PATH INTEGRAL TECHNIQUES IN MOLECULAR DYNAMICS SIMULATIONS OF OPEN BOUNDARY SYSTEMS

Dissertation
zur Erlangung des Grades eines
Doktors der Naturwissenschaften (Dr. rer. nat.)

am Fachbereich Mathematik und Informatik
der Freien Universität Berlin

vorgelegt von

Animesh Agarwal

Berlin, 2016

Erstgutachter: Prof. Dr. Christof Schütte
Zweitgutachter: Prof. Dr. Carsten Hartmann
Tag der Disputation: September 22, 2016

Publications

This thesis led to the following publications:

- "Adaptive Resolution Simulation as a Grand Canonical Molecular Dynamics Scheme: Principles, Applications, Perspectives", in: C. Clementi (Ed.), Multiscale Methods in Molecular Biophysics, Series in Computational Biophysics, CRC Press (2016), ISBN 1482225700, http://arxiv.org/pdf/1412.4540v1.pdf
- "Grand-Canonical Adaptive Resolution Centroid Molecular Dynamics: Implementation and Application", Animesh Agarwal and Luigi Delle Site, Comp. Phys. Comm., **206**, 26-34 (2016).
- "Path Integral Molecular Dynamics within the Grand Canonical-like Adaptive Resolution Technique: Simulation of Liquid Water", Animesh Agarwal and Luigi Delle Site, J. Chem. Phys., **143**, 094102 (2015).
- "Molecular Dynamics in a grand ensemble: Bergmann-Lebowitz model and adaptive resolution simulation", Animesh Agarwal, Jinglong Zhu, Carsten Hartmann, Han Wang and Luigi Delle Site, New J. Phys., **17**, 083042 (2015).
- "Chemical potential of liquids and mixtures via Adaptive Resolution Simulation", Animesh Agarwal, Han Wang, Christof Schütte, and Luigi Delle Site, J Chem. Phys., **141**, 034102 (2014).

Acknowledgements

I would like to acknowledge the invaluable support provided by Prof. Christof Schütte during the course of my thesis. Further, I express my heartfelt gratitude to my supervisor Prof. Luigi Delle Site who has been thoroughly involved with the project and has constantly kept me motivated through numerous stimulating research discussion sessions. I believe that his guidance and valuable insight has shaped me into a better researcher. I also thank Dr. Han Wang, a former post-doctoral researcher in our group, for helping me with the background required for this area of research. I would also like to thank Dr. Christoph Junghans (Los Alamos National Laboratory), Dr. Christian Krekeler (Freie Universität Berlin) and Dr. Jan Henning Peters (Freie Universität Berlin) for their valuable suggestions and help in my thesis work. They have all been a great moral support to me. It was an honour working in the Biocomputing group at Freie Universität Berlin, and I appreciate all those who were a part of my life here.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Molecular simulation approaches play an important role in studying aqueous solvation in condensed matter physics (See Ref. [69]). However, the size and time scales of the simulation are by far too small compared to those required for a complete understanding of the solvation process. Recently developed adaptive resolution multiscale techniques [22, 34, 84, 89] allow to treat only those DOF's in a region of space where process of interest takes place while allowing the rest of the system to be treated at coarser level. Such a scheme not only results in gain in computational efficiency but also paves the way for studying the physio-chemical properties in a complete and accurate way since the DOF's that are not essential to the problem and that may overshadow the underlying physics of system are, in this case, coarse-grained [49, 89]. The interesting study in this context is the treatment of hydrogen as a quantum particle (i.e. with a sizeable delocalization in space) since many of the crucial properties of water may indeed be influenced by the quantum nature of hydrogen [72]. The quantum delocalization of the hydrogen atom is treated in molecular dynamics (MD) via path-integral (PI) formalism [25, 107] where each atom is mapped onto a polymer ring containing fictitious beads and the proper statistical sampling can be obtained by averaging over the trajectories of the polymer rings; which in turn are governed by Newton's equations. Thus, PIMD can be used to determine the quantum statistical properties of the system. Compared to standard MD, the computational costs of PIMD are very high. As a consequence, studies involving large systems and long time scales that allow for an accurate calculation of statistical and thermodynamic properties as done in Refs. [49] would be prohibitive with the current computational resources. In this perspective, one can combine PIMD with the latest version of AdResS, i.e. grand-canonical AdResS [111] which is equivalent of coupling two regions with different number of effective "classical" degrees of freedom, where a relatively small high resolution (PI) region is embedded in a large reservoir containing coarse-grained (spherical) molecules. This idea implies that the high resolution system is an open system that exchanges energy and particles with the reservoir. In order to realize such a scheme, it becomes imperative to provide a mathematical/rigorous definition of equilibrium statistical properties in simulation techniques with open boundaries. Since the adaptive resolution approach is based on force interpolation and a global Hamiltonian cannot be defined [17], the principles of standard statistical mechanics can not be applied directly and have to be reformulated within the framework of adaptive resolution. In the first part of the thesis, we show the capability of GC-AdResS, with classical MD, to be a generic tool for the calculation of

thermodynamic properties (specifically, the excess chemical potential) of a system in a grand-canonical fashion, using a simple generic Weeks-Chandler-Andersen (WCA) potential in the coarse-grained reservoir. Thus, one does not need to run a full MD simulation to calculate the coarse-grained potential as is required for structural coarse-graining methods like inverted Boltzmann Inversion [96], Force matching [23] etc. We calculate the excess chemical potential for a wide variety of liquids and mixtures and compare the results with thermodynamic integration (TI) results. In the second part of this thesis, we show the correspondence between the Bergman-Lebowitz model for grand ensemble and GC-AdResS, which gives a conceptual consistency to the simulation techniques that mimic open systems and provides a theoretical definition to the equilibrium time correlation functions in systems with varying number of particles. We calculate these functions for liquid water using GC-AdResS and show that the method reproduces the reference results from full-atomistic simulations. This establishes GC-AdResS as a reliable method for the calculation of static and dynamical properties in open systems. Since the path-integral polymer rings can be interpreted in terms of classical fictitious atoms, the theoretical concepts developed in the first two parts of this thesis for classical MD are used straightforwardly for the conceptual justification of merging PI techniques with GC-AdResS. We calculate the quantum static and dynamical properties of liquid water using GC-AdResS in conjunction with three different PI techniques: path integral molecular dynamics (PIMD) [108, 118], ring polymer molecular dynamics (RPMD) [16, 31, 32] and centroid molecular dynamics (CMD) [13, 38]. The comparison with reference data is highly satisfactory and suggest that GC-AdResS may be used as a complementary method for large scale PIMD applications.

1.1 Outline of the thesis

The thesis is organised as follows: In Chapter 2, the important theoretical concepts of statistical mechanics and molecular simulations are reported that are used throughout this thesis. The adaptive resolution method (AdResS) is introduced and the theoretical evolution of the method in the last ten years has been discussed. Finally, Feynman's path integral formalism is discussed and the derivation of path integral partition function is reported. Chapter 3 deals with the calculation of structural properties and the excess chemical potential using GC-AdResS with classical MD. The method is applied to a representative class of liquids and mixtures where the coarse-grained particles interact via a generic WCA potential. Chapter 4 establishes the connection between the Bergmann-Lebowitz model for grand ensemble and GC-AdResS. The principles of this model are used to define equilibrium time correlation functions in open systems. Finally, simulation of liquid water is performed using GC-AdResS with classical MD and the equilibrium time correlation functions are calculated. The functions are compared with the functions calculated in reference MD simulations. Chapter 5 deals with the implementation of different path-integral (PI) techniques in GC-AdResS. Three different PI techniques are discussed and the technical aspects of the implementation in GC-AdResS are reported in detail. Finally, quantum static and dynamic properties of liquid water are calculated using PI-GC-AdResS. These properties are compared with the results from reference PI simulations.

Chapter 2

Theoretical Background

In this chapter, we describe all the different methods and theoretical concepts that have been used in this thesis. In the first part, we discuss the theoretical background of Molecular Dynamics (MD) and the Liouville theorem, a key theorem in statistical mechanics for MD. We describe the relevant computational aspects (integrators / thermostats / potentials) of MD that are important for the work in subsequent chapters. Since there is plethora of literature available, we just focussed on the following books for writing this chapter: *Statistical Mechanics: Theory and Molecular Simulation* [107]; *Molecular Modelling: Principles and Applications* [51]; *Understanding Molecular Simulation: From Algorithms and Applications* [26] and *Computer Simulation of Liquids* [1]. The second part of this chapter deals with coarse-grained methods and Adaptive Resolution Simulations (AdResS). We discuss the theoretical and computational aspects of the AdResS method and review the literature about the AdResS method. In the last part of this chapter, we discuss the path-integral formalism that is used to model systems where nuclear delocalization plays an important role.

2.1 Molecular Dynamics (MD) in a nutshell

In this section, we give a brief overview of MD principles and how macroscopic observables are calculated in MD [1, 26, 51, 107]. MD uses the information at the microscopic level, i.e. coordinates and momenta of the particles, and transforms this information into macroscopic data such as energy, pressure, heat capacity etc. This is done by solving Newton's equations of motions for a relatively small number of particles (compared to realistic system found in nature) ($N \sim 10^{3-5}$). The position and momenta of these N particles can be considered as points in a multidimensional phase space with $6N$ dimensions. Each point in the phase space represents a microstate. If we represent the coordinates and momenta of the i^{th} particle by $\mathbf{q}_i$ and $\mathbf{p}_i$ respectively, then the set of coordinates and momenta of N particles can be written as:

$$\mathbf{q} \equiv \mathbf{q}_1, ..., \mathbf{q}_N \tag{2.1}$$

and

$$\mathbf{p} \equiv \mathbf{p}_1, ..., \mathbf{p}_N \tag{2.2}$$

If we denote the microstate by $(\mathbf{q}, \mathbf{p})$, then the instantaneous value of any observable A can be denoted by $a(\mathbf{q}, \mathbf{p})$. The macroscopic observable A can be obtained by taking average

of $a(\mathbf{q}, \mathbf{p})$ over a long time duration [1].

$$A = \langle a(\mathbf{q}, \mathbf{p}) \rangle_{time} = \lim_{t \to \infty} \int_0^t \frac{1}{t'} a(\mathbf{q}, \mathbf{p}, t') dt' \tag{2.3}$$

where t represents a sufficiently long time. In MD simulation, t is divided into a large number of time steps (τ), where the length of each time step is $\Delta t = t/\tau$. The equations of motions are solved at each step, and the coordinates and momenta of N particles are updated. Thus, in MD the integral in Eq. 2.3 can be replaced by a summation:

$$A = \langle a(\mathbf{q}, \mathbf{p}) \rangle = \frac{1}{\tau} \sum_0^\tau A(\mathbf{q}, \mathbf{p}, t) \Delta t \tag{2.4}$$

2.2 Liouville Theorem

Liouville theorem is an important theorem in statistical mechanics that describes the time evolution of the phase space probability distribution function(See Refs. [1, 107]). In this section, we show the derivation of the Liouville equation and discuss the theoretical foundations of the ensemble theory in statistical mechanics. We also discuss the equilibrium solutions to the Liouville equation and the various ensembles that correspond to different macroscopic parameters. Finally, we discuss ergodicity and how one can calculate phase space averages using MD.

2.2.1 Ensemble theory and Liouville equation

An ensemble is defined as a collection of microstates in the phase space that have the same set of macroscopic properties [35, 101]. The probability distribution function $\rho(\mathbf{q}, \mathbf{p}, t)$ describes how the systems in the ensemble are distributed in the phase space at any given time. This function is specific to the choice of macroscopic parameters (N, V, P, T etc). Each system represents a point in the phase space which evolves in time, with no dependency on the evolution of other systems. The Liouville theorem states that $\rho(\mathbf{q}, \mathbf{p}, t)$ is conserved in time:

$$\frac{d\rho(\mathbf{q}, \mathbf{p}, t)}{dt} = 0 \tag{2.5}$$

The total time derivative of $\rho(\mathbf{q}, \mathbf{p}, t)$ can be written as:

$$\frac{d\rho(\mathbf{q}, \mathbf{p}, t)}{dt} = \frac{\partial \rho(\mathbf{q}, \mathbf{p}, t)}{\partial t} + \sum_{i=1}^{N} \dot{\mathbf{q}}_i \frac{\partial \rho(\mathbf{q}, \mathbf{p}, t)}{\partial \mathbf{q}_i} + \sum_{i=1}^{N} \dot{\mathbf{p}}_i \frac{\partial \rho(\mathbf{q}, \mathbf{p}, t)}{\partial \mathbf{p}_i} \tag{2.6}$$

A Lioville operator L is defined such that:

$$iL = \sum_{i=1}^{N} \dot{\mathbf{q}}_i \frac{\partial}{\partial \mathbf{q}_i} + \sum_{i=1}^{N} \dot{\mathbf{p}}_i \frac{\partial}{\partial \mathbf{p}_i} \tag{2.7}$$

Since total derivative of $\rho(\mathbf{q}, \mathbf{p}, t)$ is zero, Eq. 2.6 can be rewritten as:

$$\frac{\partial \rho(\mathbf{q}, \mathbf{p}, t)}{\partial t} = -(\sum_{i=1}^{N} \dot{\mathbf{q}}_i \frac{\partial \rho(\mathbf{q}, \mathbf{p}, t)}{\partial \mathbf{q}_i} + \sum_{i=1}^{N} \dot{\mathbf{p}}_i \frac{\partial \rho(\mathbf{q}, \mathbf{p}, t)}{\partial \mathbf{p}_i}) \tag{2.8}$$

The time evolution of the system can be described by Hamilton's equations [33]:

$$\frac{d\mathbf{p}_i}{dt} = -\frac{\partial \mathcal{H}(\mathbf{q},\mathbf{p})}{\partial \mathbf{q}_i} \tag{2.9}$$

and

$$\frac{d\mathbf{q}_i}{dt} = +\frac{\partial \mathcal{H}(\mathbf{q},\mathbf{p})}{\partial \mathbf{p}_i} \tag{2.10}$$

where $\mathcal{H} = \mathcal{H}(\mathbf{q},\mathbf{p})$ is the Hamiltonian of the system Therefore, once can write Eq. 2.11 as [1]:

$$\frac{\partial \rho(\mathbf{q},\mathbf{p},t)}{\partial t} = -(\sum_{i=1}^{N} \frac{\partial \mathcal{H}(\mathbf{q},\mathbf{p})}{\partial \mathbf{p}_i}\frac{\partial \rho(\mathbf{q},\mathbf{p},t)}{\partial \mathbf{q}_i} - \sum_{i=1}^{N} \frac{\partial \mathcal{H}(\mathbf{q},\mathbf{p})}{\partial \mathbf{q}_i}\frac{\partial \rho(\mathbf{q},\mathbf{p},t)}{\partial \mathbf{p}_i}) \tag{2.11}$$

which can further be represented in terms of Poisson bracket as:

$$\frac{\partial \rho(\mathbf{q},\mathbf{p},t)}{\partial t} = -\left\{\rho(\mathbf{q},\mathbf{p},t), \mathcal{H}(\mathbf{q},\mathbf{p})\right\} \tag{2.12}$$

This equation can be written in terms of Liouville operator introduced in Eq. 2.7:

$$\frac{\partial \rho(\mathbf{q},\mathbf{p},t)}{\partial t} = -iL\rho(\mathbf{q},\mathbf{p},t) \tag{2.13}$$

where $i = \sqrt[2]{-1}$. The Liouville operator is defined as:

$$iL* = \{*, \mathcal{H}\} \tag{2.14}$$

which means that the function on which Liouville operator acts gets substituted in place of "$*$" in the Poisson brackets. The formal solution to Eq. 2.13 can be written as:

$$\rho(\mathbf{q},\mathbf{p},t) = e^{-iLt}\rho(\mathbf{q},\mathbf{p},0) \tag{2.15}$$

Similarly, it can be shown that for any phase space function $a(\mathbf{p},\mathbf{q})$, the time evolution of the function can be expressed in terms of Liouville operator:

$$a(\mathbf{q},\mathbf{p},t) = e^{iLt}a(\mathbf{q},\mathbf{p}) \tag{2.16}$$

The operator e^{iLt} is referred to as the classical propagator.

2.2.2 Equilibrium Solutions of the Liouville Equation

The macroscopic variable A can be computed from ensemble averages, which is defined as [1]:

$$A = \langle A \rangle = \int \rho(\mathbf{q},\mathbf{p},t)A(\mathbf{q},\mathbf{p})d\mathbf{q}d\mathbf{p} \tag{2.17}$$

If ρ depends explicitly on time, then the observable will also depend on time. If we consider the system at equilibrium, then the macroscopic variable should be time independent. This will be possible only if the probability distribution function does not depend on time; i.e.

$$\frac{\partial \rho(\mathbf{q},\mathbf{p},t)}{\partial t} = 0 \tag{2.18}$$

Thus, Eq. 2.12 reduces to:

$$\{\rho(\mathbf{q},\mathbf{p},t),\mathcal{H}(\mathbf{q},\mathbf{p})\}=0 \tag{2.19}$$

The solution to Eq. 2.19 is that the probability distribution function is a function of H:

$$\rho(\mathbf{q},\mathbf{p})=\frac{1}{Z}F(\mathcal{H}(\mathbf{q},\mathbf{p})) \tag{2.20}$$

Here, Z is a normalization factor defined as:

$$Z=\int F(\mathcal{H}(\mathbf{q},\mathbf{p}))d\mathbf{q}d\mathbf{p} \tag{2.21}$$

Z is also known as partition function which describes the statistical properties of a system in equilibrium. It quantifies the number of microstates in the phase space that are accessible within the ensemble. This quantity is specific to the choice of macroscopic parameters used to define the ensemble. The three most commonly used ensembles are(See Ref. [35, 101]):

- Microcanonical ensemble or constant NVE ensemble: It represents the possible microstates of the system with constant number of particles (N), constant volume (V) and constant energy (E). Such a system is referred to as "isolated" system, because the system cannot exchange energy with the surroundings, and thus energy is conserved in this system. The ensemble distribution function for the microcanonical ensemble is defined as:

  $$\rho(\mathbf{q},\mathbf{p})=\frac{1}{Z}\delta(\mathcal{H}(\mathbf{q},\mathbf{p}))-E) \tag{2.22}$$

  where the partition function Z is defined as:

  $$Z=\frac{1}{N!}\frac{1}{h^{3N}}\int\delta(\mathcal{H}(\mathbf{q},\mathbf{p})-E)d\mathbf{q}d\mathbf{p} \tag{2.23}$$

  The distribution function in Eq. 2.22 selects out those states that have energy E, while discarding other states. For an isolated system in thermodynamic equilibrium, all the accessible microstates have the same probability.

- Canonical Ensemble or constant-NVT ensemble: This ensemble represents the microstates of the system which has constant number of particles (N), constant volume (V) and constant temperature (T). These conditions characterize the "closed" system, which exchanges energy with the surroundings. The surroundings can be considered as an infinite heat source, which keeps the system at a fixed temperature T, by driving energy fluctuations within the system. The ensemble distribution function for the canonical ensemble is defined as:

  $$\rho(\mathbf{q},\mathbf{p})=\frac{1}{Z}e^{\left(-\frac{\mathcal{H}(\mathbf{q},\mathbf{p})}{K_BT}\right)} \tag{2.24}$$

  where Z is:

  $$Z=\frac{1}{N!h^{3N}}\int e^{\left(-\frac{\mathcal{H}(\mathbf{q},\mathbf{p})}{K_BT}\right)}d\mathbf{q}d\mathbf{p} \tag{2.25}$$

  The distribution in Eq. 2.24 is referred to as Boltzmann distribution. In contrast to NVE system, the Hamiltonian of the system is not conserved when temperature is fixed.

- Grand-Canonical or constant-μVT ensemble: This ensemble represents the microstates of the system which has a constant chemical potential (μ), constant volume (V) and constant temperature (T). The conditions describe an "open" system, which exchanges energy and matter with the surroundings. Thus, the system is coupled to a reservoir of energy and particles, which keeps the chemical potential μ constant within the system, while driving the fluctuations in the number of particles. The grand-canonical ensemble has been used in the study of order-disorder transitions in binary substitutional alloys [97], anisotropic mass diffusion in a nanoporous medium [37], adsorption isotherms in various zeolites [102] etc. The ensemble distribution function for the grand-canonical ensemble is defined as:

$$\rho(\mathbf{q},\mathbf{p}) = \frac{1}{Z} e^{\left(-\frac{\mathcal{H}(\mathbf{q},\mathbf{p})-\mu N}{K_B T}\right)} \tag{2.26}$$

where Z is:

$$Z = \sum_N \frac{1}{N!}\frac{1}{h^{3N}} e^{\left(\frac{\mu N}{K_B T}\right)} \int e^{\left(-\frac{\mathcal{H}(\mathbf{q},\mathbf{p})}{K_B T}\right)} d\mathbf{q}d\mathbf{p} \tag{2.27}$$

2.2.3 Ergodicity

The macroscopic observables can also be computed from the phase space averages [107]:

$$A = \langle a(\mathbf{q},\mathbf{p})\rangle = \frac{1}{Z}\int a(\mathbf{q},\mathbf{p}) F(\mathcal{H}(\mathbf{q},\mathbf{p})) d\mathbf{q}d\mathbf{p} \tag{2.28}$$

where F(H($\mathbf{q}$, $\mathbf{p}$)) is the ensemble distribution function corresponding to a given set of macroscopic variables. If the system is able to visit all the microstates in the given ensemble in an infinite amount of time, then it is assumed to be ergodic. In such a scenario, the ensemble average in Eq. 2.28 is equal to the time average in Eq. 2.4.

2.3 Integrators and Periodic Boundary Conditions (PBC)

In this section, we discuss the numerical integrators used in MD and give a brief overview on the periodic boundary conditions in MD setup.

