be seamlessly employed in tandem with other methods specifically developed to reduce the computational cost of path integral simulations, as will be highlighted below.

Consider, first, the ring polymer contraction scheme by Markland and Manolopoulos [337, 338]. This approach leads to a significant decrease in numerical complexity by contracting low-frequency non-bonded and long-range interactions to only a few imaginary time slices or even a single time slice. For example, long-range interactions in PI simulations can be safely evaluated only on the ring polymer centroids, while still yielding highly accurate results [338]. Since in our methodology the centroid is well defined in both the classical and quantum regions,

the computation of long-range forces could then be seamlessly carried out at the cost of a classical simulation on the whole system. More recently, it was shown that the ring polymer contraction scheme can also be extended to *ab initio* density functional theory simulations [339, 340]. We note that reactive systems often require the use of electronic structure methods in the reactive region, and these can be combined with path integral approaches, as was done by Wang et al. [81]. Combining path integral H–AdResS with such an approach would require an *ab initio* version of the scheme, which has not yet been developed but is the subject of future work. We, therefore, defer the discussion of QM/MM approaches within H–AdResS for such a future publication.

A closely related method to the ring polymer contraction scheme is the mixed time slicing approach [341], in which different degrees of freedom, i.e. particles, are modeled with a different number of Trotter beads in order to incorporate nuclear quantum effects only for the relevant light atoms. This technique could also be combined with our work resulting, for example, in a scheme where different Trotter bead numbers are used for different atoms in the QM region while in the CL domain all path integrals would collapse and behave completely classically.

Another method for performing efficient path integral simulations was introduced by Ceriotti et al., who showed that carefully parametrized colored noise thermostats can speed up the convergence in path integral simulations by alowing the use of significantly fewer required beads, thus reducing the overall computational cost [342–345]. This scheme is compatible with our approach as well and could be utilized to reduce the computational complexity further within the quantum subregion. Finally, we mention higher order Trotter factorizations, [346] either directly sampled via Monte Carlo [347], or combined with *a posteriori* reweighting of averages [348, 349] or a cumulant expansion [350] within path integral molecular dynamics. When a fourth order factorization scheme is employed, P can be generally reduced by a factor of four while maintaining the same level of accuracy. It is important to note, however, that in dual resolution simulation approaches, this computational speedup would add to the advantage of allowing short-range or even no interactions in the low-resolution (the CL or, more generally, the coarse-grained) region, which, by itself, leads to substantial savings.

A feature that differentiates our method from those mentioned here is that it is *local*, in that it gains its overall speedup by restricting the expensive quantum treatment to a limited region in space. The other approaches, on the other hand, are *global* in the sense that they reduce the computational cost associated with the path integral computations themselves in a full path integral simulation setup. Therefore, the proposed method can be regarded as complementary to these techniques since any of the schemes that we discussed in this section can be used to treat the path-integral region in the path integral H–AdResS approach. This would reduce the overall computational complexity even further, which could prove especially useful for large simulations where only a very limited subregion requires a path integral treatment.

Finally, although we use a Monte Carlo approach to validate our scheme, it is worth pointing out that the approach can be also applied in the framework of regular path integral molecular

dynamics as well as centroid and ring polymer molecular dynamics [82–84], techniques to calculate approximate quantum dynamical properties and time-correlation functions. This is the focus of an ongoing study.

6.6. Conclusions

In conclusion, we have derived a protocol for concurrent multi-scale path integral simulations based on a rigorous, bottom-up, Hamiltonian formulation. Due to the reduced computational complexity in the classical subdomain, our method enables a computationally more efficient sampling of configurations compared to a full quantum simulation. This, in turn, allows an extension of the accessible time and length scales. Additionally, the possibility to spatially switch off and on the quantum treatment of the system makes this scheme a powerful tool to investigate the role played by nuclear delocalization in soft matter [155, 185, 187]. Furthermore, the technique is compatible with other approaches alleviating the computational complexity of path integral simulations. Future applications of the proposed approach are diverse and include, for example, simulations of the quantum grand-canonical ensemble as well as adaptive quantum-classical simulations of interface systems and biologically relevant systems such as membranes and proteins.

6.7. Acknowledgments

K. Kreis is recipient of a fellowship funded through the Excellence Initiative (DFG/GSC 266). RP, DD, and K. Kremer acknowledge funding from SFB-TRR146 of the German Research Foundation DFG. MET acknowledges support from CHE-1301314 of the National Science Foundation.

Chapter 7.

Hamiltonian Adaptive Resolution Path Integral, Ring Polymer, and Centroid Molecular Dynamics

This chapter is a draft for a research paper that describes an extension of the work presented in the previous chapter. A modified version has been published in The Journal of Chemical Physics after the submission of this thesis.

Karsten Kreis, Kurt Kremer, Raffaello Potestio, and Mark E. Tuckerman
From classical to quantum and back: Hamiltonian adaptive resolution path integral, ring polymer, and centroid molecular dynamics
J. Chem. Phys. 147, 244104 (2017)
DOI: 10.1063/1.5000701

Abstract

Path integral-based simulation methodologies play a crucial role for the investigation of nuclear quantum effects by means of computer simulations. However, these techniques are significantly more demanding than corresponding classical simulations. To reduce this numerical effort, we recently proposed a method, based on a rigorous Hamiltonian formulation, which restricts the quantum modeling to a small but relevant spatial region within a larger reservoir where particles are treated classically. In this work, we extend this idea and show how it can be implemented along with state-of-the-art path integral simulation techniques, such as ring polymer and centroid molecular dynamics, which allow the approximate calculation of both quantum statistical and quantum dynamical properties. To this end, we derive a new integration algorithm which also makes use of multiple time-stepping. The scheme is validated via adaptive classical–path-integral simulations of liquid water. Potential applications of the proposed multiresolution method are diverse and include efficient quantum simulations of interfaces as well as complex biomolecular systems such as membranes and proteins.

7.1. Introduction

Quantum delocalization of light atomic nuclei plays an important role in many soft matter systems, ranging from low temperature helium or hydrogen [76, 77, 85–87] to complex biological systems at room temperature. Examples include proton transfer in biomolecules and membranes [312–318], thermodynamics of ice [319], water adlayers on catalysts [153, 321], aqueous proton and hydroxide transport [80, 92–96], and even the structure and dynamics of bulk water [78, 88–90].

In computer simulations, nuclear quantum effects are typically modeled using Feynman's path integral (PI) formulation of quantum statistical mechanics [52, 68, 85]. The atomic nuclei are mapped onto classical ring polymers, whose beads correspond to the imaginary time slices of the PI. Based on this approach, various techniques have been developed to compute approximate quantum mechanical properties. Path integral molecular dynamics (PIMD) [52, 69–74] and path integral Monte Carlo (PIMC) [52, 73, 75] directly sample the Hamiltonian obtained after path integral quantization and can be employed to calculate time-independent quantum statistical properties. Centroid molecular dynamics (CMD) [82, 83, 351–360], following the dynamics of the ring polymers' centroids, and ring polymer molecular dynamics (RPMD) [84, 358, 359, 361–363], which is based on the evolution of the individual PI beads, additionally enable the calculation of approximate quantum dynamical properties.

However, PI-based methods are significantly more expensive than corresponding classical simulations. To overcome this, different techniques have been proposed. For example, ring polymer contraction (RPC) [337–340] makes use of the fact that long-ranged and non-bonded interactions typically do not need to be evaluated on as many PI beads as bonded interactions. A related technique is the mixed time slicing scheme [341], in which different particles are described with a different number of imaginary time slices. Other approaches include higher-order Trotter factorization [346–350] and advanced thermostating procedures based on generalized Langevin equations (GLE) [342–345]. Additionally, multiple time-stepping (MTS) techniques are frequently employed in PI simulations to decouple the computation of the expensive but slowly varying non-bonded forces and the high frequency internal motion of the ring polymers [73, 74, 364].

Most of these methods correspond to a modification of the path integral computation itself. A different approach is provided by adaptive resolution methods, which restrict the PI description to a small subregion within the simulation box and couple it with a classical model. The available computational resources can then be concentrated on the quantum (QM) subregion leading to an overall speedup compared with full QM simulations. This strategy is useful when only a small part of the overall large system actually needs to be described taking into account quantum delocalization effects, which can be the case, for example, in simulations of surfaces, membranes or the active site of a protein. One such method, based on the adaptive resolution simulation scheme (AdResS) [144, 176, 182], is the direct spatial interpolation of a classical force field with the PI-based forces obtained after quantization [177, 212, 216, 224, 240]. This method is, however, not compatible with an overall Hamiltonian description and, thus, inconsistent

with the PI formalism [219]. Nevertheless, it can in principle be used, for example, to simulate open quantum systems [228, 327].

We have recently proposed a related multiresolution quantum–classical method that, instead of interpolating forces, directly changes the "quantumness" of the particles themselves [365]. In the QM region, the ring polymers are defined as usual, while in the classical region they collapse to point-like particles, thereby effectively behaving classically. When diffusing between the different regions, the particles change their resolution on the fly. Furthermore, the number of particles in the QM region is not fixed but allowed to fluctuate. Hence, the scheme can, for example, be used to efficiently simulate a quantum grand canonical ensemble. The approach is derived in a rigorous fashion from the bottom up and is also compatible with a Hamiltonian description. When restricting the QM part to a small but relevant region in space, the scheme leads to a significant computational speedup. In our example in the previous paper, a liquid parahydrogen system, the calculation of the particle pair interactions was accelerated by a factor of up to ≈ 10. Furthermore, the approach can be combined with the previously mentioned methods for efficient PI simulations, such as RPC or GLE thermostating. Therefore, approaching the problem from a different perspective and reducing the number of PI-based interactions in the system, our method is complementary to techniques which make the PI computations themselves more efficient.

In our previous paper [365] we proposed the general Hamiltonian adaptive quantum–classical scheme, performed a simple validation of the method using a Monte Carlo algorithm to sample the hybrid Hamiltonian, and demonstrated that the approach can speed up PI-based simulations. In this follow-up article, we show how the method can be extended to perform Hamiltonian multiresolution quantum–classical CMD and RPMD simulations. To this end, we derive an MTS integration protocol suited for the proposed methodology and validate the method by adaptive quantum–classical simulations of liquid water.

Our scheme enables efficient simulations of complex systems by locally taking into account QM delocalization effects. This can be useful, for example, for interface systems and in simulations of biological objects such as membranes or proteins. Additionally, it allows an efficient implementation of the QM grand canonical ensemble and can, in principle, also be combined with quantum mechanics/molecular mechanics (QM/MM) approaches, in particular those which are based on a similar Hamiltonian interpolation scheme [169].

The paper is organized as follows: In section 7.2, we review the adaptive quantum–classical scheme proposed in our previous work and in section 7.3 we present its implementation in PIMD. In section 7.4, we discuss how to use the methodology to calculate approximate quantum dynamical quantities in the context of adaptive RPMD and CMD simulations. We describe the details of the simulations we performed for validation in section 7.5 and the results are discussed in section 7.6. In section 7.7, we summarize the article and conclude.

7.2. Quantum–classical path integrals

In quantum statistical mechanics, the partition function of a system of N interacting particles with indices α, momenta $\hat{\mathbf{p}}_\alpha$, masses m_α, kinetic energy $\hat{\mathcal{K}} = \sum_{\alpha=1}^{N} \hat{\mathbf{p}}_\alpha^2 / 2m_\alpha$ and potential energy $\hat{\mathcal{V}}$ is $Q = \mathrm{Tr}[\exp(-\beta\hat{\mathcal{H}})]$ with the inverse temperature $\beta = 1/k_b T$ and the Hamiltonian $\hat{\mathcal{H}} = \hat{\mathcal{K}} + \hat{\mathcal{V}}$. When performing PI quantization using P imaginary time slices (Trotter number P), the kinetic energy term gives rise to a configurational energy that is equivalent to the one of a classical ring polymer with P beads, which are coupled via harmonic springs (for a detailed derivation see, for example, Tuckerman [52]). This mapping from a quantum particle onto a classical polymer ring is exact in the limit $P \to \infty$. In practice, however, well converged results can be obtained for finite values of P, which typically range from 16 to 48 beads for standard PI simulations [78, 317, 328–332].

The strength of the spring constants between the beads of the ring polymers is $m\omega_P^2$ with $\omega_P = \sqrt{P}/\beta\hbar$. It is proportional to the temperature as well as the particles' masses. In other words, the rings are more collapsed the higher the temperature and the heavier the particles. The extension of the ring polymers can be interpreted as a measure of the "quantumness" of the QM particles, with classical behavior corresponding to fully collapsed and therefore localized ring polymers.

The previously proposed method for quantum–classical adaptive resolution simulations [365] is based on precisely this observation: In a PI-based formulation of quantum statistical mechanics the only role of a particle's mass is that of a spring constant. It determines how much "quantum mechanically" the particle behaves. The scheme is as follows: For each particle α we define a resolution parameter $\lambda_\alpha = \lambda(\hat{\mathbf{r}}_\alpha)$ that is a function of the particle's position $\hat{\mathbf{r}}_\alpha$. It smoothly changes from 1 in a spatially predefined QM region to 0 in a classical (CL) region via an intermediate hybrid (HY) transition region (see Fig. 7.1). Based on this resolution function, we then define a variable mass of particle α as $m_\alpha \to \mu_\alpha(\lambda_\alpha) = \lambda_\alpha m_\alpha + (1-\lambda_\alpha) M_\alpha$. Therefore, in the QM region $\mu_\alpha(1) = m_\alpha$ where m_α is the real mass of the particles while in the CL region $\mu_\alpha(0) = M_\alpha \gg m_\alpha$. The mass M_α has to be chosen large enough so that the particles with $\mu_\alpha(0) = M_\alpha$ behave essentially classically (in our previous work, we used $M_\alpha = 100\, m_\alpha$). In this way, particles in the QM region exhibit proper QM behavior, while in the CL region the polymer rings are forced to collapse to nearly point-like particles and behave classically.

In addition to variable masses, we also use the Hamiltonian adaptive resolution simulation (H-AdResS) formalism [181, 183, 184, 266], which can be employed to couple different force fields via interpolation of potential energies. A classical H-AdResS Hamiltonian H of a system of N interacting molecules reads

$$H = \mathcal{K} + \sum_{\alpha=1}^{N} \left[\lambda_\alpha V_\alpha^1 + (1-\lambda_\alpha) V_\alpha^0 + V_\alpha^{\mathrm{int}} - \Delta H(\mathbf{R}_\alpha) \right] \tag{7.1}$$

where $\mathcal{K}$ is the kinetic energy, α indexes the N particles, and $\lambda_\alpha = \lambda(\mathbf{R}_\alpha)$ is the previously defined resolution function (when assigning single resolution values λ_α to whole molecules one

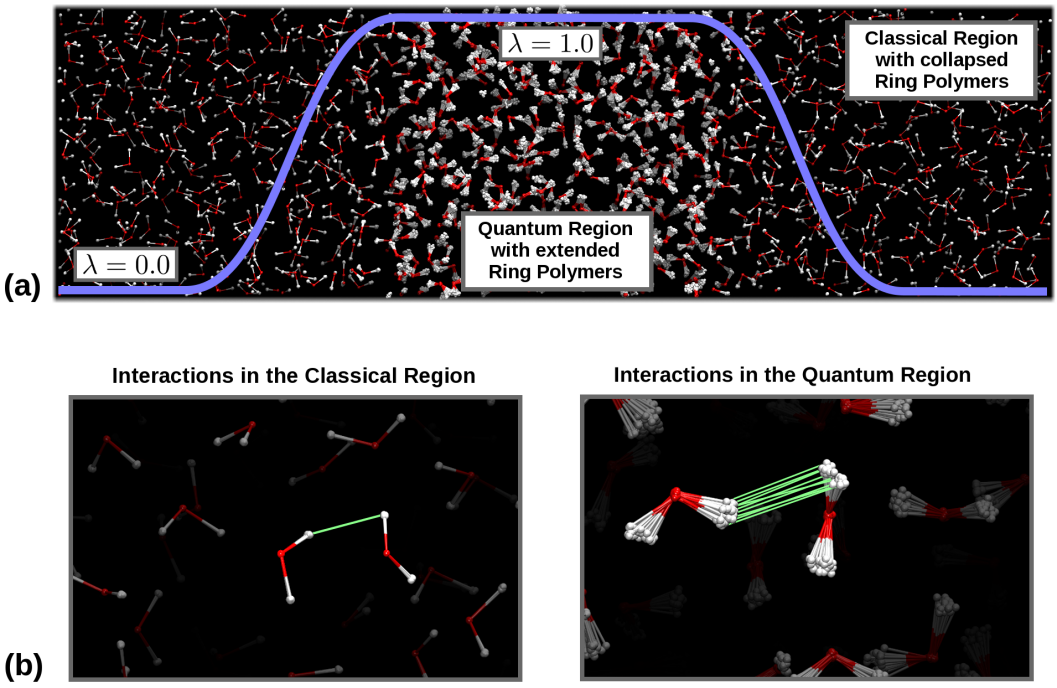

Figure 7.1.: Simulation snapshots from the adaptive quantum–classical simulations. (a) A complete box of quantum–classical water: The blue line shows the resolution function λ switching smoothly from 1 in the QM region to 0 in the CL region. In the QM region, the ring polymers, which correspond to the water atoms, are extended, modeling their quantum mechanical delocalization. In the CL region, they are collapsed to points. (b) Interacting atoms in the CL and QM regions: In the QM region, interactions between different atoms are given as the average over the P time-slices. In the CL region, in contrast, only a single calculation is required for the interaction between a pair of atoms due to the point-like structure of the particles. This alleviates the numerical effort and reproduces normal classical computational efficiency in the CL subsystem.

typically uses the molecular center of mass $\mathbf{R}_\alpha$ as reference coordinate to determine λ_α). The single-particle potentials V_α^{Res} (with $Res = 0, 1$) are the sums of all intermolecular potentials acting on particle α, properly normalized so that double counting is avoided. The term V_α^{int} represents all intramolecular interactions, such as bond and angle potentials, which are not subject to interpolation. The term ΔH, referred to as the Free Energy Compensation (FEC) [181, 183], is an external field acting in the HY transition region to eliminate the density imbalance that naturally occurs in such dual-resolution systems. Different models of the same physical system exhibit a free energy difference that needs to be neutralized in order to enforce identical thermodynamical and/or structural properties (e.g. density) everywhere in the simulation domain. The FEC levels off these free energy imbalances. H-AdResS has been used mainly to couple atomistic and coarse-grained two-body force fields [181, 183, 286, 366]. In general, however, the potentials can refer to any non-bonded interaction V_α^{Res}.

We have combined the mass-based quantum–classical interpolation with the H-AdResS scheme in order to be able to use different force fields within the QM and the CL regions. This can be advantageous, for example, when the CL region's only role is that of a particle reservoir. Then, one can use a simpler force-field in the CL region. On the other hand, when, for example, simulating a protein and only a small part of it needs to be modeled including nuclear quantum effects, one would probably resort to the same force-field everywhere in the system and just "add" the nuclear quantum effects in the relevant region with the presently discussed quantum–classical multiresolution scheme.

Combining the quantum–classical mass interpolation and the H-AdResS scheme, we can write the Hamiltonian operator of N interacting Boltzmann particles in three dimensions as [365]

$$\hat{\mathcal{H}} = \sum_{\alpha=1}^{N} \left(\frac{1}{2} \hat{\mathbf{p}}_\alpha \mu^{-1}(\hat{\mathbf{r}}_\alpha) \hat{\mathbf{p}}_\alpha + V^{\text{H-AdResS}}(\hat{\mathbf{r}}_\alpha) \right) \tag{7.2}$$

where $\mu^{-1}(\hat{\mathbf{r}}_\alpha)$ is the inverse mass operator. The potential energy term $V^{\text{H-AdResS}}(\hat{\mathbf{r}}_\alpha)$ corresponds to the interpolated H-AdResS potential energy, i.e. the term within the sum in Eq. 7.1. We have shown that PI quantization then leads to the following expression for the partition function [365]:

$$Q = \lim_{P\to\infty} \left[\prod_{k=1}^{P} \prod_{\alpha=1}^{N} \int \mathrm{d}\mathbf{r}_{\alpha,k} \left(\frac{\mu_{\alpha,k} P}{2\pi\beta\hbar^2} \right)^{\frac{3}{2}} \right] e^{-\beta V_P^\mu} \tag{7.3}$$

with

$$\begin{aligned} V_P^\mu = \sum_{k=1}^{P} \sum_{\alpha=1}^{N} \Bigg\{ & \frac{\mu_{\alpha,k}\, \omega_P^2}{2} |\mathbf{r}_{\alpha,k} - \mathbf{r}_{\alpha,k+1}|^2 \\ & + \frac{1}{P} \left[\lambda_{\alpha,k} V_{\alpha,k}^{\text{QM}} + (1-\lambda_{\alpha,k}) V_{\alpha,k}^{\text{CL}} + V_{\alpha,k}^{\text{int}} - \Delta H(\mathbf{r}_{\alpha,k}) \right] \Bigg\} \end{aligned} \tag{7.4}$$

where α indexes the different particles and k the individual Trotter beads for each of them. A resolution value $\lambda_{\alpha,k}$ is associated with each bead. We have renamed $V_{\alpha,k}^{0,1}$ to $V_{\alpha,k}^{\text{CL,QM}}$ to emphasize that $V_{\alpha,k}^{\text{CL}}$ ($V_{\alpha,k}^{\text{QM}}$) is the intermolecular potential acting in the CL (QM) region. Note that the normalization term in Eq. 7.3 depends on the position of the particles via $\mu_{\alpha,k}$. To obtain a constant normalization factor, one can transform this position dependent term to a potential energy in V_P^μ, as done in our previous article [365], and then treat it as a constant field in the hybrid region. In this work, however, we will deal with it in a different way, which we will discuss in detail later.

The above expression, Eqs. 7.3 and 7.4, is consistent with a rigorous PI quantization if

$$\left| \frac{d\mu(x)}{dx} \right| \ll \frac{2\mu(x)}{\Lambda_\mu(x)} \tag{7.5}$$

with $\Lambda_\mu(x) \equiv \sqrt{\beta\hbar^2/(P\mu(x))}$. This criterion means that the interpolation between the QM and the CL parts of the system must be sufficiently smooth such that the mass difference between

two neighboring beads on a ring polymer in the HY region is negligible. This requirement can always be satisfied by choosing a sufficiently large HY region. Note, however, that even if Eq. 7.5 is not fulfilled, the expression in Eqs. 7.3 and 7.4 corresponds to a well-defined quantum–classical simulation protocol.

The ring polymers described by the energy function V_P^μ (Eq. 7.4) are expanded in the region where the mass is small, and collapse to nearly classical point-like particles in the large-mass region. Therefore, in the CL region the interactions between different ring polymers do not need to be computed as an average over the P bead pairs as done in the QM region. Instead, due to their point-like structure, one can use only the centroid with negligible error (see Fig. 7.1). In this fashion, classical computational efficiency is regained in the CL region.

7.3. Quantum–classical path integral molecular dynamics

In our previous paper [365], we have validated the scheme introduced above via simulations of liquid parahydrogen using a basic Monte Carlo algorithm to directly sample the phase space defined by V_P^μ. A more state-of-the-art approach to the numerical evaluation of PIs is provided by PIMD, in which a thermostated dynamics is generated in phase space to sample the quantum canonical ensemble [52, 69–74]. PIMD is more easily parallelizable compared to Monte Carlo methods and therefore significantly more efficient for typical simulation setups and on multicore computer architectures. In the following, we show how the proposed quantum–classical multiresolution method can be implemented in PIMD as well as CMD and RPMD.

7.3.1. Evaluation of the adaptive mass and the resolution function on the centroids

In order to decouple the modes of the cyclic ring polymers from each other, PIMD is typically performed using staging variables [73, 75] or, more popularly, normal modes [73, 352]. In our case, we cannot transform smoothly into normal mode space, because the beads of the individual ring polymers have different masses $\mu_{\alpha,k}$ within the HY region. For typical systems like liquid water at room temperature, however, this mass difference is small as the extension of the ring polymers, measured by the root-mean-square radius of gyration r^g, is short, even in the QM region. This suggests to associate a single resolution value λ_α and a single adaptive mass value μ_α with each atomic or molecular particle α instead of with every single bead k. λ_α and μ_α can then be determined using the rings' centroid positions.

Although this corresponds only to a minor modification in Eq. 7.4, we can ask to what extent the configurational energy V_P^μ is then still compatible with formal PI quantization. To this end, we first consider the adaptive mass $\mu_{\alpha,k}$ of an individual Trotter bead within the HY region which we can approximate as

$$\mu_{\alpha,k} \approx \mu_\alpha^c + \delta x_{\alpha,k} \frac{d\mu(x_\alpha^c)}{dx}, \tag{7.6}$$

where x_α^c is the centroid coordinate along the direction of resolution change of the ring polymer α. For a setup where the resolution changes along the x-direction, this would be the centroid's x-coordinate and for a system where the QM region is spherical and the resolution changes radially, this would be the radial distance from the center. μ_α^c is the mass function evaluated at the centroid x_α^c and $\delta x_{\alpha,k}$ is the distance along the direction of resolution change between the k-th bead of ring α and its centroid x_α^c. We have $\delta x_{\alpha,k} \lesssim r^g(x_\alpha^c)$, where $r^g(x) = \sqrt{\beta\hbar^2/(4\mu(x))} \cdot \sqrt{1-1/P^2} \approx \sqrt{\beta\hbar^2/(4\mu(x))}$ is the radius of gyration of a free ring with mass $\mu(x)$ for large P. Hence, we can approximate $\mu_{\alpha,k} \approx \mu_\alpha^c$ if

$$\left|\frac{d\mu(x_\alpha^c)}{dx}\right| \ll \frac{\mu_\alpha^c}{r^g(x_\alpha^c)}, \tag{7.7}$$

or simply

$$\left|\frac{d\mu(x)}{dx}\right| \ll \frac{\mu(x)}{r^g(x)}. \tag{7.8}$$

for general x. Here, we used the free ring radius of gyration. However, for the ring polymers in typical systems the average radius of gyration differs only slightly from the free ring radius.

Eq. 7.8 is a slightly stronger criterion than the one in Eq. 7.5. This makes sense, since Eq. 7.5 essentially provides the condition under which the mass can be considered as constant between two neighboring beads, while the new criterion, Eq. 7.8, gives the condition for treating the mass constant over a whole ring. For liquid water at room temperature, Eq. 7.8 is satisfied by a hybrid region wider than $\approx 1\,\mathrm{nm}$. An even smaller hybrid region would not be desirable anyway, since the interaction cutoffs of typical interaction potentials are also of the order $\approx 1\,\mathrm{nm}$.

Next, we consider the resolution λ itself, which we would also like to treat as constant over a whole ring and let it depend only on the position of the centroid. On the one hand, λ varies between 0 and 1 and changes most steeply in the center of the HY region (see Fig. 7.1). On the other hand, an upper bound on the extension of the ring polymers is provided by the radius of gyration of the rings in the QM region, which will be denoted as r^g_{QM}. Therefore, if the change in λ, corresponding to its gradient in the center of the HY region, over a distance r^g_{QM} is much smaller than ≈ 1, we can approximate $\lambda_{\alpha,k} \approx \lambda_\alpha^c$ everywhere (as for the adaptive mass before, λ_α^c denotes the resolution function of ring α evaluated at its centroid x_α^c). This corresponds to

$$\left|\left(\left.\frac{d\lambda(x)}{dx}\right|_{x=d_{\mathrm{HY}}/2}\right)\right| r^g_{\mathrm{QM}} \ll 1, \tag{7.9}$$

where $d\lambda(x)/dx|_{x=d_{\mathrm{HY}}/2}$ denotes the gradient of λ in the center of the HY region and d_{HY} the width of the HY region. The criterion in eq. 7.9 can be easily fulfilled for typical systems such as water at room temperature, with a HY region of width $d_{\mathrm{HY}} \approx 1\,\mathrm{nm}$.

Concluding, in the following we will treat both the mass and the resolution as a constant over entire rings and write for simplicity $\mu_{\alpha,k}(x) \to \mu_\alpha(x)$ and $\lambda_{\alpha,k}(x) \to \lambda_\alpha(x)$, where we assume

that the mass and the resolution functions for an atom α have been evaluated using its centroid coordinate x_α^c and that the resulting parameters have been assigned to all beads belonging to ring α. Then we can write the partition function as

$$Q = \lim_{P\to\infty} \left[\prod_{k=1}^{P} \prod_{\alpha=1}^{N} \int \mathrm{d}\mathbf{r}_{\alpha,k} \left(\frac{\mu_\alpha P}{2\pi\beta\hbar^2} \right)^{\frac{3}{2}} \right] e^{-\beta \tilde{V}_P^\mu} \tag{7.10}$$

with

$$\begin{aligned} \tilde{V}_P^\mu &= \sum_{k=1}^{P} \sum_{\alpha=1}^{N} \left\{ \frac{\mu_\alpha\ \omega_P^2}{2} |\mathbf{r}_{\alpha,k} - \mathbf{r}_{\alpha,k+1}|^2 \right. \\ &+ \left. \frac{1}{P} \left[\lambda_\alpha V_{\alpha,k}^{\mathrm{QM}} + (1-\lambda_\alpha) V_{\alpha,k}^{\mathrm{CL}} + V_{\alpha,k}^{\mathrm{int}} - \Delta H(\mathbf{r}_\alpha^c) \right] \right\}, \end{aligned} \tag{7.11}$$

where $\mathbf{r}_\alpha^c$ denotes the centroid of ring polymer α. Note that the FEC is now also applied at the single-atom, i.e., centroid level. The criteria in Eqs. 7.8 and 7.9 quantify to what degree this partition function is still compatible with a formal, bottom-up PI quantization. The inequalities can always be fulfilled by choosing a sufficiently wide hybrid region. However, even if they are not met, the final partition function, Eqs. 7.10 and 7.11, still represents a well-defined Hamiltonian multiresolution quantum–classical simulation scheme.

7.3.2. Introducing normal modes

Now that the different beads of each ring polymer α all have the same adaptive mass μ_α, we can proceed with transforming the Cartesian coordinates into normal modes $\mathbf{u}_{\alpha,k}$ via

$$\mathbf{u}_{\alpha,k} = \sum_{j=1}^{P} \mathbf{r}_{\alpha,j} C_{jk}, \tag{7.12}$$

where, for even P, the orthogonal transformation matrix is [343]

$$C_{jk} = \begin{cases} \sqrt{1/P} & \text{if } k = 1 \\ \sqrt{2/P}\cos(2\pi jk/P) & \text{if } 2 \le k \le P/2 \\ \sqrt{1/P}\,(-1)^j & \text{if } k = P/2+1 \\ \sqrt{2/P}\sin(2\pi jk/P) & \text{if } P/2+2 \le k \le P \end{cases} \tag{7.13}$$

such that for a given ring polymer at position x:

$$\mu_\alpha(x) \sum_{k=1}^{P} |\mathbf{r}_{\alpha,k} - \mathbf{r}_{\alpha,k+1}|^2 = \mu_\alpha(x) \sum_{l=1}^{P} \xi_k \mathbf{u}_{\alpha,k}^2 \tag{7.14}$$

with

$$\xi_k = 4\sin^2\left(\frac{(k-1)\pi}{P}\right). \tag{7.15}$$

In normal mode representation, the centroid of a ring polymer α is given by the rescaled first mode coordinate, this is

$$\mathbf{r}_\alpha^c = \frac{1}{P}\sum_{k=1}^{P}\mathbf{r}_{\alpha,k} = \frac{1}{\sqrt{P}}\mathbf{u}_{\alpha,1}. \tag{7.16}$$

Therefore, the adaptive mass μ_α and the resolution function λ_α, being evaluated at the centroids, are functions of the first mode only, i.e. $\mu_\alpha(\mathbf{r}_\alpha^c) = \mu_\alpha(1/\sqrt{P}\,\mathbf{u}_{\alpha,1})$ and $\lambda_\alpha(\mathbf{r}_\alpha^c) = \lambda_\alpha(1/\sqrt{P}\,\mathbf{u}_{\alpha,1})$. To lighten the notation, though, we will drop the $1/\sqrt{P}$ factor and write simply $\mu_\alpha = \mu_\alpha(\mathbf{u}_{\alpha,1})$ and $\lambda_\alpha = \lambda_\alpha(\mathbf{u}_{\alpha,1})$.

We then obtain the following partition function

$$Q = \lim_{P\to\infty}\left[\prod_{k=1}^{P}\prod_{\alpha=1}^{N}\int \mathrm{d}\mathbf{u}_{\alpha,k}\left(\frac{\mu_\alpha(\mathbf{u}_{\alpha,1})P}{2\pi\beta\hbar^2}\right)^{\frac{3}{2}}\right] e^{-\beta\tilde{V}_P^\mu} \tag{7.17}$$

with

$$\begin{aligned}\tilde{V}_P^\mu = \sum_{k=1}^{P}\sum_{\alpha=1}^{N}\Bigg\{&\frac{1}{2}\nu_{\alpha,k}(\mathbf{u}_{\alpha,1})\omega_P^2\mathbf{u}_{\alpha,k}^2 + \frac{1}{P}\Big[\lambda_\alpha(\mathbf{u}_{\alpha,1})V_{\alpha,k}^{\mathrm{QM}}(\mathbf{u}) + \\ &+ (1-\lambda_\alpha(\mathbf{u}_{\alpha,1}))\,V_{\alpha,k}^{\mathrm{CL}}(\mathbf{u}) + V_{\alpha,k}^{\mathrm{int}}(\mathbf{u}) - \Delta H(\mathbf{u}_{\alpha,1})\Big]\Bigg\}\end{aligned} \tag{7.18}$$

and the rescaled adaptive mass $\nu_{\alpha,k}(\mathbf{u}_{\alpha,1}) = \mu_\alpha(\mathbf{u}_{\alpha,1})\,\xi_k$. For the centroid, i.e. $k = 1$, we have $\nu_{\alpha,k}(\mathbf{u}_{\alpha,1}) = 0$. In Eqs. 7.17 and 7.18, we have explicitly indicated the dependencies of the different terms on the normal modes. The notation $\mathbf{u}$ without any indices is a shorthand for the compound set of all coordinates $\mathbf{u}_{\alpha,k}$.

7.3.3. Introducing momenta

In PIMD, one usually recasts the prefactor of the partition function as Gaussian integrals over a set of variables that can be interpreted as momenta conjugate to the coordinates $\mathbf{u}_{\alpha,k}$ [52]. The energy term in the exponential can then be interpreted as a classical Hamiltonian and sampled via thermostated molecular dynamics. In our case, however, the prefactor is position dependent and, therefore, we have different options to proceed.

(a) Constant kinetic masses. As done in our previous article [365], we can write

$$\left(\frac{\mu_\alpha(\mathbf{u}_{\alpha,1})P}{2\pi\beta\hbar^2}\right)^{\frac{3}{2}} = \left(\frac{\tilde{m}P}{2\pi\beta\hbar^2}\right)^{\frac{3}{2}}\exp\left\{-\beta\left(-\frac{3}{2\beta}\log\left(\frac{\mu_\alpha(\mathbf{u}_{\alpha,1})}{\tilde{m}}\right)\right)\right\} \tag{7.19}$$

where we used an arbitrary mass $\tilde{m}$ as the reference mass scale. Then we can pull the term $-(3/2\beta)\,\log(\mu_\alpha/\tilde{m})$ into $\tilde{V}_P^\mu$ as an external field in the HY region and use the now constant prefactor to introduce a set of momenta via rephrasing the prefactor as Gaussian integrals. This

yields

$$Q = \lim_{P\to\infty} \left[\prod_{k=1}^{P} \prod_{\alpha=1}^{N} \int \mathrm{d}\mathbf{u}_{\alpha,k} \int \mathrm{d}\mathbf{p}_{\alpha,k} \right] e^{-\beta H_P^{\mathrm{ckm}}} \tag{7.20}$$

with the Hamiltonian

$$\begin{aligned} H_P^{\mathrm{ckm}} = \sum_{k=1}^{P} \sum_{\alpha=1}^{N} \Bigg\{ & \frac{\mathbf{p}_{\alpha,k}^2}{2m'_{\alpha,k}} + \frac{1}{2}\nu_{\alpha,k}\omega_P^2 \mathbf{u}_{\alpha,k}^2 - \frac{3}{2P\beta} \log \frac{\mu_\alpha(\mathbf{u}_{\alpha,1})}{\tilde{m}} + \\ & + \frac{1}{P} \left[\lambda_\alpha(\mathbf{u}_{\alpha,1}) V_{\alpha,k}^{\mathrm{QM}}(\mathbf{u}) + (1-\lambda_\alpha(\mathbf{u}_{\alpha,1}))\, V_{\alpha,k}^{\mathrm{CL}}(\mathbf{u}) + V_{\alpha,k}^{\mathrm{int}}(\mathbf{u}) - \Delta H\,(\mathbf{u}_{\alpha,1}) \right] \Bigg\}. \end{aligned} \tag{7.21}$$

As the additional logarithmic term only acts as an external field in the HY region, it can be exactly removed via the FEC function ΔH. We will denote the fictitious masses $m'_{\alpha,k}$ in the following as "kinetic" masses in contrast to $\nu_{\alpha,k}$, which we will refer to as spring masses. In principle, the set of $m'_{\alpha,k}$ can be chosen freely, as their rescaling does not affect thermodynamic averages [52].

H_P^{ckm} in Eq. 7.21 defines a classical Hamiltonian system composed of ring polymers representing the delocalized atoms. In the QM region where $\lambda_\alpha = 1$ and $\mu_\alpha = m$, the ring polymers are extended and the regular quantum mechanical behaviour is recovered. In the CL region where $\lambda_\alpha = 0$ and $\mu_\alpha = M$ the rings are collapsed to essentially point-like particles, thereby reproducing classical mechanics. The Hamiltonian gives rise to regular equations of motion that can be integrated by a symplectic integrator such as the velocity Verlet algorithm [52], with the possibility of employing multiple time-stepping.

However, let us take a closer look at the different masses in the system. While the spring masses $\nu_{\alpha,k}$ change between the CL and QM subregions of the system, the kinetic masses $m'_{\alpha,k}$ do not. We choose $m'_{\alpha,k} = m_\alpha / P$ with m_α being the real mass of atom α, since this corresponds to a realistic bead-wise approximate quantum dynamical behavior in the QM region similar to RPMD (in RPMD one usually chooses $m'_{\alpha,k} = m_\alpha$ without $1/P$, but rescales the potential energy terms by P and runs the simulation at a P-times higher temperature [84, 358, 359, 361–363]. Here, the factor of $1/P$ in the mass is equivalent to this procedure, as we perform the simulations at the actual temperature and use a Hamiltonian, Eq. 7.21, without rescaling potential energies). Therefore, in the QM region, the modes oscillate with vibration frequencies $\omega_P\sqrt{\xi_k P}$. In the CL region, however, where $\nu_{\alpha,k}$ is significantly larger than in the QM subsystem the modes oscillate faster than in the QM region by a factor of $\sqrt{M_\alpha / m_\alpha}$. For the case of $M_\alpha = 100\, m_\alpha$, this results in 10 times higher frequencies. This would require a 10 times smaller time step in the integration algorithm compared to a normal quantum simulation or compared to what would be required in the QM subregion. Although this poses no fundamental hurdle, it may slow down the simulations notably.

(b) Adaptive kinetic masses. The previous observation suggests an alternative approach: We can also directly recast the prefactor as a Gaussian integral, which includes the position

dependent mass μ_α,

$$\left(\frac{\mu_\alpha(\mathbf{u}_{\alpha,1})P}{2\pi\beta\hbar^2}\right)^{\frac{3}{2}} = \left(\frac{P^2}{4\pi^2\hbar^2}\right)^{\frac{3}{2}} \int d\boldsymbol{p}_{\alpha,k} \exp\left\{-\beta P \frac{\boldsymbol{p}_{\alpha,k}^2}{2\mu_\alpha(\mathbf{u}_{\alpha,1})}\right\}. \tag{7.22}$$

In this way, we can introduce a kinetic energy term which has adaptive kinetic masses. This leads to the construction of a Hamiltonian in which both the spring and the kinetic masses vary in the same fashion, such that the modes oscillate with the same frequencies everywhere in the quantum–classical adaptive resolution setup.

Specifically, we propose the following: Overall, we have $N \times P$ prefactors of the form

$$\left(\frac{\mu_\alpha(\mathbf{u}_{\alpha,1})P}{2\pi\beta\hbar^2}\right)^{\frac{3}{2}}, \tag{7.23}$$

one for each atom and mode. For all higher modes with $k > 1$, we transform the prefactors according to Eq. 7.22 and introduce the momentum terms in the Hamiltonian with a variable mass in the denominator. The remaining N prefactors are then treated via Eq. 7.19, and the kinetic masses for the centroid modes, which are not associated with springs since $\nu_{\alpha,1} = 0$, are chosen constant. We then obtain

$$Q = \lim_{P\to\infty} \left[\prod_{k=1}^{P}\prod_{\alpha=1}^{N} \int \mathrm{d}\mathbf{u}_{\alpha,k} \int \mathrm{d}\mathbf{p}_{\alpha,k}\right] e^{-\beta H_P^{\mathrm{akm}}} \tag{7.24}$$

with the Hamiltonian

$$\begin{aligned} H_P^{\mathrm{akm}} = \sum_{k=1}^{P}\sum_{\alpha=1}^{N} \Bigg\{ &\frac{\mathbf{p}_{\alpha,k}^2}{2\nu'_{\alpha,k}(\mathbf{u}_{\alpha,1})} + \frac{1}{2}\nu_{\alpha,k}\omega_P^2\mathbf{u}_{\alpha,k}^2 - \frac{3}{2P\beta}\log\frac{\mu_\alpha(\mathbf{u}_{\alpha,1})}{\tilde{m}} + \\ &+\frac{1}{P}\left[\lambda_\alpha(\mathbf{u}_{\alpha,1})V_{\alpha,k}^{\mathrm{QM}}(\mathbf{u}) + (1-\lambda_\alpha(\mathbf{u}_{\alpha,1}))\,V_{\alpha,k}^{\mathrm{CL}}(\mathbf{u}) + V_{\alpha,k}^{\mathrm{int}}(\mathbf{u}) - \Delta H\,(\mathbf{u}_{\alpha,1})\right]\Bigg\}, \end{aligned} \tag{7.25}$$

where $\nu'_{\alpha,k}$ is the kinetic mass of bead k of atom α. Note that a $1/P$ factor appears in front of the logarithmic term in H_P^{akm} because the term still appears in the sum over all P, although we obtain the logarithmic term only for the centroid modes. Choosing appropriate prefactors, the parameters $\nu'_{\alpha,k}$ are

$$\nu'_{\alpha,1} = \quad m_\alpha/P, \qquad k = 1, \text{ centroid mode}, \tag{7.26}$$

$$\nu'_{\alpha,k} = \quad \tilde{\nu}_\alpha = \mu_\alpha/P, \qquad k > 1, \text{ higher modes}. \tag{7.27}$$

Since the kinetic masses for the higher modes are all equal, we introduced a new abbreviation, $\tilde{\nu}_\alpha$, for them without the index k. We choose a factor $1/P$ to ensure that the approximate quantum dynamical time evolution of the centroids proceeds on the real timescale, this is, the same as in corresponding classical Newtonian dynamics. As already pointed out, this choice is equivalent to the temperature rescaling often done in ring polymer molecular dynamics

[84, 358, 359, 361–363], which we do not perform here. A further rescaling of the kinetic masses would be allowed when sampling only canonical averages [52, 73].

We can interpret the choice of the kinetic masses in the following way: While traveling from the QM to the CL via the HY region, the spring constants of the higher modes become stronger and the ring polymers collapse. Simultaneously, however, these higher modes become heavier such that they do not vibrate faster in the CL region than in the QM region, despite the stiffer springs. Their oscillation frequencies are the same everywhere in the system such that their configurations can be sampled efficiently throughout the whole system with the same time step. The centroid modes do not undergo such oscillations, as they represent only the displacements of the whole rings. Hence, their masses do not need to change across the transition from the QM to the CL part of the system and, therefore, they are chosen to be the real masses.

In fact, this approach is in line with what one would intuitively expect from a rigorous PI-based quantum–classical adaptive resolution scheme. On the one hand, considering a single atom, all higher modes, which represent the delocalization or the "quantumness" of the atom, collapse in the classical region and therefore localize the atom to a classical point particle there. On the other hand, the centroid mode, which does not correspond to any quantum properties of the atom and only represents the average position of the atom, is not changed throughout the whole system. In other words, we only adapt the internal ring behavior, i.e. the quantum properties of the atoms, while leaving its classical, more macroscopic behavior, determined via the centroid, untouched. It is only indirectly affected by the quantum–classical transition of the higher modes.

In the following we will refer to the constant kinetic mass (CKM) approach, defined by the Hamiltonian H_P^{ckm} in Eq. 7.21, as the CKM approach and to the adaptive kinetic mass (AKM) scheme, defined by the Hamiltonian H_P^{akm} in Eq. 7.25, as the AKM approach.

7.3.4. Equations of motion

We will not discuss the equations of motion obtained in the CKM approach, as they resemble a regular structure which can be integrated, for example, by a regular velocity Verlet algorithm [52]. Instead, we focus on the AKM scheme, from which the CKM approach can be obtained as a special case.

In the following, we will assume that the logarithmic term in the Hamiltonian in Eq. 7.25 has been exactly canceled by an appropriately chosen FEC function $\Delta H(\lambda)$ and therefore omit it to lighten the notation.

Then, the equations of motion are as follows:

$$\dot{\boldsymbol{u}}_{\alpha,1} = \frac{\boldsymbol{p}_{\alpha,1}}{\nu'_{\alpha,1}(\boldsymbol{u}_{\alpha,1})} = P\frac{\boldsymbol{p}_{\alpha,1}}{m_\alpha}, \qquad k = 1, \text{ centroid mode}, \tag{7.28}$$

$$\dot{\boldsymbol{u}}_{\alpha,k} = \frac{\boldsymbol{p}_{\alpha,k}}{\tilde{\nu}_\alpha(\boldsymbol{u}_{\alpha,1})} = P\frac{\boldsymbol{p}_{\alpha,k}}{\mu_\alpha(\boldsymbol{u}_{\alpha,1})}, \qquad k > 1, \text{ higher modes}, \tag{7.29}$$

$$
\begin{aligned}
\dot{\boldsymbol{p}}_{\alpha,1} = {} & \mathbf{F}^{\text{H-AdResS}}_{\alpha,1}(\mathbf{u}) \\
& - \frac{1}{P}\left[\sum_{i=1}^{P}\left\{V^{\text{QM}}_{\alpha,i}(\boldsymbol{u}) - V^{\text{CL}}_{\alpha,i}(\boldsymbol{u})\right\}\right]\boldsymbol{\nabla}_{\boldsymbol{u}_{\alpha,1}}\lambda(\boldsymbol{u}_{\alpha,1}) \qquad (\mathbf{F}^{\text{drift}}_1) \\
& - \left[\sum_{i=2}^{P}\frac{1}{2}\frac{(M_\alpha - m_\alpha)}{P\tilde{v}_\alpha(\boldsymbol{u}_{\alpha,1})^2}\boldsymbol{p}^2_{\alpha,i}\right]\boldsymbol{\nabla}_{\boldsymbol{u}_{\alpha,1}}\lambda(\boldsymbol{u}_{\alpha,1}) \qquad (\mathbf{F}^{\text{drift}}_2) \\
& - \left[\sum_{i=2}^{P}\frac{1}{2}\xi_i(m_\alpha - M_\alpha)\omega_P^2\mathbf{u}^2_{\alpha,i}\right]\boldsymbol{\nabla}_{\boldsymbol{u}_{\alpha,1}}\lambda(\boldsymbol{u}_{\alpha,1}) \qquad (\mathbf{F}^{\text{drift}}_3) \\
& + \boldsymbol{\nabla}_{\boldsymbol{u}_{\alpha,1}}\Delta H(\boldsymbol{u}_{\alpha,1}),
\end{aligned}
\qquad k = 1, \text{ centroid mode,} \tag{7.30}
$$

$$
\dot{\boldsymbol{p}}_{\alpha,k} = -\nu_{\alpha,k}(\boldsymbol{u}_{\alpha,1})\omega_P^2\mathbf{u}_{\alpha,k} + \mathbf{F}^{\text{H-AdResS}}_{\alpha,k}(\mathbf{u}), \qquad k > 1, \text{ higher modes,} \tag{7.31}
$$

where

$$
\mathbf{F}^{\text{H-AdResS}}_{\alpha,k}(\mathbf{u}) = -\frac{1}{P}\sum_{j=1}^{P}\sum_{\beta=1}^{N}\left[\lambda_\beta\boldsymbol{\nabla}_{\boldsymbol{u}_{\alpha,k}}V^{\text{QM}}_{\beta,j}(\mathbf{u}) + (1-\lambda_\beta)\,\boldsymbol{\nabla}_{\boldsymbol{u}_{\alpha,k}}V^{\text{CL}}_{\beta,j}(\mathbf{u}) + \boldsymbol{\nabla}_{\boldsymbol{u}_{\alpha,k}}V^{\text{int}}_{\beta,j}(\mathbf{u})\right]. \tag{7.32}
$$

The terms in lines 2-5 in Eq. 7.30 stem from the application of the derivative on the position dependent resolution function. The terms $\mathbf{F}^{\text{drift}}_i$ are undesired forces that act only in the hybrid region, can lead to thermodynamic imbalances in the system, and, for example, artificially push particles from one subregion of the system to the other. In accordance with earlier works using the H-AdResS scheme, we will refer to these forces as *drift forces* [181]. They need to be compensated, which can be established via the FEC, line 5 in Eq. 7.30. In fact, the latter is typically constructed to cancel their average effect [181, 183, 184].

The drift force $\mathbf{F}^{\text{drift}}_1$ comes from the potential energy interpolation and would not be present if we changed only the masses of the atoms but not the force field. $\mathbf{F}^{\text{drift}}_2$ is a result of choosing the kinetic masses of the higher modes to be adaptive and would be absent in the CKM approach. Finally, $\mathbf{F}^{\text{drift}}_3$ corresponds to the adaptive spring masses. As the resolution λ_α and the adaptive mass μ_α depend only on the centroid positions of the ring polymers, drift forces only occur in the equations of motion for the centroid. Therefore, the internal motion of the rings is not disturbed by any drift forces and can smoothly change from quantum to classical and vice versa when the atoms travel through the system. Only the translation of the rings is affected by drift forces, which can be corrected via the FEC. Note that the sums in $\mathbf{F}^{\text{drift}}_2$ and $\mathbf{F}^{\text{drift}}_3$ only run over higher modes and exclude the first one, because the centroid mode neither appears in the spring constant term in the Hamiltonian, nor is it associated with a variable mass in the kinetic energy onto which the position derivative could act.

7.3.5. Integration

To devise a suitable integration scheme for the equations of motion, Eqs. 7.28-7.31, we use the Liouville operator formalism. To lighten the notation, in the following we will drop the atom

index α, as the Liouville operators for different atoms, i.e. rings, commute and do not require further discussion.

The Liouville operator iL can be written as

$$iL = iL^{(1)} + iL^{(2)}, \tag{7.33}$$

where $iL^{(1)}$ propagates positions and $iL^{(2)}$ momenta. We further decompose

$$iL^{(1)} = iL_1^{(1)} + iL_k^{(1)}, \tag{7.34}$$
$$iL^{(2)} = iL_1^{(2)} + iL_k^{(2)}, \tag{7.35}$$

where operators $iL_1^{(i)}$ propagate the first mode and $iL_k^{(i)}$ act only on the higher order modes $k > 1$. They are

$$iL_1^{(1)} = P\frac{\mathbf{p}_1}{m}\frac{\partial}{\partial \mathbf{u}_1}, \tag{7.36}$$
$$iL_k^{(1)} = \frac{\mathbf{p}_k}{\tilde{v}(\mathbf{u}_1)}\frac{\partial}{\partial \mathbf{u}_k}, \tag{7.37}$$
$$iL_1^{(2)} = \left(\mathbf{F}_1(\mathbf{u}) + \eta(\mathbf{u}_1)\sum_{k=2}^{P}\mathbf{p}_k^2\right)\frac{\partial}{\partial \mathbf{p}_1}, \tag{7.38}$$
$$iL_k^{(2)} = \mathbf{F}_k(\mathbf{u})\frac{\partial}{\partial \mathbf{p}_k}, \tag{7.39}$$

with

$$\begin{aligned}\mathbf{F}_1(\mathbf{u}) = \mathbf{F}_1^{\text{H-AdResS}}(\mathbf{u}) &\\ - \frac{1}{P}\left[\sum_{i=1}^{P}\left\{V_i^{\text{QM}}(\mathbf{u}) - V_i^{\text{CL}}(\mathbf{u})\right\}\right]\boldsymbol{\nabla}_{\mathbf{u}_1}\lambda(\mathbf{u}_1) &\\ - \left[\sum_{i=2}^{P}\frac{1}{2}\xi_i(m-M)\omega_P^2\mathbf{u}_i^2\right]\boldsymbol{\nabla}_{\mathbf{u}_1}\lambda(\mathbf{u}_1) &\\ + \boldsymbol{\nabla}_{\boldsymbol{u}_1}\Delta H(\boldsymbol{u}_1), &\end{aligned} \tag{7.40}$$

and

$$\mathbf{F}_k(\mathbf{u}) = -\nu_k(\mathbf{u}_1)\omega_P^2\mathbf{u}_k + \mathbf{F}_k^{\text{H-AdResS}}(\mathbf{u}), \tag{7.41}$$
$$\eta(\mathbf{u}_1) = -\frac{1}{2}\frac{(M-m)}{P\tilde{v}(\mathbf{u}_1)^2}\boldsymbol{\nabla}_{\mathbf{u}_1}\lambda(\mathbf{u}_1). \tag{7.42}$$

The vector notation denotes that each Liouville operator in Eqs. 7.36-7.39 represents a set of three operators for each direction, which commute and can therefore be applied in arbitrary order.

In the first step we decompose the classical propagator $\exp\{iLt\}$ as

$$e^{iLt} = e^{iL^{(1)}t + iL^{(2)}t} = \lim_{M\to\infty} \left(e^{iL^{(2)}\frac{t}{2M}} e^{iL^{(1)}\frac{t}{M}} e^{iL^{(2)}\frac{t}{2M}} \right)^M \tag{7.43}$$

using the symmetric Trotter theorem [367, 368]. Defining the time step $\Delta t = t/M$ this yields the velocity Verlet integrator

$$e^{iL\Delta t} \approx e^{iL^{(2)}\frac{\Delta t}{2}} e^{iL^{(1)}\Delta t} e^{iL^{(2)}\frac{\Delta t}{2}}. \tag{7.44}$$

However, recalling the definitions of $iL^{(1)}$ and $iL^{(2)}$ we recognize that their constituents $iL_1^{(1)}$ and $iL_k^{(1)}$ as well as $iL_1^{(2)}$ and $iL_k^{(2)}$ do not commute. Therefore, using again the Trotter theorem, we decompose each of the propagators in Eq. 7.44 further to

$$\begin{aligned} e^{iL\Delta t} &\approx e^{iL^{(2)}\frac{\Delta t}{2}} e^{iL^{(1)}\Delta t} e^{iL^{(2)}\frac{\Delta t}{2}} \\ &\approx \left(e^{iL_k^{(2)}\frac{\Delta t}{4}} e^{iL_1^{(2)}\frac{\Delta t}{2}} e^{iL_k^{(2)}\frac{\Delta t}{4}} \right) \left(e^{iL_1^{(1)}\frac{\Delta t}{2}} e^{iL_k^{(1)}\Delta t} e^{iL_1^{(1)}\frac{\Delta t}{2}} \right) \left(e^{iL_k^{(2)}\frac{\Delta t}{4}} e^{iL_1^{(2)}\frac{\Delta t}{2}} e^{iL_k^{(2)}\frac{\Delta t}{4}} \right), \end{aligned} \tag{7.45}$$

which is correct up to second order in Δt and is suited for the integration of the system's equations of motion. Note that for the CKM approach, the adaptive mass $\nu_k'(\mathbf{u}_1)$ in $iL_k^{(1)}$, Eq. 7.37, would be constant and also the $\mathbf{p}_k^2$ term in $iL_1^{(2)}$, Eq. 7.38, would be missing. Therefore, no second decomposition level would be required and we could stick to the velocity Verlet.