2.3.1 Integrators in Molecular Dynamics

As mentioned in Section 2.1, MD algorithm involves updating the positions and velocities of all the molecules at each step. The molecules interact with each other via intermolecular forces which are derived from predefined interaction potentials. The Hamiltonian for N particles interacting with a potential U can be written as [26]:

$$\mathcal{H}(\mathbf{r},\mathbf{p}) = \sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2m_i} + U(\mathbf{r}^N) \tag{2.29}$$

where $\mathbf{r} \equiv \mathbf{r}_1, ..., \mathbf{r}_N$ denotes the cartesian coordinates of the particles. Using the relations in Eq. 2.9 and 2.10 the following microscopic equations of motion are obtained:

$$\frac{d\mathbf{r}_i}{dt} = \frac{\partial \mathcal{H}(\mathbf{r},\mathbf{p})}{\partial \mathbf{p}_i} = \frac{\mathbf{p}_i}{m} \tag{2.30}$$

for the positions and

$$\frac{d\mathbf{p}_i}{dt} = -\frac{\partial \mathcal{H}(\mathbf{r},\mathbf{p})}{\partial \mathbf{r}_i} = -\nabla_i U(\mathbf{r}) \tag{2.31}$$

for the momenta. The force acting on the molecule can be obtained from the potential energy:

$$\mathbf{F}_i(t) = -\nabla_i U(\mathbf{r}) \tag{2.32}$$

The system evolves according to these equations which are solved using a numerical integrator (see Refs. [51, 95]). The Verlet algorithm [109] is one of the many algorithms that can be used to solve these differential equations. Using a Taylor expansion with respect to the coordinates (and discarding the higher order terms), one obtains:

$$\mathbf{r}_i(t+\Delta t) \approx \mathbf{r}_i(t) + \Delta t \mathbf{v}_i(t) + \frac{\Delta t^2}{2}\frac{\mathbf{F}_i(t)}{m} \tag{2.33}$$

Similarly, one can use a Taylor expansion for $r_i(t-\Delta t)$:

$$\mathbf{r}_i(t-\Delta t) \approx \mathbf{r}_i(t) - \Delta t \mathbf{v}_i(t) + \frac{\Delta t^2}{2}\frac{\mathbf{F}_i(t)}{m} \tag{2.34}$$

Adding Eq. 2.33 and Eq. 2.34, one obtains

$$\mathbf{r}_i(t+\Delta t) \approx 2\mathbf{r}_i(t) - \mathbf{r}_i(t-\Delta t) + \frac{\Delta t^2}{2}\frac{\mathbf{F}_i(t)}{m} \tag{2.35}$$

Eq. 2.35 can be used to generate MD trajectories if one gives the initial set of positions. Since the Verlet algorithm does not explicitly involve velocities, it can be modified such that the algorithm updates the coordinates and velocities of the molecule simultaneously; the modified algorithm is called Velocity Verlet [104]. Taylor expansion of the velocity at time $t+\Delta t$ gives:

$$\mathbf{v}_i(t+\Delta t) = \mathbf{v}_i(t) + \Delta t \dot{\mathbf{v}}_i(t) + \frac{\Delta t^2}{2}\ddot{\mathbf{v}}_i(t) \tag{2.36}$$

The unknown term in Eq. 2.36 $\ddot{\mathbf{v}}(t)$ can be obtained from the Taylor expansion of $\dot{\mathbf{v}}(t+\Delta t)$:

$$\ddot{\mathbf{v}}_i(t) = \frac{\dot{\mathbf{v}}_i(t+\Delta t) - \dot{\mathbf{v}}_i(t)}{\Delta t} \tag{2.37}$$

Substituting Eq. 2.37 in Eq. 2.33, one obtains:

$$\mathbf{v}_i(t+\Delta t) = \mathbf{v}_i(t) + \frac{\Delta t}{2}(\dot{\mathbf{v}}_i(t+\Delta t) + \dot{\mathbf{v}}_i(t)) \tag{2.38}$$

Eq. 2.38 can be expressed in terms of forces:

$$\mathbf{v}_i(t+\Delta t) = \mathbf{v}_i(t) + \frac{\Delta t}{2}(\frac{\mathbf{F}_i}{m_i}(t+\Delta t) + \frac{\mathbf{F}_i}{m_i}(t)) \tag{2.39}$$

Eq. 2.33 and 2.39 are used to generate the trajectories in the Velocity-Verlet algorithm. The algorithm works in the following way:

- From the set of coordinates and velocities at time t, the coordinates at time $t+\Delta t$ are obtained.

- Next, forces are calculated at time $t + \Delta t$, using the new set of coordinates.
- Finally, velocities are updated using the new calculated forces.

There is another variant of Velocity Verlet algorithm, known as Leapfrog algorithm [1]. It uses the positions at time t and velocities at time $t - \frac{\Delta t}{2}$ to get the new positions. The system evolves according to the following sets of equations:

$$\mathbf{v}_i(t + \frac{\Delta t}{2}) = \mathbf{v}_i(t - \frac{\Delta t}{2}) + \Delta t \frac{\mathbf{F}_i(t)}{m_i} \mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \Delta t \mathbf{v}_i(t + \frac{\Delta t}{2}) \tag{2.40}$$

The alogithm works as follows:

- Forces at time t are calculated first
- Next, velocities at time $t + \frac{\Delta t}{2}$ from the velocities at time $t - \frac{\Delta t}{2}$ and forces at time t are calculated.
- Finally, positions at time $t + \Delta t$ from the positions at time t and the velocities at time $t + \frac{\Delta t}{2}$ are calculated

Velocity Verlet and Leapfrog algorithm generate identical trajectories with corresponding starting points, but if the input coordinates and velocities are same then the trajectories generated with Velocity Verlet and Leapfrog algorithm are completely different. Both the methods provide good numerical stability and are time-reversible [107].

2.3.2 Periodic Boundary Conditions

In MD, periodic boundary conditions [95] are employed which enable a simulation to be performed using a relatively small number of particles. The effect of the periodic boundary conditions is that particles experience forces as though they were in a bulk solution. The simulation volume represent the primary system that is replicated infinite times in all the directions. The particles which move out of the box after the integration step are brought back into the box from the opposite side.

2.4 Thermostats in MD

Most of the experimental setups work under the conditions of constant temperature, which means that the system can exchange energy with the surroundings. In order to mimic the experimental environment, simulations are usually performed in canonical ensemble. Even if one wants to perform microcanonical simulations, the temperature is adjusted to the desired value in the equilibration phase which requires coupling the system to a thermostat [51]. The surrounding is not explicitly modeled (at molecular level) in MD simulation as it is computationally highly expensive. Instead, the heat exchange between the system and the reservoir is taken into account implicitly by modifying the MD algorithm. There are different thermostating schemes that are routinely used in MD simulations (See Refs. [26, 107]). In this section, we will give a brief overview on some of the most common

thermostats employed in MD simulations. Since the temperature of the system is related to the kinetic energy:

$$\langle KE \rangle_{NVT} = \frac{3}{2} N K_B T, \tag{2.41}$$

the most obvious choice to attain the desired temperature is to scale the velocities at each time step by the factor $\lambda = \sqrt[2]{\frac{T_{target}}{T_{current}}}$, where T_{target} is the target temperature and $T_{current}$ is the temperature calculated at the current time step. This scheme is called "velocity-rescaling" [119]. Another way to achieve the desired temperature is by coupling the system to an external bath which acts as a source of energy [5]. The velocities are scaled at each time step by a scaling factor:

$$\lambda^2 = 1 + \frac{\Delta t}{\tau} \left(\frac{T_{target}}{T(t)} - 1 \right) \tag{2.42}$$

where $T(t)$ is the instantaneous temperature and τ is a coupling parameter that determines how much time the system takes to achieve the target temperature. The problem with the scaling methods is that they do not generate a canonical ensemble [51]. If an efficient canonical sampling is required, Anderson thermostat [2] and Nose-Hoover thermostat [39, 71] are employed in MD simulation. Anderson thermostat is based on the concept of stochastic collisions, where a particle is selected randomly and the velocity of this particle is chosen randomly from Maxwell-Boltzmann distribution. The Nose-Hoover thermostat comes under the category of extended system method, where the reservoir is considered as a part of the system, and is represented as an additional degree of freedom in the Hamiltonian of the system. In the next section, we discuss Langevin thermostat in detail since this thermostat is employed in AdResS simulations.

2.4.1 Langevin Thermostat

Based on stochastic dynamics, Langevin equations were initially formulated as a model for Brownian motion of a particle that is immersed in a solvent (See Refs. [20, 51]). Apart from the forces due to the interaction potential, there are two more forces that all the particles experience. The first force is the friction force on the particle due to the solvent which reduces the kinetic energy of the particles and the second force is due to the random fluctuations caused due to interaction with the solvent, which adds kinetic energy to the particles. The balance of the two forces makes sure that the system attains the desired temperature. The Langevin equation of motion for i^{th} particle can be written as [100]:

$$m_i \frac{d^2 \mathbf{r}_i(t)}{dt^2} = \mathbf{F}_i(\mathbf{r}_i(t)) - \gamma_i \frac{d\mathbf{r}_i(t)}{dt} m_i + R_i(t) \tag{2.43}$$

where γ_i is the friction coefficient and $R_i(t)$ is the random force. The random force is chosen such that its mean is zero and it is uncorrelated in time. Since both the friction force and random force originate from interactions with the solvent, they cannot be independent. The following equation relates the two forces [100]:

$$\langle R(t) R(t^{'}) \rangle = 2 K_B T m_i \gamma_i \delta(t - t^{'}) \tag{2.44}$$

This guarantees that the fluctuation-dissipation theorem is satisfied and the system generates NVT statistics. In MD simulations, one needs to specify the inverse friction coefficient

(τ^{-1}), which has to be chosen carefully. The value of the inverse friction coefficient determines the time that system takes to reach the desired temperature. If a strong coupling is used, then the temperature of the system quickly relaxes to the target temperature. However, dynamical properties such as equilibrium time correlation functions, would differ significantly from the realistic behaviour. This subject has been studied in Ref. [55] where gentle stochastic thermostats have been proposed to calculate the dynamical properties.

2.5 Force Fields

Force fields in molecular mechanics contain potential functions that depend on the nuclear positions which makes it possible to perform calculations of a large number of particles [51]. Most of the force fields used contain simple functional forms of the intramolecular and intermolecular potentials. In this section we discuss the most commonly used potential functions used in molecular simulations.

2.5.1 Intramolecular potentials

The intramolecular potentials or bonded potentials (See Refs. [1, 26, 51]) consist of simple terms that describe the deviation of angles and bonds between different atoms within the same molecule from the "equilibrium" values. The stretching of the bond between two atoms i and j is often modeled using a harmonic potential:

$$U_b(r_{ij}) = \frac{1}{2}k^b_{ij}(r_{ij} - bij)^2 \tag{2.45}$$

where k^b_{ij} is the force constant and b_{ij} is the equilibrium bond length. The angular potential function describes the deviation of the angle formed by three consecutive atoms i, j and k from its equilibrium value:

$$U_\theta(\theta_{ijk}) = \frac{1}{2}k^\theta_{ijk}(\theta_{ijk} - \theta^0_{ijk})^2 \tag{2.46}$$

where k^θ_{ijk} is the force constant and θ^0_{ijk} is the equilibrium bond angle. Apart from the bond-stretching and angle-bending terms, there exists a dihedral angle potential which depends on the four consecutive bonded atoms. This potential is mostly used to constrain the rotation around a bond. It depends on the torsion angle (ϕ_{ijkl}) which is the angle between the plane going through i, j and k and the plane going through j, k and l atoms. The harmonic dihedral angle potential is expressed as:

$$U_{ijkl}(\phi_{ijkl}) = \frac{1}{2}k^\phi_{ijkl}(\phi_{ijkl} - \phi^0_{ijkl})^2 \tag{2.47}$$

where k_{ijkl} is the force constant and ϕ^0_{ijkl} is the equilibrium value of the torsion angle.

2.5.2 Intermolecular interactions

The intermolecular or nonbonded interactions (See Refs. [1,26,51]) act between two particles that are not bonded to each other. They are assumed to be pairwise additive:

$$U(\mathbf{r}_1, ..., \mathbf{r}_N) = \sum_{i<j} U_{ij}(r_{ij}) \tag{2.48}$$

and symmetric:

$$\mathbf{F}_i = -\nabla_i V_{ij} = -\mathbf{F}_j \tag{2.49}$$

Nonbonded interactions can be categorized as Van der Waals interactions and electrostatic interactions. Van der Waals potential take into account the dispersive and exchange-repulsive interaction between two atoms. It consists of two components: an attractive part that varies as r^{-6} and a repulsive part that varies as r^{-12}. They can be modeled using Lennard-Jones function which has the following form [42, 43]:

$$V(r_{ij}) = 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \tag{2.50}$$

where $r_{ij} = |r_i - r_j|$ is the distance between two particles, ϵ_{ij} is the minimum of the potential and σ_{ij} represents the separation for which the energy is zero between the two particles. Lennard-Jones interactions are considered to be short-ranged since the forces due to these interactions decay faster than r^{-3}. Such interactions are usually computed by imposing a cut-off (r_c) to the potential such that the potential is assumed to be zero outside this cut-off. If the cut-off is sufficiently large, then the effects of such a truncation are negligible. The interactions are then computed using minimum image convention [95], where each particle interacts with the closest replica of the rest of the particles in the system. The particle interacts with all the other particles that are inside the cut-off radius, and the interactions with the particles outside this sphere are discarded. The cut-off method, however, produces a drift in the energy due to the discontinuity at $r = r_c$, thus it is difficult to use in NVE simulations. The electrostatic interactions are long-ranged because of their slow power-law decay ($\frac{1}{r}$), which makes them difficult to handle in MD since the simulation cell and all its periodic images must be considered. Using a spherical cut-off would create artifacts due to the movement of the charged particles in and out of the spherical surface. The most common method for treating long-range interactions is Ewald summation [1] which efficiently sums the interaction between a charged particle and all the periodic images of the simulation cell. The interaction is divided into two components: the first component corresponds to the short-range contribution which is computed in real space and the second component represents the long-range contribution which is calculated in the Fourier space. Due to the difficulties in handling long-ranged interactions in AdResS, Reaction-Field method [106] is implemented in AdResS method for computing electrostatic interactions. The method comprises of two different contributions. The first contribution consists of the the interactions of the charged particle i with all the particles inside the sphere with cut-off radius $r = r_c$. It can be written as:

$$U_1 = \sum_{i=1}^{N}\sum_{j<i}^{N}\frac{q_i q_j}{r_{ij}} \tag{2.51}$$

where q_i and q_j represents the charges on i^{th} and j^{th} atom respectively. The second contribution comes from the particles outside the cut-off radius. Their interactions is treated in an average way by considering that they form a dielectric continuum (ϵ_s) which produces a reaction field inside the sphere. The strength of the field is directly proportional

to the dipole moment within the sphere:

$$\epsilon_i = \left[\frac{2(\epsilon_s - 1)}{2\epsilon_s + 1}\right] \frac{1}{r_c^3} \sum_{k \in R} \mu_j \tag{2.52}$$

where μ_j is the dipole moment of molecule j and the sum runs over all the molecules within the sphere. The total energy from the reaction field is:

$$U_2 = -\frac{1}{2} \sum_{i=1}^{N} \mu_i \epsilon_i \tag{2.53}$$

Due to the movement of molecules in and out of the sphere, a discontinuous jump in energy takes place leading to a poor energy conservation. It implies that reaction-field method is computationally cheaper than Ewald summation, but a thermostat is needed in MD simulations.

2.6 Properties of Interest

In this thesis, we have several different physical properties to demonstrate the grand-canonical characteristic of AdResS method. However, the are some properties that have been calculated frequently in this work. These properties are radial distribution functions, equilibrium time correlation functions and excess chemical potential. In this section, we briefly discuss how these quantities are calculated in molecular simulations.

2.6.1 Radial Distribution Functions

Radial distribution functions or pair correlation functions ($g(\mathbf{r}_1, \mathbf{r}_2)$) are used to characterize the structure of the liquids. The second order distribution function ($g(\mathbf{r}_1, \mathbf{r}_2)$) in a homogenous system is defined as [107]:

$$g^{(2)}(\mathbf{r}_1, \mathbf{r}_2) = \frac{V^2(N-1)}{N Z_N} \int ... \int f(\mathcal{H}(\mathbf{r})) d\mathbf{r}_3 ... d\mathbf{r}_N \tag{2.54}$$

where $f(\mathcal{H}(\mathbf{r}))$ is the ensemble distribution function, Z_N is the partition function corresponding to a given ensemble, V is the volume, N is the number of particles and U is the potential acting between particles. This equation gives the probability of finding a particle at $\mathbf{r}_2$ given a particle is at $\mathbf{r}_1$, with respect to the ideal gas distribution. In case of spherically symmetric particles, $g^{(2)}(\mathbf{r}_1, \mathbf{r}_2)$ depends on the relative distance between two particles. Thus $g^{(2)}(\mathbf{r}_1, \mathbf{r}_2)$ can be written as $g(r)$, which can be defined as:

$$g(r) = \frac{V}{N^2} \langle \sum_i \sum_{k \neq i} \Delta(r - r_{ik}) \rangle \tag{2.55}$$

where the angular brackets represent the ensemble average. The radial distribution function of a liquid consists of a small number of peaks at short distances which decay to the value of ideal gas, which indicates that there is no long-range order in the liquids. To compute $g(r)$, the neighbours around each particle are put into concentric spheres or bins that are separated by distance Δr. The number of particles in each bin is then averaged over all the simulation time frames. The result is then normalized by the volume of each bin and the average density of atoms in the system.

2.6.2 Equilibrium time correlation functions

Molecular Dynamics enables one to correlate the value of a property at some time to the value of another property at another time; the function that stores this correlation at different t is called equilibrium time correlation function (See Refs. [1, 51, 107]). The equilibrium time correlation function between two observables A and B corresponding to the phase space functions $a(\mathbf{p},\mathbf{q})$ and $b(\mathbf{p},\mathbf{q})$ is given by [107]:

$$\begin{aligned} C_{AB}(t) = \langle a(0)b(t)\rangle &= \int d\mathbf{p}d\mathbf{q} f(\mathbf{p},\mathbf{q}) a(\mathbf{p},\mathbf{q}) e^{iLt} b(\mathbf{p},\mathbf{q}) \\ &= \int d\mathbf{p}d\mathbf{q} f(\mathbf{p},\mathbf{q}) a(\mathbf{p},\mathbf{q}) b(\mathbf{p}_t(\mathbf{p},\mathbf{q}),\mathbf{q}_t(\mathbf{p},\mathbf{q})) \end{aligned} \tag{2.56}$$

where, $a(\mathbf{p},\mathbf{q})$ and $b(\mathbf{p},\mathbf{q})$ are phase space functions corresponding to the observables A and B respectively, $a(0) = a(t=0)$ and $b(t)$ is the function at time t, $f(\mathbf{p},\mathbf{q})$ is the equilibrium distribution function and the dynamics is generated by the Liouville operator iL. $\mathbf{p}_t(\mathbf{p},\mathbf{q}), \mathbf{q}_t(\mathbf{p},\mathbf{q})$ indicates the time evolution of the positions and momenta $\mathbf{p},\mathbf{q}$. For the canonical ensemble, Eq. 4.15 can be written as:

$$C_{AB}(t) = \frac{1}{Q_N} \int d\mathbf{p}d\mathbf{q} e^{-\frac{\mathcal{H}_N(\mathbf{p},\mathbf{q})}{K_B T}} a(\mathbf{p},\mathbf{q}) b(\mathbf{p}_t(\mathbf{p},\mathbf{q}),\mathbf{q}_t(\mathbf{p},\mathbf{q})). \tag{2.57}$$

where Q_N is the Canonical partition function and $\mathcal{H}_N(\mathbf{p},\mathbf{q})$ the Hamiltonian of a system with N (constant) molecules. If the functions $a(\mathbf{p},\mathbf{q})$ and $b(\mathbf{p},\mathbf{q})$ are same, then the correlation function is called auto-correlation function.

$$C_{AA}(t) = \frac{1}{Q_N} \int d\mathbf{p}d\mathbf{q} e^{-\frac{\mathcal{H}_N(\mathbf{p},\mathbf{q})}{K_B T}} a(\mathbf{p},\mathbf{q}) a(\mathbf{p}_t(\mathbf{p},\mathbf{q}),\mathbf{q}_t(\mathbf{p},\mathbf{q})). \tag{2.58}$$

The autocorrelation function gives an indication of how the system loses the memory of its initial conditions. The characteristic time over which the trajectory $\mathbf{p}_t(\mathbf{p},\mathbf{q}), \mathbf{q}_t(\mathbf{p},\mathbf{q})$ depends on the initial condition $\mathbf{p},\mathbf{q}$ and beyond which $\mathbf{p}_t(\mathbf{p},\mathbf{q}), \mathbf{q}_t(\mathbf{p},\mathbf{q})$ does not possess any memory of its initial condition is called correlation time (t_{corr}). For times smaller than than the correlation time, $a(\mathbf{p},\mathbf{q})$ and $a(\mathbf{p},\mathbf{q})a(\mathbf{p}_t(\mathbf{p},\mathbf{q}),\mathbf{q}_t(\mathbf{p},\mathbf{q}))$ are highly correlated while for times larger than the correlation time, $a(\mathbf{p},\mathbf{q})$ and $a(\mathbf{p}_t(\mathbf{p},\mathbf{q}),\mathbf{q}_t(\mathbf{p},\mathbf{q}))$ do not have any correlation. The correlation time is highly dependent on the type of property considered. If the system is ergodic, the adequate sampling of the ensemble distribution function $e^{-\frac{\mathcal{H}_N(\mathbf{p},\mathbf{q})}{K_B T}}$ is ensured. In this case, the phase space average can be replaced by time average:

$$\begin{aligned} C_{AA}(\tau) &= \int d\mathbf{p}d\mathbf{q} e^{-\frac{\mathcal{H}_N(\mathbf{p},\mathbf{q})}{K_B T}} a(\mathbf{p},\mathbf{q}) a(\mathbf{p}_\tau(\mathbf{p},\mathbf{q}),\mathbf{q}_\tau(\mathbf{p},\mathbf{q})) \\ &= \lim_{\tau\to\infty} \frac{1}{\tau} \int dt a(\mathbf{p},\mathbf{q}) a(\mathbf{p}_{\tau+t}(\mathbf{p},\mathbf{q}),\mathbf{q}_\tau(\mathbf{p},\mathbf{q})) \end{aligned} \tag{2.59}$$

If the trajectory time is much larger than the correlation time, then the trajectory can be subdivided into a large number of segments which act as independent sampling points for the correlation functions, since the correlation function decays to zero over these segments. The length of each of these segments is slightly greater than the correlation time. Suppose there are M points in a trajectory, $i = 1...M$ and if subdivide the trajectory into segments

that consist of K points ($K << M$), then the length of each segment is $K\Delta T$, where $K\Delta t > t_{corr}$. The segments can be selected in different ways with different origin points. Thus, Eq. 2.59 can be written in discrete form:

$$C_{AA}(n\Delta t) = \sum_{m=1}^{M-n} a(\mathbf{p}_{m\Delta t}(\mathbf{p},\mathbf{q}), \mathbf{q}_{m\Delta t}(\mathbf{p},\mathbf{q}))a(\mathbf{p}_{(m+n)\Delta t}(\mathbf{p},\mathbf{q}), \mathbf{q}_{(m+n)\Delta t}(\mathbf{p},\mathbf{q})), n = 0, ..., K \tag{2.60}$$

Some correlation functions involve quantities that are averaged over all the molecules in the system, eg. velocity auto correlation function, orientational correlation function while some involve quantities that are calculated for the whole system, eg. dipole moment correlation function where the total dipole moment is the vector sum of all the individual dipoles corresponding to the molecules in the system.

2.6.3 Excess Chemical Potential

In Chapter 3, we will discuss the potentiality of AdResS method to calculate the excess chemical potential of pure liquids and solutes in different mixtures. The chemical potential is defined as the change in free energy of the system due to the addition of one particle in the system [26]. The excess chemical potential is defined as the difference between the chemical potential of a system and the chemical potential of ideal gas under the same condition. In this section, we discuss the two techniques that are generally used to calculate the chemical potential: Widom Insertion [116] and Thermodynamic Integration [105].