The interpretation of the integration scheme in Eq. 7.45 is straightforward. The first and the last term in brackets $(\cdots)$ correspond to the propagation of momenta, while the term in the middle propagates coordinates. The next decomposition level tells us how to precisely update the momenta and the coordinates. The momenta are integrated in the following way: We first propagate all higher mode momenta by a quarter step, then we update the $\mathbf{p}_k^2$ term in $iL_1^{(2)}$ using these new momenta and propagate the first mode's momentum by a half step. Next, we perform the other quarter step for the higher modes. To get the full time step, the procedure is repeated after the position update in the center of the scheme. The coordinates are updated as follows: We first propagate the first mode by a half step using $iL_1^{(1)}$. Then we update the adaptive masses in the Liouville operator $iL_k^{(1)}$ and propagate the higher modes by a full step. Finally, we integrate the first mode by another half step. The scheme requires little additional computational overhead compared to a regular velocity Verlet scheme. The number of additional operations scales only linearly with the number of particles, and the force computation, usually the numerically most demanding part of a simulation, has to be performed as usual only once after all coordinates are fully propagated.

Finally, we want to address the symplecticity of the new integrator. The Hamiltonian H_P^{akm}, Eq. 7.25, is not trivially separable into two parts, one depending only on coordinates and one containing only momenta. However, it has no term in which both the momentum and the corresponding conjugate coordinate of the same mode appear together. This is a result of our choice of the adaptive masses: The higher mode masses are position dependent, but they do not depend on their own mode coordinates but only on the centroid coordinate. However,

the centroid itself, which determines the resolution and the masses of the rings, is associated with a constant mass. Therefore, the Hamiltonian H_P^{akm} and the corresponding equations of motion still define a symplectic structure. As a consequence, also the integration scheme in Eq. 7.45 is symplectic, as it is constructed in a rigorous fashion from H_P^{akm}'s equations of motion using the Liouville operator formalism. This can also be understood considering the Liouville operators themselves: The operator $iL_k^{(1)}$, Eq. 7.37, which propagates the *higher mode coordinates* $\mathbf{u}_k$ does have an additional position dependence but only on the *centroid mode*. Hence, it can be applied as usual in a well-defined manner. Similarly, the operator $iL_1^{(2)}$, Eq. 7.38, propagating the momentum of the *centroid mode* has an additional dependence on momenta, but only on the *higher mode momenta*. Consequently, the determinant J of the time evolution matrix is 1. This can be derived easily by considering the time evolution matrices corresponding to each Liouville operator. The determinant of each of these matrices is 1. Therefore, the determinant of the overall time evolution matrix, equal to the product of the individual determinants, is also 1. The symplecticity has the practical advantage that we are able to derive an energy conserving integrator, which, in our case, is exact up to second order in time, similar to a regular velocity Verlet.

It is worth pointing out that the previous observation is in contrast with earlier works using adaptive masses [333, 369]. There, both momenta and the corresponding conjugate coordinates appear together in the same terms in the Hamiltonian. Hence, in those cases, Liouville's theorem no longer holds, the Liouville operator formalism breaks down, and symplecticity is lost.

7.3.6. Multiple time-stepping

In typical complex soft matter systems, non-bonded interactions as well as bonds, angles and dihedrals generate motion on different time scales. In PIMD, we have additionally the springs between the beads of the ring polymers onto which the quantum particles are mapped. If the kinetic masses for the higher modes are small, they vibrate strongly, which requires a small integration time step. When only sampling statistical averages, the kinetic masses can be chosen freely, for example, such that all higher modes vibrate with the same frequency. When calculating approximate quantum dynamical quantities, however, the kinetic mode masses must either correspond to the real ones, as in RPMD, or must be significantly decreased, as in CMD. This leads to an internal ring polymer dynamics which is significantly faster than the motion due to typical interatomic non-bonded or bonded potentials. Furthermore, in the CKM approach, the modes' oscillation frequencies are increased in the CL region, as the kinetic mass will be small there compared to the increased spring mass. This strongly motivates the introduction of multiple time-stepping into our integrator.

We employ the RESPA scheme [364] and decompose the force computation into three parts: one for non-bonded forces, a second for the bonds, and a third for the internal ring polymer motion. The first drift term, $\mathbf{F}_1^{\text{drift}}$, depends only on the energies associated with the non-bonded potentials and is therefore evaluated together with rest of the non-bonded forces. The second

and the third drift terms $\mathbf{F}_2^{\text{drift}}$ and $\mathbf{F}_3^{\text{drift}}$, however, depend directly on the motion of the higher modes and therefore need to be evaluated together with them. Hence, we define

$$
\begin{aligned}
iL_1^{(1)} &= P\frac{\mathbf{p}_1}{m}\frac{\partial}{\partial \mathbf{u}_1}, && (7.46)\\
iL_k^{(1)} &= \frac{\mathbf{p}_k}{\tilde{\nu}(\mathbf{u}_1)}\frac{\partial}{\partial \mathbf{u}_k}, && (7.47)\\
iL_1^{(2)} &= \left(\mathbf{F}_1^{\text{mode}}(\mathbf{u}) + \eta(\mathbf{u}_1)\sum_{k=2}^{P}\mathbf{p}_k^2\right)\frac{\partial}{\partial \mathbf{p}_1}, && (7.48)\\
iL_k^{(2)} &= \mathbf{F}_k^{\text{mode}}(\mathbf{u})\frac{\partial}{\partial \mathbf{p}_k}, && (7.49)\\
iL_q^{(3)} &= \mathbf{F}_q^{\text{int}}(\mathbf{u})\frac{\partial}{\partial \mathbf{p}_q}, && (7.50)\\
iL_q^{(4)} &= \mathbf{F}_q^{\text{nb}}(\mathbf{u})\frac{\partial}{\partial \mathbf{p}_q}, && (7.51)
\end{aligned}
$$

with

$$
\begin{aligned}
\mathbf{F}_1^{\text{mode}}(\mathbf{u}) &= -\left[\sum_{i=2}^{P}\frac{1}{2}\xi_i(m-M)\omega_P^2\mathbf{u}_i^2\right]\boldsymbol{\nabla}_{\mathbf{u}_1}\lambda(\mathbf{u}_1), && (7.52)\\
\mathbf{F}_k^{\text{mode}}(\mathbf{u}) &= -\nu_k(\mathbf{u}_1)\omega_P^2\mathbf{u}_k, && (7.53)\\
\mathbf{F}_q^{\text{int}}(\mathbf{u}) &= -\frac{1}{P}\sum_{j=1}^{P}\sum_{\beta=1}^{N}\boldsymbol{\nabla}_{\boldsymbol{u}_q}V_{\beta,j}^{\text{int}}(\mathbf{u}), && (7.54)\\
\mathbf{F}_q^{\text{nb}}(\mathbf{u}) &= -\frac{1}{P}\sum_{j=1}^{P}\sum_{\beta=1}^{N}\left[\lambda_\beta\boldsymbol{\nabla}_{\boldsymbol{u}_q}V_{\beta,j}^{\text{QM}}(\mathbf{u}) + \left(1-\lambda_\beta\right)\boldsymbol{\nabla}_{\boldsymbol{u}_q}V_{\beta,j}^{\text{CL}}(\mathbf{u})\right] && (7.55)\\
&\quad -\frac{1}{P}\left[\sum_{i=1}^{P}\left\{V_i^{\text{QM}}(\mathbf{u}) - V_i^{\text{CL}}(\mathbf{u})\right\}\right]\boldsymbol{\nabla}_{\mathbf{u}_q}\lambda(\mathbf{u}_1)\\
&\quad +\boldsymbol{\nabla}_{\boldsymbol{u}_q}\Delta H(\boldsymbol{u}_1).
\end{aligned}
$$

Note that the Liouville operators $iL_q^{(3)}$ and $iL_q^{(4)}$ do not need to be split into centroid and higher terms, as these commute in this case. Hence, for the sake of brevity, we have subsumed both parts and changed to the index q, which includes all $1 \leq q \leq P$. Finally, we obtain the following RESPA multiple time-stepping scheme:

$$
\begin{aligned}
e^{iL\Delta t} \approx e^{iL^{(4)}\frac{\Delta t}{2}}\Bigg\{e^{iL^{(3)}\frac{\delta t}{2}}\Bigg[\left(e^{iL_k^{(2)}\frac{dt}{4}}e^{iL_1^{(2)}\frac{dt}{2}}e^{iL_k^{(2)}\frac{dt}{4}}\right)\left(e^{iL_1^{(1)}\frac{dt}{2}}e^{iL_k^{(1)}dt}e^{iL_1^{(1)}\frac{dt}{2}}\right)\times\\
\times\left(e^{iL_k^{(2)}\frac{dt}{4}}e^{iL_1^{(2)}\frac{dt}{2}}e^{iL_k^{(2)}\frac{dt}{4}}\right)\Bigg]^n e^{iL^{(3)}\frac{\delta t}{2}}\Bigg\}^N e^{iL^{(4)}\frac{\Delta t}{2}},
\end{aligned}
\tag{7.56}
$$

where $\Delta t = N\cdot\delta t = N\cdot n\cdot dt$. The internal ring vibrations as well as the drift terms depending on these higher ring modes are integrated with the shortest time step dt. The intramolecular

but interatomic bonds and angles are integrated with a distinct, medium time step δt, and the intermolecular non-bonded interactions as well as the corresponding drift terms are integrated with the largest time step Δt. The whole integration may be carried out in normal mode space, although in practice the interatomic forces are computed in real space and then transformed into mode space.

7.3.7. Langevin thermostating

To generate a canonical ensemble we need to couple the system to a thermostat. We resort to a Langevin thermostat, as Langevin equation-based frameworks have been shown to be favorable in PIMD and RPMD simulations and can be used to optimize sampling efficiency [342–345]. As the focus of this work is not advanced thermostating, however, we use a simple white noise Langevin thermostat without memory instead of, for example, a GLE approach. The implementation follows the BAOAB method by Leimkuhler and Matthews [370], which provides high configurational sampling accuracy. Within the proposed multiple time-stepping scheme this yields

$$\begin{aligned} e^{iL\Delta t} \approx e^{iL^{(4)}\frac{\Delta t}{2}} \Bigg\{ e^{iL^{(3)}\frac{\delta t}{2}} \Bigg[\left(e^{iL_k^{(2)}\frac{dt}{4}} e^{iL_1^{(2)}\frac{dt}{2}} e^{iL_k^{(2)}\frac{dt}{4}} \right) \left(e^{iL_1^{(1)}\frac{dt}{4}} e^{iL_k^{(1)}\frac{dt}{2}} e^{iL_1^{(1)}\frac{dt}{4}} \right) \times \\ \times e^{iL^{\text{Langevin}}dt} \left(e^{iL_1^{(1)}\frac{dt}{4}} e^{iL_k^{(1)}\frac{dt}{2}} e^{iL_1^{(1)}\frac{dt}{4}} \right) \left(e^{iL_k^{(2)}\frac{dt}{4}} e^{iL_1^{(2)}\frac{dt}{2}} e^{iL_k^{(2)}\frac{dt}{4}} \right) \Bigg]^n e^{iL^{(3)}\frac{\delta t}{2}} \Bigg\}^N e^{iL^{(4)}\frac{\Delta t}{2}}, \end{aligned} \tag{7.57}$$

with the action of the Langevin Liouville operator iL^{Langevin} on mode i

$$e^{iL^{\text{Langevin}}dt}\,\mathbf{u}_i = \mathbf{u}_i, \tag{7.58}$$

$$e^{iL^{\text{Langevin}}dt}\,\mathbf{p}_i = \mathbf{p}_i\, e^{-\gamma dt} + \sqrt{\nu_i'(\mathbf{u}_1) k_b T (1 - e^{-2\gamma dt})}\,\mathbf{R}(t). \tag{7.59}$$

γ is the friction parameter, T the temperature, k_B Boltzmann's constant, and $R(t)$ are independent and identically distributed normal random numbers with mean 0, variance 1, and $\langle \mathbf{R}(t)\mathbf{R}(t')\rangle = \delta(t - t')$. This thermostating method can also be adapted such that each mode is thermostated with a different optimized friction constant, as done in the path integral Langevin equation (PILE) scheme by Ceriotti et al. [343, 363].

Using the integration scheme of Eq. 7.57, we can perform efficient adaptive quantum–classical PIMD simulations with either the AKM or the CKM approach. It is derived in a rigorous fashion from a symplectic Hamiltonian and is also consistent with PI quantization, provided that the criteria in Eqs. 7.8 and 7.9 are satisfied. It is computationally advantageous over full-quantum simulations, because in the CL region all forces between interacting ring polymers can be approximated by a single calculation between the centroids. Furthermore, because the ring polymers are collapsed in the CL region and interact classically, the integration of the internal motion, i.e. of the higher modes, can be stopped and the rings can be frozen in this part of the system. In the CKM case or for full-quantum systems, the algorithm reduces to a

regular velocity Verlet scheme with multiple time-stepping and Langevin thermostating.

The derived quantum–classical multiresolution scheme can be combined with other optimization techniques for PI simulations. For example, the non-bonded forces could also in the QM region be evaluated based on fewer than P beads, using the RPC scheme by Markland and Manolopoulos [337, 338]. Alternatively, instead of white noise Langevin thermostating, we could make use of a colored noise thermostat. It was shown by Ceriotti et al. that a carefully parametrized PI GLE can lower the number of Trotter beads required for converged quantum behavior [342–345]. As our approach reduces the overall computational effort of a PI simulation by restricting the QM region of the system, it is complementary to these methods, which reduce the numerical complexity of the PI interactions themselves.

7.4. Approximate quantum dynamics

In PIMD, we only measure quantum statistical properties. In the following, we will discuss how our integration scheme can be extended to allow for multiresolution quantum–classical CMD and RPMD with only minor changes.

7.4.1. Quantum–classical centroid molecular dynamics

Centroid molecular dynamics (CMD) is a method for the calculation of real-time quantum correlation functions in the short-time limit [82, 83, 351–360]. It is based on the notion that approximate quantum dynamical properties can be calculated from the time evolution of the centroid subject to the potential of mean force generated by the ring polymer. Formally, this potential is obtained by integration over all possible ring configurations with constrained centroid position. This would be not just computationally expensive but practically intractable. The idea of CMD is to adiabatically decouple the internal fluctuations of the ring polymers from the centroid motion. By rescaling the higher mode kinetic masses ($k > 1$) with a sufficiently small adiabadicity parameter $0 < \gamma_{\mathrm{CMD}}^2 < 1$, such that $\nu_k' \to \gamma_{\mathrm{CMD}}^2 \nu_k'$, the higher modes can be forced to evolve significantly faster than the centroid. Thereby, the centroid potential of mean force of the ring polymer is generated "on the fly" during the simulation. It has been shown, however, that in practice a partial adiabatic decoupling is sufficient for most applications [358]. In addition to the mass rescaling, the higher modes alone are coupled to thermostats such that the centroid dynamics remains Newtonian.

The previously described implementation of CMD, i.e. the removal of the thermostat from the centroid and the kinetic mass rescaling, can also be done easily in our quantum–classical multiresolution scheme. In practice, one would typically be interested in the quantum dynamics only in the QM region. Hence, it suffices to only remove the centroid thermostat in this region. Then we could measure approximate quantum dynamical properties in the QM region while the classical part would still behave as in the canonical ensemble and could serve, for example, as a particle reservoir for the QM region. Another relevant scenario is the simulation of a complex biomolecule like a protein. In this case, an overall large simulation box would be

required to preserve the structure and the solvating environment of the system, although we may want to probe the dynamics only in a specific subregion, such as near the protein's active site.

7.4.2. Quantum–classical ring polymer molecular dynamics

An alternative approach to the calculation of approximate quantum dynamics is provided by ring polymer molecular dynamics (RPMD) [84, 358, 359, 361, 362]. In the normal RPMD approach, the kinetic masses are chosen to be the real physical masses (as we did above also for PIMD) and no thermostats are used, such that the ring polymer evolution is completely Newtonian. In comparison to CMD, RPMD uses the whole chain to approximate quantum time correlation functions. However, the internal ring fluctuations can lead to artifacts when measuring, for example, vibrational spectra [371]. To overcome this deficiency, Rossi et al. recently proposed a thermostated ring polymer molecular dynamics (TRPMD) approach [363] that can be interpreted as an intermediate method between normal RPMD and CMD. In TRPMD, the kinetic masses are also chosen to be the real physical masses and measurements are performed based on the whole chain. However, as in CMD, Langevin thermostats are attached to all higher modes $k > 1$. Provided the thermostats are adjusted carefully, TRPMD avoids both the spurious resonances in the vibrational spectra and also the curvature problem of CMD [371], while retaining the appealing properties of RPMD. An ideal choice for the Langevin friction parameters in TRPMD is given by the PILE scheme [343, 363]. In the PILE method, each higher mode $k > 1$ is thermostated with a different optimized coupling constant γ_k based on the mode vibration frequency as $\gamma_k = \omega_P \sqrt{\xi_k P}$.

Just as with CMD, TRPMD simulations can be easily run with our quantum–classical PI scheme and the corresponding integrator, Eq. 7.57. We only need to adapt the thermostats on the different modes accordingly and remove the thermostat on the centroid.

Note that in the AKM approach the kinetic masses will change in the CL region. However, in this part of the simulation box the spring masses also have different values and we are typically not interested in the dynamics anyway. Thus, one may also reintroduce the centroid thermostat in this outer region.

7.5. Simulations

To validate the proposed adaptive resolution PIMD approach, we implement it in the ESPResSo++ molecular simulation package [1] and perform adaptive resolution simulations of liquid water. Nuclear quantum effects in liquid water have been thoroughly investigated and shown to be important for an accurate description of its structure and dynamics [78, 88–90]. Hence, water is an ideal test case for the method.

7.5.1. Water system

We consider a system of 918 water molecules in a slab-shaped box with dimensions $L_x = 6.92\,\mathrm{nm}$, $L_y = L_z = 2.0\,\mathrm{nm}$ ($33.168\,\mathrm{molecules/nm^3}$), and periodic boundary conditions in all directions. The resolution changes along the X-direction, the full width of the QM region is set to $d_{\mathrm{QM}} = 2.0\,\mathrm{nm}$, and the width of the adjacent HY regions to $d_{\mathrm{HY}} = 1.5\,\mathrm{nm}$. The resolution function is given by a squared cosine, commonly used in adaptive resolution simulations [155, 184, 193, 212, 365, 366, 372]. We perform the simulations at a temperature of $300\,K$ and use a Trotter number $P = 32$, as this has been shown to provide well-converged results for most dynamical and structural water properties [78, 88, 89]. A simulation snapshot of the system is presented in Fig. 7.1 (a).

To model the water, we use a force field that was recently developed by Fritsch et al. specifically for PI simulations of bulk liquid water [78]. It is parametrized from *ab initio* density functional theory calculations using the force matching [102–104] and iterative Boltzmann inversion methods [106]. All interactions are mapped onto a set of short-ranged tabulated potentials and no explicit charges are present. Separate potentials are provided for the non-bonded O-O, O-H, H-H interactions, for the O-H bond, and for the H-O-H angle. An additional bonded potential is applied between the two H-atoms of the same molecule. This force field describes the structural and dynamical properties of liquid water at $300\,K$ and at a density of 1.1 $\mathrm{g/cm^2}$ very well. Furthermore, it is very efficient in simulations, since it is purely short-ranged with an interaction cutoff of 7.8 Å. We have chosen this potential for its numerical efficiency and its suitability for PI simulations, and we note that the derived adaptive resolution methodology can also be applied for analytic potentials as well as those that include charges.

In order to collapse the ring polymers in the CL region, we choose $M_\alpha = 100\,m_\alpha$ for all particles α. Because of their point-like structure we only use the centroids to calculate non-bonded and bonded interactions between atoms in the CL region (see Fig. 7.1 (b)). Furthermore, we stop the integration of the higher modes in the CL region, i.e., we freeze the internal degrees of freedom of the ring polymers. Note that the setup satisfies the criteria in Eqs. 7.8 and 7.9 and can therefore be considered to be consistent with formal path integral quantization.

We run simulations using both the AKM and CKM approaches, although we focus on the AKM method, for which we have derived a non-standard integrator and which allows larger time steps. In all simulations, the kinetic mass of the centroids is given by Eq. 7.26, which corresponds to using the real mass. For the AKM simulations, we choose the kinetic masses of the higher modes according to Eq. 7.27. As argued already, this corresponds to using the real masses in the QM region, which facilitates a realistic ring polymer time evolution and therefore allows the calculation of approximate quantum dynamical properties from RPMD simulations. In the CL region, the higher mode masses are increased as explained previously. For the CKM simulations, we choose the masses in a similar way, although they remain constant over all simulation domains. For CMD simulations, we introduce an additional rescaling of the higher modes' kinetic masses with an adiabadicity parameter $\gamma^2_{\mathrm{CMD}} = 0.05$. Note that we do not rescale the kinetic masses of the higher modes with the eigenvalues of the normal mode

transformation, as it is often done in CMD [358]. We also keep the additional $1/P$ factor in the kinetic masses, which we introduced earlier to ensure dynamics on the correct time scale. Therefore, the ring polymers' higher modes vibrate with frequencies $\omega_P\sqrt{\xi_k P/\gamma^2_{\mathrm{CMD}}}$.

To enforce the correct temperature, we couple all modes to white noise Langevin thermostats. The centroid mode is thermostated with a friction constant $\gamma = 2.0\,\mathrm{ps}^{-1}$, except in CMD and TRPMD simulations, where no thermostat is applied on it. For the higher modes k, we employed the PILE scheme by Ceriotti et al. [343, 363] and used frictions $\gamma_k = \omega_P\sqrt{\xi_k P}$ that are proportional to the modes' vibration frequencies (for CMD simulations, we used $\gamma_k = \omega_P\sqrt{\xi_k P/\gamma^2_{\mathrm{CMD}}}$). The PILE method leads to optimized sampling and can also be applied in the context of TRPMD simulations.

The derived adaptive quantum–classical simulation method allows to not only change the "quantumness" of the particles, but also their non-bonded interaction potentials. This has been demonstrated in our previous paper in simulations of liquid parahydrogen [365]. Here, we perform the majority of the validation simulations changing only the quantumness of the particles, using the same interaction potential in both the QM and the CL region. Nevertheless, we also test the scheme using a different potential in the CL region, a purely repulsive Weeks-Chandler-Andersen (WCA) potential [274] of the form

$$V^{\mathrm{CL}}(r) = \begin{cases} 4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} + \frac{1}{4}\right] & : r \leq R_{\mathrm{c}} \\ 0 & : r > R_{\mathrm{c}} \end{cases} \tag{7.60}$$

with $\epsilon = k_B T$, $\sigma = 0.25\,\mathrm{nm}$, and $R_{\mathrm{c}} = 2^{\frac{1}{6}}\sigma = 0.28\,\mathrm{nm}$. The potential acts only between the oxygen atoms. Note that the intramolecular bonded interactions are kept in the CL region to prevent the molecules from disintegrating.

7.5.2. Setups

We perform simulations employing the following setups:

- **Setup 1:** We use adaptive kinetic masses and the same interaction potentials in both regions. Applying thermostats to all modes, we calculate various structural properties of the water in the QM region. Additionally, we remove the thermostat from the centroid and use TRPMD to calculate several dynamical quantities.

- **Setup 2:** The same as setup 1, except the kinetic masses of the higher modes are rescaled with the adiabadicity parameter γ^2_{CMD}. Then, we calculate the dynamical properties via CMD.

- **Setup 3:** The AKM method is applied as in setup 1, but the WCA potential is employed to model intermolecular interactions in the CL region. In this scenario, we only validate the coupling by calculating density profiles as well as profiles of the atomistic radii of gyration.

- **Setup 4:** We switch to constant kinetic masses and employ the same interaction potentials in both regions. As in setup 3, we only validate the coupling and calculate density profiles as well as profiles of the atomistic radii of gyration.

Additionally, we perform full-quantum and full-classical ($P = 1$) reference simulations without any interpolation between different particle masses or interaction potentials. All simulation parameters, including the box dimensions, are the same as for the adaptive simulations. The only exception is the friction constant of the Langevin thermostat, which is set to $10.0\,\mathrm{ps}^{-1}$ in the full-classical simulations.

The time steps used in the simulation setups are presented in Tab. 7.1. For the full-classical

Setup	Δt	δt	dt
#1: AKM method, same potentials, TRPMD	2.0 fs	0.5 fs	0.05 fs
#2: AKM method, same potentials, CMD	0.4 fs	0.1 fs	0.01 fs
#3: AKM method, WCA potential in CL region, TRPMD	2.0 fs	0.5 fs	0.05 fs
#4: CKM method, same potentials, TRPMD	1.0 fs	0.1 fs	0.00625 fs

Table 7.1.: Time steps for the quantum–classical adaptive resolution PIMD, RPMD and CMD simulations and for the reference calculations.

and the full-quantum reference simulations we use the same time steps as in the corresponding adaptive resolution setups. The time steps in the table refer to the ones used in equilibration simulations, during the derivation of the free energy correction and the thermodynamic force (see next section), as well as during all other simulations sampling statistical averages. For the calculation of dynamic quantities in setup 1 we reduce all time steps to the same ones as used in the CMD simulations in setup 2. We do this for two reasons: On the one hand, our implementation of the integration scheme allows one to print out positions or velocities only after a full step Δt. Hence, this large time step needs to be short enough to allow a fine sampling when calculating, for example, velocity autocorrelation functions. On the other hand, we want to avoid artifacts resulting from the use of different time steps when comparing CMD to TRPMD. Note, however, that only few and very short simulations need to be run with this modification. The majority of simulations use the time steps in Tab. 7.1.

In general, all time steps are chosen to be as large as possible but still sufficiently small to accurately sample phase space, retain an acceptable level of energy-conservation in microcanonical test simulations, and generate the correct temperature in simulations in the canonical ensemble. The time steps we find to work well seem reasonable: In classical simulations, updating the regular non-bonded forces every 1-2 fs is a frequent choice [270, 366, 372–375], while the vibration frequency for the bonds and angles in water demands a time step of around 0.5 fs [78, 376]. The vibration frequency of the springs between the PI beads is yet higher, requiring an even smaller time step. Furthermore, CMD simulations are known to require particularly small time steps, as the internal ring polymer motion is strongly accelerated. A similar effect is observed in simulations with the CKM approach. In this case, the internal

motion of collapsed rings is also significantly enhanced (we mentioned that in the CL region, the rings are frozen. However, the rings are already strongly collapsed at the outer parts of the HY region, where a full integration of the internal motion is still necessary to accommodate the gradual collapse and extension of the rings). Therefore, already at this point it becomes clear that the AKM scheme is better suited for the proposed adaptive quantum–classical simulation protocol than the naive CKM method. We want to stress, however, that finding optimal time steps is not the primary goal of this work and that there is certainly room for further fine-tuning.

For the calculation of all structural quantities and statistical averages we run simulations of duration 200 ps, if not otherwise indicated. Additionally, we perform short 2 ps runs during which we calculate velocity autocorrelation functions and vibrational spectra. We also measure hydrogen bond population fluctuations, which is done in simulations of duration 32 ps. In all cases we start from equilibrated configurations, run 10 independent simulations, and average the results.

7.5.3. Free energy corrections

To correct for the thermodynamic imbalance between the low-mass QM and the high-mass CL region, we apply a free energy correction (FEC) ΔH [181, 183, 184, 266, 365, 366]. We derive the FEC via Kirkwood thermodynamic integration (KTI) between the fully CL ($\lambda = 0$) and the fully QM ($\lambda = 1$) system and we calculate the averages

1. $$\left\langle \frac{1}{NP} \sum_{\alpha=1}^{N} \sum_{i=1}^{P} \left\{ V_{\alpha,i}^{\mathrm{QM}}(\boldsymbol{u}) - V_{\alpha,i}^{\mathrm{CL}}(\boldsymbol{u}) \right\} \right\rangle_{\lambda}, \tag{7.61}$$

2. $$\left\langle \frac{1}{N} \sum_{\alpha=1}^{N} \sum_{i=2}^{P} \frac{1}{2} \frac{(M_\alpha - m_\alpha)}{P\tilde{v}_\alpha(\boldsymbol{u}_{\alpha,1})^2} \boldsymbol{p}_{\alpha,i}^2 \right\rangle_{\lambda}, \tag{7.62}$$

3. $$\left\langle \frac{1}{N} \sum_{\alpha=1}^{N} \sum_{i=2}^{P} \frac{1}{2} \xi_i (m_\alpha - M_\alpha) \omega_P^2 \mathbf{u}_{\alpha,i}^2 \right\rangle_{\lambda}, \tag{7.63}$$

as well as the pressure $p(\lambda)$ for a set of 101 λ's along the integration path from $\lambda = 0$ to $\lambda = 1$. The KTI is run in a smaller box of dimensions $L_x = L_y = L_z = 3.0\,\mathrm{nm}$. All other simulation parameters are as explained above, except for the thermostat friction of the centroid mode, which was set to $10\,\mathrm{ps}^{-1}$ to achieve rapid equilibration after changing λ. We start the KTI from an equilibrated system at $\lambda = 0$ and we perform for each λ a short 0.3 ps equilibration run (for setup 2 only 0.12 ps due to the short time step) and another 1.5 ps run (for setup 2 only 0.6 ps) during which we take measurements. From these results we construct the FEC ΔH to cancel the averages of the drift forces, Eq. 7.30, and the pressure difference between the subsystems. Since calculating the FEC via KTI is an approximate method to correct for the thermodynamic imbalance, we refine the FEC using the thermodynamic force (TF) scheme [182, 365]. The TF is an iterative approach that directly constructs a correction force in the HY region from the distorted density profile along the direction of resolution change in order to flatten the density throughout the system. Each TF iteration consists of a 50 ps equilibration run (10 ps for setup

2 and 15 ps for setup 4) and a 150 ps production run (30 ps for setup 2 and 20 ps for setup 4) during which we sample the density. We perform 20 iterations for each setup.

Although we calculate the quantities in Eqs. 7.61-7.63 separately for the oxygen and hydrogen atoms, we determine global molecular pressures instead of species-wise partial pressures. After constructing the correction force from the pressure $p(\lambda)$, we distribute it between oxygen and hydrogens proportionally to their masses. Finally, the FEC is applied on the atomistic level based on the atom's centroid positions.

Both the derivation of the FEC via KTI and the iterative correction are well-established methods for achieving a smooth coupling in adaptive resolution simulations. See, for example, Refs. [181–184, 266, 365, 366] for further technical details.

7.6. Results

7.6.1. Structure

We first investigate the structural properties of the adaptive quantum–classical water systems. Fig. 7.2 shows the density profiles along the x-direction of the four setups without correcting for the thermodynamic imbalance (green curves), with FEC but without iterative refinement (blue curves), and with FEC including the iterative refinement via TF (red curves). Without any corrections the density is strongly distorted. Applying the non-iterative FEC significantly improves the coupling between the regions, although the density in the QM region is still slightly too low for setups 1-3 and much too low for setup 4. This can be expected, as the non-iterative FEC is an approximate method and since statistical inaccuracies can occur during its derivation via KTI. Refining the FEC with the iterative TF technique, we obtain flat density profiles for all setups, except setup 4, for which significant deviations in the HY region remain. Note that for setup 4, which uses the CKM scheme, we were not able to run stable simulations without any compensation. In this case, the drift forces are so strong that all molecules are immediately pushed to one subregion. In comparison, the AKM approach works much better and requires a more moderate FEC.

The derivation of the FEC via KTI and several iterations of TF may seem cumbersome. However, both the KTI as well as the TF iterations can be run using simulation setups that are much smaller than the actual system. For large applications this step will likely take significantly less time than the simulation of the complete system. Additionally, more advanced approaches have recently been developed that efficiently calculate the FEC on the fly during the simulation of the full system or a representative smaller one [231]. Note that a FEC or a similar compensation force is required in all adaptive resolution methods that allow a free exchange of particles between subregions that feature different thermodynamics [181–183, 216, 224, 231]. All results reported below are calculated in setups in which the refined FEC is applied.

Fig. 7.3 presents the radii of gyration of the ring polymers corresponding to the water's oxygen and hydrogen atoms as a function of their position along the x-direction. In the QM region, the radii of gyration perfectly match with those from full-quantum reference simulations,

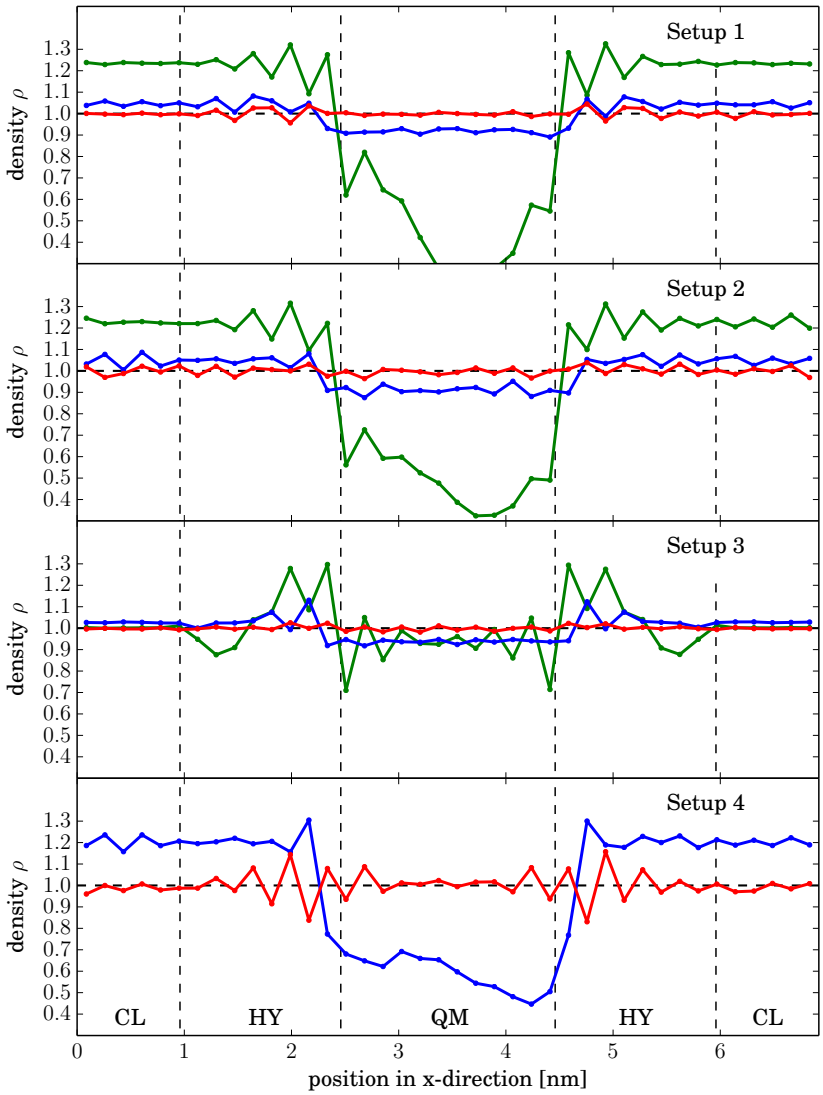

Figure 7.2.: Normalized density profiles ρ in the quantum–classical adaptive resolution simulations of liquid water for the four different setups without FEC (green), with KTI-based FEC but without iterative refinement (blue), and with FEC including iterative refinement via TF (red). For setup 4, we are not able to run stable simulations without any FEC (setup 1: AKM, same potentials, TRPMD. Setup 2: AKM, same potentials, CMD. Setup 3: AKM, WCA potential in CL region, TRPMD. Setup 4: CKM, same potentials, TRPMD).

while in the CL region they drop by $\approx 90\%$ (also see Fig. 7.1). Therefore, the molecules exhibit their full-blown "quantumness" in the QM region, while in the CL region the ring polymers shrink to nearly point-like particles and behave classically. To collapse the ring polymers even further, one would simply need to choose a heavier particle mass M_α in the CL region. The radius of gyration in the CL region is approximately proportional to $1/\sqrt{M_\alpha}$. Note that the

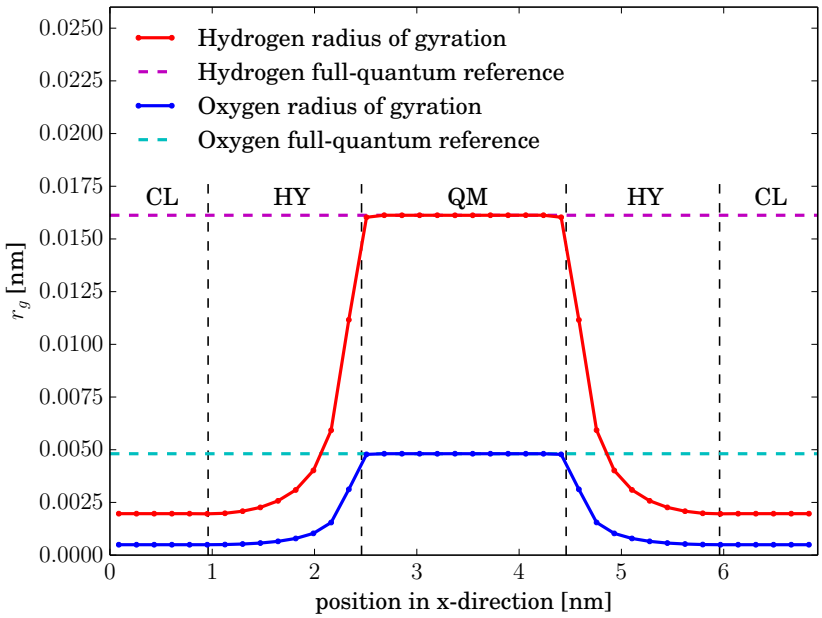

Figure 7.3.: Radii of gyration of the ring polymers corresponding to oxygen and hydrogen atoms as a function of the rings' position along the x-direction. The magenta and cyan lines correspond to the radii of gyration of atoms in full-quantum reference simulations.

data in Fig. 7.3 correspond to system 1 and that the other setups show the exact same behavior.

Using setup 1, we also calculate the water's radial distribution functions (RDFs) and the tetrahedral order parameter q_{tet} within the QM region (Fig. 7.4). We left a small buffer of 0.25 nm at the interface to the HY region and considered the inner 1.5 nm of the QM region in order to avoid artifacts by molecules at the outer edges of the QM region that interact strongly with molecules in the HY region. For a molecule i, q_{tet} is given by

$$q_{\text{tet}} = 1 - \frac{3}{8} \sum_{j=1}^{3} \sum_{k=j+1}^{4} \left(\cos(\theta_{j,k}) + \frac{1}{3} \right)^2 . \tag{7.64}$$

The indices j and k run over i's four nearest neighbor molecules and the angle $\theta_{j,k}$ is formed by the oxygen atoms of molecules i, j, and k with i in the center. The order parameter q_{tet} is defined such that it is 1 when the molecule forms a perfect tetrahedron with its four nearest neighbors and on average 0 for an ideal gas. The RDFs and q_{tet} in the QM region of the adaptive quantum–classical water systems perfectly match the results from full-quantum reference simulations. Consistent with previous work [78], we do not find any quantum effects for the tetrahedral order parameter q_{tet}. We conclude that the PI-based water structure in the QM region is undisturbed by the coupling to the CL particle reservoir.

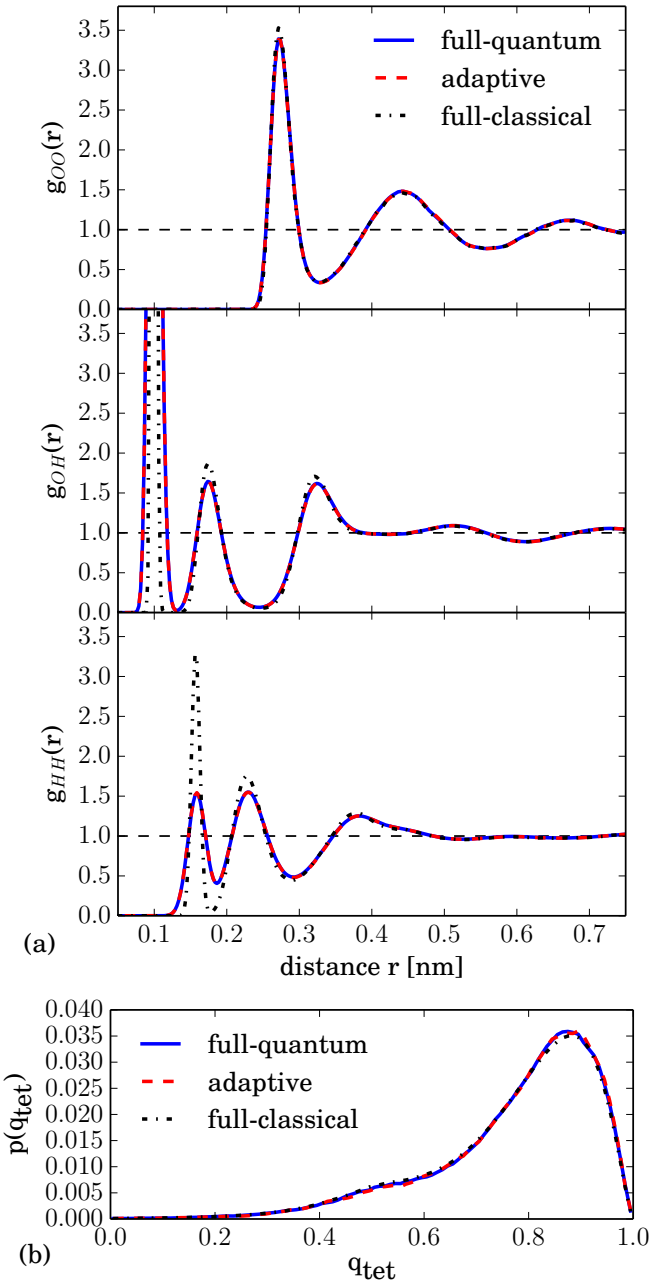

Figure 7.4.: (a) (Bead-bead) RDFs of the quantum–classical adaptive resolution simulations calculated in the QM region, and of full-quantum and full-classical ($P = 1$) reference simulations. (b) Same for the tetrahedral order parameter q_{tet}.

7.6.2. Dynamics

We also probe the dynamics in the inner QM region of the adaptive quantum–classical water systems. First, we calculate the vibrational spectrum from the water molecules' velocity

autocorrelation function. We do this both via TRPMD (setup 1) and CMD (setup 2) and compare the results to full-quantum and full-classical ($P = 1$) reference simulations (Fig. 7.5). The vibrational dynamics in the QM region perfectly reproduces the full-quantum reference data, both for CMD and TRPMD. While CMD and TRPMD give similar results, the classical system shows blue shifts in the H-O-H bending and O-H stretching modes. The spectra also agree with the results from Fritsch et al. [78].

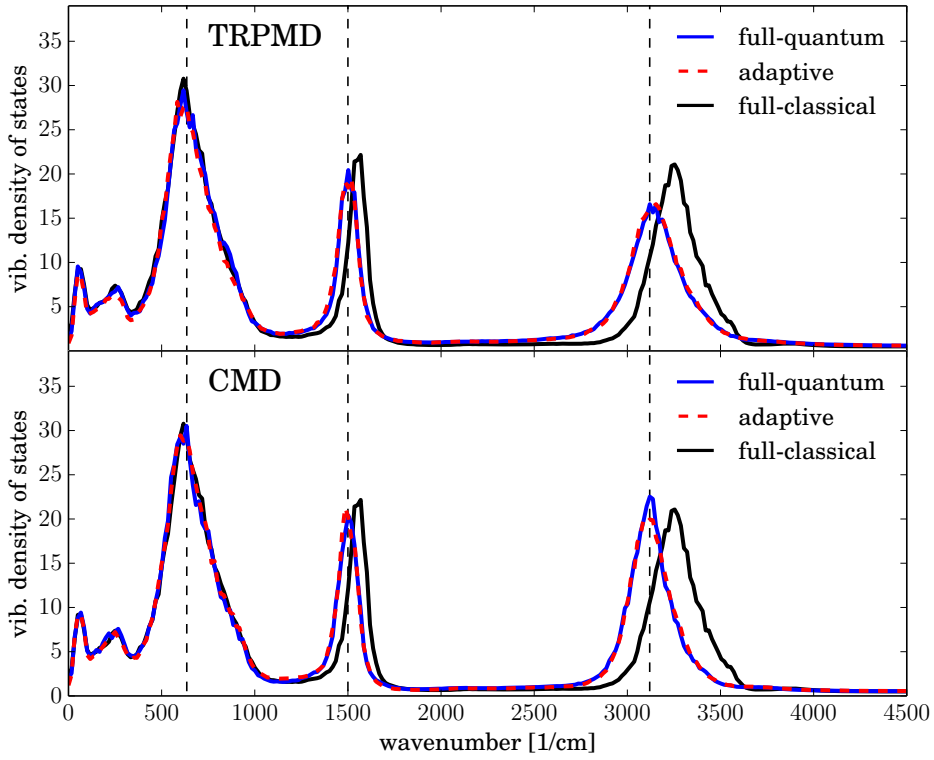

Figure 7.5.: Vibrational density of states in the QM region of the quantum–classical adaptive resolution simulations of liquid water and in full-quantum and full-classical reference simulations, calculated using TRPMD and CMD. The dashed vertical lines indicate the diffusion mode, the H-O-H bending mode, and the O-H stretching mode.

As hydrogen bonds play a critical role in the behavior of water [377–379], we additionally assess the hydrogen bonding kinetics of water in the QM region. The breaking and forming of hydrogen bonds can be characterized by the correlation function

$$C(t) = \frac{\langle h(0)h(t)\rangle}{\langle h\rangle}, \tag{7.65}$$

which measures fluctuations in the hydrogen bond populations throughout the system [380, 381]. The hydrogen bond population operator $h(t)$ is 1, if a particular pair of molecules is hydrogen bonded and 0 otherwise ($\langle h\rangle$ denotes the average of $h(t)$). We consider two molecules to be hydrogen bonded if the distance between their oxygen atoms is $< 3.5\,\text{Å}$ and the angle between

the O-O axis and one of the O-H bonds is $< 30°$. Based on Eq. 7.65, we can determine the hydrogen bond relaxation rate $k(t)$ as

$$k(t) = -\frac{\mathrm{d}C(t)}{\mathrm{dt}}. \tag{7.66}$$

The quantity $-k(t)$ can be interpreted as the average rate of change of hydrogen bonds that are broken at time t later. It has been widely used in studies of the hydrogen bond kinetics in liquid water [380–385].

Since $C(t)$ is defined to be either strictly 1 or 0, it cannot be simply averaged over the ring polymers' beads. Therefore, we employ the centroids for measuring the hydrogen bonds and calculate $C(t)$ via CMD using setup 2. We also perform full-quantum and full-classical reference simulations. The results are shown in Fig. 7.6. The hydrogen bonding kinetics in the QM region of the adaptive system reproduces the full-quantum reference within the statistical error. We conclude that the hydrogen bond kinetics in the QM region of the adaptive simulations is well preserved. Furthermore, we observe no quantum effects. The classical and the quantum system behave the same in their hydrogen bonding dynamics.

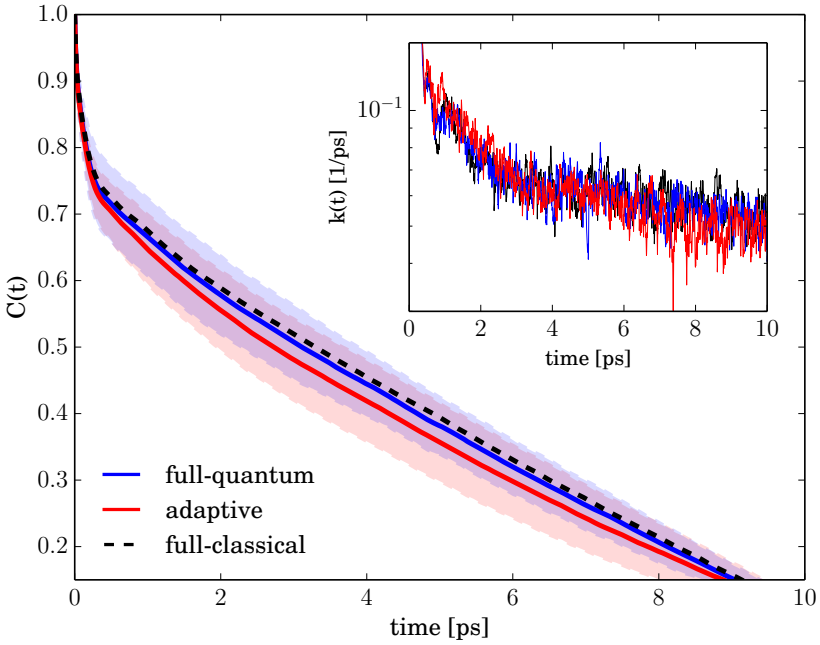

Figure 7.6.: Hydrogen bond population fluctuations characterized via the correlation function $C(t)$ in the QM region of the quantum–classical adaptive resolution simulations and in full-quantum and full-classical reference simulations. The shaded regions indicate the standard deviations of the data corresponding to the full-quantum and adaptive simulations. Inset: average hydrogen bond relaxation rates k(t) in a semi-log plot.

We conclude that not only the water structure but also the PI-based dynamics in the QM

region is unaffected by the coupling to the CL domain. Importantly, we have shown that one can apply both CMD and TRPMD in the proposed quantum–classical adaptive resolution simulation scheme.

7.6.3. Particle fluctuations

It is important that the proposed method allows for a free flow of particles through the HY region without any barriers. The QM region must behave as if embedded in an overall QM environment. To test this, we label all molecules that reside at the beginning of a simulation in the inner QM region, leaving a buffer of 0.25 nm. We then track how many of the labeled particles remain in the inner QM region after time t, and compare this to a full-quantum reference simulation in which we label and track all molecules in a similar subregion of the system. Note that we keep the thermostat on the ring polymers' centroid modes for these simulations. The results are presented in Fig. 7.7 (a) and show that the particles diffuse out of the QM region in the adaptive setup in a similar fashion as in the full-quantum system.

Additionally, we measure the particle number fluctuations in the inner QM region (Fig. 7.7 (b)). The fluctuations match the full-quantum reference nearly perfectly. Note that the data for the adaptive system in Fig. 7.7 correspond to setup 1. All other setups show similar behavior.

The results indicate that the HY region allows a free exchange of molecules between the CL and QM regions and that the QM region exchanges particles with its environment as if embedded in a full-quantum environment. Considering the complexity of the setup, the application of a correction force in the HY region, and the different structure and thermodynamics in the CL and QM subsystems, this is non-trivial. Because of the free flow of particles and the correct particle number fluctuations in the QM region, the scheme can, for example, be used for efficient simulations of open quantum systems.

7.7. Discussion and conclusions

We have proposed and validated a concurrent multiscale method for Hamiltonian adaptive resolution molecular dynamics simulations using the PI formalism. The scheme is based on a position-dependent particle mass, which controls the extension and collapse of the polymer rings. In the QM region, where the particles have their real masses, the ring polymers are extended, while in the CL region, where the mass is increased, the ring polymers collapse to point-like particles. Therefore, the interaction becomes classical and the dynamics obeys classical Newtonian mechanics in the CL region. The particles freely diffuse between the two regions and change their description on the fly. The method allows a more efficient evaluation of forces and energies in the CL domain, which leads to a speedup compared to full PI simulations. Importantly, we provide criteria that quantify to what extent such an adaptive PI setup is consistent with a bottom-up PI quantization. We want to point out that this differentiates our approach from related methodologies which are based on a direct interpolation of the forces corresponding to a classical and a PI system [177, 212, 216, 224, 240]. These techniques do not

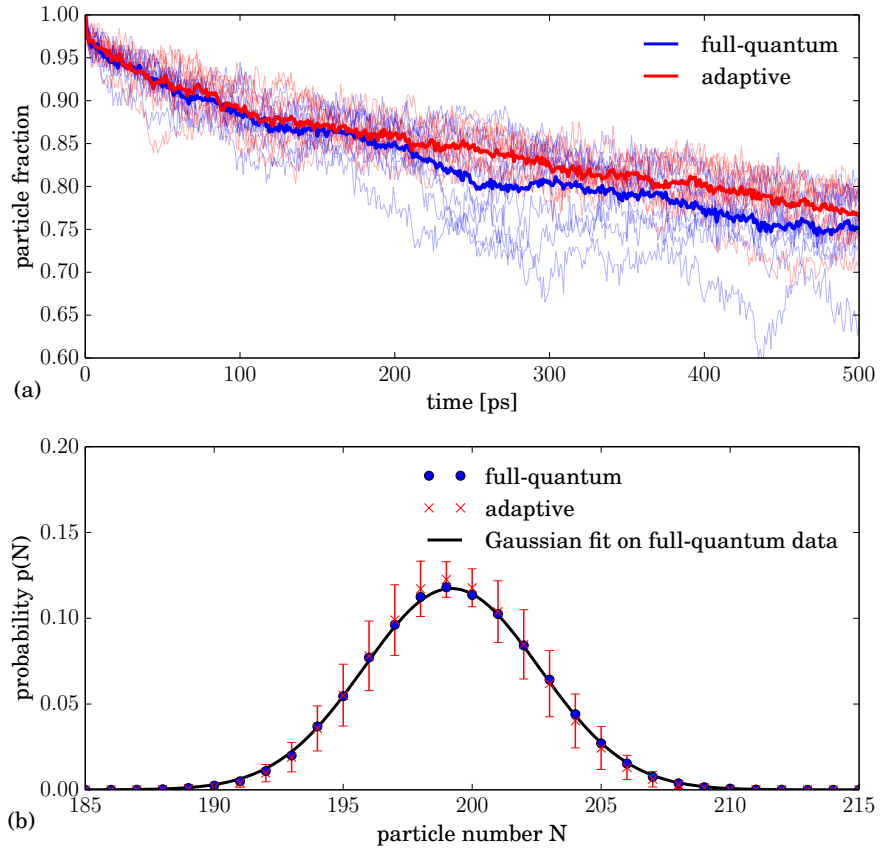

Figure 7.7.: (a) Fraction of molecules that resided at the beginning of the simulation in the inner quantum region and still stay there after time t for an adaptive system (setup 1) and for a full-quantum reference system, considering the same subregion. The thick lines denote the average from 10 simulations (thin lines). (b) Particle number probability distribution of the inner part of the quantum region in the adaptive simulations and of the same volume in a reference full-quantum simulation. The black curve is a Gaussian fit to the latter ($\mu = 199.2$, $\sigma = 3.4$). The error bars denote the standard deviation of the adaptive simulation data. The statistical error of the full-quantum data is similar.

allow a Hamiltonian description of the system [219], which, however, is the basis for a bottom-up PI treatment in the first place. Our scheme aims at overcoming this limitation. It allows both adaptive PIMD simulations sampling quantum statistical averages as well as quantum–classical RPMD and CMD, which enable us to calculate approximate quantum dynamical quantities and time correlation functions. Finally, the method allows one not only to selectively turn on and off nuclear quantum effects in different regions but also to change the intermolecular interaction potential. In this way, one can use a more efficient, possibly coarse-grained model in the CL region. This would be useful, for example, when the CL domain only serves as a particle reservoir.