Widom Insertion

The canonical ensemble partition function can be expressed in terms of coordinates as:

$$Q(N,V,T) = \frac{V^N}{N!\lambda^{3N}} \int d\mathbf{r}^N e^{-\beta U(\mathbf{r}_1,...,\mathbf{r}_N)} \tag{2.61}$$

where V is the volume, λ is the thermal wavelength and U is the interaction potential. In the thermodynamic limit, the chemical potential is given by:

$$\mu = -k_B T \log\left(\frac{Q(N+1,V,T)}{Q(N,V,T)}\right) \tag{2.62}$$

Substituting $Q(N+1,V,T)$ and $Q(N,V,T)$ in Eq. 2.62, we get

$$\mu = -K_B T \log\left(\frac{V}{\lambda^3(N+1)}\right) - K_B T \log\left(\frac{\int d\mathbf{r}^{N+1} e^{-\beta U(\mathbf{r}_1,...,\mathbf{r}_{N+1})}}{\int d\mathbf{r}^N e^{-\beta U(\mathbf{r}_1,...,\mathbf{r}_N)}}\right) = \mu_{id} + \mu_{ex} \tag{2.63}$$

The total chemical potential can be separated into two parts: one is the ideal gas contribution (μ_{id}) and other is the excess contribution (μ_{ex}). Since μ_{id} can be evaluated analytically, one evaluated μ_{ex} from the molecular simulations. The potential energy of the system with $N+1$ particle can be subdivided into the potential energy of N particle system and interaction energy of $(N+1)^{th}$ particle with all the other particles (ΔU).

$$\Delta U = U(\mathbf{r}^{N+1}) - U(\mathbf{r}^N) \tag{2.64}$$

The excess chemical potential can be expressed in terms of ΔU:

$$\mu_{ex} = -K_B T \int d\mathbf{r}^{N+1} \langle e^{-\beta \Delta U} \rangle_N \tag{2.65}$$

where $\langle . \rangle$ denotes the canonical averaging over N particle phase space. The ensemble average can be computed using Metropolis Monte Carlo scheme. The algorithm works as follows: A Monte-Carlo simulation of N particles is performed. A random particle is generated at frequent intervals and $e^{\beta \Delta U}$ is evaluated for this random particle at position $\mathbf{r}^{N+1}$. This quantity is then divided by number of times the particle was generated. To achieve good and reliable statistics, many insertions are performed. It should be noted that the random particle is considered virtual and does not remain in the system since the trial confirmations are not accepted.

Thermodynamic Integration

The thermodynamic integration method calculates the free energy difference between two systems (A and B) with the same macroscopic properties but with different potentials (U_A and U_B). If the free energy of one of these systems is known, then one can compute the free energy of the other system. The free energy difference is integrated along a reversible path from A to B. At any state between A and B, the potential is assumed to depend linearly on a coupling parameter λ

$$U(\lambda) = (1 - \lambda) U_A + \lambda U_B \tag{2.66}$$

If $\lambda = 0$, then the potential of the state is equal to the potential of the system A (whose free energy is known), and if $\lambda = 1$, then the potential is equal to the potential of the system B. The free energy (F) can be expressed in terms of the partition function as:

$$F(\lambda) = -K_B T \log Q(N, V, T, \lambda) \tag{2.67}$$

where $Q(N, V, T, \lambda) = \frac{V^N}{N! \lambda^{3N}} \int d\mathbf{r} e^{-\beta U(\lambda)}$ is the partition function corresponding to the system with potential $U(\lambda)$. The derivative of $F(\lambda)$ with respect to λ can be written as:

$$\frac{\partial F(\lambda)}{\partial \lambda} = -\frac{1}{K_B T Q(N, V, T, \lambda)} \frac{\partial Q(N, V, T, \lambda)}{\partial \lambda} = \frac{\int d\mathbf{r} \left(\frac{\partial U(\lambda)}{\partial \lambda} \right) e^{-K_B T U(\lambda)}}{\int d\mathbf{r} e^{-K_B T U(\lambda)}} = \langle \frac{\partial U(\lambda)}{\partial \lambda} \rangle_\lambda \tag{2.68}$$

Thus, one can write $\frac{\partial F(\lambda)}{\partial \lambda}$ in terms of ensemble average which can be evaluated using Monte Carlo or molecular dynamics simulation. If we integrate Eq. 2.68, then the free energy difference can be written as:

$$F(\lambda = 1) - F(\lambda = 0) = \int_{\lambda=0}^{\lambda=1} d\lambda \langle \frac{\partial U(\lambda)}{\partial \lambda} \rangle_\lambda \tag{2.69}$$

This integral can be evaluated by choosing a large number of values of λ, and adding the ensemble averages calculated at each value of λ. In order to calculate the excess chemical potential, two systems are considered. In the first system, there are $N-1$ particles and one test particle that is not interacting with the other particles, while in the second system, the

test particle is fully coupled to rest of the $N-1$ particles. The excess chemical potential is the difference of the free energy of the two systems, which can be evaluated using the procedure described before. The test particle is called "λ particle", and its potential is expressed in terms of λ as done in Eq. 2.66. The direct use of the potential $U(\lambda)$ leads to very large values of $\langle\frac{\partial U(\lambda)}{\partial\lambda}\rangle_\lambda$ when the particle suddenly starts appearing or disappearing (λ close to 0 or 1). This numerical singularity is avoided by using a soft-core potential [26].

2.7 Coarse-Graining technique

Multiscale modeling refers to the hierarchy of simulation approaches that are employed to simulate the system at different scales (See Refs. [78, 89]). An appropriate simulation scheme is chosen to simulate the system for a given length and time scale. Coarse-grained (CG) methods are usually employed when one needs to access very large length and time scales. In this section, we will discuss structure-based coarse-graining where one represents a group of atoms in a molecule by a single "superatom". Since the number of explicit degrees of freedom are reduced, MD simulations of large systems that require large computational resources may be performed, which otherwise would have been prohibitive. Since several internal frequencies are eliminated in the coarse-grained model, one can choose large time steps, which makes it possible to access large time scales. While creating a coarse-grained representation, one has to formulate a mapping function that maps the full-atomistic system on coarse-grained superatoms. This implies the derivation of new coarse-grained potentials for the superatoms such that the CG system reproduces the properties of the reference full-atomistic system [68]. It should be noted that several properties cannot be reproduced in the CG system since they may depend strongly on the atomistic details.

2.7.1 Mapping atomistic degrees of freedom on a superatom

A mapping function is required to construct the coarse-grained representation of the full-atomistic model. It takes as input the coordinates of the full-atomistic system and maps it to a unique coarse-grained representation. Suppose $\mathbf{r} = \mathbf{r}_{AT}$ represents the coordinates of the the molecules in the full-atomistic system and $\mathbf{R} = \mathbf{r}_{CG}$ denote the coordinates of the molecule in the coarse-grained system, then the mapping function M has the form [70]:

$$\mathbf{R} = M(\mathbf{r}) \tag{2.70}$$

In all the AdResS simulations performed in this thesis, superatoms are defined at the center-of-mass (COM) of the molecule in the full-atomistic representation. Thus, the coarse-grained coordinates are given by:

$$\mathbf{R}_i^{CG} = \frac{\sum m_j \mathbf{r}_j}{\sum m_j} \tag{2.71}$$

The summation is done over all the atoms in molecule and the atomistic masses are denoted by m_j. It is preferred that large molecules such as polymers are represented by more than one coarse-grained superatom. In case of multiple CG atoms, the bonded interactions are present between the neighbouring superatoms as well.

2.7.2 Calculation of the coarse-grained potential

The potential between the CG superatoms is derived in such a way that the CG system reproduces some of the properties of the reference full-atomistic system. There are various methods in literature that are used to obtain the CG potential. Most of the methods require simulation of a very small representative atomistic system and the information is transferred from this simulation to the CG level. The 'optimal' CG potential is the one which is not pairwise additive and can take any functional form. Furthermore, the properties obtained from the CG simulation with the 'optimal' potential would be the same as in the full-atomistic system. This potential can be obtained by matching the configurational probability for $\mathbf{R}$ in the CG phase space to that in the atomistic phase space [70]:

$$P_{CG}(\mathbf{R}) = P_{AT}(\mathbf{R}) \tag{2.72}$$

where $P_{AT}(\mathbf{R})$ is defined as:

$$P_{AT}(\mathbf{R}) = \int P_{AT}(\mathbf{r})\delta[M(\mathbf{r}) - \mathbf{R}]d\mathbf{r} \tag{2.73}$$

The delta function separates out all the configurations in the atomistic phase space $\mathbf{r}$ that are mapped onto the coarse-grained configuration $\mathbf{R}$. For canonical ensemble, one obtains:

$$\frac{e^{-\beta U_{CG}(\mathbf{R})}}{Z_{CG}} = \int \frac{e^{-\beta U_{AT}(\mathbf{R})}}{Z_{AT}}\delta(M(\mathbf{r}) - \mathbf{R})d\mathbf{r} \tag{2.74}$$

Taking the logarithm on both sides, one obtains:

$$U_{CG}(\mathbf{R}) = -\beta \log \int e^{-\beta U_{AT}(\mathbf{r})}\delta(M(\mathbf{r}) - \mathbf{R})d\mathbf{r} + constant \tag{2.75}$$

The potential $U_{CG}(\mathbf{R})$ is the potential of mean force or the free energy surface that corresponds to the reduced degrees of freedom $\mathbf{R}$. As mentioned before, $U_{CG}(\mathbf{R})$ is not a pairwise sum of interactions between the superatoms. Infact, it depends on all the CG degrees of freedom and hence it can not be used in the simulations as it would be computationaly very expensive. To obtain an effective two-body potential one approximates the potential of mean force using a variety of methods. Some of the methods are Iterative Boltzmann Inversion [96], Pressure correction [112], Force matching [23], Inverse Monte Carlo [61], Relative Entropy minimization [63] etc. The choice of the method depends on the properties that the CG model should reproduce.

2.8 Adaptive Resolution Simulation (AdResS)

The Adaptive Resolution Simulation (AdResS) method [27,89,111] comes under the branch of concurrent coupling methods [18], where different scales are treated within the same simulation system simultaneously. The space is divided into different subregions where each subregion has a different molecular representation. The particles move between different regions and change their resolution accordingly while preserving the thermodynamic properties of the region of interest (high resolution region). The difference with methods like QM/MM [15] is that the free exchange of molecules from one subregion to another

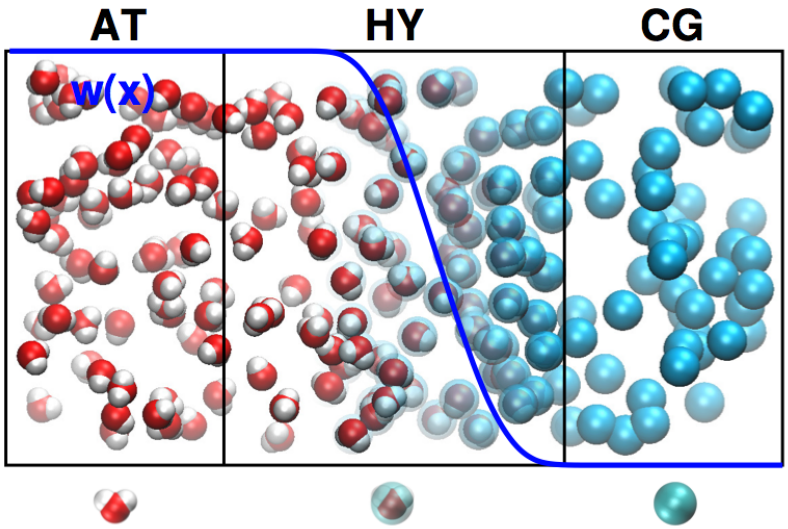

Figure 2.1: Pictorial representation of the adaptive box containing water in different molecular representation. The region on the left, is the low resolution region (coarse-grained), the central part is the transition (hybrid) region Δ, where the switching function $w(x)$ is defined, and the region on the right, is the high resolution region (atomistic).

subregion is allowed; this implies that the particle density fluctuations are not artificially suppressed. In this part of the chapter, we describe the theoretical concepts of the AdResS method, that were developed over the course of ten years, starting from an empirical intuitive idea to the grand-canonical like setup of the latest version.

2.8.1 Theoretical concepts of AdResS

The main components of the original AdResS is idea are based on an empirical principle:

- The simulation box is divided in three regions, one is represented by atomistic resolution, another one is represented by coarse-grained resolution and in between, there is a third region where the molecules have a mixed resolution, where both atomistic and coarse-grained degrees of freedom are present (see Figure 2.1).

- Molecules in the different regions are coupled through a spatial interpolation formula on the forces:

$$\mathbf{F}_{\alpha\beta} = w(X_\alpha)w(X_\beta)\mathbf{F}^{Atom}_{\alpha\beta} + [1 - w(X_\alpha)w(X_\beta)]\mathbf{F}^{CG}_{\alpha\beta} \tag{2.76}$$

  where α and β indicate two molecules, $\mathbf{F}^{Atom}$ is the force derived from the atomistic force field and $\mathbf{F}^{CG}$ from the corresponding coarse-grained potential, X is the x coordinate of the center of mass of the molecule and w is an interpolating function which smoothly goes from 0 to 1 (or vice versa) in the interface region, (Δ), where the lower resolution is slowly transformed (according to w) in the high resolution (or vice versa), as illustrated in Fig.2.1. The most common weighting function that has

been used in AdResS is:

$$w(x) = \begin{cases} 1 & x < d_{AT} \\ cos^2\left[\frac{\pi}{2(d_\Delta)}(x - d_{AT})\right] & d_{AT} < x < d_{AT} + d_\Delta \\ 0 & d_{AT} + d_\Delta < x \end{cases}$$

where, d_{AT} and d_Δ are size of atomistic and hybrid regions respectively and x is the x-coordinate of the center of mass of the molecule. According to Eq. 2.76, the molecules that are present in the atomistic region interact with forces derived from the atomistic potential, the coarse-grained molecules interact via forces derived from the coarse-grained potential while rest of the molecules interact according to some space dependent hybrid interpolation of atomistic and coarse-grained forces. This implies that as the molecule moves from the atomistic to the coarse-grained region, the force on its atomistic degrees of freedom slowly diminishes, while the coarse-grained forces start dominating. The same process takes place in reverse order when a coarse-grained molecule enters into the atomistic zone. The technical aspects of Eq. 2.76 can be found in Ref. [44,114]. The smooth interpolation of the forces allows for the smooth transition of a molecule from atomistic dynamics to coarse-grained dynamics and vice versa. Since the transition region is chosen to be as small as possible, the perturbation of the dynamics in atomistic and coarse-grained region due to the change in resolution is expected to be negligible. Therefore, the behaviour of the molecules in the atomistic and coarse-grained regions is statistically close to the molecules in a full-atomistic or full coarse-grained simulation.

- As the molecule moves from atomistic region to the coarse-grained region, it slowly acquires the vibrational and rotational DOF's, and loses the same DOF's while moving in the opposite direction. Thus, a molecules needs to acquire (or remove) the free energy per DOF so that it equilibrates with the environment once it has changed its resolution. This is achieved by coupling the whole system to a thermostat that ensures overall thermodynamic equilibrium.

The setup gave accurate results when applied to toy systems [85, 87, 88] and later it was successfully tested to simulate physio-chemical systems [49, 64, 92]. The coarse-grained model in these simulations were derived from full-atomistic systems such that it reproduces the COM-COM radial distribution functions at the same thermodynamic conditions [96, 98]. However, in these systems, it was seen that there was a systematic discrepancy between the reference particle number density and the particle number density in the transition region. The discrepancy in the transition region implies that there is a small systematic error in the densities in the atomistic and coarse-grained regions as well. The natural question that arises is why such a discrepancy is present and whether it is possible to modify the method to control and improve the accuracy in the transition region. In such a scenario, one would be able to improve the accuracy in the high resolution (atomistic) region. This would require a deeper understanding of the thermodynamics and statistical mechanics of AdResS systems, while considering the following aspects:

- AdResS is based on the force interpolation and it has been shown that the force cannot be derived from a potential [17].

- Since the global Hamiltonian cannot be defined, the standard statistical mechanics cannot be used to define the thermodynamic properties and the laws of statistical mechanics must be redefined within the realm of adaptive resolution systems.

The hybrid resolution does not have any physical meaning per se; infact, a system of N water molecules must have the same properties of water irrespective of the molecular representation, i.e. the macroscopic properties (ρ, T and P) must be same in the different regions. In the next sections, we describe how various thermodynamic properties are balanced in the different regions of AdResS system.

Temperature in transition region

In AdResS system, the temperature is well defined in the atomistic and coarse-grained region via the equipartition theorem, as the average kinetic energy divided by the total number of DOFs

$$T_{Atom} = \frac{2\langle K_{Atom}\rangle}{n_{Atom}} \tag{2.77}$$

$$T_{CG} = \frac{2\langle K_{CG}\rangle}{n_{CG}} \tag{2.78}$$

where $\langle K_{Atom}\rangle$ and $\langle K_{CG}\rangle$ are the kinetic energies in the atomistic and coarse-grained region, and n_{Atom} and n_{CG} are the DOF's in the atomistic and coarse-grained region. However, such a definition does not apply in the transition region since the temperature in the transition region is defined as:

$$T_{\Delta} = \frac{2\,\langle K_{\Delta}\rangle}{n_{\Delta}} \tag{2.79}$$

where $n_{\Delta} = n_{Atom}w(x) + n_{CG}(1 - w(x))$, with n the number of DOFs, molecules would have if they had full atomistic resolution. One cannot define K_{Δ} since DOF's are partially removed/reintroduced as the position of the particle changes. It should be noted that in the context of adaptive resolution simulations, the calculation of average kinetic energy implies that each removed/reintroduced DOF in the transition region does not contribute. In fact in the integral of $\langle K\rangle$, the dimensionality of the changing resolution goes from 0 (a coarse-grained molecule) to 1 (full-atomistic molecule). Such cases can be described in an exact way via the mathematical formalism of integration in fractional dimensions, which implies an extra multiplicative term in the integral of $\langle K\rangle$ [90, 91]. The final result shows that the temperature at each point x in the transition region is calculated as:

$$T_{\Delta} = \frac{2\,\langle K_{\Delta}\rangle}{n_{\Delta}(x)} \tag{2.80}$$

Moreover, since $\langle K_{\Delta}\rangle$ is proportional to $n_{\Delta}(x)$, the ratio $\frac{\langle K_{\Delta}\rangle}{n_{\Delta}(x)}$ is a constant value which is the target temperature of the full-atomistic reference system. It was shown numerically that with this definition of the temperature in transition region, the system was kept at the correct thermodynamic state during the simulation. This was the first essential step in the direction of having a deeper understanding and justification of the basic algorithm. Once the procedure for the calculation of the temperature was proven satisfactory, then the development moved towards the question as to why in the transition region a small, but systematic, discrepancy for the particle density is observed in all the simulations.

Balancing the chemical potential

If one interfaces three open systems where each system consists of a different molecular specie, then the different species would remain in equilibrium if their chemical potentials are same. In AdResS method, molecules do not belong to different species, they just have different representation and are free to move between different regions. The density in such an adaptive system would remain constant if the chemical potential characterizing each resolution is same as that of the reference full atomistic system. Let us assume the chemical potential μ_{AT} in the atomistic region, μ_{CG} in the coarse-grained region and $\mu(x) = \mu_{\Delta}(x) = \mu(w(x))$ in the transition region. In order to have a thermodynamic equilibrium and a uniform density across the whole box, it would be sufficient to have $\mu(w(x)) = \mu_{AT}, \forall x$. This target can be achieved by imposing a force proportional to the gradient of the difference between the full atomistic chemical potential and the space dependent chemical potential [80]:

$$\mathbf{F}(x_{\Delta}) \approx \nabla[\mu_{Atom} - \mu(x)] \tag{2.81}$$

This force acts on the center of mass of the molecules in the transition region, and disappears in both AT and CG regions. One can use thermodynamic integration or Widom insertion methods to calculate the chemical potential in the AT and CG systems. The problem arises with the calculation of chemical potential in the transition region since both these techniques require the definition of a Hamiltonian. This problem was circumvented by dividing the transition region in several small regions that are characterized by a constant value of w. Then each region with the fixed w was treated separately, i.e. an auxiliary simulation was performed where the molecules in the box interact with AdResS force with w fixed in Eq. 2.76. In such a scenario, it is possible to define a Hamiltonian and calculate the chemical potential $\mu(\hat{w})$ corresponding to $w = \hat{w}$. Thus, one has a discrete set of $\mu(\hat{w})$ (for each $\hat{w}$) and this can then be fit to $\mu(w(x))$. This is an approximate (although numerically accurate) way to determine $\mu(x)$ in the transition region. One can then calculate the force $\mathbf{F}_{th} \approx \nabla[\mu_{Atom} - \mu(x)]$, called "thermodynamic force" [80] and apply it to the center of mass of molecules in the transition region, which would ensure a uniform density in the box, while reproducing the thermodynamic and structural properties in the atomistic region. Even in the absence of the global Hamiltonian, the basic properties were reproduced accurately. From the conceptual point of view, this study gave insights into how the principles of standard thermodynamics can be used to interpret the process of changing resolution. From the practical point of view, however, the calculation of the thermodynamic force was expensive as it required a large number of extra calculations. Therefore, further conceptual developments were stimulated which are reported in the next section.

Thermodynamics and Statistical mechanics of a generic open system

In the approach described in the previous section, it was required that the coarse-grained model is characterized by the same pressure as that of atomistic system. This is done by adding a pressure correction term to the coarse-grained potential derived by the inversion of molecule-molecule radial distribution function. A more generalized scheme was devised that imposes a uniform density profile across the box, where a specific coarse-grained model

is not required. This scheme was formalized on the principles of standard thermodynamics. Since AdResS system consists of open systems that are interfaced with each other, the quantity of interest is the grand potential $\Omega = PV$. The volume of each subsystem is a fixed quantity, as a result the balance of pressure implies equilibrium between the atomistic and coarse-grained region. Thus, in the transition region, the conditions are defined such that the pressure balance in the atomistic and coarse-region regions is assured. The thermodynamic force was reformulated in terms of pressure balance [27]:

$$p_{Atom} + \rho_0 \int_{\Delta} \mathbf{F}_{th}(\mathbf{r}) d\mathbf{r} = p_{CG}, \tag{2.82}$$

where p_{Atom} is the target pressure of the atomistic region, p_{CG} is the pressure of the coarse-grained region, ρ_0 is the target molecular density of the reference full atomistic system. The thermodynamic force $\mathbf{F}_{th}$ is calculated via an iterative procedure as a function of the molecular density in the transition region. The iterative formulas at the i-th step is:

$$\mathbf{F}_{th}^{i+1}(\mathbf{r}) = \mathbf{F}_{th}^{i}(\mathbf{r}) - \frac{1}{\rho_0 \kappa_T^2} \nabla \rho^i(\mathbf{r}), \tag{2.83}$$

where κ_T the isothermal compressibility. The force is converged once the density in the transition region is equal to the density in the transition region. This formulation implies a grand-canonical like setup where each region functions as a reservoir for other region and the transition region acts a filter that ensures thermodynamic equilibrium in different regions. Due to the limited size of the reservoir, the approach cannot be defined as grand-canonical. Nevertheless, this approach was satisfactory from the theoretical and computational point of view [27, 67], and paves the path for further development towards a proper grand-canonical setup [27, 114], which will be discussed in more detail in Chapter 3.

2.9 Path Integral Formalism

The quantum delocalization of light atoms is treated in MD via path-integral formalism developed by Richard Feynman [25, 36, 107]. In this section, we will briefly discuss the derivation of path integral partition function, which forms the basis for the three techniques: path integral molecular dynamics (PIMD) [108, 118], ring polymer molecular dynamics (RPMD) [16, 31, 32] and centroid molecular dynamics (CMD) [13, 38], which will be discussed in Chapter 5. These techniques are used to calculate equilibrium statistical and dynamical properties in systems where nuclear delocalization plays a major role. Light nuclei such as hydrogen exhibit quantum character at low temperature; the thermal de Broglie wavelength for hydrogen at room temperature is 0.1 nm which is of the same order as the average bond length in a water molecule. This verifies that hydrogen atom is delocalized in space and that hydrogen atoms at low temperatures must not be treated as classical point particles. Therefore, hydrogen atoms must be treated within the realm of quantum mechanics since many of the crucial properties of water may be influenced by the non classical nature of this element.

2.9.1 Derivation of path-integral partition function

A more detailed derivation of path-integrals can be found in Ref. [107]. In this section, we show how one obtains a quantized Hamiltonian starting from the Hamiltonian for a single particle moving under the influence of an external potential U, which is written as:

$$\hat{\mathcal{H}} = \frac{\hat{p}^2}{2m} + U(\hat{(}x)) = \hat{K} + \hat{U} \tag{2.84}$$

where $\hat{K}$ and $\hat{U}$ are the non-commuting kinetic and potential energy operators. ($[\hat{K}, \hat{U}] \neq 0$). The coordinate space matrix elements of the density matrix are given by:

$$\rho(x, x^{'}) = \langle x^{'} | e^{-\beta \mathcal{H}} | x \rangle \tag{2.85}$$

where x and $x^{'}$ represent the initiation and detection points between which the particle moves undetected. The matrix elements can be computed by taking the sum of all the possible paths between x and $x^{'}$. Since $\hat{K}$ and $\hat{U}$ do not commute, there is no way to evaluate $\exp(-\beta H)$ directly. To circumvent this problem, Trotter theorem is used. For two operators A and B that do not commute, Trotter theorem can be written as:

$$e^{\lambda(\hat{A}+\hat{B})} = \lim_{P \to \infty} [e^{\frac{\lambda \hat{A}}{2P}} e^{\frac{\lambda \hat{B}}{P}} e^{\frac{\lambda \hat{A}}{2P}}]^P \tag{2.86}$$

where P is called Trotter number. Using Trotter theorem for $\hat{K}$ and $\hat{U}$, one obtains:

$$e^{-\beta(\hat{K}+\hat{U})} = \lim_{P \to \infty} [e^{-\frac{\beta \hat{U}}{2P}} e^{-\frac{\beta \hat{K}}{P}} e^{-\frac{\beta \hat{U}}{2P}}]^P \tag{2.87}$$

Substituting Eq. 2.87 in Eq. 2.85, one obtains:

$$\rho(x, x^{'}) = \lim_{P \to \infty} \langle x^{'} | [e^{-\frac{\beta \hat{U}}{2P}} e^{-\frac{\beta \hat{K}}{P}} e^{-\frac{\beta \hat{U}}{2P}}]^P | x \rangle \tag{2.88}$$

If one defines an operator $\hat{\Omega}$, such that:

$$\hat{\Omega} = e^{-\frac{\beta \hat{U}}{2P}} e^{-\frac{\beta \hat{K}}{P}} e^{-\frac{\beta \hat{U}}{2P}} \tag{2.89}$$

then Eq. 2.88 can be written as:

$$\rho(x, x^{'}) = \lim_{P \to \infty} \langle x^{'} | \hat{\Omega}^P | x \rangle = \langle x^{'} | \hat{\Omega}^P | x \rangle = \lim_{P \to \infty} \langle x^{'} | \hat{\Omega}\hat{\Omega}\hat{\Omega}\hat{\Omega}...\hat{\Omega} | x \rangle \tag{2.90}$$

If one introduces an identity operator $\hat{I} = \int dx |x\rangle\langle x|$ between each factor of $\hat{\Omega}$, then $P-1$ insertions of the identity operator are required. This will generate a $P-1$ dimensional integral for the density matrix.

$$\rho(x, x^{'}) = \lim_{P \to \infty} \int dx_2 ... dx_P \times \langle x^{'} | \Omega | x_P \rangle \langle x_P | \Omega | x_{P-1} \rangle \langle x_{P-1} | ... | x_2 \rangle \langle x_2 | \Omega | x \rangle \tag{2.91}$$

One needs to calculate the matrix elements of $\hat{\Omega}$ to evaluate the integral in Eq. 2.91.

$$\langle x_{k+1} \mid \hat{\Omega} \mid x_k \rangle = \langle x_{k+1} \mid e^{-\frac{\beta \hat{U}}{2P}} e^{-\frac{\beta \hat{K}}{P}} e^{-\frac{\beta \hat{U}}{2P}} \mid x_k \rangle \tag{2.92}$$

Since the potential energy operator depends on the coordinates, the eigen values of the operator $e^{-\frac{\beta\hat{U}}{2P}}$ is $e^{-\frac{\beta U(x_k)}{2P}}$. Thus, the expression $\langle x_{k+1} \mid e^{-\frac{\beta\hat{U}}{2P}} e^{-\frac{\beta\hat{K}}{P}} e^{-\frac{\beta\hat{U}}{2P}} \mid x_k\rangle$ can be written as:

$$\langle x_{k+1} \mid e^{-\frac{\beta\hat{U}}{2P}} e^{-\frac{\beta\hat{K}}{P}} e^{-\frac{\beta\hat{U}}{2P}} \mid x_k\rangle = e^{-\frac{\beta U(x_{k+1})}{2P}} \langle x_{k+1} \mid e^{-\frac{\beta\hat{K}}{P}} \mid x_k\rangle e^{-\frac{\beta U(x_k)}{2P}} \tag{2.93}$$

In order to evaluate the eigenvalues of the momentum operator, one uses an identity operator containing momentum eigen states.

$$I = \int dp \mid p\rangle\langle p \mid \tag{2.94}$$

Thus, the matrix element can be written as:

$$\langle x_{k+1} \mid e^{-\frac{\beta\hat{K}}{P}} \mid x_k\rangle = \int dp \langle x_{k+1} \mid p\rangle\langle p \mid e^{-\frac{\beta\hat{K}}{P}} \mid x_k\rangle \tag{2.95}$$

The kinetic energy operator acts on the momentum eigen state from the left. Thus, Eq. 2.95 can be expressed as:

$$\langle x_{k+1} \mid e^{-\frac{\beta\hat{K}}{P}} \mid x_k\rangle = \int e^{-\frac{\beta p^2}{2mP}} \langle x_{k+1} \mid p\rangle\langle p \mid x_k\rangle \tag{2.96}$$

The relation between the position and momentum eigenstate is given by:

$$\langle x \mid p\rangle = \frac{1}{\sqrt{2\pi\hbar}} e^{\frac{ipx}{\hbar}} \tag{2.97}$$

Substituting Eq. 2.97 in Eq. 2.96, one obtains:

$$\langle x_{k+1} \mid e^{-\frac{\beta\hat{K}}{P}} \mid x_k\rangle = \frac{1}{2\pi\hbar} \int dp e^{\frac{ip(x_{k+1}-x_k)}{\hbar}} e^{-\frac{\beta p^2}{2mP}} \tag{2.98}$$

which can also be formulated as:

$$\langle x_{k+1} \mid e^{-\frac{\beta\hat{K}}{P}} \mid x_k\rangle = \frac{1}{2\pi\hbar} \int dp e^{-\frac{\beta}{2mP}\left[p-\left(\frac{imP}{\beta\hbar}(x_{k+1}-x_k)\right)\right]^2} \times e^{-\frac{mP}{2\beta\hbar^2}(x_{k+1}-x_k)^2} \tag{2.99}$$

The second exponential term in the integral is devoid of any momentum variables and the first term can be written as a gaussian integral, if one makes a change of variables $\tilde{p} = p - \frac{imP(x_{k+1}-x_k)}{\beta\hbar}$. Thus, the above equation takes the following form:

$$\langle x_{k+1} \mid e^{-\frac{\beta\hat{K}}{P}} \mid x_k\rangle = \left(\frac{mP}{2\pi\beta\hbar^2}\right)^{1/2} e^{-\frac{mP}{2\beta\hbar^2}(x_{k+1}-x_k)^2} \tag{2.100}$$

Substituting Eq. 2.100 in density matrix, one obtains:

$$\begin{aligned}\rho(x, x') = \lim_{P\to\infty} \left(\frac{mP}{2\pi\beta\hbar^2}\right) \int dx_2 ... dx_P \\ \times \exp\left(-\frac{1}{\hbar}\sum_{k=1}^{P}\left[\frac{mP}{2\beta\hbar^2}(x_{k+1}-x_k)^2 + \frac{\beta\hbar}{2P}(U(x_{k+1})+U(x_k))\right]\right)\Bigg|_{x_1=x}^{x_{P+1}=x'}\end{aligned} \tag{2.101}$$

In this equation, the end points of the path x_1 and x_{P+1} are fixed at x and $x^{'}$ respectively. This is also know as the discretized path integral representation of the density matrix in the limit $P \to \infty$. The partition function of a single quantum particle is given by:

$$Z = Tr[\rho] = Tr[\exp(-\beta\hat{\mathcal{H}})] \tag{2.102}$$

The partition function can be evaluated by integrating the diagonal elements of the density matrix:

$$Z = \int_0^L dx \langle x \mid e^{-\beta\hat{\mathcal{H}}} \mid x \rangle = \int_0^L dx \rho(x,x) \tag{2.103}$$

where the particle is restricted within an interval $[0, L]$. To compute the integral in Eq. 2.103, the diagonal elements of the density matrix can be obtained by assigning $x_1 = x_{P+1} = x$ in Eq. 2.101. Therefore, one can also perform the integration over x_1 in place of x in Eq 2.103. Thus, one obtains:

$$Z = \lim_{P\to\infty} \left(\frac{mP}{2\pi\beta\hbar^2}\right)^{P/2} \int_{D(L)} dx_1 ... dx_P \times \exp\left(-\frac{1}{\hbar}\sum_{k=1}^{P}\left[-\frac{mP}{2\beta\hbar}(x_{k+1}-x_k)^2 + \frac{\beta\hbar}{P}U(x_k)\right]\right)\Bigg|_{x_1=x}^{x_{P+1}=x^{'}} \tag{2.104}$$

which can further be expressed as the classical configurational partition function:

$$Z = \lim_{P\to\infty} \left(\frac{mP}{2\pi\beta\hbar^2}\right)^{P/2} \int_{D(L)} dx_1 ... dx_P e^{-\beta\phi(x_1 ... x_P)} \tag{2.105}$$

where the effective classical Hamiltonian is given by:

$$\phi(x_1 ... x_P) = \sum_{k=1}^{P}\left[\frac{1}{2}m\omega_P^2(x_{k+1}-x_k)^2 + \frac{1}{P}U(x_k)\right] \tag{2.106}$$

where $\omega_P = \frac{\sqrt{P}}{\beta\hbar}$ and $x_{P+1} = x_1$. The expression in Eq. 2.105 is known as the discretized path integral representation of the partition function. It resembles the partition function of a polymer ring with P beads with harmonic nearest neighbour couplings where the strength of harmonic coupling is denoted by ω_P. The potential acting on each bead is reduced by a factor $\frac{1}{P}$. The extension of the above results to a system of N particles is not trivial since the symmetry of the physical states must be included as well, and this gives rise to exchange terms. For Fermionic systems, the calculation of the exchange terms is numerically problematic; the problem is termed as Fermi Sign problem. For Bosonic systems, however, the sign problem does not exist and the exchange terms can be numerically solved. In the present study, we completely ignore the exchange effects which is the case for liquids at room temperature. The particles treated in this work do not follow a quantum statistics (fermions or bosons), thus they are semiclassical. The path-integral formalism is used to model only the nuclear quantum effects (zero-point energy and delocalization) in this thesis. Consider the Hamiltonian of N particles in d-dimensions:

$$\hat{\mathcal{H}} = \sum_{i=1}^{N}\frac{\hat{p}_i^2}{2m} + U(\hat{\mathbf{r}}_1, ..., \hat{\mathbf{r}}_N) \tag{2.107}$$

If the exchange effects are excluded, then each particle is represented by a polymer ring, and the quantum partition function takes the form:

$$Z = \lim_{P\to\infty} \prod_{i=1}^{N} \left(\frac{m_i P}{2\pi\beta\hbar^2}\right)^{dP/2} \int \prod_{i=1}^{N} dr_i^{(1)} ... dr_i^{(P)} \\ \times \exp\left(-\sum_{k=1}^{P}\left[\sum_{i=1}^{N} \frac{m_i P}{2\beta\hbar^2}(r_i^{(k+1)} - r_i^{(k)})^2 + \frac{\beta}{P} U(\mathbf{r_1^{(k)}}, ..., \mathbf{r_N^{(k)}})\right]\right)_{\mathbf{r}_i^{(P+1)}=\mathbf{r}_i^{(1)}} \tag{2.108}$$

According to the partition function in Eq. 2.108, the beads with the same index k will interact with each other while other cross-interactions will not be considered. This partition function can be sampled using Monte-Carlo or Molecular Dynamics methods. In this thesis, we will use MD methods as sampling tool. It must be noted that the standard Path Integral Molecular Dynamics (PIMD) algorithms are used to calculate the quantum equilibrium properties. The calculation of quantum dynamics properties entail numerical difficulties. There are approximate methods for computing real-time quantum correlation functions. The methods are Ring Polymer Molecular Dynamics (RPMD) and Centroid Molecular dynamics (CMD). All these approaches will be discussed in detail in Chapter 5.

2.10 Conclusions

In this chapter, we have covered the various theoretical concepts and computational methods that will be used throughout the rest of this thesis. In the next chapter, we discuss the latest version of AdResS (the development of which was one of this work), i.e. grand-canonical AdResS (GC-AdResS) and how it can be used as a tool to calculate the excess chemical potential.

Chapter 3

AdResS as a tool to efficiently calculate the excess chemical potential of liquids and mixtures

3.1 Introduction

Molecular Simulation techniques such as Widom insertion particle method (IPM) [117] or Thermodynamic Integration (TI) [105] have been regularly used to calculate excess chemical potential of complex molecular systems. IPM is computationally very demanding, and for complex molecules, the time taken for convergence is rather large. TI is computationally convenient, but it is specifically designed to calculate the chemical potential, which may not be the optimal choice if one wants to calculate other properties within the same simulation framework. It was suggested in Ref. [111], that the chemical potential can also be calculated by using AdResS in its Grand-Canonical like formulation (GC-AdResS) [27,111,114]. However, no analytical derivation of the formula to evaluate the chemical potential was given, and the method was only applied to liquid water. In this part of the thesis, we give an analytical form to the formula, and apply it to various liquids and mixtures of particular relevance in natural science, and compare our results with those obtained from TI. We further comment on the practical applicability of GC-AdResS as a tool to calculate the chemical potential, pointing out its computational convenience compared to methods like IPM and TI.

3.2 Grand-Canonical formalism of AdResS

In this section, we summarize the theoretical foundations behind the grand canonical formalism of AdResS developed in work previous to the work of this thesis. Next we discuss the formalism developed within this work on how AdResS can be used to calculate the excess chemical potential in molecular systems.

3.2.1 From AdResS to GC-AdResS

The formulation of AdResS in terms of grand-canonical ensemble emerged in Refs. [111, 114]. A well known principle of standard statistical mechanics tells us that when the ref-

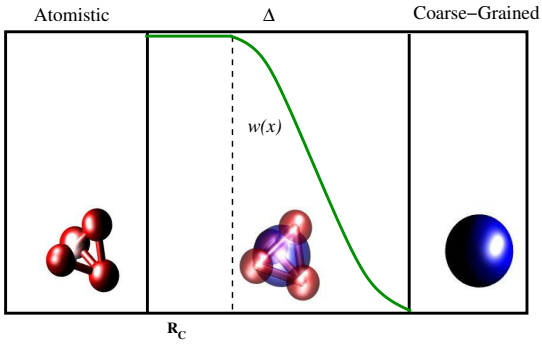

Figure 3.1: Schematic representation of the adaptive box, showing the coarse-grained region (the region on the right) , the hybrid region (the central part) Δ, where the switching function $w(x)$ (in green) is defined, and the high resolution atomistic region (the region on the left). It should be noted that in GC-AdResS, the definition of $w(x)$ is modified. The transition region is extended by an amount $\boldsymbol{R}_c$, which is the cut-off radius of atomistic interactions, so that the molecules in the atomistic region interact with other molecules via well defined atomistic interactions. This allows to write an exact Hamiltonian for the atomistic region, which is essential if one wants to treat the system in a Grand-Canonical fashion.

erence atomistic system (i.e. a large full atomistic simulation) reaches the thermodynamic limit, a subsystem of a large system samples the grand-canonical ensemble [50]. Therefore, it is interesting to investigate if phase space variables $\boldsymbol{x}_{\mathrm{AT}}$ and the number of molecules in the atomistic region N_{AT} of the adaptive system are subject to grand-canonical distribution, which implies numerically verification of the following formula:

$$p(\boldsymbol{x}_{\mathrm{AT}}, N_{\mathrm{AT}}) \approx \frac{1}{\mathcal{Z}} \exp\left\{\beta\mu^*_{\mathrm{AT}} N_{\mathrm{AT}} - \beta\mathcal{H}^{\mathrm{AT}}(\boldsymbol{x}_{\mathrm{AT}})\right\} \tag{3.1}$$

where μ^*_{AT} is the chemical potential of the reference atomistic system at the desired equilibrium, $\mathcal{Z}$ is the grand canonical partition function, $\mathcal{H}^{\mathrm{AT}}$ is the *atomistic* Hamiltonian, and β denotes the inverse temperature $1/(k_B T)$. In order to have a fair comparison with the reference atomistic system, one has to consider the thermodynamic limit in AdResS, which corresponds to large enough atomistic region, a relatively thin transition region and a coarse-grained region, which acts as a large reservoir; this implies the assumption that $N_{\mathrm{CG}} \gg 1$, $N_{\mathrm{AT}} \gg 1$, $N_{\mathrm{CG}} \gg N_{\mathrm{AT}} \gg N_{\Delta}$ (N_{CG} is the number of molecules in the coarse-grained region and N_{Δ} is the number of molecules in the hybrid region). Regarding the accuracy of AdResS, the probability distributions in AdResS should be compared with that of a full-atomistic system. Therefore the full-atomistic reference system serves as a good comparison for the accuracy of AdResS as a Grand-Canonical setup. The analysis of Eq. 3.1 within the grand-canonical setup is rather complex, therefore, the problem was circumvented through various steps, the first of which consists of splitting the distribution as:

$$p(\boldsymbol{x}_{\mathrm{AT}}, N_{\mathrm{AT}}) = p(\boldsymbol{x}_{\mathrm{AT}}|N_{\mathrm{AT}})\, p(N_{\mathrm{AT}}) \tag{3.2}$$

The first term in the R.H.S. of Eq. 3.2 can be written as:

$$p(\boldsymbol{x}_{\mathrm{AT}}|N_{\mathrm{AT}}) = \sum_{N_\Delta} \int \mathrm{d}\boldsymbol{x}_\Delta \, p(\boldsymbol{x}_{\mathrm{AT}}|N_{\mathrm{AT}}; \boldsymbol{x}_\Delta, N_\Delta) \cdot p(\boldsymbol{x}_\Delta, N_\Delta|N_{\mathrm{AT}}) \tag{3.3}$$

The conditional distribution is the Boltzmann distribution:

$$p(\boldsymbol{x}_{\mathrm{AT}}|N_{\mathrm{AT}}; \boldsymbol{x}_\Delta, N_\Delta) \propto \exp\{-\beta \mathcal{H}^{\mathrm{AT}}_{N_{\mathrm{AT}}}(\boldsymbol{x}_{\mathrm{AT}}; \boldsymbol{x}_\Delta, N_\Delta)\} \tag{3.4}$$

where

$$\mathcal{H}^{\mathrm{AT}}_{N_{\mathrm{AT}}}(\boldsymbol{x}_{\mathrm{AT}}; \boldsymbol{x}_\Delta, N_\Delta) = \sum_{i=1}^{N_{\mathrm{AT}}} \frac{1}{2} m_i \boldsymbol{v}_i^2 + \sum_{i<j}^{N_{\mathrm{AT}}} U^{\mathrm{AT}}(\boldsymbol{r}_i - \boldsymbol{r}_j) + \sum_{i=1}^{N_{\mathrm{AT}}} \sum_{j=N_{\mathrm{AT}}+1}^{N_{\mathrm{AT}}+N_\Delta} U^{\mathrm{AT}}(\boldsymbol{r}_i - \boldsymbol{r}_j). \tag{3.5}$$

For an exact definition of $\mathcal{H}^{\mathrm{AT}}_{N_{\mathrm{AT}}}$, a technical modification was proposed in the transition region and of the corresponding definition of the switching function $w(x)$. In Ref. [114], the switching function was modified in such way that it allows the separation of atomistic interactions from coarse-grained and hybrid interactions. This was made possible by extending the transition region into the atomistic region, by an amount corresponding to the cut-off of molecule-molecule interaction. In this way, a molecule in the atomistic region in AdResS interacts with other molecules through atomistic interactions. Figure 5.1 shows the modified system with the extended transition region. It should be noted that probability density in Eq. 3.4 is identical to that of the reference atomistic system. Therefore according to Eq. 3.3, in order to have a high accuracy for the distribution $p(\boldsymbol{x}_{\mathrm{AT}}|N_{\mathrm{AT}})$, one should have high accuracy of the distribution $p(\boldsymbol{x}_\Delta, N_\Delta|N_{\mathrm{AT}})$. Ideally, this distribution must be the same as in the reference system, however, it was seen that the density distribution in the hybrid region deviates from the reference density [80]. Thus, one can define proper conditions for the AdResS system, which ensure satisfactory accuracy of $p(\boldsymbol{x}_\Delta, N_\Delta|N_{\mathrm{AT}})$. The first necessary condition that assures a first order accuracy (usually sufficient in MD) is that the density profile of the hybrid region is identical to the reference density profile:

$$\rho_\Delta = \rho_{\mathrm{AT}} \tag{3.6}$$

The second necessary condition is that the two body probability density, i.e. g(r) in the hybrid region should be same as in the reference full atomistic system.