To implement our methodology in a molecular dynamics framework, a kinetic energy term

needs to be introduced into the configurational energy obtained from PI quantization. At this point, our approach can be implemented in two different ways: The kinetic masses in this kinetic energy term can be chosen to be either constant throughout the whole system (CKM) or they can vary in a way similar to the particle masses that control the springs between the PI beads (AKM). The CKM approach results in a simpler scheme that can be integrated with a standard velocity Verlet integrator. However, it leads to strong thermodynamic imbalances between the CL and the QM regions and requires very small time steps due to the accelerated vibrations of the ring polymers in the CL region. On the other hand, the AKM method requires a more sophisticated integration scheme due to the position-dependent kinetic masses. We derived an integrator which is tailored to the problem, employs multiple time-stepping, and allows a symplectic integration of the equations of motion. This scheme also facilitates time steps which are much larger than in the CKM protocol and which are similar to those used in normal PI simulations. The AKM method also enables a smoother connection of the CL and QM systems, requiring a milder correction force in the HY coupling region.

The new integrator may appear complicated, but requires little additional overhead in practice. In molecular dynamics simulations, most time is typically spent for non-bonded force calculations and for inter-processor communication. However, these two tasks do not need to be performed more often than in a standard velocity Verlet integrator. In its essence, our methodology elegantly decouples the change of the particles' "quantumness", which is connected only to the higher modes and requires an additional decomposition step in the inner loop of the integrator, from the interatomic and intermolecular interactions, which are related to the more expensive bonded and non-bonded force calculations. In fact, it is only the masses of the higher modes that are position-dependent, while the masses of the centroid modes are constant. In our implementation the integration is performed in normal mode space. Only before the calculation of the bonded and non-bonded forces, the particles' real positions are updated and the force calculation is performed in real space. Afterwards, the forces are transformed back to normal modes. Therefore, the innermost loop of our integrator, in which the additional decomposition step occurs, does not require any inter-processor communication. When applying the methodology on systems in which only a very small part of the simulation domain is modeled quantum mechanically the additional overhead will be negligible compared to the gain in computational efficiency over a similar full QM system. We did not perform a detailed study of the speedup, though, as this depends in practice on a large number of factors, such as the system at hand, the scheme's implementation, the parallelization methodology, and the load balancing protocol (in highly parallelized simulations that employ many CPUs a suitable load balancing method that allows to concentrate computational resources in the QM region is crucial). Nonetheless, we have shown already in our previous paper that the speedup can be significant [365]. Provided the QM subsystem is small, the interatomic force computations can be accelerated by a factor of 10 or more. Note that the presented adaptive resolution method can of course also be used in setups with different geometrical arrangements of the QM and CL regions compared to the one in this article. A typical example would be

a small spherical QM domain positioned at an area of particular interest within a large CL system.

The proposed adaptive PI simulation scheme gains its efficiency by restricting the QM region to a small but relevant region in space and treating the rest of the system with a more efficient classical model. This is in contrast to other approaches that aim to alleviate the computational cost of PI simulations by modifying the PI calculations themselves, such as RPC [337, 338], higher order Trotter factorizations [346], or advanced thermostating techniques [342–345]. Our method is complementary to these approaches and could be combined with them. For example, one could apply RPC to further reduce the numerical effort in the QM region or use a colored noise instead of a white-noise Langevin thermostat, which would allow one to employ less Trotter beads. One could also further improve the multiple time-stepping and tailor it to the investigated systems.

The applications of the proposed methodology are diverse. The scheme is useful whenever only a small subdomain of an overall large system needs to be described including PIs. This can be the case, for example, in biomolecular systems, in which the study of nuclear quantum effects has gained significant interest [80, 81, 92–96, 312–319]. Biological systems are often complicated and quantum delocalization plays an important role usually only in a small part of the system, such as the active site of proteins [81]. Our multiscale method could be used to describe the active site quantum mechanically and an efficient classical model could be employed for the rest of the system, in the same spirit of QM/MM approaches but at a different level of "quantumness". This would allow an extension of the accessible length and time scales compared to full path integral simulations. Similar applications of the scheme are simulations of interfaces or membranes. The possibility to selectively switch on and off the nuclear quantum effects in different regions also allows one to investigate the locality of quantum properties. One can ask, for example, how much quantum mechanically modeled environment is required to support the quantum mechanical features in a certain subregion [185, 187, 374, 386]. This would not be possible in bulk PI simulations in a straightforward way. Furthermore, the method enables an efficient simulation of a quantum grand canonical ensemble: a QM region can be coupled to a large particle reservoir, which itself is described classically. Yet another interesting possible application of our methodology is its combination with the aforementioned QM/MM techniques, which concurrently couple *ab initio* and classical empirical force fields. Recently, Boereboom et al. [169] proposed an adaptive QM/MM method based on the Hamiltonian adaptive resolution scheme. The latter is also used in the PI-based adaptive resolution scheme presented in this article. In fact, although the interatomic potentials employed in this work are empirical force fields, the forces and energies could also come from *ab initio* calculations. Therefore, one could combine our approach with the one from Boereboom et al. and construct a Hamiltonian adaptive QM/MM scheme that also incorporates a multiscale treatment of PIs.

Finally, we would like to point out that the derived concurrent multiscale PI simulation methodology has been implemented in the ESPResSo++ package [1] and is publicly available.

Chapter 8.

Conclusions

In this chapter, we conclude the thesis. In section 8.1, we present a brief summary of the main results, and in section 8.2 we provide an outlook and describe ongoing work as well as possible directions for future research on adaptive resolution simulations and their applications.

8.1. Summary

While the introductory part of this thesis, chapter 1, provided an overview over relevant molecular simulation methodologies and introduced the adaptive resolution simulation scheme, the chapters 2-7 contain the scientific findings established in this work.

Chapter 2—Coupling to an ideal gas

In chapter 2, we coupled liquid water to an ideal gas of non-interacting particles and showed that it is possible to connect two dynamically and structurally very different systems with each other without significantly affecting the accuracy of the high-resolution region. We found that great care needs to be taken when deriving the correction forces acting in the coupling region. Since an ideal gas does not require any pair-force calculations, we expect a great speedup in adaptive ideal gas–water simulations compared to all-atom systems. We ran adaptive resolution simulations with different ratios of AT to CG region volumes and demonstrated that for large systems, the force computation can be accelerated by a factor of 3 or higher. Coupling a high-resolution region to an ideal gas, as demonstrated here, may also find useful applications as an alternative to standard grand-canonical simulations [387, 388], since particle insertion in the ideal gas region is trivial. For example, Mukherji and Kremer showed that an efficient particle exchange in the CG domain allows one to control the chemical potential in adaptive resolution simulations of mixtures [175]. Note, however, that they employed a CG potential based on iterative Boltzmann inversion [106] instead of an ideal gas.

Chapter 3—Unifying force-based and energy-based AdResS

In chapter 3, we derived a unified framework for force-based and energy-based AdResS simulations. Starting from the H–AdResS Hamiltonian, we demonstrated that the drift force, when its

average is removed by the free energy correction, is equivalent to a non-Markovian noise. Therefore, it is exactly this "noise" that guarantees the energy conservation in energy-based AdResS compared to force-based AdResS. We also constructed the corresponding memory-dependent friction kernel from the drift force's autocorrelation function and defined a generalized Langevin thermostat with memory. Applying this Langevin scheme on the H–AdResS Hamiltonian such that the thermostat's noise term and the drift force cancel out, we obtained the corresponding force-based AdResS approach with a non-Markovian friction that exactly removes the excess heat generated in the hybrid region. These findings demonstrate that the force-based AdResS approach can be directly derived from H–AdResS by applying a colored noise thermostat that is based on the drift force.

Chapter 4—Adaptive resolution simulations and the relative entropy

In chapter 4, we built on the results from chapter 3, showing that the amplitude of the drift force's autocorrelation function at mixed resolution λ is proportional to the derivative of the relative entropy between the Boltzmann distributions of the system at resolution λ and the CG system. This suggests that minimizing the relative entropy between the AT and CG potentials helps reduce both the drift force in H–AdResS as well as the energy drift in AdResS. We validated that this is the case, using liquid water as a model system. Additionally, we derived expressions that demonstrate that the excess heat produced by individual particles in the hybrid region in force-based AdResS scales with the width of the hybrid region d_{HY} with $1/d_{\mathrm{HY}}^2$ while the overall energy drift scales with $1/d_{\mathrm{HY}}$. This was also validated by numerical simulations. The findings do not only provide a clearer picture of the inner workings of adaptive resolution simulations but they also provide a framework for setting up AdResS simulations in an optimal fashion. On the one hand, one can, for example, use the relative entropy based coarse-graining method by Shell and coworkers to derive CG potentials that perform well in adaptive settings [108, 282–284]. On the other hand, the results provide a guide on how to tune the width of the hybrid region to reduce the undesired drift force in H–AdResS or the energy drift in force-based AdResS.

Chapter 5—Towards arbitrary and self-adjusting geometries

In chapter 5, we devised a scheme to overcome the previous geometrical constraints in adaptive resolution simulations. Associating overlapping and moving spherical high-resolution regions with several particles in the system, we constructed atomistic regions with arbitrary geometry. This, however, complicates the application of correction forces in the HY region, which may attain complex shapes. We focused on the force-based approach and derived a robust method that enables the successful application of the thermodynamic force even in scenarios with highly distorted AT and HY regions. The scheme even allows to freely adapt the geometry during the simulation to follow, for example, conformational transitions of macromolecules. The technique was validated by simulations of a folding peptide and it was shown that any introduced disturbances are negligible. Additionally, it was made available to the public and

implemented in the ESPResSo++ software package [1]. In conclusion, the new methodology represents a significant improvement to the AdResS framework, allowing efficient simulations with complex and adaptive high-resolution region geometries.

Chapter 6—Rigorous classical/path integral adaptive resolution simulations

In chapter 6, we derived a path integral-based adaptive resolution method that allows a clean coupling of classical force fields and path integral-based models. The scheme is derived from the bottom up, is Hamiltonian by construction, and is, opposed to previous work, consistent with a formal path integral quantization procedure. The central idea is to force the ring polymers to collapse to point-like particles in the classical region by introducing a position-dependent mass, allowing for an efficient evaluation of pair forces and energies. We validated the proposed approach by quantum–classical adaptive resolution simulations of liquid parahydrogen, sampling the adaptive Hamiltonian with a Markov chain Monte Carlo algorithm. Most importantly, we demonstrated that the technique leads to a significant speedup compared to all-quantum simulations. Finally, our method is also compatible with other protocols, such as ring polymer contraction [337, 338], colored-noise thermostats [342–345], or higher order Trotter factorizations [346], to further speed up path integral simulations. Its applications include quantum–classical adaptive resolution simulations of biological or chemical systems as well as the efficient simulation of quantum grand canonical ensembles.

Chapter 7—Hamiltonian adaptive resolution path integral molecular dynamics

Chapter 7 is based on the results of the previous chapter. We demonstrated how the proposed path integral-based quantum–classical adaptive resolution method can be implemented in a molecular dynamics framework and introduced a kinetic energy term into the configurational energy obtained from path integral quantization. Additionally, we switched to a description of the system in terms of normal modes. In this picture, the higher modes correspond to the vibrations and the extension of the ring polymers and hence determine the "quantumness" of the particles, while the centroid mode is associated with the displacement of the ring polymers as a whole. Denoting the masses in the kinetic energy term as kinetic masses, we choose the kinetic masses of the higher modes to be similarly adaptive as the masses associated with the spring constants between the beads of the ring polymers. The centroid mode's kinetic mass, however, is kept constant throughout the system. In this way, only the particles' "quantumness" changes, while their macroscopic behavior is the same everywhere. We then derived a tailored multiple time-stepping integrator which allows a symplectic integration of the resulting equations of motion. We validated the method on adaptive resolution simulations of liquid water and we showed that it is also possible to perform ring polymer and centroid molecular dynamics, which allow the calculation of approximate quantum dynamics and quantum time correlation functions, in this framework. The methodology was implemented in the ESPResSo++ package [1] and is available to the public.

8.2. Outlook

In this thesis, we investigated the what and how of adaptive resolution simulations and provided insights into the connection of AdResS and H–AdResS. We also devised strategies for setting them up in an optimal fashion. Additionally, we developed algorithms that enhance the current AdResS methodology. Although many challenges remain, our results pave the way for novel applications of the AdResS framework, while also clarifying directions for further technical developments.

Simulations of biological macromolecules, i.e., proteins, have received considerable attention as a particularly useful application of the AdResS scheme [155, 188, 190]. Among other reasons, a large portion of the computational cost of these simulations is due to solvent molecules which may have little impact on the properties of interest when residing far away from the protein. In this thesis, we have developed tools geared towards such applications, including a method for flexible and self-adjusting geometries of the high-resolution region. The scheme is particularly suited for the simulation of complex biological systems that undergo conformational transitions, such a proteins. Our work enables efficient AdResS simulations of larger and geometrically complex proteins that do not fit well in simple spherical or cylindrical atomistic regions. Importantly, our method allows efficient protein folding studies in adaptive resolution setups. A variety of other systems, such as fluctuating membranes and interfaces as well as aggregating particles, could also benefit from a geometrically flexible adaptive modeling. Expanding upon this idea, the method may even be combined with other recent technical developments in AdResS and not only the solvent molecules could be modeled with an adaptive resolution and in a geometrically flexible fashion but also the protein itself. In fact, the work by Fogarty et al. [190], which employs two different protein models within a single adaptive resolution framework, is the first step into the direction of AdResS simulations using a dual-resolution description also for the protein itself. However, the parts of the protein that are described differently are fixed throughout the simulation while only the surrounding water is modeled in an adaptive fashion. This indicates another challenge: To date, AdResS has mostly been used for the interpolation of non-bonded force fields mapping individual solvent molecules like water onto single CG beads. In order to more flexibly simulate proteins in adaptive resolution it would be advantageous to be able to smoothly interpolate also different bonded (and angular, etc.) interactions. In related adaptive resolution methods [389, 390], this issue has been addressed with advanced backmappings from the CG to the AT level. In AdResS, it was recently shown that four grouped water molecules can be described as CG supramolecules which consist of three bonded CG beads [215]. This aspect needs to be investigated in more detail to enable the interpolation of more complex AT and CG force fields.

Most work on biological systems with AdResS consists of validating the preservation of basic structural and dynamic quantities in the atomistic region. But what about properties that are not as simple to measure? For example, can we also perform free energy calculations using the AdResS method? A next step will be to combine AdResS with techniques such as thermodynamic integration.

Another area for future research is the application of the derived quantum–classical coupling scheme in more complicated situations and on more complex systems. For example, one could describe only the active site of a protein including quantum effects, while treating the rest atomistically. One could even attempt to couple a third resolution by representing the solvent far from the protein with a CG description. This could be particularly useful when studying, for example, protein–ligand binding. Using the methodology derived in chapter 7, we could even calculate approximate quantum dynamical properties at the active site, at which nuclear quantum effects have been shown to play an important role [81]. One could also combine the adaptive quantum–classical simulation method with adaptive QM/MM schemes. This may lead to a clean adaptive QM/MM methodology that allows an efficient inclusion of path integrals. An adaptive QM/MM technique that would be particularly suited for this is one that was recently proposed by Boereboom et al. [169], which is also based on H–AdResS. In the long term, such technical advancements could pave the way to efficient multi-resolution simulations of large biological systems spanning several levels of resolution.

In chapter 2, we studied the adaptive coupling of liquid water with an ideal gas of non-interacting particles and found that it is possible to smoothly connect these very different systems. Building on the work by Mukherji and Kremer [174, 175], one could exploit these findings and use an ideal gas as CG model in adaptive resolution simulations of liquid mixtures. The trivial insertion and deletion of particles in the ideal gas region would allow an exact control of the concentration and the chemical potential in the system. This could be useful in the context of free energy calculations in a semi-grand canonical ensemble.

There are several more technical questions in adaptive resolution simulations that are still open. For example, we have mentioned already that, in principle, AdResS and H–AdResS could be generalized to also allow the interpolation of non-bonded 3-body force-fields or even higher order interactions. This needs to be investigated and tested in practice. Furthermore, the adaptive resolution methodology has been mainly developed with the idea to gain a computational speedup from the reduced number of non-bonded interaction sites in the CG region. This neglects another key aspect of CG models. Their computational efficiency compared to atomistic systems not only stems from the reduced number of degrees of freedom. Additionally, the CG energy landscape is usually strongly smoothened out which leads to much softer interactions. In AdResS, ongoing work is attempting to exploit this by using multiple time-stepping algorithms that evaluate CG forces significantly less often than the AT interactions, in order to gain additional computational speedup.

Even the most elaborate and advanced molecular simulation algorithm is useless without an efficient implementation in an optimized software package. This is particularly true for adaptive resolution methods whose main purpose is to run fast simulations and that, if not programmed correctly, may lead to a significant numerical overhead. For example, the idea of AdResS relies on the possibility to flexibly shift and distribute computational resources between different parts of the simulation box and to concentrate them in the high-resolution region. This is only possible with advanced schemes for domain decomposition and load balancing. There

currently exists no simulation package that has implemented all available adaptive resolution methods and algorithms and that also offers the kind of optimized environment that makes them run efficiently. An important step into the direction of an optimal implementation of the complete AdResS methodology is its implementation in the ESPResSo++ package [1], although also in ESPResSo++ there is still room for improvement in terms of speed and computational efficiency. This is possibly one of the most pressing open challenges in the context of AdResS and H–AdResS. If these hurdles are overcome, however, adaptive resolution techniques have the potential to open the door to molecular simulations of complex multiscale systems, in particular in the biological and chemical area, that are otherwise impossible and they may provide new insights into important open questions in biology and chemistry.

Appendix A.

Contributions

The chapters 2–6 have been published as scientific research papers and chapter 7 is a draft for another paper soon to be submitted for publication. These works are the results of fruitful collaborations with colleagues from the Max Planck Institute for Polymer Research, New York University, and the University of California, Davis. Here, we lay out in detail the individual contributions.

Chapter 2

The simulation setups were conceived by Aoife Fogarty, Karsten Kreis, and Raffaello Potestio. Karsten Kreis implemented the energy-based H–AdResS approach into ESPResSo++ and ran all H–AdResS simulations, while Aoife Fogarty performed the simulations of the force-based AdResS systems. The data analysis was conducted by Aoife Fogarty and Karsten Kreis and the paper was written by Karsten Kreis, Aoife Fogarty and Raffaello Potestio. Kurt Kremer contributed important ideas and support.

Chapter 3

Davide Donadio and Raffaello Potestio initiated the project. Karsten Kreis, Davide Donadio and Raffaello Potestio jointly derived the theory and designed the simulation setups. Karsten Kreis implemented the non-Markovian generalized Langevin friction into ESPResSo++, performed all simulations and evaluated, together with Raffaello Potestio, the results. The paper was written by Raffaello Potestio and Karsten Kreis with critical support from Davide Donadio. Kurt Kremer pushed the project with ideas and discussions.

Chapter 4

Raffaello Potestio had the initial idea of introducing the relative entropy in the adaptive resolution scheme. Karsten Kreis and Raffaello Potestio jointly derived the final theoretical framework. The simulations were designed, run and evaluated by Karsten Kreis. Also the paper was written by Karsten Kreis, incorporating critical comments from Raffaello Potestio.

Chapter 5

Karsten Kreis initiated the project and derived the algorithm to interpolate the thermodynamic force and to construct arbitrarily shaped high-resolution regions formed by different overlapping spheres. Karsten Kreis also implemented the methodology into ESPResSo++. The simulation setups were designed, run, and evaluated jointly by Aoife Fogarty and Karsten Kreis. More specifically, Karsten Kreis focused on the water systems, while Aoife Fogarty provided important expertise on the peptide system. Also the paper was written jointly by Aoife Fogarty and Karsten Kreis. Raffaello Potestio and Kurt Kremer supported the work with ideas and discussions as well as critical feedback.

Chapter 6

The methodology was derived jointly by Karsten Kreis, Mark Tuckerman and Raffaello Potestio. The simulation setups were conceived by Karsten Kreis and Raffaello Potestio. Karsten Kreis and Raffaello Potestio also developed the Monte Carlo code. More precisely, Raffaello Potestio provided a simple version of the code, which was used for another project before, and Karsten Kreis extended it for the quantum–classical adaptive Monte Carlo simulations. All simulations were run and evaluated by Karsten Kreis. The paper was written by Karsten Kreis, Raffaello Potestio and Mark Tuckerman with critical feedback from Davide Donadio. Kurt Kremer and Davide Donadio provided insightful ideas and help.

Chapter 7

The methodology and the integration scheme was derived by Karsten Kreis with crucial help from Mark Tuckerman, who supervised the first part of the project. The implementation of the approach was done by Karsten Kreis and also the simulation setups were conceived, run, and evaluated by Karsten Kreis. The draft was written by Karsten Kreis with critical feedback from Raffaello Potestio, who also supported the work with ideas and discussions. A crucial part of the project was performed during a research at New York University in the group of Mark Tuckerman. The contact with Mark Tuckerman was initiated by Kurt Kremer.

Appendix B.

Abbreviations

Throughout this thesis several abbreviations are used for the sake of appropriate readability. Here is an overview, sorted alphabetically.

Abbreviation	**Meaning**
AdResS	Adaptive resolution simulation scheme
AKM	Adaptive kinetic mass
AT	Atomistic
CG	Coarse-grained
CKM	Constant kinetic mass
CL	Classical
CMD	Centroid molecular dynamics
DFT	Density functional theory
EB	Energy-based
FB	Force-based
FEC	Free energy compensation / free energy correction
FM	Forcematching
GLE	Generalized Langevin equation
H–AdResS	Hamiltonian adaptive resolution simulation scheme
HY	Hybrid
IBI	Iterative Boltzmann inversion
IMC	Inverse Monte Carlo
KTI	Kirkwood thermodynamic integration
MC	Monte Carlo
MD	Molecular dynamics
MTS	Multiple time-stepping
NVE	Microcanonical ensemble
NVT	Canonical ensemble
PDF	Probability density function
PI	Path integral
PILE	Path integral Langevin equation

Abbreviation	**Meaning**
PIMC	Path integral Monte Carlo
PIMD	Path integral molecular dynamics
PMF	Potential of mean force
QM	Quantum
QM/MM	Quantum mechanics/molecular mechanics
RDF	Radial distribution function
RE	Relative entropy
RPC	Ring polymer contraction
RPMD	Ring polymer molecular dynamics
TF	Thermodynamic force
TRPMD	Thermostated ring polymer molecular dynamics
VACF	Velocity autocorrelation function
WCA	Weeks-Chandler-Anderson

Bibliography

[1] J. D. Halverson, T. Brandes, O. Lenz, A. Arnold, S. Bevc, V. Starchenko, K. Kremer, T. Stuehn, and D. Reith, "ESPResSo++: A modern multiscale simulation package for soft matter systems," *Computer Physics Communications*, vol. 184, no. 4, pp. 1129–1149, 2013.

[2] J. R. Pasta, S. M. Ulam, and E. Fermi, "Studies of Nonlinear Problems," in *Collected Papers of Enrico Fermi* (E. Segrè, ed.), vol. 2, pp. 978–988, Chicago: Univ. of Chicago Press, 1965.

[3] N. J. Zabusky and M. D. Kruskal, "Interaction of "Solitons" in a Collisionless Plasma and the Recurrence of Initial States," *Physical Review Letters*, vol. 15, no. 6, pp. 240–243, 1965.

[4] B. J. Alder and T. E. Wainwright, "Studies in Molecular Dynamics. I. General Method," *Journal of Chemical Physics*, vol. 31, no. 2, p. 459, 1959.

[5] A. Rahman, "Correlations in the Motion of Atoms in Liquid Argon," *Physical Review*, vol. 136, no. 2A, pp. A405–A411, 1964.

[6] D. W. Borhani and D. E. Shaw, "The future of molecular dynamics simulations in drug discovery," *Journal of Computer-Aided Molecular Design*, vol. 26, no. 1, pp. 15–26, 2012.

[7] P. Deglmann, A. Schäfer, and C. Lennartz, "Application of quantum calculations in the chemical industry—An overview," *International Journal of Quantum Chemistry*, vol. 115, no. 3, pp. 107–136, 2015.

[8] T. Bereau, D. Andrienko, and K. Kremer, "Research Update: Computational materials discovery in soft matter," *APL Materials*, vol. 4, no. 5, p. 053101, 2016.

[9] R. Olivares-Amaya, C. Amador-Bedolla, J. Hachmann, S. Atahan-Evrenk, R. S. Sánchez-Carrera, L. Vogt, and A. Aspuru-Guzik, "Accelerated computational discovery of high-performance materials for organic photovoltaics by means of cheminformatics," *Energy & Environmental Science*, vol. 4, no. 12, pp. 4849–4861, 2011.

[10] P. Kordt, J. J. M. Van Der Holst, M. Al Helwi, W. Kowalsky, F. May, A. Badinski, C. Lennartz, and D. Andrienko, "Modeling of Organic Light Emitting Diodes: From Molecular to Device Properties," *Advanced Functional Materials*, vol. 25, no. 13, pp. 1955–1971, 2015.

[11] M. Schwarze, W. Tress, B. Beyer, F. Gao, R. Scholz, C. Poelking, K. Ortstein, A. A. Günther, D. Kasemann, D. Andrienko, and K. Leo, "Band structure engineering in organic semiconductors," *Science.*, vol. 352, no. 6292, pp. 1446–1449, 2016.

[12] S. H. White and W. C. Wimley, "Membrane Protein Folding and Stability: Physical Principles.," *Annual review of biophysics and biomolecular structure*, vol. 28, pp. 319–365, 1999.

[13] L. R. Forrest and M. S. Sansom, "Membrane simulations: Bigger and better?," *Current Opinion in Structural Biology*, vol. 10, no. 2, pp. 174–181, 2000.

[14] G. S. Ayton and G. A. Voth, "Systematic multiscale simulation of membrane protein systems," *Current Opinion in Structural Biology*, vol. 19, no. 2, pp. 138–144, 2009.

[15] M. Deserno, "Mesoscopic membrane physics: Concepts, simulations, and selected applications," *Macromolecular Rapid Communications*, vol. 30, no. 9-10, pp. 752–771, 2009.

[16] P. J. Stansfeld and M. S. Sansom, "Molecular Simulation Approaches to Membrane Proteins," *Structure*, vol. 19, no. 11, pp. 1562–1572, 2011.

[17] V. Daggett, "Protein Folding-Simulation," *Chemical Reviews*, vol. 106, no. 5, pp. 1898–1916, 2006.

[18] Y. Chen, F. Ding, H. Nie, A. W. Serohijos, S. Sharma, K. C. Wilcox, S. Yin, and N. V. Dokholyan, "Protein folding: Then and now," *Archives of Biochemistry and Biophysics*, vol. 469, no. 1, pp. 4–19, 2008.

[19] P. L. Freddolino, C. B. Harrison, Y. Liu, and K. Schulten, "Challenges in protein-folding simulations," *Nature Physics*, vol. 6, no. 10, pp. 751–758, 2010.

[20] T. J. Lane, D. Shukla, K. A. Beauchamp, and V. S. Pande, "To milliseconds and beyond: challenges in the simulation of protein folding," *Current Opinion in Structural Biology*, vol. 23, no. 1, pp. 58–65, 2013.

[21] D. E. Shaw, K. J. Bowers, E. Chow, M. P. Eastwood, D. J. Ierardi, J. L. Klepeis, J. S. Kuskin, R. H. Larson, K. Lindorff-Larsen, P. Maragakis, M. A. Moraes, R. O. Dror, S. Piana, Y. Shan, B. Towles, J. K. Salmon, J. P. Grossman, K. M. Mackenzie, J. A. Bank, C. Young, M. M. Deneroff, and B. Batson, "Millisecond-Scale Molecular Dynamics Simulations on Anton," in *Proceedings of the Conference on High Performance Computing, Networking, Storage and Analysis (SC09)*, pp. 1–11, 2009.