$$g_\Delta(r) = g_{\mathrm{AT}}(r) \tag{3.7}$$

The first condition is fulfilled by the thermodynamic force, while the second condition can be fulfilled by a force which corrects the RDF's in the hybrid region. It is also possible to raise necessary conditions for the accuracy of the distribution $p(N_{\mathrm{AT}})$, in Eq. 3.2. The necessary condition for the first order accuracy is the balance of the chemical potential:

$$\mu_{\mathrm{CG}} - \mu_{\mathrm{AT}} = \omega_0 \tag{3.8}$$

where μ_{CG} and μ_{AT} are the chemical potential in the coarse-grained and atomistic region respectively, and ω_0 is the work done in the transition region. This relation holds automatically with the introduction of the thermodynamic force, which uniformizes the density across the simulation box. The second order accuracy is achieved when the isothermal compressibility of the coarse-grained model is equal to the atomistic compressibility:

$$\kappa_{\mathrm{CG}} = \kappa_{\mathrm{AT}} \tag{3.9}$$

3.2.2 Calculation of Excess Chemical Potential

Here, we outline the procedure to calculate the chemical potential, which is an essential thermodynamic quantity in a grand-canonical ensemble. In Ref. [111], it was shown that the chemical potential of the atomistic and coarse-grained regions are related by the following formula:

$$\mu_{\mathrm{CG}} = \mu_{\mathrm{AT}} + \omega_{\mathrm{th}} + \omega_{\mathrm{Q}} \tag{3.10}$$

where μ_{CG} and μ_{AT} are the chemical potential in the coarse-grained and atomistic regions respectively, $\omega_{\mathrm{th}} = \int_{\Delta} \boldsymbol{F}_{\mathrm{th}}(\boldsymbol{r}) d\boldsymbol{r}$ is the work done by the thermodynamic force in the transition region, ω_{Q} is the heat provided by the thermostat, in order to thermalize the inserted/removed degrees of freedom, as the molecule enters the atomistic region or vice versa. ω_{Q} is composed of two parts: ω_{extra} and ω_{DOF}. ω_{extra} compensates the dissipation of energy due to non-conservative interactions in the transition region, and ω_{DOF} is related to the equilibration of the reinserted/removed degrees of freedom (rotational and vibrational). From the equipartition theorem, $\omega_{\mathrm{DOF}} = \frac{1}{2} k_B T$, per degree of freedom. The calculation of ω_{extra} is not direct, and a procedure, which involves auxiliary Hamiltonian approach was proposed in Ref. [111]. In this approach, the coarse-grained and atomistic potentials are interpolated, instead of forces as in the standard AdResS. Next, it was imposed that the Hamiltonian system must have the same thermodynamic equilibrium as the force-interpolation AdResS, which is achieved by introducing a thermodynamic force, which keeps the density of particles the same across the system, at the target temperature. The equilibrium is the same in both standard and auxiliary Hamiltonian AdResS. The change of resolution is provided by ($\boldsymbol{F}_{\mathrm{th}}$, ω_{DOF}, ω_{extra}) and ($\boldsymbol{F}^{\mathrm{H}}_{\mathrm{th}}$, ω_{DOF}) in the standard and auxiliary Hamiltonian AdResS respectively; it follows that the difference between the work of the original thermodynamic force and work of the thermodynamic force in the auxiliary Hamiltonian AdResS should be equal to ω_{extra}. It was also shown numerically for liquid water, that $\omega_{\mathrm{extra}} = \int_{\Delta} \nabla w(\boldsymbol{r}) \langle w(U^{\mathrm{AT}} - U^{\mathrm{CG}}) \rangle_{\boldsymbol{r}} d\boldsymbol{r}$, where U_{AT} and U_{CG} are the atomistic and coarse-grained potential. It was numerically verified that the ensemble average remains same irrespective of the type of simulation approach (force interpolation or potential interpolation) one may wish to use. In the next section, it is shown that the above formula is exact at the first order w.r.t. probability distribution of the system. Thus, ω_{Q} can be calculated during the equilibration phase within the GC-AdResS code. However, the quantity of interest is not the total chemical potential but the excess chemical potential, which is defined by subtracting the kinetic contribution from the total chemical potential, i.e., $\mu^{exc} = \mu - \mu^{kin}$. Substituting it in Eq. 4.14, we get:

$$\mu^{exc}_{\mathrm{CG}} - \mu^{exc}_{\mathrm{AT}} = \omega_{\mathrm{th}} + \omega_{\mathrm{extra}} + \omega_{\mathrm{DOF}} - \Delta\mu^{kin} \tag{3.11}$$

It should be noted here that, the kinetic contribution from the center-of-mass (COM) is same for the atomistic and coarse-grained molecules, thus the fourth term on the R.H.S. in the above equation is automatically removed. Also, the kinetic part of μ_{AT} due to the rotational and vibrational degrees of freedom corresponds to ω_{DOF}, which can be calculated by hand removing $\frac{1}{2} k_B T$ for each degree of freedom. However, this is not required as in the current version of AdResS in GROMACS, the removed degrees of freedom are kept as phantom variables and are thermally equilibrated [44]. As a consequence, the heat provided by the thermostat for this part is same for the atomistic and coarse-grained molecule, and

is automatically removed. Finally, one can calculate the chemical potential of the coarse-grained system using standard methods such as IPM or TI, which for very simple models (e.g. spherical molecules) requires small computational cost. The final expression for calculating the excess chemical potential is:

$$\mu^{\mathrm{ex}}_{\mathrm{AT}} = \mu^{\mathrm{ex}}_{\mathrm{CG}} - \int_{\Delta} F_{th}(\boldsymbol{r})d\boldsymbol{r} - \int_{\Delta} \nabla_{\boldsymbol{r}} w(\boldsymbol{r}) \langle w(U^{\mathrm{AT}} - U^{\mathrm{CG}})\rangle_{\boldsymbol{r}} d\boldsymbol{r} \tag{3.12}$$

3.3 Quantification of the work of the thermostat

In this section, we derive the formula for work done by the thermostat ω_{extra}, for single as well as multi component systems.

3.3.1 Single Component

If we consider a potential coupling between the atomistic and coarse-grained regions, based on the spatial interpolation of the potential in the atomistic and coarse-grained region, then the total potential energy of the system can be written as (similar to the force interpolation expression in standard AdResS):

$$U = \sum_{i<j} w(\boldsymbol{r}_i) w(\boldsymbol{r}_j) U^{\mathrm{AT}}_{i,j} + \sum_{i<j} [1 - w(\boldsymbol{r}_i) w(\boldsymbol{r}_j)] U^{\mathrm{CG}}_{i,j}, \tag{3.13}$$

where $U^{\mathrm{AT}}_{i,j}$ is the potential in the atomistic region and U^{CG}_{ij} is the potential acting in the coarse-grained region between molecules i and j, given by:

$$U^{\mathrm{AT}}_{i,j} = \sum_{\alpha\in i}\sum_{\beta\in j} U^{\mathrm{AT}}(\boldsymbol{r}_\alpha - \boldsymbol{r}_\beta), \quad U^{\mathrm{CG}}_{i,j} = U^{\mathrm{CG}}(\boldsymbol{r}_i - \boldsymbol{r}_j), \tag{3.14}$$

where α and β denote the atom indexes of the corresponding molecule. The COM of the molecule is defined as:

$$\boldsymbol{r}_i = \sum_{\alpha\in i} \frac{m_\alpha}{\sum_{\alpha\in i} m_\alpha} \boldsymbol{r}_\alpha, \tag{3.15}$$

where m_α is the mass of atom α of molecule i. The potential interpolation was referred to as auxiliary Hamiltonian approach in the previous section. The force derived from Eq. 3.13 is given by:

$$\boldsymbol{F}_{i,j} = w(\boldsymbol{r}_i) w(\boldsymbol{r}_j) \boldsymbol{F}^{\mathrm{AT}}_{i,j} + [1 - w(\boldsymbol{r}_i) w(\boldsymbol{r}_j)] \boldsymbol{F}^{\mathrm{CG}}_{i,j} - \nabla_{\boldsymbol{r}} w(\boldsymbol{r}_i) w(\boldsymbol{r}_j) (U^{\mathrm{AT}}_{i,j} - U^{\mathrm{CG}}_{i,j}). \tag{3.16}$$

All the properties of auxiliary Hamiltonian approach will be denoted by using a superscript 'H'. The force of changing representation (from atomistic to coarse-grained and vice versa) for a single molecule is defined by:

$$\boldsymbol{F}_{\mathrm{rep},i} = \sum_{j} \nabla_{\boldsymbol{r}} w(\boldsymbol{r}_i) w(\boldsymbol{r}_j) (U^{\mathrm{AT}}_{i,j} - U^{\mathrm{CG}}_{i,j}). \tag{3.17}$$

The thermodynamic variables (number of molecules N, Volume V and temperature T) in the atomistic, hybrid and coarse-grained region are denoted by (N_1, V_1, T), (N_2, V_2, T)

and (N_3, V_3, T) respectively. The transition region is assumed to be much smaller than atomistic and coarse-grained region. Thus it is reasonable to assume that:

$$V \approx V_1 + V_3 \tag{3.18}$$

$$N \approx N_1 + N_3 \tag{3.19}$$

We further assume that the system is in the thermodynamic limit and the molecules are short-range correlated, i.e. the range of any interaction is slightly smaller compared to the size of the transition region. The thermodynamic force in the standard AdResS ($\boldsymbol{F}_{\mathrm{th}}$) and the auxiliary Hamiltonian AdResS ($\boldsymbol{F}^{\mathrm{H}}_{\mathrm{th}}$) ensure that the system has a flat density across the box, i.e.,

$$\rho_{\Delta} = \rho_{\mathrm{AT}} = \rho_{\mathrm{CG}} = \rho_0 \tag{3.20}$$

$$\rho^{\mathrm{H}}_{\Delta} = \rho_{\mathrm{AT}} = \rho_{\mathrm{CG}} = \rho_0 \tag{3.21}$$

where ρ_0 is the equilibrium number density of the system defined by $\rho_0 = N/V$. The work done by thermodynamic force balances the grand potential between the atomistic and coarse-grained regions, which leads to:

$$p_{\mathrm{AT}} = p_{\mathrm{CG}} - \rho_0\, \omega_{\mathrm{th}} \tag{3.22}$$

where ω_{th} represents the work done by the thermodynamic force $\boldsymbol{F}_{\mathrm{th}}$. In the auxiliary Hamiltonian approach, the third term on R.H.S in Eq. 3.16 is not symmetric w.r.t. particle exchange, thus Newton's third law is violated and the total momentum is not conserved anymore. As a result of this, the pressure equation is no more valid, and has to be derived again for the consistency with Eq. 3.16. In the following, we will report how the equation of equilibrium is now obtained. We assume that the resolution of the system changes only along the x-direction. The transition region can be assumed to be an ideal "piston", which is given a displacement of ΔL towards the coarse-grained region. The displacement ΔL should be infinitesimal, i.e. much smaller than than the transition region. This implies that there is an infinitesimal increment of volume ΔV and an infinitesimal decrement of volume ΔV in the atomistic and coarse-grained regions respectively, where $\Delta V = \Delta L \cdot S$, and S is the cutting surface area. As the displacements of the molecules are infinitesimal, it can be safely assumed that the resolution of the molecules does not change under a displacement of ΔL. Thus, the approximate change in free energy is:

$$\begin{aligned}\Delta A \approx\, & A_{\mathrm{AT}}(N_1, V_1 + \Delta V, T) - A_{\mathrm{AT}}(N_1, V_1, T) + A_{\mathrm{CG}}(N_3, V_3 - \Delta V, T) - A_{\mathrm{CG}}(N_3, V_3, T) \\ & + \int_{\Delta} d\boldsymbol{x}\, \rho_0 V_1 \cdot [\, \boldsymbol{F}^{\mathrm{H}}_{\mathrm{th}}(\boldsymbol{r}) - \langle \boldsymbol{F}_{\mathrm{rep}}(\boldsymbol{r}) \rangle \,] - \int_{\Delta} d\boldsymbol{x}\, \rho_0 (V_1 + \Delta V) \cdot [\, \boldsymbol{F}^{\mathrm{H}}_{\mathrm{th}}(\boldsymbol{r}) - \langle \boldsymbol{F}_{\mathrm{rep}}(\boldsymbol{r}) \rangle \,]\end{aligned} \tag{3.23}$$

where A_{AT} and A_{CG} are the free energies in the atomistic and coarse-grained region respectively. Since the resolution changes only along the x-direction, F_{th} and $\langle \boldsymbol{F}_{\mathrm{rep}}(\boldsymbol{r}) \rangle$ depend on x component of the center of mass of the molecule. This can be generalized to resolution change along any direction, by replacing x by $\boldsymbol{r}$. Eq. 3.23 is justified under the hypothesis that the system is in thermodynamic limit, and the interactions are short range correlated, with the interaction range comparable to the size of the transition region. The

fifth term on R.H.S. in Eq. 3.23 denotes the work done by the ideal piston. This term is made of two parts, first is the work done by the thermodynamic force, and second is the work done by the force of changing representation which does not vanish as it violates Newton's action-reaction law. From the physical point of view, this term represents the work done by the thermodynamic force and the force of changing representation to bring the molecules across the transition region Since the number of particles in the atomistic region after moving the piston would be $\rho_0.(V_1 + \Delta V)$, there is an extra work done by the thermodynamic force and the force of changing representation to bring $\rho_0.\Delta V$ extra molecules across the transition region. This work is denoted by the last term on R.H.S. in Eq. 3.23. This equation can be further simplified as:

$$\Delta A \approx A_{\mathrm{AT}}(N_1, V_1 + \Delta V, T) - A_{\mathrm{AT}}(N_1, V_1, T) + A_{\mathrm{CG}}(N_3, V_3 - \Delta V, T) - A_{\mathrm{CG}}(N_3, V_3, T) + \int_{\Delta} d\boldsymbol{r}\, \rho_0 S \cdot \Delta L \cdot \left[\boldsymbol{F}^{\mathrm{H}}_{\mathrm{th}}(\boldsymbol{r}) - \left\langle \boldsymbol{F}_{\mathrm{rep}}(\boldsymbol{r}) \right\rangle \right] \tag{3.24}$$

As the first two terms in Eq. (3.16) are based on pairwise interactions, they do not contribute to the difference in energy; in fact the total work of these forces is zero. From Eq. 3.24, one can deduce:

$$\Delta A \approx -p_{\mathrm{AT}} \Delta V + p_{\mathrm{CG}} \Delta V - \rho_0 \Delta V (\omega^{\mathrm{H}}_{\mathrm{th}} - \omega_{\mathrm{rep}}), \tag{3.25}$$

where $w^{\mathrm{H}}_{\mathrm{th}}$ is the integral of the thermodynamic force $\boldsymbol{F}^{\mathrm{H}}_{\mathrm{th}}$, and ω_{rep} is the work of changing representation, which can be written as:

$$\omega_{\mathrm{rep}} = \int_{\Delta} d\boldsymbol{r} \left\langle \boldsymbol{F}_{\mathrm{rep}}(\boldsymbol{r}) \right\rangle = \int_{\Delta} d\boldsymbol{r}\, \nabla_{\boldsymbol{r}} w(\boldsymbol{r}) \left\langle w(\boldsymbol{r}') \left[\sum_{\alpha,\beta} U^{\mathrm{AT}}(\boldsymbol{r}_\alpha - \boldsymbol{r}'_\beta) - U^{\mathrm{CG}}(\boldsymbol{r} - \boldsymbol{r}') \right] \right\rangle_{\boldsymbol{r}';\boldsymbol{r}} \tag{3.26}$$

Here, $\langle \cdot \rangle$ denotes the ensemble average, where the average is performed over all the possible positions of the second molecule $\boldsymbol{r}'$ keeping the position of the first molecule $\boldsymbol{r}$ fixed. The equilibrium volume in the atomistic region maximizes the free energy in the thermodynamic limit, i.e., $\Delta A/\Delta V = 0$. Hence, we obtain:

$$p_{\mathrm{CG}} - p_{\mathrm{AT}} = \rho_0 (\omega^{\mathrm{H}}_{\mathrm{th}} - \omega_{\mathrm{rep}}). \tag{3.27}$$

As the pressure difference between the atomistic and coarse-grained region in the standard AdResS and auxiliary Hamiltonian AdResS is the same, we get the following relation by comparing the Eq. 3.27 with Eq. 3.22:

$$\omega_{\mathrm{rep}} = \omega^{\mathrm{H}}_{\mathrm{th}} - \omega_{\mathrm{th}} \tag{3.28}$$

The above equation relates the thermodynamic force of auxiliary Hamiltonian AdResS with thermodynamic force of GC-AdResS. The chemical potential difference between the different resolutions in GC-AdResS is given by:

$$\mu_{\mathrm{CG}} - \mu_{\mathrm{AT}} = \omega_{\mathrm{th}} + \omega_{\mathrm{DOF}} + \omega_{\mathrm{extra}} \tag{3.29}$$

Using the same arguments, the chemical potential difference between the atomistic and coarse-grained resolutions in auxiliary Hamiltonian AdResS is given by:

$$\mu_{\mathrm{CG}} - \mu_{\mathrm{AT}} = \omega^{\mathrm{H}}_{\mathrm{th}} + \omega_{\mathrm{DOF}} \tag{3.30}$$

By comparing (3.29) with (3.30), we have the relation:

$$\omega_{\text{extra}} = \omega_{\text{th}}^{\text{H}} - \omega_{\text{th}}, \tag{3.31}$$

which also relates the thermodynamic force of the auxiliary Hamiltonian AdResS and GC-AdResS. It has been shown in Ref. [111], that the ensemble average of $\boldsymbol{F}_{\text{rep}}$ is equivalent in both GC-AdResS and auxiliary Hamiltonian AdResS. The complete formula for the calculation of excess chemical potential in GC-AdResS becomes:

$$\mu_{\text{AT}}^{\text{ex}} = \mu_{\text{CG}}^{\text{ex}} - \int_{\Delta} d\boldsymbol{r}\boldsymbol{F}_{\text{th}}(\boldsymbol{r}) - \int_{\Delta} d\boldsymbol{r}\,\nabla_{\boldsymbol{r}} w(\boldsymbol{r})\,\langle w(\boldsymbol{r}')\,[\sum_{\alpha,\beta} U^{\text{AT}}(\boldsymbol{r}_\alpha - \boldsymbol{r}'_\beta) - U^{\text{CG}}(\boldsymbol{r} - \boldsymbol{r}')]\rangle_{\boldsymbol{r}';\boldsymbol{r}} \tag{3.32}$$

3.3.2 Extension to Multi Component systems

We assume that there are two components present in the system A and B, with total number of molecules of each component N^{A} and N^{B}, respectively. We denote the number of molecules of type A by N_1^{A}, N_2^{A} and N_3^{A} in the atomistic, transition, and coarse-grained regions respectively and similarly of type B by N_1^{B}, N_2^{B} and N_3^{B}. Under the assumption that the transition region is very small compared to the atomistic and coarse-grained regions, we have the following:

$$V \approx V_1 + V_3 \tag{3.33}$$

$$N \approx N^{\text{A}} + N^{\text{B}} \tag{3.34}$$

$$N^{\text{A}} \approx N_1^{\text{A}} + N_3^{\text{A}} \tag{3.35}$$

$$N^{\text{B}} \approx N_1^{\text{B}} + N_3^{\text{B}} \tag{3.36}$$

The thermodynamic force is applied to each component, which ensures uniform density of a particular component throughout the whole simulation box. Thus, we have:

$$\rho_{\Delta}^{\text{A}} = \rho_{\text{AT}}^{\text{A}} = \rho_{\text{CG}}^{\text{A}} = \rho_0^{\text{A}} \tag{3.37}$$

$$\rho_{\Delta}^{\text{B}} = \rho_{\text{AT}}^{\text{B}} = \rho_{\text{CG}}^{\text{B}} = \rho_0^{\text{B}} \tag{3.38}$$

Similar to Eq. (3.22), for mixtures, we have the following relation:

$$p_{\text{CG}} - p_{\text{AT}} = \rho_0^{\text{A}}\omega_{\text{th}}^{\text{A}} + \rho_0^{\text{B}}\omega_{\text{th}}^{\text{B}} \tag{3.39}$$

Using the relation between the thermodynamic force in GC-AdResS and auxiliary Hamiltonian AdResS, we get the following:

$$p_{\text{CG}} - p_{\text{AT}} = \rho_0^{\text{A}}(\omega_{\text{th}}^{\text{A,H}} - \omega_{\text{rep}}^{\text{A}}) + \rho_0^{\text{B}}(\omega_{\text{th}}^{\text{B,H}} - \omega_{\text{rep}}^{\text{B}}), \tag{3.40}$$

where work of changing representation for different components is given by:

$$\omega_{\text{rep}}^{\text{A}} = \omega_{\text{rep}}^{\text{AA}} + \omega_{\text{rep}}^{\text{AB}} = \int_{\Delta} d\boldsymbol{r}\langle \boldsymbol{F}_{\text{rep}}^{\text{AA}}(\boldsymbol{r})\rangle + \int_{\Delta} d\boldsymbol{r}\langle \boldsymbol{F}_{\text{rep}}^{\text{AB}}(\boldsymbol{r})\rangle, \tag{3.41}$$

$$\omega_{\text{rep}}^{\text{B}} = \omega_{\text{rep}}^{\text{BA}} + \omega_{\text{rep}}^{\text{BB}} = \int_{\Delta} d\boldsymbol{r}\langle \boldsymbol{F}_{\text{rep}}^{\text{BA}}(\boldsymbol{r})\rangle + \int_{\Delta} d\boldsymbol{r}\langle \boldsymbol{F}_{\text{rep}}^{\text{BB}}(\boldsymbol{r})\rangle. \tag{3.42}$$

where $\omega_{\rm rep}^{\rm AA}$ is work of changing representation for molecule of type A, due to interactions with molecule of type A only, while $\omega_{\rm rep}^{\rm AB}$ denotes the work of changing representation, due to interactions with molecule of type B. Similarly, one can define $\omega_{\rm rep}^{\rm BA}$ and $\omega_{\rm rep}^{\rm BB}$. The explicit expressions are:

$$\langle \boldsymbol{F}_{\rm rep}^{\rm AA}(\boldsymbol{r})\rangle = \nabla_{\boldsymbol{r}} w(\boldsymbol{r})\, \langle w(\boldsymbol{r}')[\, \sum_{\alpha,\beta} U_{\rm AA}^{\rm AT}(\boldsymbol{r}_\alpha - \boldsymbol{r}'_\beta) - U_{\rm AA}^{\rm CG}(\boldsymbol{r}-\boldsymbol{r}')\,]\rangle_{\boldsymbol{r}',{\rm A};\boldsymbol{r},{\rm A}} \tag{3.43}$$

$$\langle \boldsymbol{F}_{\rm rep}^{\rm AB}(\boldsymbol{r})\rangle = \nabla_{\boldsymbol{r}} w(\boldsymbol{r})\, \langle w(\boldsymbol{r}')[\, \sum_{\alpha,\beta} U_{\rm AB}^{\rm AT}(\boldsymbol{r}_\alpha - \boldsymbol{r}'_\beta) - U_{\rm AB}^{\rm CG}(\boldsymbol{r}-\boldsymbol{r}')\,]\rangle_{\boldsymbol{r}',{\rm B};\boldsymbol{r},{\rm A}} \tag{3.44}$$

$$\langle \boldsymbol{F}_{\rm rep}^{\rm BA}(\boldsymbol{r})\rangle = \nabla_{\boldsymbol{r}} w(\boldsymbol{r})\, \langle w(\boldsymbol{r}')[\, \sum_{\alpha,\beta} U_{\rm BA}^{\rm AT}(\boldsymbol{r}_\alpha - \boldsymbol{r}'_\beta) - U_{\rm BA}^{\rm CG}(\boldsymbol{r}-\boldsymbol{r}')\,]\rangle_{\boldsymbol{r}',{\rm A};\boldsymbol{r},{\rm B}} \tag{3.45}$$

$$\langle \boldsymbol{F}_{\rm rep}^{\rm BB}(\boldsymbol{r})\rangle = \nabla_{\boldsymbol{r}} w(\boldsymbol{r})\, \langle w(\boldsymbol{r}')[\, \sum_{\alpha,\beta} U_{\rm BB}^{\rm AT}(\boldsymbol{r}_\alpha - \boldsymbol{r}'_\beta) - U_{\rm BB}^{\rm CG}(\boldsymbol{r}-\boldsymbol{r}')\,]\rangle_{\boldsymbol{r}',{\rm B};\boldsymbol{r},{\rm B}} \tag{3.46}$$

where, $U_{\rm AB}^{\rm AT}$ represents the interaction potential in the atomistic region between molecules of type A and B and $U_{\rm AB}^{\rm CG}$ is the coarse-grained potential between molecules of type A and B. Similarly, all other interaction terms can be defined. Notation $\langle\cdot\rangle_{\boldsymbol{r}',{\rm B};\boldsymbol{r},{\rm A}}$ denotes the ensemble average is performed w.r.t. position of molecule B, keeping the position of A molecule fixed at $\boldsymbol{r}$. (The same convention holds for the other combinations of indices $\boldsymbol{r}', \boldsymbol{r}, {\rm A}, {\rm B}$). Thus, $\langle \boldsymbol{F}_{\rm rep}^{\rm AB}(\boldsymbol{r})\rangle$ represents the force acting on molecule of type A, due to the interactions with molecules of B. From Eq. (3.39), (3.40), we get:

$$\rho_0^{\rm A}(\omega_{\rm th}^{\rm A,H} - \omega_{\rm th}^{\rm A} - \omega_{\rm rep}^{\rm AA} - \omega_{\rm rep}^{\rm AB}) + \rho_0^{\rm B}(\omega_{\rm th}^{\rm B,H} - \omega_{\rm th}^{\rm B} - \omega_{\rm rep}^{\rm BA} - \omega_{\rm rep}^{\rm BB}) = 0 \tag{3.47}$$