[22] R. O. Dror, M. Ø. Jensen, D. W. Borhani, and D. E. Shaw, "Exploring atomic resolution physiology on a femtosecond to millisecond timescale using molecular dynamics simulations," *Journal of General Physiology*, vol. 135, no. 6, pp. 555–562, 2010.

[23] R. O. Dror, R. M. Dirks, J. P. Grossman, H. Xu, and D. E. Shaw, "Biomolecular simulation: a computational microscope for molecular biology.," *Annual Review of Biophysics*, vol. 41, pp. 429–452, 2012.

[24] D. E. Shaw, J. P. Grossman, J. A. Bank, B. Batson, J. A. Butts, J. C. Chao, M. M. Deneroff, R. O. Dror, A. Even, C. H. Fenton, A. Forte, J. Gagliardo, G. Gill, B. Greskamp, C. R. Ho, D. J. Ierardi, L. Iserovich, J. S. Kuskin, R. H. Larson, T. Layman, L. S. Lee, A. K. Lerer, C. Li, D. Killebrew, K. M. Mackenzie, S. Y. H. Mok, M. A. Moraes, R. Mueller, L. J. Nociolo, J. L. Peticolas, T. Quan, D. Ramot, J. K. Salmon, D. P. Scarpazza, U. B. Schafer, N. Siddique, C. W. Snyder, J. Spengler, P. T. P. Tang, M. Theobald, H. Toma, B. Towles, B. Vitale, S. C. Wang, and C. Young, "Anton 2: Raising the bar for performance and programmability in a special-purpose molecular dynamics supercomputer," in *Proceedings of the Conference on High Performance Computing, Networking, Storage and Analysis (SC14)*, pp. 41–53, 2014.

[25] H. Lin and D. G. Truhlar, "QM/MM: What have we learned, where are we, and where do we go from here?," *Theoretical Chemistry Accounts*, vol. 117, no. 2, pp. 185–199, 2007.

[26] K. Burke, "Perspective on density functional theory," *Journal of Chemical Physics*, vol. 136, p. 150901, 2012.

[27] R. O. Jones, "Density functional theory: Its origins, rise to prominence, and future," *Reviews of Modern Physics*, vol. 87, no. 3, pp. 897–923, 2015.

[28] R. Pariser and R. G. Parr, "A Semi–Empirical Theory of the Electronic Spectra and Electronic Structure of Complex Unsaturated Molecules. I.," *Journal of Chemical Physics*, vol. 21, no. 3, pp. 466–471, 1953.

[29] R. Pariser and R. G. Parr, "A Semi–Empirical Theory of the Electronic Spectra and Electronic Structure of Complex Unsaturated Molecules. II," *Journal of Chemical Physics*, vol. 21, no. 5, pp. 767–776, 1953.

[30] J. A. Pople, "Electron interaction in unsaturated hydrocarbons," *Transactions of the Faraday Society*, vol. 49, pp. 1375–1385, 1953.

[31] R. G. Parr and W. Yang, "Density-Functional Theory of the Electronic Structure of Molecules," *Annual Review of Physical Chemistry*, vol. 46, pp. 701–728, 1995.

[32] P. Hohenberg and W. Kohn, "Inhomogeneous Electron Gas," *Physical Review*, vol. 136, no. 3B, pp. B864–B871, 1964.

[33] S. F. Sousa, P. A. Fernandes, and M. J. Ramos, "General Performance of Density Functionals," *Journal of Physical Chemistry A*, vol. 111, no. 42, pp. 10439–10452, 2007.

[34] D. Marx and J. Hutter, *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods*. Cambridge: Cambridge University Press, 2009.

[35] G. I. Csonka, A. Ruzsinszky, J. Tao, and J. P. Perdew, "Energies of organic molecules and atoms in density functional theory," *International Journal of Quantum Chemistry*, vol. 101, no. 5, pp. 506–511, 2005.

[36] R. P. Muller, A. E. Mattsson, and C. L. Janssen, "Calculation of chemical reaction energies using the AM05 density functional," *Journal of Computational Chemistry*, vol. 31, no. 9, pp. 1860–1863, 2010.

[37] B. Baumeier, J. Kirkpatrick, and D. Andrienko, "Density-functional based determination of intermolecular charge transfer properties for large-scale morphologies.," *Physical Chemistry Chemical Physics*, vol. 12, no. 36, pp. 11103–11113, 2010.

[38] J. Wu and Z. Li, "Density-Functional Theory for Complex Fluids," *Annual Review of Physical Chemistry*, vol. 58, pp. 85–112, 2007.

[39] B. Baumeier, F. May, C. Lennartz, and D. Andrienko, "Challenges for in silico design of organic semiconductors," *Journal of Materials Chemistry*, vol. 22, pp. 10971–10976, 2012.

[40] W. J. Hehre, L. Radom, P. v. R. Schleyer, and J. Pople, *AB INITIO Molecular Orbital Theory*. New York: John Wiley and Sons, Inc., 1986.

[41] D. R. Hartree, "The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part I. Theory and Methods," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 24, no. 1, pp. 89–110, 1928.

[42] V. Fock, "Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems," *Zeitschrift für Physik*, vol. 61, no. 1, pp. 126–148, 1930.

[43] P. Echenique and J. L. Alonso, "A mathematical and computational review of Hartree-Fock SCF methods in quantum chemistry," *Molecular Physics*, vol. 105, no. 23-24, pp. 3057–3098, 2007.

[44] R. J. Bartlett, "Many-Body Perturbation Theory and Coupled Cluster Theory for Electron Correlation in Molecules," *Annual Review of Physical Chemistry*, vol. 32, pp. 359–401, 1981.

[45] R. J. Bartlett, "Coupled-cluster approach to molecular structure and spectra: a step toward predictive quantum chemistry," *Journal of Physical Chemistry*, vol. 93, no. 5, pp. 1697–1708, 1989.

[46] C. J. Cramer, *Essentials of Computational Chemistry: Theories and Models*. Chichester, England: John Wiley & Sons, Inc., 2004.

[47] P. H. Acioli, "Review of quantum Monte Carlo methods and their applications," *Journal of Molecular Structure: THEOCHEM*, vol. 394, no. 2-3, pp. 75–85, 1997.

[48] J. Kolorenč and L. Mitas, "Applications of quantum Monte Carlo methods in condensed systems," *Reports on Progress in Physics*, vol. 74, no. 2, p. 026502, 2011.

[49] W. Thiel, "Semiempirical quantum-chemical methods," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 4, no. 2, pp. 145–157, 2014.

[50] A. S. Christensen, T. Kubař, Q. Cui, and M. Elstner, "Semiempirical quantum mechanical methods for noncovalent interactions for chemical and biochemical applications," *Chemical Reviews*, vol. 116, no. 9, pp. 5301–5337, 2016.

[51] D. Frenkel and B. Smit, *Understanding Molecular Simulation*. San Diego: Academic Press, 2002.

[52] M. E. Tuckerman, *Statistical Mechanics: Theory and Molecular Simulation*. New York: Oxford University Press, 2010.

[53] P. E. M. Lopes, O. Guvench, and A. D. MacKerell Jr., "Current Status of Protein Force Fields for Molecular Dynamics Simulations," in *Molecular Modeling of Proteins* (A. Kukol, ed.), pp. 47–71, Springer New York, 2015.

[54] W. C. Swope, H. C. Andersen, P. H. Berens, and K. R. Wilson, "A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters," *Journal of Chemical Physics*, vol. 76, no. 1, pp. 637–649, 1982.

[55] S. Miyamoto and P. A. Kollman, "SETTLE: an analytical version of the SHAKE and RATTLE algorithm for rigid water models," *Journal of Computational Chemistry*, vol. 13, no. 8, pp. 952–962, 1992.

[56] 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*, vol. 81, no. 8, pp. 3684–3690, 1984.

[57] G. Bussi, D. Donadio, and M. Parrinello, "Canonical sampling through velocity rescaling," *Journal of Chemical Physics*, vol. 126, no. 1, p. 014101, 2007.

[58] S. Nosé, "A unified formulation of the constant temperature molecular dynamics methods," *Journal of Chemical Physics*, vol. 81, no. 1, pp. 511–519, 1984.

[59] W. G. Hoover, "Canonical dynamics: Equilibrium phase-space distributions," *Physical Review A*, vol. 31, no. 3, pp. 1695–1697, 1985.

[60] G. J. Martyna, M. L. Klein, and M. Tuckerman, "Nosé–Hoover chains: The canonical ensemble via continuous dynamics," *Journal of Chemical Physics*, vol. 97, no. 4, pp. 2635–2643, 1992.

[61] P. Langevin, "Sur la théorie du mouvement brownien," *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences*, vol. 146, pp. 530–533, 1908.

[62] T. Schneider and E. Stoll, "Molecular-dynamics study of a three-dimensional one-component model for distortive phase transitions," *Physical Review B*, vol. 17, pp. 1302–1322, 1978.

[63] W. Eckhardt, A. Heinecke, R. Bader, M. Brehm, N. Hammer, H. Huber, H.-G. Kleinhenz, J. Vrabec, H. Hasse, M. Horsch, M. Bernreuther, C. W. Glass, C. Niethammer, A. Bode, and H.-J. Bungartz, "591 TFLOPS Multi-trillion Particles Simulation on SuperMUC," in *Supercomputing: 28th International Supercomputing Conference, ISC 2013, Leipzig, Germany, June 16-20, 2013. Proceedings*, (Berlin, Heidelberg), pp. 1–12, Springer Berlin Heidelberg, 2013.

[64] K. Kremer and G. S. Grest, "Molecular dynamics (MD) simulations for polymers," *Journal of Physics: Condensed Matter*, vol. 2, no. S, pp. SA295–SA298, 1990.

[65] K. Kremer and G. S. Grest, "Dynamics of entangled linear polymer melts: A molecular-dynamics simulation," *Journal of Chemical Physics*, vol. 92, no. 8, p. 5057, 1990.

[66] K. Kremer and F. Müller-Plathe, "Multiscale Problems in Polymer Science: Simulation Approaches," *MRS Bulletin*, vol. 26, no. March, pp. 205–210, 2001.

[67] M. Karplus and A. McCammon, "Molecular dynamics simulations of biomolecules," *Nature Structural Biology*, vol. 9, no. 9, pp. 646–652, 2002.

[68] R. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals*. New York: McGraw-Hill, 1965.

[69] D. Chandler and P. G. Wolynes, "Exploiting the isomorphism between quantum theory and classical statistical mechanics of polyatomic fluids," *Journal of Chemical Physics*, vol. 74, no. 7, p. 4078, 1981.

[70] M. Parrinello and A. Rahman, "Study of an F center in molten KCl," *Journal of Chemical Physics*, vol. 80, no. 2, pp. 860–867, 1984.

[71] R. W. Hall and B. J. Berne, "Nonergodicity in path integral molecular dynamics," *Journal of Chemical Physics*, vol. 81, no. 8, p. 3641, 1984.

[72] B. J. Berne and D. Thirumalai, "On the Simulation of Quantum Systems: Path Integral Methods," *Annual Review of Physical Chemistry*, vol. 37, pp. 401–424, 1986.

[73] M. E. Tuckerman, B. J. Berne, G. J. Martyna, and M. L. Klein, "Efficient molecular dynamics and hybrid Monte Carlo algorithms for path integrals," *Journal of Chemical Physics*, vol. 99, no. 4, pp. 2796–2808, 1993.

[74] G. J. Martyna, A. Hughes, and M. E. Tuckerman, "Molecular dynamics algorithms for path integrals at constant pressure," *Journal of Chemical Physics*, vol. 110, no. 7, pp. 3275–3290, 1999.

[75] E. L. Pollock and D. M. Ceperley, "Simulation of quantum many-body systems by path-integral methods," *Physical Review B*, vol. 30, no. 5, pp. 2555–2568, 1984.

[76] D. Scharf, G. J. Martyna, and M. L. Klein, "Structure and energetics of fluid parahydrogen," *Fizika Nizkikh Temperatur*, vol. 19, no. 5, pp. 516–519, 1993.

[77] T. Lindenau, M. L. Ristig, K. A. Gernoth, J. Dawidowski, and F. J. Bermejo, "The physics of liquid parahydrogen," *International Journal of Modern Physics B*, vol. 20, no. 30n31, pp. 5035–5046, 2006.

[78] S. Fritsch, R. Potestio, D. Donadio, and K. Kremer, "Nuclear quantum effects in water: A multiscale study," *Journal of Chemical Theory and Computation*, vol. 10, no. 2, pp. 816–824, 2014.

[79] D. Marx and M. Parrinello, "Ab initio path integral molecular dynamics: Basic ideas," *Journal of Chemical Physics*, vol. 104, no. 11, pp. 4077–4082, 1996.

[80] D. Marx, M. E. Tuckerman, J. Hutter, and M. Parrinello, "The nature of the hydrated excess proton in water," *Nature*, vol. 397, no. 6720, pp. 601–604, 1999.

[81] L. Wang, S. D. Fried, S. G. Boxer, and T. E. Markland, "Quantum delocalization of protons in the hydrogen-bond network of an enzyme active site," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 111, no. 52, pp. 18454–18459, 2014.

[82] J. Cao and G. A. Voth, "The formulation of quantum statistical mechanics based on the Feynman path centroid density. I. Equilibrium properties," *Journal of Chemical Physics*, vol. 100, no. 7, pp. 5093–5105, 1994.

[83] J. Cao and G. A. Voth, "The formulation of quantum statistical mechanics based on the Feynman path centroid density. II. Dynamical properties," *Journal of Chemical Physics*, vol. 100, no. 7, pp. 5106–5117, 1994.

[84] I. R. Craig and D. E. Manolopoulos, "Quantum statistics and classical mechanics: Real time correlation functions from ring polymer molecular dynamics," *Journal of Chemical Physics*, vol. 121, no. 8, pp. 3368–3373, 2004.

[85] D. M. Ceperley, "Path integrals in the theory of condensed helium," *Reviews of Modern Physics*, vol. 67, no. 2, pp. 279–355, 1995.

[86] R. T. Azuah, W. G. Stirling, H. R. Glyde, M. Boninsegni, P. E. Sokol, and S. M. Bennington, "Condensate and final-state effects in superfluid 4 He," *Physical Review B*, vol. 56, no. 22, pp. 14620–14630, 1997.

[87] P. Nozières and D. Pines, *The theory of quantum liquids*. Advanced book classics, Cambridge, MA: Perseus, 1999.

[88] J. Morrone and R. Car, "Nuclear Quantum Effects in Water," *Physical Review Letters*, vol. 101, no. 1, p. 017801, 2008.

[89] F. Paesani, S. Yoo, H. J. Bakker, and S. S. Xantheas, "Nuclear Quantum Effects in the Reorientation of Water," *Journal of Physical Chemistry Letters*, vol. 1, no. 15, pp. 2316–2321, 2010.

[90] M. Ceriotti, J. Cuny, M. Parrinello, and D. E. Manolopoulos, "Nuclear quantum effects and hydrogen bond fluctuations in water," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, no. 39, pp. 15591–15596, 2013.

[91] M. Ceriotti, W. Fang, P. G. Kusalik, R. H. McKenzie, A. Michaelides, M. A. Morales, and T. E. Markland, "Nuclear Quantum Effects in Water and Aqueous Systems: Experiment, Theory, and Current Challenges," *Chemical Reviews*, vol. 116, no. 13, pp. 7529–7550, 2016.

[92] U. W. Schmitt and G. A. Voth, "The Computer Simulation of Proton Transport in Water," *Journal of Chemical Physics*, vol. 111, no. 20, pp. 9361–9381, 1999.

[93] M. E. Tuckerman, D. Marx, and M. Parrinello, "The nature and transport mechanism of hydrated hydroxide ions in aqueous solution.," *Nature*, vol. 417, no. 6892, pp. 925–929, 2002.

[94] Y. Wu, H. Chen, F. Wang, F. Paesani, and G. A. Voth, "An Improved Multistate Empirical Valence Bond Model for Aqueous Proton Solvation and Transport," *Journal of Physical Chemistry B*, vol. 112, no. 2, pp. 467–482, 2008.

[95] T. C. Berkelbach, H. S. Lee, and M. E. Tuckerman, "Concerted Hydrogen-Bond Dynamics in the Transport Mechanism of the Hydrated Proton: A First-Principles Molecular Dynamics Study," *Physical Review Letters*, vol. 103, no. 23, p. 238302, 2009.

[96] D. Marx, A. Chandra, and M. E. Tuckerman, "Aqueous Basic Solutions: Hydroxide Solvation, Structural Diffusion, and Comparison to the Hydrated Proton," *Chemical Reviews*, vol. 110, no. 4, pp. 2174–2216, 2010.

[97] C. Peter and K. Kremer, "Multiscale simulation of soft matter systems–from the atomistic to the coarse-grained level and back," *Soft Matter*, vol. 5, pp. 4357–4366, 2009.

[98] S. C. L. Kamerlin, S. Vicatos, A. Dryga, and A. Warshel, "Coarse-Grained (Multiscale) Simulations in Studies of Biophysical and Chemical Systems," *Annual Review of Physical Chemistry*, vol. 62, pp. 41–64, 2011.

[99] S. Takada, "Coarse-grained molecular simulations of large biomolecules," *Current Opinion in Structural Biology*, vol. 22, no. 2, pp. 130–137, 2012.

[100] S. Riniker, J. R. Allison, and W. F. van Gunsteren, "On developing coarse-grained models for biomolecular simulation: a review," *Physical Chemistry Chemical Physics*, vol. 14, no. 36, p. 12423, 2012.

[101] W. G. Noid, "Perspective: Coarse-grained models for biomolecular systems," *Journal of Chemical Physics*, vol. 139, no. 9, p. 090901, 2013.

[102] F. Ercolessi and J. B. Adams, "Interatomic Potentials from First-Principles Calculations: The Force-Matching Method," *Europhysics Letters (EPL)*, vol. 26, no. 8, pp. 583–588, 1994.

[103] S. Izvekov and G. A. Voth, "Multiscale coarse graining of liquid-state systems," *Journal of Chemical Physics*, vol. 123, no. 13, p. 134105, 2005.

[104] S. Izvekov and G. A. Voth, "Effective force field for liquid hydrogen fluoride from ab initio molecular dynamics simulation using the force-matching method," *Journal of Physical Chemistry B*, vol. 109, no. 14, pp. 6573–6586, 2005.

[105] A. P. Lyubartsev and A. Laaksonen, "Calculation of effective interaction potentials from radial distribution functions: A reverse Monte Carlo approach," *Physical Review E*, vol. 52, no. 4, pp. 3730–3737, 1995.

[106] D. Reith, M. Pütz, and F. Müller-Plathe, "Deriving effective mesoscale potentials from atomistic simulations," *Journal of Computational Chemistry*, vol. 24, no. 13, pp. 1624–1636, 2003.

[107] T. E. de Oliveira, P. A. Netz, K. Kremer, C. Junghans, and D. Mukherji, "C -IBI: Targeting cumulative coordination within an iterative protocol to derive coarse-grained models of (multi-component) complex fluids," *Journal of Chemical Physics*, vol. 144, no. 17, p. 174106, 2016.

[108] M. S. Shell, "The relative entropy is fundamental to multiscale and inverse thermodynamic problems," *Journal of Chemical Physics*, vol. 129, no. 14, p. 144108, 2008.

[109] J. F. Rudzinski and W. G. Noid, "Coarse-graining entropy, forces, and structures," *Journal of Chemical Physics*, vol. 135, no. 21, p. 214101, 2011.

[110] S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. de Vries, "The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations," *Journal of Physical Chemistry B*, vol. 111, no. 27, pp. 7812–7824, 2007.

[111] S. J. Marrink and D. P. Tieleman, "Perspective on the Martini model," *Chemical Society Reviews*, vol. 42, no. 16, pp. 6801–6822, 2013.

[112] M. Levitt and A. Warshel, "Computer simulation of protein folding," *Nature*, vol. 253, no. 5494, pp. 694–698, 1975.

[113] M. Levitt, "A simplified representation of protein conformations for rapid simulation of protein folding," *Journal of Molecular Biology*, vol. 104, no. 1, pp. 59–107, 1976.

[114] T. Head-Gordon and S. Brown, "Minimalist models for protein folding and design," *Current Opinion in Structural Biology*, vol. 13, no. 2, pp. 160–167, 2003.

[115] S. Matysiak and C. Clementi, "Mapping folding energy landscapes with theory and experiment," *Archives of Biochemistry and Biophysics*, vol. 469, no. 1, pp. 29–33, 2008.

[116] V. Tozzini, "Multiscale modeling of proteins," *Accounts of Chemical Research*, vol. 43, no. 2, pp. 220–230, 2010.

[117] V. Tozzini, "Minimalist models for proteins: a comparative analysis," *Quarterly Reviews of Biophysics*, vol. 43, no. 3, pp. 333–371, 2010.

[118] C. Wu and J.-E. Shea, "Coarse-grained models for protein aggregation," *Current Opinion in Structural Biology*, vol. 21, no. 2, pp. 209–220, 2011.

[119] M. G. Saunders and G. A. Voth, "Coarse-Graining Methods for Computational Biology," *Annual Review of Biophysics*, vol. 42, pp. 73–93, 2013.

[120] G. Brannigan, L. C. L. Lin, and F. L. H. Brown, "Implicit solvent simulation models for biomembranes," *European Biophysics Journal*, vol. 35, no. 2, pp. 104–124, 2006.

[121] M. Müller, K. Katsov, and M. Schick, "Biological and synthetic membranes: What can be learned from a coarse-grained description?," *Physics Reports*, vol. 434, no. 5-6, pp. 113–176, 2006.

[122] M. Venturoli, M. Sperotto, M. Kranenburg, and B. Smit, "Mesoscopic models of biological membranes," *Physics Reports*, vol. 437, no. 1-2, pp. 1–54, 2006.

[123] B. J. Reynwar, G. Illya, V. a. Harmandaris, M. M. Müller, K. Kremer, and M. Deserno, "Aggregation and vesiculation of membrane proteins by curvature-mediated interactions.," *Nature*, vol. 447, no. 7143, pp. 461–464, 2007.

[124] W. Tschöp, K. Kremer, J. Batoulis, T. Bürger, and O. Hahn, "Simulation of polymer melts. I. Coarse-graining procedure for polycarbonates," *Acta Polymerica*, vol. 49, no. 2-3, pp. 61–74, 1998.

[125] W. Tschöp, K. Kremer, O. Hahn, J. Batoulis, and T. Bürger, "Simulation of polymer melts. II. From coarse-grained models back to atomistic description," *Acta Polymerica*, vol. 49, no. 2-3, pp. 75–79, 1998.

[126] F. Müller-Plathe, "Coarse-Graining in Polymer Simulation: From the Atomistic to the Mesoscopic Scale and Back," *ChemPhysChem*, vol. 3, no. 9, pp. 754–769, 2002.

[127] T. Spyriouni, C. Tzoumanekas, D. Theodorou, F. Müller-Plathe, and G. Milano, "Coarse-Grained and Reverse-Mapped United-Atom Simulations of Long-Chain Atactic Polystyrene Melts: Structure, Thermodynamic Properties, Chain Conformation, and Entanglements," *Macromolecules*, vol. 40, no. 10, pp. 3876–3885, 2007.

[128] D. Fritz, V. A. Harmandaris, K. Kremer, and N. F. A. van der Vegt, "Coarse-Grained Polymer Melts Based on Isolated Atomistic Chains: Simulation of Polystyrene of Different Tacticities," *Macromolecules*, vol. 42, no. 19, pp. 7579–7588, 2009.

[129] D. Fritz, K. Koschke, V. A. Harmandaris, N. F. A. van der Vegt, and K. Kremer, "Multiscale modeling of soft matter: scaling of dynamics.," *Physical Chemistry Chemical Physics*, vol. 13, no. 22, pp. 10412–10420, 2011.

[130] J. T. Padding and W. J. Briels, "Systematic coarse-graining of the dynamics of entangled polymer melts: the road from chemistry to rheology," *Journal of physics. Condensed matter : an Institute of Physics journal*, vol. 23, no. 23, p. 233101, 2011.

[131] G. D'Adamo, A. Pelissetto, and C. Pierleoni, "Coarse-graining strategies in polymer solutions," *Soft Matter*, vol. 8, no. 19, p. 5151, 2012.

[132] H. A. Karimi-Varzaneh, N. F. A. van der Vegt, F. Müller-Plathe, and P. Carbone, "How good are coarse-grained polymer models? A comparison for atactic polystyrene," *ChemPhysChem*, vol. 13, no. 15, pp. 3428–3439, 2012.

[133] H. A. Karimi-Varzaneh and F. Müller-Plathe, "Coarse-Grained Modeling for Macromolecular Chemistry," *Top. Curr. Chem.*, vol. 207, p. 295, 2012.

[134] Z. Li, X. Bian, B. Caswell, and G. E. Karniadakis, "Construction of dissipative particle dynamics models for complex fluids via the Mori-Zwanzig formulation.," *Soft Matter*, vol. 10, no. 43, pp. 8659–72, 2014.

[135] Z. Li, X. Bian, X. Li, and G. E. Karniadakis, "Incorporation of memory effects in coarse-grained modeling via the Mori-Zwanzig formalism," *Journal of Chemical Physics*, vol. 143, no. 24, p. 243128, 2015.

[136] T. Head-Gordon and F. H. Stillinger, "An orientational perturbation theory for pure liquid water," *Journal of Chemical Physics*, vol. 98, no. 4, pp. 3313–3327, 1993.

[137] V. Molinero and E. B. Moore, "Water modeled as an intermediate element between carbon and silicon," *Journal of Physical Chemistry B*, vol. 113, no. 13, pp. 4008–4016, 2009.

[138] L. Darré, M. R. MacHado, P. D. Dans, F. E. Herrera, and S. Pantano, "Another Coarse Grain Model for Aqueous Solvation: WAT FOUR?," *Journal of Chemical Theory and Computation*, vol. 6, no. 12, pp. 3793–3807, 2010.

[139] M. Jochum, D. Andrienko, K. Kremer, and C. Peter, "Structure-based coarse-graining in liquid slabs," *Journal of Chemical Physics*, vol. 137, no. 6, p. 064102, 2012.

[140] M. S. Shell, "Systematic coarse-graining of potential energy landscapes and dynamics in liquids," *Journal of Chemical Physics*, vol. 137, no. 8, p. 084503, 2012.

[141] M. Guenza, "Thermodynamic consistency and other challenges in coarse-graining models," *European Physical Journal Special Topics*, vol. 224, no. 12, pp. 2177–2191, 2015.

[142] J. Tu, G. H. Yeoh, and C. Liu, *Computational Fluid Dynamics*. Burlington: Butterworth-Heinemann, 2008.

[143] J. N. Reddy, *An Introduction to the Finite Element Method*. New York: McGraw-Hill, 2005.

[144] M. Praprotnik, L. Delle Site, and K. Kremer, "Multiscale simulation of soft matter: from scale bridging to adaptive resolution.," *Annual Review of Physical Chemistry*, vol. 59, pp. 545–571, 2008.

[145] L. Delle Site, "What is a multiscale problem in molecular dynamics?," *Entropy*, vol. 16, no. 1, pp. 23–40, 2014.

[146] R. Potestio, C. Peter, and K. Kremer, "Computer Simulations of Soft Matter: Linking the Scales," *Entropy*, vol. 16, no. 8, pp. 4199–4245, 2014.

[147] K. A. Henzler-Wildman, M. Lei, V. Thai, S. J. Kerns, M. Karplus, and D. Kern, "A hierarchy of timescales in protein dynamics is linked to enzyme catalysis," *Nature*, vol. 450, no. 7171, pp. 913–916, 2007.

[148] K. Johnston and V. Harmandaris, "Hierarchical Multiscale Modeling of Polymer–Solid Interfaces: Atomistic to Coarse-Grained Description and Structural and Conformational Properties of Polystyrene–Gold Systems," *Macromolecules*, vol. 46, no. 14, pp. 5741–5750, 2013.

[149] Q. H. Zeng, A. B. Yu, and G. Q. Lu, "Multiscale modeling and simulation of polymer nanocomposites," *Progress in Polymer Science*, vol. 33, no. 2, pp. 191–269, 2008.

[150] J. Zhou, I. F. Thorpe, S. Izvekov, and G. A. Voth, "Coarse-Grained Peptide Modeling Using a Systematic Multiscale Approach," *Biophysical Journal*, vol. 92, no. 12, pp. 4289–4303, 2007.

[151] G. S. Ayton, W. G. Noid, and G. A. Voth, "Multiscale modeling of biomolecular systems: in serial and in parallel," *Current Opinion in Structural Biology*, vol. 17, no. 2, pp. 192–198, 2007.

[152] P. Sherwood, B. R. Brooks, and M. S. Sansom, "Multiscale methods for macromolecular simulations," *Current Opinion in Structural Biology*, vol. 18, no. 5, pp. 630–640, 2008.

[153] W. Li, H. Yoshii, N. Hori, T. Kameda, and S. Takada, "Multiscale methods for protein folding simulations," *Methods*, vol. 52, no. 1, pp. 106–114, 2010.

[154] P. Ortoleva, A. Singharoy, and S. Pankavich, "Hierarchical multiscale modeling of macromolecules and their assemblies," *Soft Matter*, vol. 9, pp. 4319–4335, 2013.

[155] A. C. Fogarty, R. Potestio, and K. Kremer, "Adaptive resolution simulation of a biomolecule and its hydration shell: Structural and dynamical properties," *Journal of Chemical Physics*, vol. 142, no. 19, p. 195101, 2015.

[156] A. Warshel and M. Levitt, "Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme," *Journal of Molecular Biology*, vol. 103, no. 2, pp. 227–249, 1976.

[157] J. Gao and X. Xia, "A priori evaluation of aqueous polarization effects through Monte Carlo QM-MM simulations," *Science*, vol. 258, no. 5082, pp. 631–635, 1992.

[158] M. Svensson, S. Humbel, R. D. J. Froese, T. Matsubara, S. Sieber, and K. Morokuma, "ONIOM: A Multilayered Integrated MO + MM Method for Geometry Optimizations and Single Point Energy Predictions. A Test for Diels-Alder Reactions and Pt(P(t-Bu)3)2 + H2 Oxidative Addition," *Journal of Physical Chemistry*, vol. 100, no. 50, pp. 19357–19363, 1996.

[159] P. Carloni, U. Rothlisberger, and M. Parrinello, "The Role and Perspective of Ab Initio Molecular Dynamics in the Study of Biological Systems," *Accounts of Chemical Research*, vol. 35, no. 6, pp. 455–464, 2002.

[160] J. Gao, "Methods and Applications of Combined Quantum Mechanical and Molecular Mechanical Potentials," in *Reviews in Computational Chemistry* (K. B. Lipkowitz and D. B. Boyd, eds.), vol. 7, pp. 119–185, John Wiley & Sons, Inc., 2007.

[161] H. M. Senn and W. Thiel, "QM/MM Methods for Biomolecular Systems," *Angewandte Chemie - International Edition*, vol. 48, no. 7, pp. 1198–1229, 2009.

[162] F. H. Wallrapp and V. Guallar, "Mixed quantum mechanics and molecular mechanics methods: Looking inside proteins," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 1, no. 2, pp. 315–322, 2011.

[163] M. W. van der Kamp and A. J. Mulholland, "Combined Quantum Mechanics/Molecular Mechanics (QM/MM) Methods in Computational Enzymology," *Biochemistry*, vol. 52, no. 16, pp. 2708–2728, 2013.

[164] E. Brunk and U. Rothlisberger, "Mixed quantum mechanical/molecular mechanical molecular dynamics simulations of biological systems in ground and electronically excited states," *Chemical Reviews*, vol. 115, no. 12, pp. 6217–6263, 2015.

[165] A. Heyden, H. Lin, and D. G. Truhlar, "Adaptive Partitioning in Combined Quantum Mechanical and Molecular Mechanical Calculations of Potential Energy Functions for Multiscale Simulations," *Journal of Physical Chemistry B*, vol. 111, no. 9, pp. 2231–2241, 2007.

[166] R. E. Bulo, B. Ensing, J. Sikkema, and L. Visscher, "Toward a practical method for adaptive QM/MM simulations," *Journal of Chemical Theory and Computation*, vol. 5, no. 9, pp. 2212–2221, 2009.

[167] N. Bernstein, C. Várnai, I. Solt, S. A. Winfield, M. C. Payne, I. Simon, M. Fuxreiter, and G. Csányi, "QM/MM simulation of liquid water with an adaptive quantum region," *Physical Chemistry Chemical Physics*, vol. 14, no. 2, pp. 646–56, 2012.

[168] P. Fleurat-Lessard, C. Michel, and R. E. Bulo, "Energy extrapolation schemes for adaptive multi-scale molecular dynamics simulations," *Journal of Chemical Physics*, vol. 137, no. 7, p. 074111, 2012.

[169] J. M. Boereboom, R. Potestio, D. Donadio, and R. E. Bulo, "Toward Hamiltonian Adaptive QM/MM: Accurate Solvent Structures Using Many-Body Potentials," *Journal of Chemical Theory and Computation*, vol. 12, no. 8, pp. 3441–3448, 2016.

[170] G. Csányi, T. Albaret, M. C. Payne, and A. De Vita, ""Learn on the Fly": A Hybrid Classical and Quantum-Mechanical Molecular Dynamics Simulation," *Physical Review Letters*, vol. 93, no. 17, p. 175503, 2004.

[171] Y. Zhao, C. Wang, Q. Peng, and G. Lu, "Error analysis and applications of a general QM/MM approach," *Computational Materials Science*, vol. 50, no. 2, pp. 714–719, 2010.

[172] X. Zhang, G. Lu, and W. Curtin, "Multiscale quantum/atomistic coupling using constrained density functional theory," *Physical Review B*, vol. 87, no. 5, p. 054113, 2013.

[173] J. R. Kermode, L. Ben-Bashat, F. Atrash, J. J. Cilliers, D. Sherman, and A. De Vita, "Macroscopic scattering of cracks initiated at single impurity atoms," *Nature Communications*, vol. 4, p. 2441, 2013.

[174] D. Mukherji, N. F. A. van der Vegt, K. Kremer, and L. Delle Site, "Kirkwood-buff analysis of liquid mixtures in an open boundary simulation," *Journal of Chemical Theory and Computation*, vol. 8, no. 2, pp. 375–379, 2012.

[175] D. Mukherji and K. Kremer, "Coil–Globule–Coil Transition of PNIPAm in Aqueous Methanol: Coupling All-Atom Simulations to Semi-Grand Canonical Coarse-Grained Reservoir," *Macromolecules*, vol. 46, no. 22, pp. 9158–9163, 2013.

[176] M. Praprotnik, L. Delle Site, and K. Kremer, "Adaptive resolution molecular-dynamics simulation: Changing the degrees of freedom on the fly," *Journal of Chemical Physics*, vol. 123, no. 22, p. 224106, 2005.

[177] A. B. Poma and L. Delle Site, "Classical to Path-Integral Adaptive Resolution in Molecular Simulation: Towards a Smooth Quantum-Classical Coupling," *Physical Review Letters*, vol. 104, no. 25, p. 250201, 2010.

[178] B. Ensing, S. O. Nielsen, P. B. Moore, M. L. Klein, and M. Parrinello, "Energy conservation in adaptive hybrid atomistic/coarse-grain molecular dynamics," *Journal of Chemical Theory and Computation*, vol. 3, no. 3, pp. 1100–1105, 2007.

[179] A. Heyden and D. G. Truhlar, "Conservative Algorithm for an Adaptive Change of Resolution in Mixed Atomistic/Coarse-Grained Multiscale Simulations," *Journal of Chemical Theory and Computation*, vol. 4, no. 2, pp. 217–221, 2008.

[180] J. H. Park and A. Heyden, "Solving the equations of motion for mixed atomistic and coarse-grained systems," *Molecular Simulation*, vol. 35, no. 10, pp. 962–973, 2009.

[181] R. Potestio, S. Fritsch, P. Español, R. Delgado-Buscalioni, K. Kremer, R. Everaers, and D. Donadio, "Hamiltonian adaptive resolution simulation for molecular liquids," *Physical Review Letters*, vol. 110, no. 10, p. 108301, 2013.

[182] S. Fritsch, S. Poblete, C. Junghans, G. Ciccotti, L. Delle Site, and K. Kremer, "Adaptive resolution molecular dynamics simulation through coupling to an internal particle reservoir," *Physical Review Letters*, vol. 108, no. 17, p. 170602, 2012.

[183] R. Potestio, P. Español, R. Delgado-Buscalioni, R. Everaers, K. Kremer, and D. Donadio, "Monte carlo adaptive resolution simulation of multicomponent molecular liquids," *Physical Review Letters*, vol. 111, no. 6, p. 060601, 2013.

[184] P. Español, R. Delgado-Buscalioni, R. Everaers, R. Potestio, D. Donadio, and K. Kremer, "Statistical mechanics of Hamiltonian adaptive resolution simulations," *Journal of Chemical Physics*, vol. 142, no. 6, p. 064115, 2015.

[185] B. P. Lambeth, C. Junghans, K. Kremer, C. Clementi, and L. Delle Site, "Communication: On the locality of Hydrogen bond networks at hydrophobic interfaces," *Journal of Chemical Physics*, vol. 133, no. 22, p. 221101, 2010.

[186] M. Praprotnik, S. Poblete, and K. Kremer, "Statistical Physics Problems in Adaptive Resolution Computer Simulations of Complex Fluids," *Journal of Statistical Physics*, vol. 145, no. 4, pp. 946–966, 2011.

[187] S. Fritsch, C. Junghans, and K. Kremer, "Structure formation of toluene around C60: Implementation of the adaptive resolution scheme (AdResS) into GROMACS," *Journal of Chemical Theory and Computation*, vol. 8, no. 2, pp. 398–403, 2012.

[188] J. Zavadlav, M. N. Melo, S. J. Marrink, and M. Praprotnik, "Adaptive resolution simulation of an atomistic protein in MARTINI water," *Journal of Chemical Physics*, vol. 140, no. 5, p. 054114, 2014.

[189] J. Zavadlav, R. Podgornik, and M. Praprotnik, "Adaptive Resolution Simulation of a DNA Molecule in Salt Solution," *Journal of Chemical Theory and Computation*, vol. 11, no. 10, pp. 5035–5044, 2015.

[190] A. C. Fogarty, R. Potestio, and K. Kremer, "A multi-resolution model to capture both global fluctuations of an enzyme and molecular recognition in the ligand-binding site," *Proteins: Structure, Function, and Bioinformatics*, vol. 84, no. 12, pp. 1902–1913, 2016.

[191] J. Zavadlav, R. Podgornik, M. N. Melo, S. J. Marrink, and M. Praprotnik, "Adaptive resolution simulation of an atomistic dna molecule in martini salt solution," *European Physical Journal Special Topics*, vol. 225, no. 8, pp. 1595–1607, 2016.

[192] F. Stanzione and A. Jayaraman, "Hybrid Atomistic and Coarse-Grained Molecular Dynamics Simulations of Polyethylene Glycol (PEG) in Explicit Water," *Journal of Physical Chemistry B*, vol. 120, no. 17, pp. 4160–4173, 2016.

[193] J. H. Peters, R. Klein, and L. Delle Site, "Simulation of macromolecular liquids with the adaptive resolution molecular dynamics technique," *Physical Review E*, vol. 94, no. 2, p. 023309, 2016.

[194] J. Sablić, M. Praprotnik, and R. Delgado-Buscalioni, "Open boundary molecular dynamics of sheared star-polymer melts," *Soft Matter*, vol. 12, pp. 2416–2439, 2016.

[195] S. Kohlhoff, P. Gumbsch, and H. F. Fischmeister, "Crack propagation in b.c.c. crystals studied with a combined finite-element and atomistic model," *Philosophical Magazine A*, vol. 64, no. 4, pp. 851–878, 1991.

[196] F. F. Abraham, J. Q. Broughton, N. Bernstein, and E. Kaxiras, "Spanning the length scales in dynamic simulation," *Computers in Physics*, vol. 12, no. 6, pp. 538–46, 1998.

[197] F. Abraham, J. Broughton, N. Bernstein, and E. Kaxiras, "Spanning the continuum to quantum length scales in a dynamic simulation of brittle fracture," *Europhysics Letters (EPL)*, vol. 44, no. 6, pp. 783–787, 1998.

[198] J. Broughton, F. Abraham, N. Bernstein, and E. Kaxiras, "Concurrent coupling of length scales: Methodology and application," *Physical Review B*, vol. 60, no. 4, pp. 2391–2403, 1999.

[199] E. B. Tadmor, R. Phillips, and M. Ortiz, "Mixed Atomistic and Continuum Models of Deformation in Solids," *Langmuir*, vol. 12, no. 19, pp. 4529–4534, 1996.

[200] V. B. Shenoy, R. Miller, E. B. Tadmor, R. Phillips, and M. Ortiz, "Quasicontinuum Models of Interfacial Structure and Deformation," *Physical Review Letters*, vol. 80, no. 4, pp. 742–745, 1998.

[201] V. B. Shenoy, R. Miller, E. B. Tadmor, D. Rodney, R. Phillips, and M. Ortiz, "An adaptive finite element approach to atomic-scale mechanics - The quasicontinuum method," *Journal of the Mechanics and Physics of Solids*, vol. 47, no. 3, pp. 611–642, 1999.

[202] R. E. Rudd and J. Q. Broughton, "Concurrent Coupling of Length Scales in Solid State Systems," *Physica Status Solidi B*, vol. 217, no. 1, pp. 251–291, 2000.

[203] J. Rottler, S. Barsky, and M. O. Robbins, "Cracks and crazes: on calculating the macroscopic fracture energy of glassy polymers from molecular simulations.," *Physical Review Letters*, vol. 89, no. 14, p. 148304, 2002.

[204] S. P. Xiao and T. Belytschko, "A bridging domain method for coupling continua with molecular dynamics," *Computer Methods in Applied Mechanics and Engineering*, vol. 193, no. 17-20, pp. 1645–1669, 2004.

[205] X. B. Nie, S. Y. Chen, W. N. E, and M. O. Robbins, "A continuum and molecular dynamics hybrid method for micro- and nano-fluid flow," *Journal of Fluid Mechanics*, vol. 500, no. 2004, pp. 55–64, 2004.

[206] G. Lu, E. B. Tadmor, and E. Kaxiras, "From electrons to finite elements: A concurrent multiscale approach for metals," *Physical Review B*, vol. 73, no. 2, p. 024108, 2006.

[207] X. Nie, S. Chen, and M. O. Robbins, "Hybrid continuum-atomistic simulation of singular corner flow," *Physics of Fluids*, vol. 16, no. 10, pp. 3579–3591, 2004.

[208] X. Nie, M. O. Robbins, and S. Chen, "Resolving singular forces in cavity flow: Multiscale modeling from atomic to millimeter scales," *Physical Review Letters*, vol. 96, no. 13, p. 134501, 2006.

[209] M. Praprotnik, S. Matysiak, L. Delle Site, K. Kremer, and C. Clementi, "Adaptive resolution simulation of liquid water," *Journal of Physics: Condensed Matter*, vol. 19, no. 29, p. 292201, 2007.

[210] S. Matysiak, C. Clementi, M. Praprotnik, K. Kremer, and L. Delle Site, "Modeling diffusive dynamics in adaptive resolution simulation of liquid water," *Journal of Chemical Physics*, vol. 128, no. 2, p. 024503, 2008.

[211] H. Wang, C. Schütte, and L. Delle Site, "Adaptive Resolution Simulation (AdResS): A Smooth Thermodynamic and Structural Transition from Atomistic to Coarse Grained Resolution and Vice Versa in a Grand Canonical Fashion," *Journal of Chemical Theory and Computation*, vol. 8, no. 8, pp. 2878–2887, 2012.

[212] R. Potestio and L. Delle Site, "Quantum locality and equilibrium properties in low-temperature parahydrogen: A multiscale simulation study," *Journal of Chemical Physics*, vol. 136, no. 5, p. 054101, 2012.

[213] S. Bevc, C. Junghans, K. Kremer, and M. Praprotnik, "Adaptive resolution simulation of salt solutions," *New Journal of Physics*, vol. 15, p. 105007, 2013.

[214] J. Zavadlav, M. N. Melo, A. V. Cunha, A. H. De Vries, S. J. Marrink, and M. Praprotnik, "Adaptive resolution simulation of MARTINI solvents," *Journal of Chemical Theory and Computation*, vol. 10, no. 6, pp. 2591–2598, 2014.

[215] J. Zavadlav, M. N. Melo, S. J. Marrink, and M. Praprotnik, "Adaptive resolution simulation of polarizable supramolecular coarse-grained water models," *Journal of Chemical Physics*, vol. 142, no. 24, p. 244118, 2015.

[216] A. Agarwal and L. Delle Site, "Path integral molecular dynamics within the grand canonical-like adaptive resolution technique: Simulation of liquid water," *Journal of Chemical Physics*, vol. 143, no. 9, p. 094102, 2015.

[217] 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*, vol. 102, no. 13, p. 5451, 1995.

[218] M. Praprotnik, L. Delle Site, and K. Kremer, "Adaptive resolution scheme for efficient hybrid atomistic-mesoscale molecular dynamics simulations of dense liquids," *Physical Review E*, vol. 73, no. 6, p. 066701, 2006.

[219] L. Delle Site, "Some fundamental problems for an energy-conserving adaptive-resolution molecular dynamics scheme," *Physical Review E*, vol. 76, no. 4, p. 047701, 2007.

[220] M. Praprotnik, K. Kremer, and L. Delle Site, "Adaptive molecular resolution via a continuous change of the phase space dimensionality," *Physical Review E*, vol. 75, no. 1, p. 017701, 2007.

[221] M. Praprotnik, K. Kremer, and L. Delle Site, "Fractional dimensions of phase space variables: a tool for varying the degrees of freedom of a system in a multiscale treatment," *Journal of Physics A: Mathematical and Theoretical*, vol. 40, no. 15, pp. F281–F288, 2007.

[222] S. Poblete, M. Praprotnik, K. Kremer, and L. Delle Site, "Coupling different levels of resolution in molecular simulations," *Journal of Chemical Physics*, vol. 132, no. 11, p. 114101, 2010.

[223] M. Praprotnik, S. Poblete, L. Delle Site, and K. Kremer, "Comment on "Adaptive Multiscale Molecular Dynamics of Macromolecular Fluids"," *Physical Review Letters*, vol. 107, no. 9, p. 99801, 2011.

[224] A. Agarwal and L. Delle Site, "Grand-Canonical Adaptive Resolution Centroid Molecular Dynamics: Implementation and application," *Computer Physics Communications*, vol. 206, pp. 26–34, 2016.

[225] F. H. Stillinger, H. Sakai, and S. Torquato, "Statistical mechanical models with effective potentials: Definitions, applications, and thermodynamic consequences," *Journal of Chemical Physics*, vol. 117, no. 1, pp. 288–296, 2002.

[226] A. A. Louis, "Beware of density dependent pair potentials," *Journal of Physics: Condensed Matter*, vol. 14, no. 40, pp. 9187–9206, 2002.

[227] G. D'Adamo, A. Pelissetto, and C. Pierleoni, "Predicting the thermodynamics by using state-dependent interactions," *Journal of Chemical Physics*, vol. 138, no. 23, p. 234107, 2013.

[228] H. Wang, C. Hartmann, C. Schütte, and L. Delle Site, "Grand-canonical-like molecular-dynamics simulations by using an adaptive-resolution technique," *Physical Review X*, vol. 3, no. 1, p. 011018, 2013.

[229] P. A. Netz, R. Potestio, and K. Kremer, "Adaptive resolution simulation of oligonucleotides," *Journal of Chemical Physics*, vol. 145, no. 23, p. 234101, 2016.

[230] J. G. Kirkwood, "Statistical Mechanics of Fluid Mixtures," *Journal of Chemical Physics*, vol. 3, no. 5, p. 300, 1935.

[231] M. Heidari, R. Cortes-Huerto, D. Donadio, and R. Potestio, "Accurate and general treatment of electrostatic interaction in hamiltonian adaptive resolution simulations," *European Physical Journal Special Topics*, vol. 225, no. 8, pp. 1505–1526, 2016.

[232] P. P. Ewald, "Die Berechnung optischer und elektrostatischer Gitterpotentiale," *Annalen der Physik*, vol. 369, no. 3, pp. 253–287, 1921.

[233] R. W. Hockney and J. W. Eastwood, *Computer Simulation Using Particles*. New York: McGraw-Hill, 1981.

[234] T. Darden, D. York, and L. Pedersen, "Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems," *Journal of Chemical Physics*, vol. 98, no. 12, p. 10089, 1993.

[235] L. Onsager, "Electric Moments of Molecules in Liquids," *Journal of the American Chemical Society*, vol. 58, no. 8, pp. 1486–1493, 1936.

[236] J. A. Barker and R. O. Watts, "Monte Carlo studies of the dielectric properties of water-like models," *Molecular Physics*, vol. 26, no. 3, pp. 789–792, 1973.

[237] 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*, vol. 108, no. 24, pp. 10220–10230, 1998.

[238] D. Wolf, P. Keblinski, S. R. Phillpot, and J. Eggebrecht, "Exact method for the simulation of Coulombic systems by spherically truncated, pairwise r^{-1} summation," *Journal Of Chemical Physics*, vol. 110, no. 17, pp. 8254–8282, 1999.

[239] C. J. Fennell and J. D. Gezelter, "Is the Ewald summation still necessary? Pairwise alternatives to the accepted standard for long-range electrostatics," *Journal of Chemical Physics*, vol. 124, no. 23, p. 234104, 2006.

[240] A. B. Poma and L. Delle Site, "Adaptive resolution simulation of liquid para-hydrogen: testing the robustness of the quantum-classical adaptive coupling," *Physical Chemistry Chemical Physics*, vol. 13, pp. 10510–10519, 2011.

[241] R. Delgado-Buscalioni, K. Kremer, and M. Praprotnik, "Concurrent triple-scale simulation of molecular liquids," *Journal of Chemical Physics*, vol. 128, no. 11, p. 114110, 2008.

[242] R. Delgado-Buscalioni, K. Kremer, and M. Praprotnik, "Coupling atomistic and continuum hydrodynamics through a mesoscopic model: Application to liquid water," *Journal of Chemical Physics*, vol. 131, no. 24, p. 244107, 2009.

[243] R. Delgado-Buscalioni, J. Sablić, and M. Praprotnik, "Open boundary molecular dynamics," *European Physical Journal Special Topics*, vol. 224, no. 12, pp. 2331–2349, 2015.

[244] J. Zavadlav, S. J. Marrink, and M. Praprotnik, "Adaptive Resolution Simulation of Supramolecular Water: The Concurrent Making, Breaking, and Remaking of Water Bundles," *Journal of Chemical Theory and Computation*, vol. 12, no. 8, pp. 4138–4145, 2016.

[245] G. S. Grest and K. Kremer, "Molecular dynamics simulation for polymers in the presence of a heat bath," *Physical Review A*, vol. 33, no. 5, pp. 3628–3631, 1986.

[246] K. Kremer, G. S. Grest, and I. Carmesin, "Crossover from rouse to reptation dynamics: A molecular-dynamics simulation," *Physical Review Letters*, vol. 61, no. 5, pp. 566–569, 1988.

[247] L. Yelash, M. Müller, W. Paul, and K. Binder, "How well can coarse-grained models of real polymers describe their structure? The case of polybutadiene," *Journal of Chemical Theory and Computation*, vol. 2, no. 3, pp. 588–597, 2006.

[248] J. A. McCammon and M. Karplus, "Internal motions of antibody molecules," *Nature*, vol. 268, no. 5622, pp. 765–766, 1977.

[249] M. Karplus and J. A. McCammon, "Protein structural fluctuations during a period of 100 ps," *Nature*, vol. 277, no. 5697, p. 578, 1979.

[250] P. Raiteri, A. Laio, F. L. Gervasio, C. Micheletti, and M. Parrinello, "Efficient reconstruction of complex free energy landscapes by multiple walkers metadynamics," *Journal of Physical Chemistry B*, vol. 110, no. 8, pp. 3533–3539, 2006.

[251] H. Lou and R. I. Cukier, "Molecular Dynamics of Apo-Adenylate Kinase: A Principal Component Analysis," *Journal of Physical Chemistry B*, vol. 110, no. 25, pp. 12796–12808, 2006.

[252] K. Arora and C. L. Brooks, "Large-scale allosteric conformational transitions of adenylate kinase appear to involve a population-shift mechanism," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 47, pp. 18496–18501, 2007.

[253] F. Pontiggia, A. Zen, and C. Micheletti, "Small- and large-scale conformational changes of adenylate kinase: a molecular dynamics study of the subdomain motion and mechanics," *Biophysical journal*, vol. 95, no. 12, pp. 5901–12, 2008.

[254] M. M. Tirion and D. Ben-Avraham, "Normal mode analysis of G-actin," *Journal of Molecular Biology*, vol. 230, no. 1, pp. 186–95, 1993.

[255] M. M. Tirion, "Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis," *Physical Review Letters*, vol. 77, no. 9, pp. 1905–1908, 1996.

[256] I. Bahar, A. R. Atilgan, and B. Erman, "Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential.," *Folding & Design*, vol. 2, no. 3, pp. 173–181, 1997.

[257] C. Micheletti, P. Carloni, and A. Maritan, "Accurate and Efficient Description of Protein Vibrational Dynamics: Comparing Molecular Dynamics and Gaussian Models," *Proteins: Structure, Function and Genetics*, vol. 55, no. 3, pp. 635–645, 2004.

[258] R. Potestio, F. Pontiggia, and C. Micheletti, "Coarse-Grained Description of Protein Internal Dynamics: An Optimal Strategy for Decomposing Proteins in Rigid Subunits," *Biophysical Journal*, vol. 96, no. 12, pp. 4993–5002, 2009.

[259] C. Globisch, V. Krishnamani, M. Deserno, and C. Peter, "Optimization of an Elastic Network Augmented Coarse Grained Model to Study CCMV Capsid Deformation," *PLoS ONE*, vol. 8, no. 4, p. e60582, 2013.

[260] K. Kremer, "Computer simulations in soft matter science," in *Soft and Fragile Matter: Nonequilibrium Dynamics, Metastability and Flow* (M. E. Cates and M. R. Evans, eds.), Scottish Graduate Series, pp. 145–184, Taylor & Francis, 2000.

[261] N. F. A. van der Vegt, C. Peter, and K. Kremer, "Structure-Based Coarse- and Fine-Graining in Soft Matter Simulations," in *Coarse-Graining of Condensed Phase and Biomolecular Systems* (G. A. Voth, ed.), pp. 379–398, CRC Press, Taylor and Francis Group, 2008.

[262] C. Hijón, P. Español, E. Vanden-Eijnden, and R. Delgado-Buscalioni, "Mori-Zwanzig formalism as a practical computational tool.," *Faraday Discussions*, vol. 144, no. 1, pp. 301–322; discussion 323–345, 467–481, 2010.

[263] W. G. Noid, "Systematic methods for structurally consistent coarse-grained models," *Methods in Molecular Biology*, vol. 924, pp. 487–531, 2013.

[264] M. Praprotnik, L. Delle Site, and K. Kremer, "A macromolecule in a solvent: Adaptive resolution molecular dynamics simulation," *Journal of Chemical Physics*, vol. 126, no. 13, p. 134902, 2007.

[265] A. Agarwal, H. Wang, C. Schütte, and L. Delle Site, "Chemical potential of liquids and mixtures via adaptive resolution simulation," *Journal of Chemical Physics*, vol. 141, no. 3, p. 034102, 2014.

[266] K. Kreis, D. Donadio, K. Kremer, and R. Potestio, "A unified framework for force-based and energy-based adaptive resolution simulations," *EPL (Europhysics Letters)*, vol. 108, no. 3, p. 30007, 2014.

[267] B. Mukherjee, L. Delle Site, K. Kremer, and C. Peter, "Derivation of Coarse Grained Models for Multiscale Simulation of Liquid Crystalline Phase Transitions," *Journal of Physical Chemistry B*, vol. 116, no. 29, pp. 8474–8484, 2012.