It should be specified that we do not have $\omega_{\rm rep}^{\rm BA} = \omega_{\rm rep}^{\rm AB}$, even though $U_{\rm AB}^{\rm AT} = U_{\rm BA}^{\rm AT}$ and $U_{\rm AB}^{\rm CG} = U_{\rm BA}^{\rm CG}$. Let us denote the work done by the transition region on molecules of type A and B by $\omega_0^{\rm A}$ and $\omega_0^{\rm B}$, respectively. The chemical potential balance equation for the atomistic and coarse-grained regions, derived in Ref. [111], can easily be extended to multi-component systems.

$$\mu_{\rm AT}^{\rm A}(N_1^{\rm A}, N_1^{\rm B}, V_1, T) = \mu_{\rm CG}^{\rm A}(N_3^{\rm A}, N_3^{\rm B}, V_3, T) - \omega_0^{\rm A} \tag{3.48}$$

$$\mu_{\rm AT}^{\rm B}(N_1^{\rm A}, N_1^{\rm B}, V_1, T) = \mu_{\rm CG}^{\rm B}(N_3^{\rm A}, N_3^{\rm B}, V_3, T) - \omega_0^{\rm B} \tag{3.49}$$

Here, $\mu_{\rm AT}^{\rm A}$ is the excess free energy of solvation when molecule of type A is inserted into the infinitely large A–B mixture. For GC-AdResS, one can write:

$$\mu_{\rm CG}^{\rm A} - \mu_{\rm AT}^{\rm A} = \omega_{\rm th}^{\rm A} + \omega_{\rm DOF}^{\rm A} + \omega_{\rm extra}^{\rm A} \tag{3.50}$$

$$\mu_{\rm CG}^{\rm B} - \mu_{\rm AT}^{\rm B} = \omega_{\rm th}^{\rm B} + \omega_{\rm DOF}^{\rm B} + \omega_{\rm extra}^{\rm B} \tag{3.51}$$

where $\omega_{\rm extra}^{\rm A}$ is the energy dissipation due to molecule of type A that changes resolution. Similar definition holds for $\omega_{\rm extra}^{\rm B}$. The energy dissipation can be further written as:

$$\omega_{\rm extra}^{\rm A} = \omega_{\rm extra}^{\rm AA} + \omega_{\rm extra}^{\rm AB} \tag{3.52}$$

$$\omega_{\rm extra}^{\rm B} = \omega_{\rm extra}^{\rm BA} + \omega_{\rm extra}^{\rm BB} \tag{3.53}$$

$\omega_{\rm extra}^{\rm AB}$ is the energy dissipation of a molecule of type A produced by non-conservative interactions with molecules of type B. Similarly $\omega_{\rm extra}^{\rm AA}$ is the energy dissipation of a molecule

A produced by non-conservative interactions between molecules of type A and A only. The definitions are similar for $\omega^{\mathrm{BA}}_{\mathrm{extra}}$ and $\omega^{\mathrm{BB}}_{\mathrm{extra}}$. It should be noted that, we do not have $\omega^{\mathrm{BA}}_{\mathrm{extra}} = \omega^{\mathrm{AB}}_{\mathrm{extra}}$ in general. In accordance with the auxiliary Hamiltonian approach, for a single component, the chemical potential difference between the atomistic and coarse-grained regions can be written as:

$$\mu^{\mathrm{A}}_{\mathrm{CG}} - \mu^{\mathrm{A}}_{\mathrm{AT}} = \omega^{\mathrm{A,H}}_{\mathrm{th}} + \omega^{\mathrm{A}}_{\mathrm{DOF}} \tag{3.54}$$

$$\mu^{\mathrm{B}}_{\mathrm{CG}} - \mu^{\mathrm{B}}_{\mathrm{AT}} = \omega^{\mathrm{B,H}}_{\mathrm{th}} + \omega^{\mathrm{B}}_{\mathrm{DOF}} \tag{3.55}$$

By using Eq. (3.50), (3.52) and (3.54), we have

$$\omega^{\mathrm{AA}}_{\mathrm{extra}} + \omega^{\mathrm{AB}}_{\mathrm{extra}} = \omega^{\mathrm{A,H}}_{\mathrm{th}} - \omega^{\mathrm{A}}_{\mathrm{th}}. \tag{3.56}$$

Using Eq. (3.51), (3.53) and (3.55), we have

$$\omega^{\mathrm{BA}}_{\mathrm{extra}} + \omega^{\mathrm{BB}}_{\mathrm{extra}} = \omega^{\mathrm{B,H}}_{\mathrm{th}} - \omega^{\mathrm{B}}_{\mathrm{th}}. \tag{3.57}$$

By inserting Eq. (3.56) and (3.57) into Eq. (3.47), we have

$$\rho^{\mathrm{A}}_0(\omega^{\mathrm{AA}}_{\mathrm{extra}} + \omega^{\mathrm{AB}}_{\mathrm{extra}} - \omega^{\mathrm{AA}}_{\mathrm{rep}} - \omega^{\mathrm{AB}}_{\mathrm{rep}}) + \rho^{\mathrm{B}}_0(\omega^{\mathrm{BA}}_{\mathrm{extra}} + \omega^{\mathrm{BB}}_{\mathrm{extra}} - \omega^{\mathrm{BA}}_{\mathrm{rep}} - \omega^{\mathrm{BB}}_{\mathrm{rep}}) = 0 \tag{3.58}$$

The intuitive physical conclusion from the above equation is $\omega^{\mathrm{AA}}_{\mathrm{extra}} = \omega^{\mathrm{AA}}_{\mathrm{rep}}$, because these two terms exclusively involve A–A interactions, thus its related energy dissipation. The same is true for B–B interaction: $\omega^{\mathrm{BB}}_{\mathrm{extra}} = \omega^{\mathrm{BB}}_{\mathrm{rep}}$. The physical meaning of $\omega^{\mathrm{AB}}_{\mathrm{extra}}$, $\omega^{\mathrm{AB}}_{\mathrm{rep}}$, $\omega^{\mathrm{BA}}_{\mathrm{extra}}$ and $\omega^{\mathrm{BA}}_{\mathrm{rep}}$, leads to $\omega^{\mathrm{AB}}_{\mathrm{extra}} = \omega^{\mathrm{AB}}_{\mathrm{rep}}$, and $\omega^{\mathrm{BA}}_{\mathrm{extra}} = \omega^{\mathrm{BA}}_{\mathrm{rep}}$. It follows that for component A, the excess chemical potential difference is:

$$\mu^{\mathrm{A,ex}}_{\mathrm{CG}} - \mu^{\mathrm{A,ex}}_{\mathrm{AT}} = \int_{\Delta} \boldsymbol{F}^{\mathrm{A}}_{\mathrm{th}}(\boldsymbol{r})d\boldsymbol{r} + \int_{\Delta} \langle \boldsymbol{F}^{\mathrm{AA}}_{\mathrm{rep}}(\boldsymbol{r})\rangle d\boldsymbol{r} + \int_{\Delta} \langle \boldsymbol{F}^{\mathrm{AB}}_{\mathrm{rep}}(\boldsymbol{r})\rangle d\boldsymbol{r} \tag{3.59}$$

3.4 Numerical Results

In this section, we report the technical details of all the simulations, for reproducibility of our results. We report the excess chemical potential for a variety of liquids and mixtures that are important in (bio)chemistry and natural science as well as the structural properties of two mixtures: TBA/water and urea/water. We further compare the efficiency of GC-AdResS as a tool to compute the chemical potential with popular MD approach such as TI and IPM.

3.4.1 Simulation Setup

The potential energy function and the force field parameters for all the molecules were taken from GROMOS53A6 parameter set. Liquid water was described by the SPC model [4], methanol was described by the model developed by Walser et al [110], urea by the model described in [103], tert-butyl alcohol by the parameter set of [54] and DMSO was described by the model given by Geerke et.al. [28]. For liquid methanol simulations, GROMOS43A1 parameter set was used, as it was shown to be more accurate for calculating excess chemical potential of methanol [29]. The resolution of the molecules is assumed to change only

System	$\epsilon(kJ/mol)$	$\sigma(nm)$
methane	0.65	0.40
ethane	0.20	0.50
propane	0.65	0.55
methanol	0.65	0.40
DMSO	0.30	0.50
methanol in methanol/water	0.65	0.40
methane in methane/water	0.65	0.40
urea in urea/water	0.65	0.40
ethane in ethane/water	0.65	0.45
TBA in TBA/water	0.65	0.60
DMSO in DMSO/water	0.65	0.50
TBA in TBA/DMSO	0.40	0.60
DMSO in TBA/DMSO	0.30	0.50

Table 3.1: Parameters for WCA potential in the coarse-grained region

in the x-direction in all AdResS simulations performed in this work. To obtain a converged thermodynamic force, and subsequently a flat density, we performed 30 iterations for each system. Each iteration comprised of 200 ps of equilibration followed by 200 ps of data collection. All the simulations were performed at 298 K except simulations of liquid methane and ethane which were performed at 111.66 K and 184.52 K respectively. As it was discussed in [111], there is no requirement of a coarse-grained model that resembles the structural and thermodynamic properties of a full atomistic model. It was shown numerically that the proper exchange of energy and molecules was independent from the molecular model used in the coarse-grained region, showing the convenience of GC-AdResS. In this work, generic WCA potential was used in the coarse-grained region. The interaction potential between the coarse-grained particles is given by:

$$U(r) = 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right] + \epsilon, \quad r \leq 2^{1/6}\sigma. \tag{3.60}$$

The parameters σ and ϵ were chosen such that the radial distribution function of center of mass of particles reproduces a liquid structure. For water molecule, the parameters used in this study are $\epsilon = 0.65$ kJ/mol and $\sigma = 0.30$ nm. Table 3.1 shows the WCA parameters for other molecules that we have studied. For interactions between solute and solvent, σ values were obtained by averaging over the individual parameters. The solute-solvent ϵ is kept same as the solute-solute ϵ. To obtain the chemical potential of coarse-grained component, insertion particle method was used, where a trajectory of 8 ns was obtained and the coordinates were written after every 0.4 ps. The insertions of the molecule were performed 4,000,000 times in each frame at random locations and with random orientations of the molecule. The excess chemical potential value was calculated by averaging over the last ten iterations after the thermodynamic force had converged and the statistical uncertainty was determined by the standard deviation in the data. The excess chemical potential of the solute or the excess free energy of solvation was calculated

System	N_{solute}	$N_{solvent}$	System size (nm^3)	AT + HY region (nm^3)
water	—	13824	$30.2 \times 3.8 \times 3.8$	$14.6 \times 3.8 \times 3.8$
methane	—	2000	$9.0 \times 3.7 \times 3.6$	$6.0 \times 3.7 \times 3.6$
ethane	—	2000	$12.0 \times 3.9 \times 3.7$	$7.0 \times 3.9 \times 3.7$
propane	—	1433	$10.0 \times 4.5 \times 4.5$	$7.0 \times 4.5 \times 4.5$
methanol	—	4000	$12.0 \times 4.6 \times 4.5$	$7.4 \times 4.6 \times 4.5$
DMSO	—	1500	$15.0 \times 3.6 \times 3.3$	$7.0 \times 3.6 \times 3.3$
methanol/water	128	12672	$29.5 \times 3.7 \times 3.7$	$14.6 \times 3.7 \times 3.7$
methane/water	40	6960	$10.0 \times 4.8 \times 4.7$	$7.0 \times 4.8 \times 4.7$
urea/water	50	2500	$9.7 \times 2.9 \times 2.8$	$6.8 \times 2.9 \times 2.8$
ethane/water	40	6960	$10.0 \times 4.7 \times 4.6$	$7.0 \times 4.7 \times 4.6$
TBA/water ($x_{\mathrm{TBA}} = 0.001$)	40	39960	$50.1 \times 5.8 \times 4.3$	$7.0 \times 5.8 \times 4.3$
TBA/water ($x_{\mathrm{TBA}} = 0.02$)	80	4400	$10.0 \times 3.6 \times 4.2$	$7.0 \times 3.6 \times 4.2$
TBA/water ($x_{\mathrm{TBA}} = 0.04$)	180	4300	$10.0 \times 4.3 \times 3.7$	$7.0 \times 4.3 \times 3.7$
TBA/water ($x_{\mathrm{TBA}} = 0.12$)	538	3942	$10.0 \times 4.4 \times 4.6$	$7.0 \times 4.4 \times 4.6$
TBA/water ($x_{\mathrm{TBA}} = 0.16$)	717	3763	$10.0 \times 4.9 \times 4.6$	$7.0 \times 4.9 \times 4.6$
TBA/water ($x_{\mathrm{TBA}} = 0.20$)	896	3584	$12.0 \times 4.6 \times 4.5$	$7.0 \times 4.6 \times 4.5$
DMSO/water	50	4950	$12.0 \times 4.0 \times 3.3$	$7.0 \times 4.0 \times 3.3$
TBA/DMSO	80	4400	$10.0 \times 7.3 \times 7.2$	$7.0 \times 7.3 \times 7.2$

Table 3.2: Simulation details of AdResS and full-atom systems.

using the thermodynamic integration (TI) approach, introduced in Section 2.6.3. We computed the excess chemical potential using a two-stage approach as described in [66], first coupling van der Waals interactions to transform the non-interacting molecule into a partially-interacting uncharged molecule, then coupling Coulomb interactions from an uncharged interacting molecule to fully-interacting molecule. The resulting free energy ΔG_{final} is the sum of ΔG values obtained from the two procedures,

$$\Delta G_{final} = \Delta G_{ele} + \Delta G_{vdw} \tag{3.61}$$

where ΔG_{vdw} is the free energy change associated with introducing the van der Waals interactions and ΔG_{ele} is the free energy change associated with introducing Coulomb interactions. We evaluated the above integral for 21 values of λ (evenly spaced between 0 and 1) in both the procedures. At each value of λ, first a steepest descent energy minimization was performed followed by 200 ps of NPT equilibration and 400 ps of data collection under constant volume and temperature conditions, in accordance with AdResS simulations. During the van der Waals coupling, soft-core interactions were used with soft-core parameters $\alpha_{LJ} = 0.5$, $\sigma = 0.3$ and the power of λ in soft-core equation was taken as 1. Free energy estimates and the errors were calculated through Bennet's acceptance ratio method (BAR) [3]. For both AdResS and full-atom simulations, same system size was used. Table 5.2 gives a detailed summary of each system studied.

In all the simulations, a leap-frog stochastic dynamics integrator with a time step of 2 fs and an inverse friction coefficient of 0.1 ps was used. All bond-lengths were constrained using the LINCS algorithm. For liquid water, methanol, methanol/water, methane/water

and ethane/water a cut-off radius of 0.9 nm was used for van der Waals and Coulomb interactions, while for rest of the systems, a cut-off radius of 1.4 nm was used. For the TBA/water system, the chemical potential converges at cut-off 0.9 nm for mole-fraction $x_{\mathrm{TBA}} = 0.02$. Since it would be too expensive to do the convergence tests for all concentrations, we simply use a large cut-off 1.4 nm for the concentration dependency study of TBA/water. Electrostatic interactions were calculated using the reaction-field term [106] with a dielectric permittivity 54 for urea in SPC water [103], 64.8 for TBA in SPC water [54], 61 for other solutes in SPC water, 19 for methanol and 46 for DMSO as the solvent [29].

3.4.2 Chemical Potential Results

We have calculated μ^{ex} for a large variety of liquids and mixtures, with different molecular characteristics, eg. alkane liquids such as methane, ethane and propane, hydrophobic solvation in alkane/water mixtures, hydrophilic solvation in TBA/water and urea/water mixture, other liquids such as methanol and DMSO and non-aqueous mixture such as TBA/DMSO. Some systems such as urea/water are commonly used as co-solvent of biological molecules and TBA/water plays an important role in modern technology. Table 3.3 shows the chemical potential values obtained using GC-AdResS, compared with the value obtained using TI at the same concentrations. In Ref. [111], it was shown that the GC-AdResS reproduces μ^{ex} of liquid water obtained from IPM, however the computational cost of IPM was rather large, thus we have not computed μ^{ex} using IPM in this work. The results show that the agreement between GC-AdResS and TI values is highly satisfactory, with the maximum deviation in the values being 10%. This validates the robustness of GC-AdResS in describing the essential thermodynamics of the system correctly. We have also compared the values with those available in literature. The concentration of solute in the mixtures we have treated in GC-AdResS is slightly higher than those in Ref. [29, 54], but the chemical potential does not change significantly in the dilute regime, and has been verified for one relevant system (TBA/water), which will be discussed later on in this section. Figure 3.2 and 3.3 provide complementary information to Table 3.3. In GC-AdResS, the electrostatic interactions are computed using the reaction-field method, and the cut-off radius may play an important role in some of the systems investigated. Moreover, in TI, long range dispersion corrections are taken into account while calculating the excess energy of solvation (or excess chemical potential), while in GC-AdResS, we do not consider long range dispersion corrections. Hence we studied the behaviour μ^{ex} for DMSO in DMSO/water mixture as a function of cut-off radius. It can be seen from Figure 3.2, that as the value of cut-off increases, the chemical potential obtained from GC-AdResS simulations improves systematically and converges to the value obtained using TI at cut-off = 1.5 nm. It must be mentioned here that in MD simulations, 1.4 nm is taken as the standard cut-off radius, and results are highly satisfactory for this value. Figure 3.3 shows the behaviour of μ^{ex} as a function of concentration of TBA in TBA/water mixture. This study again demonstrates the robustness of GC-AdResS in predicting the correct thermodynamics in a very dilute regime. It can be seen that the agreement between GC-AdResS and TI values is remarkable, and at a very dilute concentration, GC-AdResS value matches perfectly with the experiment value. Moreover, the trend of the chemical potential values as a function of concentration is similar to that reported in Ref. [54].

Liquid component	Mole fraction of solute	GC-AdResS	TI	Experiment
methane	–	-4.6 ± 0.1	-5.2 ± 0.1	–
ethane	–	-8.2 ± 0.3	-8.8 ± 0.1	–
propane	–	-8.5 ± 0.1	-9.5 ± 0.2	–
methanol	–	-20.1 ± 0.1	-20.6 ± 0.4	-20.5 [29]
DMSO	–	-32.2 ± 0.3	-34.7 ± 0.7	-32.2 [79]
methanol in water	0.01	-18.1 ± 0.2	-19.7 ± 0.2	–
methane in water	0.006	9.1 ± 0.1	8.5 ± 0.2	–
urea in water	0.02	-56.1 ± 0.6	-58.2 ± 0.5	-57.8 ± 2.5 [103]
ethane in water	0.006	7.2 ± 0.2	7.4 ± 0.3	–
TBA in water	0.001	-19.5 ± 0.3	-20.8 ± 0.6	-19.0 [54]
DMSO in water	0.01	-31.4 ± 0.5	-33.2 ± 0.3	–
TBA in DMSO	0.02	-24.8 ± 0.4	-24.0 ± 0.5	–

Table 3.3: The excess chemical potential (in kJ/mol) of different liquids and mixtures calculated in GC-AdResS and TI. In some cases, experiment values are also reported at the same concentrations. For methanol we compare our results with those available in the literature using the same force field. The chemical potential of solvent is not reported above as it does not change significantly in a dilute regime.

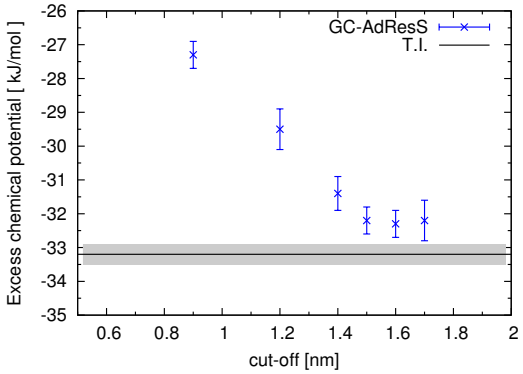

Figure 3.2: Excess chemical potential of Dimethyl sulfoxide (DMSO) in water as a function of cut-off radius calculated using GC-AdResS. The value obtained from thermodynamic integration calculation is also shown, with a grey region indicating the standard deviation. This value was calculated using a cut-off radius of 1.4 nm. It was seen that the value does not change significantly if the cut-off radius is varied.

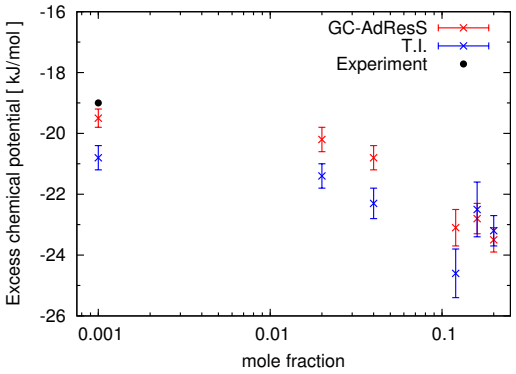

Figure 3.3: Excess chemical potential of tert-butyl alcohol (TBA) in water for different concentrations (in semi-logarithmic scale), calculated using GC-AdResS. The results are compared with thermodynamic integration values. At mole-fraction $x_{\mathrm{TBA}} = 0.001$, the experiment value is shown.

3.4.3 Structural Properties

As discussed previously, in addition to the calculation of μ^{ex} on-the-fly, one can also determine various thermodynamic and structural properties in a single GC-AdResS simulation to further strengthen the robustness of the method. The action of thermodynamic force and thermostat ensures that the density in the hybrid region is sufficiently close to the reference density. Figure 3.4 shows the molecular density of TBA and water in TBA/water mixture in the hybrid region, compared with the reference density. The results show that the largest difference is below 20% and the average difference is below 10%, which is good enough to ensure that there are no significant artifacts on the molecular density in the explicit region due to the interpolation of forces in the hybrid region. Figure 3.5 shows the molecular density of urea and water in urea/water mixture. In this case too, the average deviation of the molecular density is below 10%. Figure 3.6 and 3.7 show various radial distribution functions calculated in the explicit region in GC-AdResS and reference full-atomistic simulation for two systems: TBA/water and urea/water. The agreement between GC-AdResS and full-atomistic results is highly satisfactory. It must be noted that the systems studied here represent the worst case scenario of the technical setup in GC-AdResS, with a very small atomistic region (0.9 nm) and coarse-grained region (0.5 nm), and a relatively large transition region (2.7 nm). Even under such extreme technical conditions, the local distribution functions as well as relevant thermodynamic quantity such as μ^{ex} are well reproduced.

3.4.4 Efficiency

We compare the time taken to run a full GC-AdResS simulation to the time for a thermodynamic integration calculation for different systems at different concentrations of TBA in water to check the numerical efficiency of our approach. The total time required for a GC-AdResS simulation consists of the time taken to obtain a converged thermodynamic force and the time taken to obtain the coarse-grained chemical potential. The total time is obtained by adding the time taken to complete the TI procedure at each value of λ. TI

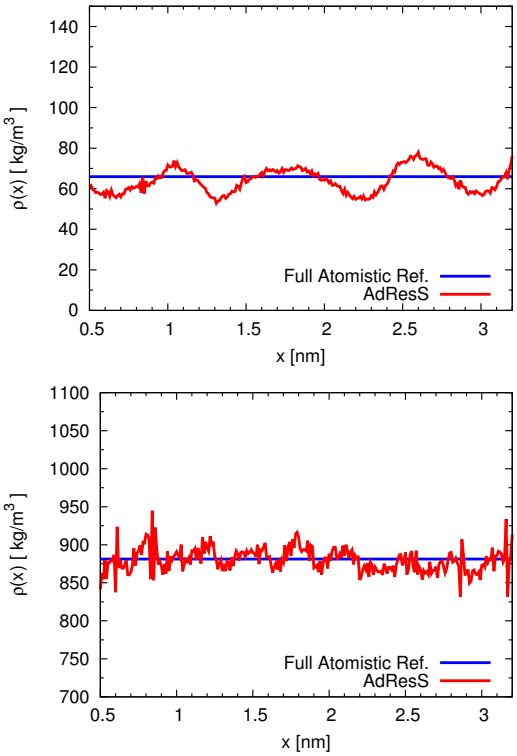

Figure 3.4: Top: Molecular density profile in Δ for TBA/water mixture; Bottom, the same plot for water. The action of the thermodynamic force and that of the thermostat leads to the largest deviation of 20% and an average difference of 10% from the reference all atomistic average density; which is numerically satisfactory. The mole-fraction is $x_{\mathrm{TBA}} = 0.02$, and the cut-off radius is 0.9 nm.

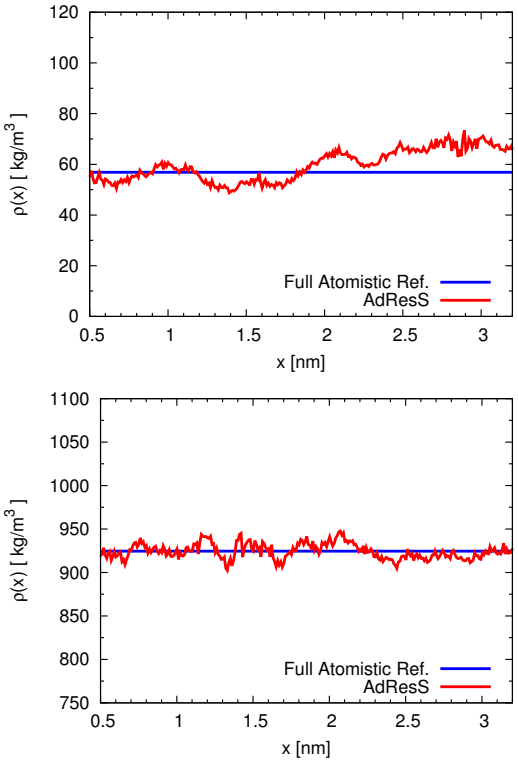

Figure 3.5: Top: Molecular density profile in Δ for urea/water mixture; Bottom, the same plot for water. The action of the thermodynamic force and that of the thermostat leads to the largest deviation of 20% and an average difference of 10% from the reference all atomistic average density; which is numerically satisfactory. The mole-fraction is $x_{\mathrm{TBA}} = 0.02$, and the cut-off radius is 0.9 nm.

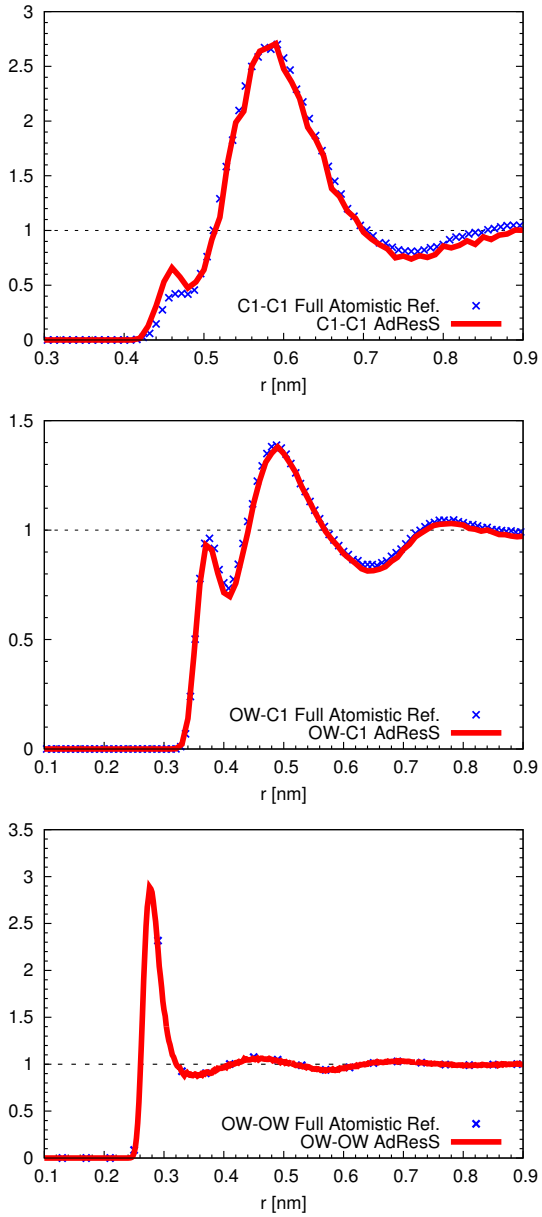

Figure 3.6: Top: g(r) for TBA-TBA ; Middle: g(r) for TBA-water; Bottom: g(r) for water-water. Red: the results of the AdResS simulation. Blue: the results of a full-atomistic reference simulation. Note: $g(r)$ is calculated only in the atomistic region. The mole-fraction is $x_{\mathrm{TBA}} = 0.02$, and the cut-off radius is 0.9 nm.

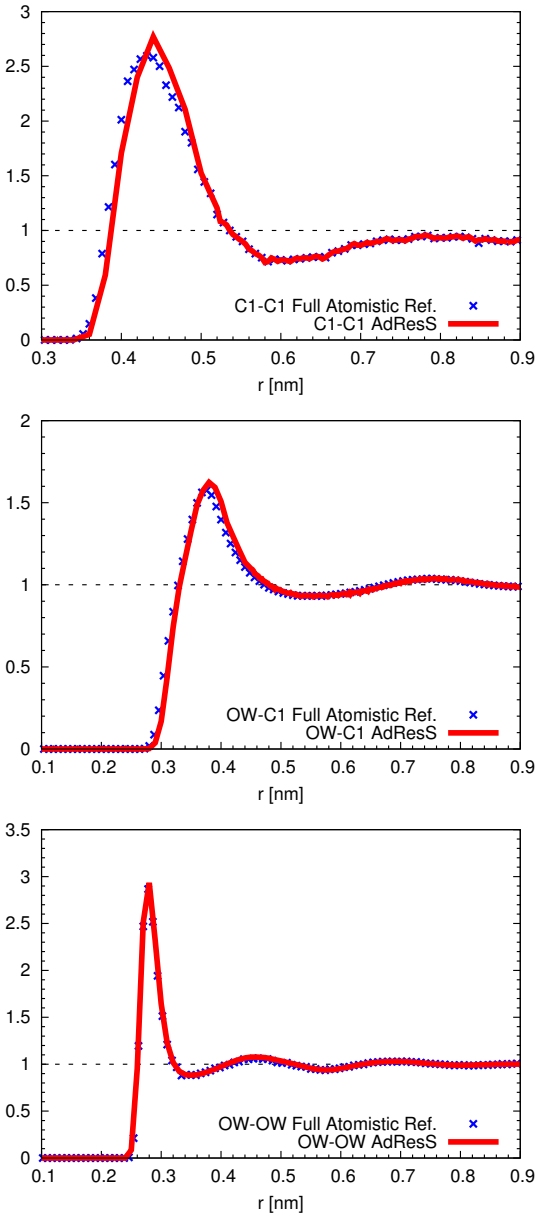

Figure 3.7: Top: g(r) for Urea-Urea; Middle: g(r) for Urea-water; Bottom: g(r) for water-water. Red: the results of the AdResS simulation. Blue: the results of a full-atomistic reference simulation. Note: $g(r)$ is calculated only in the atomistic region. The mole-fraction is $x_{\mathrm{UREA}} = 0.02$, and the cut-off radius is 0.9 nm.

procedure is preformed in two stages first the van der Waals interactions are coupled and then electrostatic interactions are coupled. At each stage, we use 21 equally distributed values of λ. Thus, in total 42 simulations are performed to calculate each TI chemical potential value. In AdResS, the time for thermodynamic force convergence largely depends on the initial guess. We start with a randomly chosen initial guess for system with the highest mole fraction of TBA. Since we use a random thermodynamic force, the number of iterations taken for convergence is large. For all the other systems, we use the converged thermodynamic force obtained from the first system as an initial guess, which makes the convergence faster in all the other systems. Table 5.1 shows the number of iterations required for the thermodynamic force convergence in GC-AdResS and total time required for GC-AdResS and TI calculation. The advantage of GC-AdResS over TI is that we get two values of excess chemical potential for both solute and solvent in a single calculation, while in TI, the whole process has to be repeated to get the excess chemical potential of other component. For very dilute systems ($x_{\mathrm{TBA}} = 0.001$), however, one has to take a very large systems in GC-AdResS. The time taken to obtain a converged thermodynamic force increases significantly in this case, and hence TI is always a better choice at such low concentrations, where much smaller systems can be simulated, at the same mole fraction.

	GC-AdResS		TI
x_{TBA}	No. of iterations	Time(hrs.)	Time(hrs.)
0.200	20	52.4	30.8
0.160	5	15.1	31.0
0.120	5	16.3	29.9
0.040	8	27.2	26.7
0.020	8	28.8	26.6
0.001*	20	252	202

Table 3.4: Time required for a full GC-AdResS and thermodynamic integration (TI) calculation for tert-butyl-alcohol(TBA) in water at different mole-fractions. The "*" indicates that a larger system was used in order to simulate a very dilute system ($x_{\mathrm{TBA}} = 0.001$). All the simulations were performed on a workstation that has two Quad-Core AMD Opteron(tm) 2376 Processors.

3.5 Current Computational Convenience: Comparison with other methods

An important question arising from the previous discussion is whether or not GC-AdResS provides a computationally convenient approach to calculate μ^{ex} compared to TI or Insertion Particle (IP) approaches. The current scenario does not give any straightforward answers, but the work is a first step towards a slightly more positive answer for future. In the fastest version of AdResS implemented in GROMACS-4.5, a speedup factor 4 was obtained compared to full-atomistic simulations using the same code. In this case, using GC-AdResS would be more convenient as one can not only calculate μ^{ex}, but also various structural and thermodynamical quantities. However, in GROMACS-4.6, the performance

of atomistic simulations (specifically SPC/E water) have improved considerably, while AdResS implementation has not been optimized fully. At the current stage, a speedup factor of 2 to 3 is obtained for large systems (30000 molecules, except SPC/E water) in AdResS simulations compared to atomistic simulations. Thus, TI would be computationally less demanding than GC-AdResS in this case. Another point that must be addressed for a fair comparison between TI and GC-AdResS is the following: TI has an advantage of using a single molecule as a solute, which acts as the minor component in a very dilute solution. However in GC-AdResS, one must consider a true mixture, with certain number of molecules of minor component present in the liquid of major component. At low concentrations, GC-AdResS requires much larger systems than those required in TI (for reference, see Table 5.2). Because of the low density of the solute component, thermodynamic force convergence requires long runs, as the density profile fluctuates highly. Thus, for very dilute systems, TI is more preferable to GC-AdResS, however, if one wants to calculate other properties too, then (optimized) GC-AdResS would be more convenient. At high concentrations, GC-AdReS is more preferable to TI for calculating both the chemical potential as well as general properties of the mixture because of the faster convergence of the thermodynamic force acting on the minor component. Another advantage of GC-AdResS over TI is that the chemical potential of both the components can be calculated in a single simulation, while one needs to run two separate simulations in order to get the chemical potential of both the components. The results reported in the previous section are encouraging, however, at the current state of the code, there is no clear convenience in using GC-AdResS over TI in calculating μ^{ex}. It should be noted that the major motivation of this work was to extend the theoretical concepts of GC-AdResS and show the robustness of the method in calculating important thermodynamic property such as μ^{ex}. In future, with all the technical optimizations in the AdResS code, we may have a more definitive answer as to where GC-AdResS stands in comparison to other methods.

3.6 Conclusions

In this chapter, we have given the analytical derivation of the formula to calculate excess chemical potential in GC-AdResS, and shown the accuracy and robustness of GC-AdResS in calculating the excess chemical potential for a representative class of complex liquids and mixtures. In every AdResS simulation, during the initial equilibration process, where the thermodynamic force is being calculated, one can automatically obtain the excess chemical potential. One may further calculate various other structural and thermodynamical properties within the same simulation. Thus, one does not need to run specific simulations in order to calculate μ^{ex}, which highlights one advantage of GC-AdResS over other computational methods that are used to compute μ^{ex}. Thus, GC-AdResS has evolved into a reliable multiscale technique, which upon further computational optimizations may become an effective tool to calculate μ^{ex}. Since this chapter firmly establishes GC-AdResS as a robust method to calculate thermodynamic and structural properties in open boundary systems, in the following chapter, we will deal with dynamical properties (equilibrium time correlation functions). We will give a theoretical definition of equilibrium time correlation functions in open boundary systems using the so-called Bergmann-Lebowitz model and we will calculate these functions for liquid water.

Chapter 4

Molecular Dynamics in open boundary systems: Bergmann-Lebowitz model and Adaptive Resolution Simulation (AdResS)

4.1 Introduction

Real life systems as well as experimental setups are likely (in most of the cases) to exchange matter and energy with their surroundings. Such systems fall under the class of systems, known as "open boundary systems" [94]. The theoretical description of open boundary systems is not trivial, and the theorems of statistical mechanics for systems with fixed number of particles, cannot be applied anymore and have to be modified accordingly. Molecular dynamics techniques with open boundary systems did not succeed as expected, due to their reduced efficiency when compared to standard MD simulations with fixed number of particles [60]. However, grand canonical adaptive resolution simulations (GC-AdResS) have led to the development of a technique, which allows for an efficient exchange of particles between the system and an external environment. In comparison to other algorithms with varying number of particles, GC-AdResS is highly efficient and flexible. This flexibility allows for the calculation of various equilibrium properties such as time correlation functions, some of which, because of varying N, require a theoretical redefinition compared to the standard canonical MD studies. In order to define the equilibrium time correlation functions in open boundary systems, it becomes necessary that we revisit the theoretical concepts that were developed five to six decades ago for statistical mechanics of open boundary systems [7, 52, 53]. In this part of the thesis, we will refer to open boundary systems which exchange energy and matter with the surroundings as grand ensemble systems; the grand-canonical ensemble is one particular case of grand ensemble. The aim of this chapter is: (a) discussion of theoretical aspects of open boundary MD simulations and a brief overview about the development of algorithms with varying number of particles in MD; (b) review of the so-called Bergman-Lebowitz (BL) model [7, 52, 53] for open boundary systems (c) discussion on the conceptual consistency of GC-AdResS with the BL model (d) definition of equilibrium time correlation functions in open boundary systems, and calculation of these functions for liquid water in GC-AdResS and reference full-atomistic simulations. The need to approach more complex systems, which are characterized by a realistic exchange of energy and matter with the surroundings is becoming

a guiding principle in the development of open boundary molecular dynamics systems [76], and this work represents only a starting point.

4.2 Background

In this section, we give a general overview on the theoretical concepts developed about the statistical mechanics of open boundary systems and discuss the general features of different MD techniques with varying number of particles reported in literature.

4.2.1 Statistical mechanics of open boundary systems

In literature [7, 21, 52, 53, 75], most of the ideas involving open boundary systems focus on coupling a system to a reservoir of energy, or non-equilibrium scenarios, such as transport of matter with an external source. Exchange of matter with the surroundings is usually limited to simple (formal) extensions of the case of heat exchange and heat flow [21, 52, 53]. However, while the exchange of heat has played an important role in MD simulations, the system is coupled to an external reservoir of energy which makes the system achieve its desired temperature and avoids the need of very large systems, as required in NVE simulations [107]. The idea of open boundary systems that automatically implies exchange of matter did not succeed in MD. Thus, concepts developed in Refs. [7, 52, 53] regarding the exchange of matter were not implemented in numerical tools for calculations. However, it has been shown that GC-AdResS is a highly efficient grand canonical-like MD technique. It becomes imperative to rediscover and develop the work in Refs. [7, 52, 53] further so that quantities which need a redefinition (where N is varying) such as equilibrium time correlation functions, can have a rigorous definition in GC-AdResS. In this work we discuss how in some approaches the microscopic description of the reservoir and its coupling to the system is not explicitly required; such a scenario is close enough to the GC-AdResS implementation. The basic principle of the approaches where an explicit coupling term in the Hamiltonian is not required, is that a small system is in contact with a large reservoir and the objective is to extract the (thermo) dynamical laws of small systems from the microscopic equation of the global system, i.e. small system and the reservoir. The Liouville equation of the global system is an ideal starting point; however the variables in the reservoir are macroscopic variables, that are considered to be averages over the microscopic states and do not explicitly enter into the evolution of the small system. The general hypothesis at the core of such models is that the small system is influenced by the reservoir only through its intensive properties [52]. The essential idea is that the intensive properties of the reservoir are constants of motion, even though its extensive properties change. As a consequence, the small system is governed by a self-consistent dynamical evolution, which does not contain any time dependent function of the reservoir. In this regard, we will discuss the grand-ensemble model proposed by Bergmann and Lebowitz for open boundary systems in Section 4.3, which (compared to other models) is conceptually solid and practical enough to be implemented in MD code.

4.2.2 Molecular Dynamics of subsystems with a varying number of particles

Grand Canonical MD simulations have mostly been intended as the tool to calculate the excess chemical potential following the Widom insertion [117] or the thermodynamic integration [105] techniques. These methods describe the effect of inserting/removing one molecule from a system. They are computationally expensive and the calculation of the excess chemical potential is the only aim of such studies. Hybrid MC/MD approach is an extension to these methods, where the dynamical evolution of the MD system is interfaced with MC moves which insert or remove molecules, and the system is equilibrated locally before the next MD step is initiated (see discussion in Ref. [24] and references therein). This approach is computationally highly demanding as each insertion/removal has costs of the order of those of Widom-like techniques for the calculation of chemical potential. Petitt and collaborators [10, 11, 41, 60, 115] initiated the development of grand-canonical like molecular dynamics. This method is based on the introduction of an additional variable 's', a fractional number that depends on the degree of presence of an additional molecule. An extended Hamiltonian which describes the evolution of '$N + s$' particles is then derived. It has been shown that this approach was not efficient when applied to liquid water [48]. The method was further improved by Eslani and Muller-Plathe [24]. In this method, the idea of fractional particles is conceptually appealing, but the extra computational costs and the numerical stability of the algorithm have restricted its popularity. As anticipated before, with the increasing success of multiscale simulations and the development of concurrent MD methods, a new generation of grand-canonical MD algorithms have come to forefront [89]. Such a category is that of adaptive resolution techniques, where molecules in the different regions are coupled via space-dependent forces [86, 111, 114], Hamiltonians [22, 84] or Lagrangians [34]. These algorithms can be easily converted to a grand-canonical MD scheme if the following requirements are fulfilled: (a) The reservoir is large enough to ensure physically realistic particle number fluctuations (b) The high resolution region is big enough to consist statistically relevant number of molecules. These methods are computationally superior to open boundary MD techniques of the previous generations. From this point of view, they represent a realistic pathway to future open boundary MD simulations, and in particular for those studies where the physics of subsystems is highly relevant.

4.3 Bergmann-Lebowitz model and Liouville equation for open boundary systems

In this section, we will discuss the Bergmann-Lebowitz (BL) model and report the Liouville equation for open boundary systems where the number of particles is not constant in time. In a seminal paper, Bergmann and Lebowitz [52] (see also Ref. [53]) present a general model for a many-particle system that is interacting with different reservoirs of energy and particles. For simplicity and consistency with our grand-canonical like MD setup, we will only consider the case of a single reservoir. The assumption in the BL model is that the interaction between the system and the reservoir is impulsive. Because of this assumption, it can be stated that the state of the reservoir state does not depend on the previous history of the system at the beginning of the interaction. Since there is no further interaction, the time dependence of the reservoir can be completely ignored. If

one specifies the distribution of the reservoir before the interaction with the system, then the effect of the reservoir on the system can be completely described; thus, one does not need to deal with the microscopic details of the reservoir. The interaction of the system with the reservoir component produces a discontinuous transition of a system from a state with N particles ($X_N^{'}$) to one with M particles (X_M). The change in the system state can be described by a stochastic kernel, $K_{MN}(X_M, X_N^{'})$. $K_{MN}(X_M, X_N^{'})$ represents the contingent probability that, in an infinitesimally small time interval, the system at $X_N^{'}$ makes a transition to X_M as a result of the interaction with the reservoir component. The evolution of the distribution function $\rho(X_M, M, t)$ in the system phase space is governed by the extended Liouville equation which we will name as the Bergmann-Lebowitz Liouville equation:

$$\frac{\partial \rho(X_M, M, t)}{\partial t} + \{\rho(X_M, M, t), \mathcal{H}(X_M)\} \tag{4.1}$$
$$= \sum_{N=0}^{\infty} \int dX_N^{'} [K_{MN}(X_M, X_N^{'})\rho(X_N^{'}, N, t) - K_{NM}(X_N^{'}, X_M)\rho(X_M, M, t)]$$

where $\mathcal{H}(X_M)$ is the Hamiltonian of the system corresponding to the point X_M and $\{*, *\}$ are the canonical Poisson brackets. The natural motion of the system and its interaction with the reservoir are responsible for the evolution of the distribution function $\rho(X_M, M, t)$ in the phase space of the system. If the condition of flux balance is satisfied (necessary and sufficient condition):

$$\sum_{N=0}^{\infty} \int [e^{-\beta \mathcal{H}(X_N^{'}) + \beta \mu N} K_{MN}(X_M, X_N^{'}) - K_{NM}(X_N^{'}, X_M) e^{-\beta \mathcal{H}(X_M) + \beta \mu M}] dX_N^{'} = 0 \tag{4.2}$$

then the stationary solution is the grand-canonical distribution given by:

$$\rho_M(X_M, M) = \frac{1}{Q} e^{-\beta \mathcal{H}_M(X_M) + \beta \mu M} \tag{4.3}$$

where β the Boltzmann factor of the system at a given temperature T, μ is the chemical potential and $Q = \sum_{M=0}^{\infty} e^{\beta \mu M} \int e^{-\beta \mathcal{H}_M(X_M)} dX_M$. In such a scenario, the BL Liouville equation represented in Eq. 4.1 is reduced to the standard Liouville equation, with the number of particles being a stochastic variable:

$$\frac{\partial \rho(X_M, M, t)}{\partial t} = \{\mathcal{H}(X_M), \rho(X_M, M, t)\}. \tag{4.4}$$

A rigorous mathematical formulation of Liouville theorem and operator for grand-ensemble systems is discussed in Ref. [19]

4.4 Correspondance between Bergmann-Lebowitz model and GC-AdResS

In this section we analyze the correspondence between BL model and GC-AdResS. The section is divided into four parts: In the first part, we will discuss the technical and conceptual features of GC-AdResS and relate them directly to the hypothesis/requirements

of the theoretical derivation of BL model. In the second part, we will show numerically why one can relate the Hamiltonian in the atomistic region of GC-AdResS to the Hamiltonian prescribed in the BL model. The third part focusses on the transition kernel in BL model and how one can interpret it within the GC-AdResS protocol. In the final part, we give a conceptual solidity to the calculation of equilibrium time correlation functions in GC-AdResS using the principles of BL model.