[268] B. Mukherjee, C. Peter, and K. Kremer, "Dual translocation pathways in smectic liquid crystals facilitated by molecular flexibility," *Physical Review E*, vol. 88, no. 1, p. 010502, 2013.

[269] H. Wang, C. Junghans, and K. Kremer, "Comparative atomistic and coarse-grained study of water: What do we lose by coarse-graining?," *European Physical Journal E*, vol. 28, no. 2, pp. 221–229, 2009.

[270] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, "The Missing Term in Effective Pair Potentials," *Journal of Physical Chemistry*, vol. 91, no. 24, pp. 6269–6271, 1987.

[271] V. Ruehle, C. Junghans, A. Lukyanov, K. Kremer, and D. Andrienko, "Versatile object-oriented toolkit for coarse-graining applications," *Journal of Chemical Theory and Computation*, vol. 5, no. 12, pp. 3211–3223, 2009.

[272] G. J. Wagner and W. K. Liu, "Coupling of atomistic and continuum simulations using a bridging scale decomposition," *Journal of Computational Physics*, vol. 190, no. 1, pp. 249–274, 2003.

[273] D. E. Jiang and E. A. Carter, "First principles assessment of ideal fracture energies of materials with mobile impurities: Implications for hydrogen embrittlement of metals," *Acta Materialia*, vol. 52, no. 16, pp. 4801–4807, 2004.

[274] J. D. Weeks, D. Chandler, and H. C. Andersen, "Role of Repulsive Forces in Determining the Equilibrium Structure of Simple Liquids," *Journal of Chemical Physics*, vol. 54, no. 12, pp. 5237–5247, 1971.

[275] R. Zwanzig, "Nonlinear generalized Langevin equations," *Journal of Statistical Physics*, vol. 9, no. 3, pp. 215–220, 1973.

[276] F. Illuminati, M. Patriarca, and P. Sodano, "Classical and quantum dissipation in non-homogeneous environments," *Physica A: Statistical Mechanics and its Applications*, vol. 211, no. 4, pp. 449–464, 1994.

[277] R. F. Fox, "The generalized Langevin equation with Gaussian fluctuations," *Journal of Mathematical Physics*, vol. 18, no. 12, p. 2331, 1977.

[278] A. D. Baczewski and S. D. Bond, "Numerical integration of the extended variable generalized Langevin equation with a positive Prony representable memory kernel," *Journal of Chemical Physics*, vol. 139, no. 4, p. 044107, 2013.

[279] F. R. Krajewski and M. Parrinello, "Linear scaling electronic structure calculations and accurate statistical mechanics sampling with noisy forces," *Physical Review B*, vol. 73, no. 4, p. 041105, 2006.

[280] C. Attaccalite and S. Sorella, "Stable liquid hydrogen at high pressure by a novel Ab Initio molecular-dynamics calculation," *Physical Review Letters*, vol. 100, no. 11, p. 114501, 2008.

[281] R. L. McGreevy and L. Pusztai, "Reverse Monte Carlo Simulation: A New Technique for the Determination of Disordered Structures," *Molecular Simulation*, vol. 1, no. 6, pp. 359–367, 1988.

[282] A. Chaimovich and M. S. Shell, "Anomalous waterlike behavior in spherically-symmetric water models optimized with the relative entropy," *Physical Chemistry Chemical Physics*, vol. 11, no. 12, pp. 1901–1915, 2009.

[283] A. Chaimovich and M. S. Shell, "Relative entropy as a universal metric for multiscale errors," *Physical Review E*, vol. 81, no. 6, p. 060104(R), 2010.

[284] A. Chaimovich and M. S. Shell, "Coarse-graining errors and numerical optimization using a relative entropy framework.," *Journal of Chemical Physics*, vol. 134, no. 9, p. 094112, 2011.

[285] C. F. Abrams, "Concurrent dual-resolution Monte Carlo simulation of liquid methane," *Journal of Chemical Physics*, vol. 123, no. 23, p. 234101, 2005.

[286] K. Kreis, A. C. Fogarty, K. Kremer, and R. Potestio, "Advantages and challenges in coupling an ideal gas to atomistic models in adaptive resolution simulations," *European Physical Journal Special Topics*, vol. 224, no. 12, p. 2289, 2015.

[287] S. Y. Mashayak, M. N. Jochum, K. Koschke, N. R. Aluru, V. Rühle, and C. Junghans, "Relative entropy and optimization-driven coarse-graining methods in VOTCA," *PLoS ONE*, vol. 10, no. 7, p. e0131754, 2015.

[288] M. J. Abraham, D. van der Spoel, E. Lindahl, B. Hess, and the GROMACS development team, *GROMACS User Manual version 5.1.2*, 2016.

[289] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl, "GRGMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation," *Journal of Chemical Theory and Computation*, vol. 4, no. 3, pp. 435–447, 2008.

[290] L. Larini, L. Lu, and G. A. Voth, "The multiscale coarse-graining method. VI. Implementation of three-body coarse-grained potentials," *Journal of Chemical Physics*, vol. 132, no. 16, p. 164107, 2010.

[291] D. Ben-Amotz and G. Stell, "Reformulation of Weeks-Chandler-Andersen Perturbation Theory Directly in Terms of a Hard-Sphere Reference System," *Journal of Physical Chemistry B*, vol. 108, pp. 6877–6882, 2004.

[292] L. Filion, R. Ni, D. Frenkel, and M. Dijkstra, "Simulation of nucleation in almost hard-sphere colloids: The discrepancy between experiment and simulation persists," *Journal of Chemical Physics*, vol. 134, no. 13, p. 134901, 2011.

[293] N. a. Mahynski, A. Z. Panagiotopoulos, D. Meng, and S. K. Kumar, "Stabilizing colloidal crystals by leveraging void distributions," *Nature Communications*, vol. 5, p. 4472, 2014.

[294] R. L. Henderson, "A uniqueness theorem for fluid pair correlation functions," *Physics Letters A*, vol. 49, no. 3, pp. 197–198, 1974.

[295] H. C. Andersen, D. Chandler, and J. D. Weeks, "Roles of Repulsive and Attractive Forces in Liquids: The Equilibrium Theory of Classical Fluids," *Advances in Chemical Physics*, vol. 34, pp. 105–156, 1976.

[296] V. Daggett and A. Fersht, "The present view of the mechanism of protein folding," *Nature Reviews Molecular Cell Biology*, vol. 4, no. 6, pp. 497–502, 2003.

[297] M. A. Ward and T. K. Georgiou, "Thermoresponsive polymers for biomedical applications," *Polymers*, vol. 3, no. 3, pp. 1215–1242, 2011.

[298] O. M. Szklarczyk, N. S. Bieler, P. H. Hünenberger, and W. F. van Gunsteren, "Flexible boundaries for multiresolution solvation: An algorithm for spatial multiscaling in molecular dynamics simulations," *Journal of Chemical Theory and Computation*, vol. 11, no. 11, pp. 5447–5463, 2015.

[299] J. A. Wagoner and V. S. Pande, "Finite domain simulations with adaptive boundaries: Accurate potentials and nonequilibrium movesets," *Journal of Chemical Physics*, vol. 139, no. 23, p. 234114, 2013.

[300] J. Wang, P. Cieplak, and P. A. Kollman, "How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules?," *Journal of Computational Chemistry*, vol. 21, no. 12, pp. 1049–1074, 2000.

[301] R. Raucci, G. Colonna, G. Castello, and S. Costantini, "Peptide folding problem: A molecular dynamics study on polyalanines using different force fields," *International Journal of Peptide Research and Therapeutics*, vol. 19, no. 2, pp. 117–123, 2013.

[302] P. Palenčár and T. Bleha, "Molecular dynamics simulations of the folding of poly(alanine) peptides," *Journal of Molecular Modeling*, vol. 17, no. 9, pp. 2367–2374, 2011.

[303] V. Daggett, P. a. Kollman, and I. D. Kuntz, "A molecular dynamics simulation of polyalanine: An analysis of equilibrium motions and helix-coil transitions," *Biopolymers*, vol. 31, no. 9, pp. 1115–1134, 1991.

[304] M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeerschd, E. Zurek, and G. R. Hutchison, "Avogadro: an advanced semantic chemical editor, visualization, and analysis platform," *Journal of Cheminformatics*, vol. 4, no. 8, p. 17, 2012.

[305] M. W. Mahoney and W. L. Jorgensen, "Diffusion constant of the TIP5P model of liquid water," *Journal of Chemical Physics*, vol. 114, no. 1, pp. 363–366, 2001.

[306] D. Laage, G. Stirnemann, F. Sterpone, and J. T. Hynes, "Water jump reorientation: From theoretical prediction to experimental observation," *Accounts of Chemical Research*, vol. 45, no. 1, pp. 53–62, 2012.

[307] J. F. Rudzinski and W. G. Noid, "Bottom-Up Coarse-Graining of Peptide Ensembles and Helix-Coil Transitions," *Journal of Chemical Theory and Computation*, vol. 11, no. 3, pp. 1278–1291, 2015.

[308] E. K. Bradley, J. F. Thomason, F. E. Cohen, P. A. Kosen, and I. D. Kuntz, "Studies of synthetic helical peptides using circular dichroism and nuclear magnetic resonance," *Journal of Molecular Biology*, vol. 215, no. 4, pp. 607–622, 1990.

[309] S. Hovmöller, T. Zhou, and T. Ohlson, "Conformations of amino acids in proteins," *Acta Crystallographica Section D: Biological Crystallography*, vol. 58, no. 5, pp. 768–776, 2002.

[310] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling, "Comparison of multiple Amber force fields and development of improved protein backbone parameters," *Proteins: Structure, Function and Genetics*, vol. 65, no. 3, pp. 712–725, 2006.

[311] Y. Levy and J. N. Onuchic, "Water mediation in protein folding and molecular recognition," *Annual Review of Biophysics and Biomolecular Structure*, vol. 35, no. 1, pp. 389–415, 2006.

[312] A. Y. Smirnov, L. G. Mourokh, and F. Nori, "Electrostatic models of electron-driven proton transfer across a lipid membrane," *Journal of Physics: Condensed Matter*, vol. 23, no. 23, p. 234101, 2011.

[313] L. J. C. Jeuken, R. J. Bushby, and S. D. Evans, "Proton transport into a tethered bilayer lipid membrane," *Electrochemistry Communications*, vol. 9, no. 4, pp. 610–614, 2007.

[314] T. H. Haines, "Do sterols reduce proton and sodium leaks through lipid bilayers?," *Progress in Lipid Research*, vol. 40, no. 4, pp. 299–324, 2001.

[315] P.-O. Löwdin, "Proton Tunneling in DNA and its Biological Implications," *Reviews of Modern Physics*, vol. 35, no. 3, pp. 724–732, 1963.

[316] R. Rein and F. E. Harris, "Proton Tunneling in Radiation-Induced Mutation," *Science*, vol. 146, no. 3644, pp. 649–650, 1964.

[317] A. Pérez, M. E. Tuckerman, H. P. Hjalmarson, and O. A. Von Lilienfeld, "Enol tautomers of Watson-Crick base pair models are metastable because of nuclear quantum effects," *Journal of the American Chemical Society*, vol. 132, no. 33, pp. 11510–11515, 2010.

[318] D. Jacquemin, J. Zúñiga, A. Requena, and J. P. Céron-Carrasco, "Assessing the importance of proton transfer reactions in DNA," *Accounts of Chemical Research*, vol. 47, no. 8, pp. 2467–2474, 2014.

[319] B. Pamuk, J. M. Soler, R. Ramírez, C. P. Herrero, P. W. Stephens, P. B. Allen, and M. V. Fernández-Serra, "Anomalous Nuclear Quantum Effects in Ice," *Physical Review Letters*, vol. 108, no. 19, p. 193003, 2012.

[320] X. Z. Li, M. I. J. Probert, A. Alavi, and A. Michaelides, "Quantum nature of the proton in water-hydroxyl overlayers on metal surfaces," *Physical Review Letters*, vol. 104, no. 6, p. 066102, 2010.

[321] Y. Nagata, R. E. Pool, E. H. G. Backus, and M. Bonn, "Nuclear Quantum Effects Affect Bond Orientation of Water at the Water-Vapor Interface," *Physical Review Letters*, vol. 109, no. 22, p. 226101, 2012.

[322] M. P. Gelfand and M. E. Fisher, "Finite-size effects in fluid interfaces," *Physica A: Statistical Mechanics and its Applications*, vol. 166, no. 1, pp. 1–74, 1990.

[323] A. Aguado, W. Scott, and P. A. Madden, "Molecular dynamics simulations of the liquid-vapor interface of a molten salt. II. Finite size effects and comparison to experiment," *Journal of Chemical Physics*, vol. 115, no. 18, p. 8612, 2001.

[324] F. Schmitz, P. Virnau, and K. Binder, "Logarithmic finite-size effects on interfacial free energies: Phenomenological theory and Monte Carlo studies," *Physical Review E*, vol. 90, no. 1, p. 012128, 2014.

[325] J. K. Hwang and a. Warshel, "How Important Are Quantum-Mechanical Nuclear Motions in Enzyme Catalysis," *Journal of the American Chemical Society*, vol. 118, no. 47, pp. 11745–11751, 1996.

[326] H. Engel, D. Doron, A. Kohen, and D. T. Major, "Momentum distribution as a fingerprint of quantum delocalization in enzymatic reactions: Open-chain path-integral simulations of model systems and the hydride transfer in dihydrofolate reductase," *Journal of Chemical Theory and Computation*, vol. 8, no. 4, pp. 1223–1234, 2012.

[327] A. Agarwal, J. Zhu, C. Hartmann, H. Wang, and L. Delle Site, "Molecular dynamics in a grand ensemble: Bergmann-Lebowitz model and adaptive resolution simulation," *New Journal of Physics*, vol. 17, no. 8, p. 083042, 2015.

[328] M. E. Tuckerman and D. Marx, "Heavy-atom skeleton quantization and proton tunneling in "intermediate-barrier" hydrogen bonds," *Physical Review Letters*, vol. 86, no. 21, pp. 4946–4949, 2001.

[329] S. Habershon, T. E. Markland, and D. E. Manolopoulos, "Competing quantum effects in the dynamics of a flexible water model," *Journal of Chemical Physics*, vol. 131, no. 2, p. 024501, 2009.

[330] L. Wang, M. Ceriotti, and T. E. Markland, "Quantum fluctuations and isotope effects in ab initio descriptions of water," *Journal of Chemical Physics*, vol. 141, no. 10, p. 104502, 2014.

[331] D. M. Wilkins, D. E. Manolopoulos, and L. X. Dang, "Nuclear quantum effects in water exchange around lithium and fluoride ions," *Journal of Chemical Physics*, vol. 142, no. 6, p. 064509, 2015.

[332] T. F. Miller III and D. E. Manolopoulos, "Quantum diffusion in liquid para-hydrogen from ring-polymer molecular dynamics," *Journal of Chemical Physics*, vol. 122, no. 18, p. 184503, 2005.

[333] R. Potestio, "Computer simulation of particles with position-dependent mass," *European Physical Journal B*, vol. 87, no. 10, p. 245, 2014.

[334] J. W. Leachman, R. T. Jacobsen, S. G. Penoncello, and E. W. Lemmon, "Fundamental equations of state for parahydrogen, normal hydrogen, and orthohydrogen," *Journal of Physical and Chemical Reference Data*, vol. 38, no. 3, pp. 721–748, 2009.

[335] I. F. Silvera and V. V. Goldman, "The isotropic intermolecular potential for H2 and D2 in the solid and gas phases," *Journal of Chemical Physics*, vol. 69, no. 9, p. 4209, 1978.

[336] I. F. Silvera, "The solid molecular hydrogens in the condensed phase: Fundamentals and static properties," *Reviews of Modern Physics*, vol. 52, no. 2, pp. 393–452, 1980.

[337] T. E. Markland and D. E. Manolopoulos, "An efficient ring polymer contraction scheme for imaginary time path integral simulations," *Journal of Chemical Physics*, vol. 129, no. 2, p. 024105, 2008.

[338] T. E. Markland and D. E. Manolopoulos, "A refined ring polymer contraction scheme for systems with electrostatic interactions," *Chemical Physics Letters*, vol. 464, no. 4-6, pp. 256–261, 2008.

[339] O. Marsalek and T. E. Markland, "Ab initio molecular dynamics with nuclear quantum effects at classical cost: Ring polymer contraction for density functional theory," *Journal of Chemical Physics*, vol. 144, no. 5, p. 054112, 2016.

[340] C. John, T. Spura, S. Habershon, and T. D. Kühne, "Quantum ring-polymer contraction method: Including nuclear quantum effects at no additional computational cost in comparison to ab initio molecular dynamics," *Physical Review E*, vol. 93, no. 4, p. 043305, 2016.

[341] R. P. Steele, J. Zwickl, P. Shushkov, and J. C. Tully, "Mixed time slicing in path integral simulations," *Journal of Chemical Physics*, vol. 134, no. 7, p. 074112, 2011.

[342] M. Ceriotti, G. Bussi, and M. Parrinello, "Nuclear quantum effects in solids using a colored-noise thermostat," *Physical Review Letters*, vol. 103, no. 3, p. 020601, 2009.

[343] M. Ceriotti, M. Parrinello, T. E. Markland, and D. E. Manolopoulos, "Efficient stochastic thermostatting of path integral molecular dynamics," *Journal of Chemical Physics*, vol. 133, no. 12, p. 124104, 2010.

[344] M. Ceriotti, D. E. Manolopoulos, and M. Parrinello, "Accelerating the convergence of path integral dynamics with a generalized Langevin equation," *Journal of Chemical Physics*, vol. 134, no. 8, p. 084104, 2011.

[345] M. Ceriotti and D. E. Manolopoulos, "Efficient First-Principles Calculation of the Quantum Kinetic Energy and Momentum Distribution of Nuclei," *Physical Review Letters*, vol. 109, no. 10, p. 100604, 2012.

[346] M. Takahashi and M. Imada, "Monte Carlo calculation of quantum systems. II. Higher order correction," *Journal of the Physical Society of Japan*, vol. 53, no. 11, pp. 3765–3769, 1984.

[347] K. Suzuki, M. Tachikawa, and M. Shiga, "Efficient ab initio path integral hybrid Monte Carlo based on the fourth-order Trotter expansion: Application to fluoride ion-water cluster," *Journal of Chemical Physics*, vol. 132, no. 14, p. 144108, 2010.

[348] S. Jang, S. Jang, and G. A. Voth, "Applications of higher order composite factorization schemes in imaginary time path integral simulations," *Journal of Chemical Physics*, vol. 115, no. 17, pp. 7832–7842, 2001.

[349] A. Pérez and M. E. Tuckerman, "Improving the convergence of closed and open path integral molecular dynamics via higher order Trotter factorization schemes," *Journal of Chemical Physics*, vol. 135, no. 6, p. 064104, 2011.

[350] I. Poltavsky and A. Tkatchenko, "Modeling quantum nuclei with perturbed path integral molecular dynamics," *Chemical Science*, vol. 7, no. 2, pp. 1368–1372, 2016.

[351] J. Cao and G. A. Voth, "The formulation of quantum statistical mechanics based on the Feynman path centroid density. III. Phase space formalism and analysis of centroid molecular dynamics," *Journal of Chemical Physics*, vol. 101, no. 7, pp. 6157–6167, 1994.

[352] J. Cao and G. A. Voth, "The formulation of quantum statistical mechanics based on the Feynman path centroid density. IV. Algorithms for centroid molecular dynamics," *Journal of Chemical Physics*, vol. 101, no. 7, pp. 6168–6183, 1994.

[353] J. Cao and G. A. Voth, "The formulation of quantum statistical mechanics based on the Feynman path centroid density. V. Quantum instantaneous normal mode theory of liquids," *Journal of Chemical Physics*, vol. 101, no. 7, pp. 6184–6192, 1994.

[354] J. Cao and G. A. Voth, "Semiclassical approximations to quantum dynamical time correlation functions," *Journal of Chemical Physics*, vol. 104, no. 1, pp. 273–285, 1996.

[355] G. J. Martyna, "Adiabatic path integral molecular dynamics methods. I. Theory," *Journal of Chemical Physics*, vol. 104, no. 5, pp. 2018–2027, 1996.

[356] J. Cao and G. J. Martyna, "Adiabatic path integral molecular dynamics methods. II. Algorithms," *Journal of Chemical Physics*, vol. 104, no. 5, p. 2028, 1996.

[357] S. Jang and G. A. Voth, "A derivation of centroid molecular dynamics and other approximate time evolution methods for path integral centroid variables," *Journal of Chemical Physics*, vol. 111, no. 6, pp. 2371–2384, 1999.

[358] T. D. Hone, P. J. Rossky, and G. A. Voth, "A comparative study of imaginary time path integral based methods for quantum dynamics," *Journal of Chemical Physics*, vol. 124, no. 15, p. 154103, 2006.

[359] A. Pérez, M. E. Tuckerman, and M. H. Müser, "A comparative study of the centroid and ring-polymer molecular dynamics methods for approximating quantum time correlation functions from path integrals," *Journal of Chemical Physics*, vol. 130, no. 18, p. 184105, 2009.

[360] E. A. Polyakov, A. P. Lyubartsev, and P. N. Vorontsov-Velyaminov, "Centroid molecular dynamics: Comparison with exact results for model systems," *Journal of Chemical Physics*, vol. 133, no. 19, p. 194103, 2010.

[361] B. J. Braams and D. E. Manolopoulos, "On the short-time limit of ring polymer molecular dynamics," *Journal of Chemical Physics*, vol. 125, no. 12, p. 124105, 2006.

[362] S. Habershon, D. E. Manolopoulos, T. E. Markland, and T. F. Miller III, "Ring-polymer molecular dynamics: quantum effects in chemical dynamics from classical trajectories in an extended phase space," *Annual Review of Physical Chemistry*, vol. 64, pp. 387–413, 2013.

[363] M. Rossi, M. Ceriotti, and D. E. Manolopoulos, "How to remove the spurious resonances from ring polymer molecular dynamics," *Journal of Chemical Physics*, vol. 140, no. 23, p. 234116, 2014.

[364] M. E. Tuckerman, B. J. Berne, and G. J. Martyna, "Reversible multiple time scale molecular dynamics," *Journal of Chemical Physics*, vol. 97, no. 3, pp. 1990–2001, 1992.

[365] K. Kreis, M. E. Tuckerman, D. Donadio, K. Kremer, and R. Potestio, "From Classical to Quantum and Back: A Hamiltonian Scheme for Adaptive Multiresolution Classical/Path-Integral Simulations," *Journal of Chemical Theory and Computation*, vol. 12, no. 7, pp. 3030–3039, 2016.

[366] K. Kreis and R. Potestio, "The relative entropy is fundamental to adaptive resolution simulations," *Journal of Chemical Physics*, vol. 145, no. 4, p. 044104, 2016.

[367] H. F. Trotter, "On the product of semi-groups of operators," *Proceedings of the American Mathematical Society*, vol. 10, pp. 545–551, 1959.

[368] G. Strang, "On the Construction and Comparison of Difference Schemes," *SIAM Journal on Numerical Analysis*, vol. 5, no. 3, pp. 506–517, 1968.

[369] W. R. Rudnicki, B. Lesyng, and S. C. Harvey, "Lagrangian molecular dynamics using selected conformational degrees of freedom, with application to the pseudorotation dynamics of furanose rings," *Biopolymers*, vol. 34, no. 3, pp. 383–392, 1994.

[370] B. Leimkuhler and C. Matthews, "Rational construction of stochastic numerical methods for molecular sampling," *Applied Mathematics Research eXpress*, vol. 2013, no. 1, pp. 34–56, 2013.

[371] A. Witt, S. D. Ivanov, M. Shiga, H. Forbert, and D. Marx, "On the applicability of centroid and ring polymer path integral molecular dynamics for vibrational spectroscopy," *Journal of Chemical Physics*, vol. 130, no. 19, p. 194510, 2009.

[372] K. Kreis, R. Potestio, K. Kremer, and A. C. Fogarty, "Adaptive Resolution Simulations with Self-Adjusting High-Resolution Regions," *Journal of Chemical Theory and Computation*, vol. 12, no. 8, pp. 4067–4081, 2016.

[373] Y. Wu, H. L. Tepper, and G. A. Voth, "Flexible simple point-charge water model with improved liquid-state properties," *Journal of Chemical Physics*, vol. 124, no. 2, p. 024503, 2006.

[374] A. C. Fogarty and D. Laage, "Water Dynamics in Protein Hydration Shells: The Molecular Origins of the Dynamical Perturbation," *Journal of Physical Chemistry B*, vol. 118, no. 28, pp. 7715–7729, 2014.

[375] M. Salvalaglio, C. Perego, F. Giberti, M. Mazzotti, and M. Parrinello, "Molecular-dynamics simulations of urea nucleation from aqueous solution," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 112, no. 1, pp. E6–E14, 2015.

[376] K. Toukan and A. Rahman, "Molecular-dynamics study of atomic motions in water," *Physical Review B*, vol. 31, pp. 2643–2648, 1985.

[377] D. Eisenberg and W. Kauzmann, *The Structure and Properties of Water*. New York: Oxford University Press, 1969.

[378] H. E. Stanley and N. Ostrowzky, *Correlations and Connectivity, Geometric Aspects of Physics, Chemistry and Biology*. Dordrecht: Kluwer Academic, 1990.

[379] Teixeira, J., "The physics of liquid water," *Journal de Physique IV*, vol. 3, no. C1, pp. 163–169, 1993.

[380] A. Luzar and D. Chandler, "Hydrogen-bond kinetics in liquid water," *Nature*, vol. 379, no. 6560, pp. 55–57, 1996.

[381] A. Luzar and D. Chandler, "Effect of Environment on Hydrogen Bond Dynamics in Liquid Water," *Physical Review Letters*, vol. 76, no. 6, pp. 928–931, 1996.

[382] A. Luzar, "Resolving the hydrogen bond dynamics conundrum," *Journal of Chemical Physics*, vol. 113, no. 23, pp. 10663–10675, 2000.

[383] A. Luzar, "Extent of inter-hydrogen bond correlations in water. Temperature effect," *Chemical Physics*, vol. 258, no. 2-3, pp. 267–276, 2000.

[384] I. Benjamin, "Hydrogen Bond Dynamics at Water/Organic Liquid Interfaces," *Journal of Physical Chemistry B*, vol. 109, no. 28, pp. 13711–13715, 2005.

[385] N. Winter, J. Vieceli, and I. Benjamin, "Hydrogen-Bond Structure and Dynamics at the Interface between Water and Carboxylic Acid-Functionalized Self-Assembled Monolayers," *Journal of Physical Chemistry B*, vol. 112, no. 2, pp. 227–231, 2008.

[386] A. Agarwal, C. Clementi, and L. Delle Site, "Path integral-GC-AdResS simulation of a large hydrophobic solute in water: a tool to investigate the interplay between local microscopic structures and quantum delocalization of atoms in space," *Physical Chemistry Chemical Physics*, 2017, Advance Article.

[387] D. J. Adams, "Chemical potential of hard-sphere fluids by Monte Carlo methods," *Molecular Physics*, vol. 28, no. 5, pp. 1241–1252, 1974.

[388] D. J. Adams, "Grand canonical ensemble Monte Carlo for a Lennard-Jones fluid," *Molecular Physics*, vol. 29, no. 1, pp. 307–311, 1975.

[389] B. Ensing and S. O. Nielsen, "Multiscale Molecular Dynamics and the Reverse Mapping Problem," in *Trends in Computational Nanomechanics: Transcending Length and Time Scales* (T. Dumitrica, ed.), pp. 25–59, Dordrecht: Springer Netherlands, 2010.

[390] S. O. Nielsen, R. E. Bulo, P. B. Moore, and B. Ensing, "Recent progress in adaptive multiscale molecular dynamics simulations of soft matter," *Physical Chemistry Chemical Physics*, vol. 12, no. 39, pp. 12401–12414, 2010.