4.4.1 GC-AdResS formalism

The theoretical and computational formalism of GC-AdResS has already been discussed in detail in Chapter 3. Here, we will only discuss those features that are relevant to this Chapter. In the original AdResS setup [114], the thermostat acts over the whole system; in this work the idea has been extended further and in order to fulfil the requirements of the BL model for the calculation of the equilibrium time correlation functions, we have constructed a setup in which the thermostat is coupled to the reservoir and AT region is not subjected to the action of the thermostat; this particular setup will be referred to as 'local-thermostat'. Figure 5.1 shows the schematic representation of the original AdResS setup and 'local-thermostat' AdResS setup used for the calculation of dynamic properties. In Ref. [111], it was shown that in the coarse-grained region, one can use a generic liquid of spheres where the only requirement is that it has the same molecular density as the reference density. This is in accordance with the basic principle of BL model that we need only to know the distribution of the reservoir and not its microscopic details.

4.4.2 Mapping the Hamiltonian of the Atomistic region in GC-AdResS to the Bergmann-Lebowitz Hamiltonian

Consider a molecule i at position $\boldsymbol{r}_i$ in the AT region of AdResS; we have $w(\boldsymbol{r}_i) = 1$. The corresponding force can be divided into two parts, one is the force generated by the interaction of molecule i with molecules in the AT region:

$$\boldsymbol{F}_{i,j} = \boldsymbol{F}^{\mathrm{AT}}_{i,j}, \forall j \in \mathrm{AT} \tag{4.5}$$

and the other is the force generated by interaction with molecules in the reservoir:

$$\boldsymbol{F}_{i,j} = w(\boldsymbol{r}_j)\boldsymbol{F}^{\mathrm{AT}}_{i,j} + [1 - w(\boldsymbol{r}_j)]\boldsymbol{F}^{\mathrm{CG}}_{i,j}, \forall j \in \Delta + CG. \tag{4.6}$$

The total force acting on the ith molecule due to interaction with molecules in AT region is:

$$\boldsymbol{F}_i = \sum_{j \neq i} \boldsymbol{F}^{\mathrm{AT}}_{i,j} \tag{4.7}$$

This force can be expressed in terms of the gradient of the atomistic potential:

$$\boldsymbol{F}_i = \sum_{j \neq i} \nabla_i U_{\mathrm{AT}} \tag{4.8}$$

where ∇_i is the gradient w.r.t. molecule i. Eq. 4.6 represents the action of molecules in the reservoir on molecule i, i.e. an external force. The microscopic Hamiltonian of the system, embedded in the reservoir can not be defined due to system-reservoir coupling in Eq.4.6.

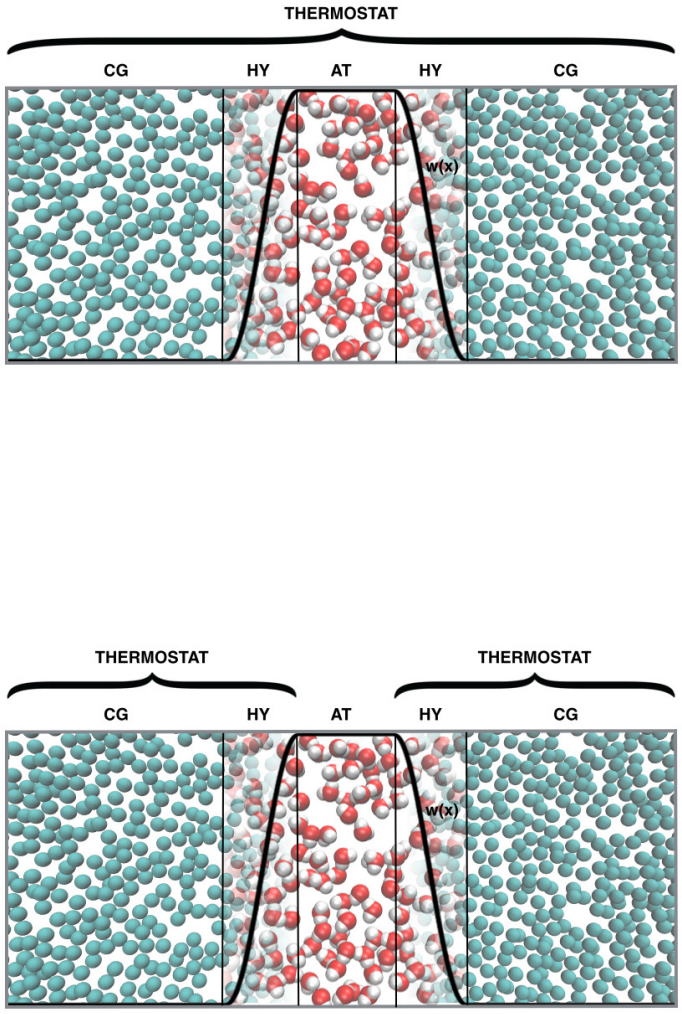

Figure 4.1: Schematic representation of the AdResS scheme; CG indicates the coarse-grained region, AT the atomistic region and HY the hybrid region where atomistic and coarse-grained forces are interpolated via a space-dependent, slowly varying, function $w(x)$. Top, the standard set up with the thermostat that acts globally on the whole system. Bottom, the "local" thermostat technique employed in this work where the thermostat acts in the HY and CG regions.

This restricts a straightforward correspondence between the Hamiltonian in the BL model ($\mathcal{H}_M$) and the Hamiltonian of the AT region in AdResS ($\mathcal{H}_{AT}$). However, in this work, we want to advocate that AdResS can be mapped to the BL framework and the numerical evidence will be provided later in this section. The non-integrable dynamics in the HY region can be considered as a boundary effect that can be incorporated in the definition of transition kernel. Further elaborating on this point, we will rewrite Eq. 4.6 as:

$$\boldsymbol{F}_i = \sum_{j\in\Delta+\mathrm{CG}} [w(\boldsymbol{r}_j)\boldsymbol{F}^{\mathrm{AT}}_{i,j} + [1-w(\boldsymbol{r}_j)]\boldsymbol{F}^{\mathrm{CG}}_{i,j}] = \sum_{j\in\Delta+\mathrm{CG}} [w(\boldsymbol{r}_j)\nabla_i U_{AT} + [1-w(\boldsymbol{r}_j)]\nabla_i U_{CG}]. \tag{4.9}$$

The net force on the ith particle can be considered as a non-local gradient field that is instantaneously produced by the external field generated by the molecules in the reservoir. Thus, the energy of the ith molecule at time $t > 0$ associated with Eq. 4.9 can be written as:

$$W^i_{AT-RES}(t) = \sum_{j\in\Delta+\mathrm{CG}} [w(\boldsymbol{r}_j)U^{ij}_{AT} + [1-w(\boldsymbol{r}_j)]U^{ij}_{CG}], \tag{4.10}$$

where $U^{ij}_{\cdot}$ represents the interaction energy between molecule at position $\mathbf{r}_i$ and molecule at position $\mathbf{r}_j$. The total energy of coupling is then defined as:

$$W_{AT-RES}(t) = \sum_{i\in AT} W^i_{AT-RES}(t)\,. \tag{4.11}$$

The total energy in Eq. 4.11 should be compared to the amount of energy, W_{AT-AT}, corresponding to the interaction between molecules of the AT region only: $W_{AT-AT}(t) = \sum_{i<j} U^{ij}_{AT}; i,j \in AT$. If

$$\frac{|W_{AT-AT}(t)| - |W_{AT-RES}(t)|}{|W_{AT-AT}(t)|} \approx 1; \forall t \tag{4.12}$$

then it seems reasonable to approximate the total energy of the atomistic system by the Hamiltonian of the AT region:

$$\mathcal{H}_{AT} \approx \mathcal{H}_{AT-AT}\,. \tag{4.13}$$

which corresponds to the Hamiltonian H_M of the BL model. For all the practical purposes, Eq. 4.12 holds true when the HY region is very small as compared to the AT region. However, Eq. 4.12 may not hold for the systems used routinely in AdResS simulations, where the size of the HY region is comparable to the size of the AT region. Figure 4.2 shows $W_{AT-AT}(t)$ and $W_{AT-RES}(t)$ for a system of 5000 molecules (about 450 in the AT region) that represents the worst case scenario. It can be seen that $W_{AT-AT}(t)$ is at least one order of magnitude larger than $W_{AT-RES}(t)$, so that the modeling error in terms of equilibrium expectation values that arises from replacing $\mathcal{H}_M$ of the BL model by $\mathcal{H}_{AT-AT}$ is about 10%. We have also performed simulations of a system where the technical setup was close to the thermodynamic limit (100000 molecules, with 20000 molecules in the AT region). In this case, the relative energy $W_{AT-RES}(t)/W_{AT-AT}(t)$ is less than 1%. Thus, microscopic characteristics of the AT system can be fully specified by assuming $\mathcal{H}_M = \mathcal{H}_{AT-AT}$

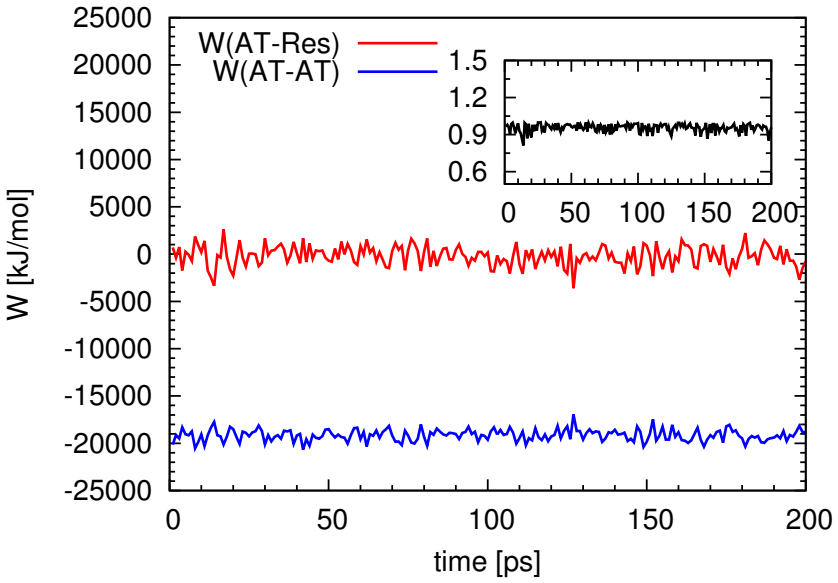

Figure 4.2: Potential energy of the subsystem only as a function of time, $W_{AT-AT}(t)$ compared to the energy associated with the interaction between subsystem and reservoir, $W_{AT-RES}(t)$; the former is at least one order of magnitude larger than the latter. Inset: The relative interaction between the AT region and the reservoir as a function of time : $\frac{|W_{AT-AT}(t)|-|W_{AT-RES}(t)|}{|W_{AT-AT}(t)|}$, it can be seen that the contribution of the system-reservoir interaction is atmost 10%.

4.4.3 Interpretation of Bergmann-Lebowitz transition kernel within the GC-AdResS framework

The reservoir in GC-AdResS influences the dynamics in the AT region via three different components: (a) the thermostat (b) thermodynamic force and (c) coupling force. As already discussed in the Chapter 3, the role of the thermostat is to provide thermal stability to the reservoir, and as a consequence, to the system. According to the BL model, no explicit knowledge of the phase space of the reservoir is required. This allows in GC-AdResS the use of stochastic thermostat in the reservoir, which is independent of the *apriori* knowledge of the microscopic structure of the reservoir and of the thermodynamic force ($\boldsymbol{F}_{th}$) in order to equilibrate the system to the thermodynamic state desired. The thermodynamic force converges once the density across the HY region becomes flat. This requires an exchange of particles between the AT and CG regions. The effect of the thermodynamic force is that the number of particles in the AT region vary such that the average number density is equal to the reference density. Thus, $\boldsymbol{F}_{th}$ imposes the stationary distribution of the reservoir at the first order ($\rho(x)$), which is independent of the interaction between the reservoir and the system. This condition is equivalent to the main requirement that the reservoir needs to fulfil in the BL model. The calculation of the thermodynamic force corresponds to the equilibration process in GC-AdResS and the chemical potential (μ_{AT}) is then automatically determined according to the equation:

$$\mu_{CG} = \mu_{AT} + \omega_{th} + \omega_{Q}, \tag{4.14}$$

where $\omega_{\mathrm{th}} = \int_{\Delta} \boldsymbol{F}_{\mathrm{th}}(\boldsymbol{r}) d\boldsymbol{r}$ and $\omega_{\mathrm{Q}} = \int_{\Delta} \nabla w(\boldsymbol{r}) \langle w(U^{\mathrm{AT}} - U^{\mathrm{CG}}) \rangle_{\boldsymbol{r}} d\boldsymbol{r} + \omega_{gas}$, ω_{gas} is the chemical potential in absence of intermolecular interactions and $\langle \cdot \rangle_{\boldsymbol{r}}$ indicates the conditional equilibrium average for fixed AT configurations. Eq. 4.14 is the necessary condition that GC-AdResS should fulfil in order to have a grand-canonical like distribution; i.e. to satisy Eq. 4.2, and as discussed above, this condition is imposed by the thermodynamic force. It follows that if the reservoir in the GC-AdResS has the identical insertion/deletion behaviour (Eq. 4.2), then the chemical potential in the AT region must be same. This implies that the condition of Eq. 4.2 in the BL model corresponds to the Eq. 4.14 in GC-AdResS. As shown in Section 4.4.2, the coupling force does not have a major energetic contribution to the interaction energy in the AT region. However, it consists of strong repulsive interactions that prevents molecules entering in the AT region to overlap with the molecules already present in AT region; which would produce numerical singularities that would automatically stop the simulation. Thus, the coupling force can be conceptually neglected as far as the construction of the transition kernel is concerned. However, it plays an important role as it imposes collision-avoidance between particles in the AT region and the reservoir in a robust and numerically efficient way. Although, a rigorous derivation of the BL kernel with GC-AdResS framework cannot be given, we have described how the different force contributions correspond to some of the properties of the kernel. The explicit mathematical formulation of the kernel in terms of the forces is beyond the scope of this work.

4.4.4 Equilibrium time correlation functions in GC-AdResS

The general definition of equilibrium time correlation function, $C_{AB}(t)$ between the two physical observables, A and B is [107]:

$$\begin{aligned} C_{AB}(t) = \langle a(0)b(t)\rangle &= \int d\mathbf{p}d\mathbf{q} f(\mathbf{p},\mathbf{q}) a(\mathbf{p},\mathbf{q}) e^{iLt} b(\mathbf{p},\mathbf{q}) \\ &= \int d\mathbf{p}d\mathbf{q} f(\mathbf{p},\mathbf{q}) a(\mathbf{p},\mathbf{q}) b(\mathbf{p}_t(\mathbf{p},\mathbf{q}),\mathbf{q}_t(\mathbf{p},\mathbf{q})) \end{aligned} \tag{4.15}$$

where, $a(\mathbf{p},\mathbf{q})$ and $b(\mathbf{p},\mathbf{q})$ are phase space functions corresponding to the observables A and B respectively, $a(0) = a(t=0)$ and $b(t)$ is the function at time t, $f(\mathbf{p},\mathbf{q})$ is the equilibrium distribution function and the dynamics is generated by the Liouville operator iL. $\mathbf{p}_t(\mathbf{p},\mathbf{q}), \mathbf{q}_t(\mathbf{p},\mathbf{q})$ indicates the time evolution of the positions and momenta $\mathbf{p},\mathbf{q}$. For the canonical ensemble, Eq. 4.15 can be written as:

$$C_{AB}(t) = \frac{1}{Q_N} \int d\mathbf{p}d\mathbf{q} e^{-\frac{\mathcal{H}_N(\mathbf{p},\mathbf{q})}{kT}} a(\mathbf{p},\mathbf{q}) b(\mathbf{p}_t(\mathbf{p},\mathbf{q}),\mathbf{q}_t(\mathbf{p},\mathbf{q})). \tag{4.16}$$

where Q_N is the Canonical partition function and $\mathcal{H}_N(\mathbf{p},\mathbf{q})$ the Hamiltonian of a system with N (constant) molecules. According to Eq. 4.16, $C_{AB}(t)$ can be computed by calculating $a(\mathbf{p},\mathbf{q})$ and $b(\mathbf{p}_t(\mathbf{p},\mathbf{q}),\mathbf{q}_t(\mathbf{p},\mathbf{q}))$ along the MD trajectory and then averaging over all the data. The trajectories must be long enough to ensure the basic requirements of ergodicity. In this case the dynamics generated by the Liouvillian operator is well defined, since the Liouville operator is well defined by the Hamiltonian of N molecules:

$$iL = \sum_{j=1}^{N} \left[\frac{\partial \mathcal{H}}{\partial \mathbf{p}_j} \frac{\partial}{\partial \mathbf{q}^j} - \frac{\partial \mathcal{H}}{\partial \mathbf{q}^j} \frac{\partial}{\partial \mathbf{p}_j} \right] = \{*, \mathcal{H}\} \tag{4.17}$$

For grand-canonical ensemble, Eq. 4.16 can be extended as:

$$C_{AB}(t) = \frac{1}{Q_{GC}} \sum_N \int d\mathbf{p}_N d\mathbf{q}_N e^{-\frac{[\mathcal{H}_N(\mathbf{p}_N,\mathbf{q}_N)-\mu N]}{kT}} a(\mathbf{p}_N,\mathbf{q}_N) b(\mathbf{p}_t(\mathbf{p}_N,\mathbf{q}_N),\mathbf{q}_t(\mathbf{p}_N,\mathbf{q}_N)). \tag{4.18}$$

where Q_{GC} is the Grand-Canonical Partition function, μ the chemical potential and N the number of particles (now varying in time) of the system. The question that arises is how to calculate the quantity $b(\mathbf{p}_t(\mathbf{p}_N,\mathbf{q}_N),\mathbf{q}_t(\mathbf{p}_N,\mathbf{q}_N))$. Suppose a system at time '0' has evolved from its initial condition and at any given time 't' has different number of molecules $N^{'}$ from its initial state. The correspondence between BL model and GC-AdResS enables to interpret the quantity $b(\mathbf{p}_t(\mathbf{p}_N,\mathbf{q}_N),\mathbf{q}_t(\mathbf{p}_N,\mathbf{q}_N))$ in numerical simulation, since the evolution of the system from $(\mathbf{p}_N,\mathbf{q}_N)$ to $(\mathbf{p}_t,\mathbf{q}_t)$ with $N^{'}$ molecules is given by Eq. 4.4. This leads to the following definition of the equilibrium time correlation functions for numerical simulations in GC-AdResS: "if a molecule at time t_o leaves the AT region at time t, then its contribution to the correlation function outside the time window $[t_o, t]$ is neglected". This is fully in accordance with the principle of BL model, which states that the microscopic identity of the molecule is lost once it enters into the reservoir.

4.5 Numerical Results

In this section, we report the results for SPC/E water [6] at room temperature. The section is divided into four parts: In the first part, we report the technical details of all the simulations performed in this work. The second part reports the various thermodynamic quantities calculated in this work. In the third part, we calculate the static properties with the aim of verifying numerically that GC-AdResS reproduces results of a natural grand-canonical system. The final part deals with the calculation of equilibrium time correlation functions. In Section 4.4.4, we have already discussed how to define the Liouville operator in the atomistic region, and how to count correlations when a molecule exits/enters the atomistic region. We will first show that GC-AdResS gives the same results as those of an open subsystem in a full atomistic NVE simulation. Next, we will show that by increasing the size of the atomistic region, results systematically converge to those of full NVE simulations, where the correlation functions are calculated over the whole system.

4.5.1 Technical Setup

All simulations are performed by home-modified GROMACS [93], and the thermodynamic force in AdResS simulations is obtained using VOTCA [98] package. The system contains 5000 water molecules and the dimensions of the system are $14.6 \times 3.2 \times 3.2\ nm^3$. In AdResS simulations, the resolution of the molecules changes only in the x-direction as depicted in Figure 5.1. Three different atomistic regions are used in AdResS simulations, whose sizes are $0.6 \times 3.2 \times 3.2\ nm^3$, $1.2 \times 3.2 \times 3.2\ nm^3$ and $4.8 \times 3.2 \times 3.2\ nm^3$ (this latter being a worst-case scenario for the reservoir, still results are highly satisfactory). The size of the hybrid region is kept same in all the three cases $2.9 \times 3.2 \times 3.2\ nm^3$. The remaining system contains coarse-grained particles, which interact via generic WCA potential given by Eq. 3.60: The parameters σ and ϵ in the current simulations are 0.30 nm and 0.65 kJ/mol respectively. The time step used in the simulations is 0.002 ps, and the coordinates and velocities are recorded after every 10 time steps, i.e. 0.02 ps. All simulations are performed at room temperature 298 K. The coarse-grained and the hybrid region in the AdResS system are coupled to a Langevin thermostat, whose time scale is 0.1 ps. The reaction field method [57, 106] is used for calculating the electrostatic interactions in the system, with dielectric constant $\epsilon_{RF} = \infty$, as this tends to give good energy conservation. The "switch" cut-off method is used to treat the Van der Waals interactions. The cut-off radius for interactions is 1.2 nm. For a 1 ns full atomistic simulation (without any thermostat), the total energy obtained is -195846 kJ/mol and the drift is just 11.4 kJ/mol, which is less than 0.01%. The dynamical results from this micro-canonical ensemble are compared with results from AdResS simulations. All the dynamical properties are computed from equilibrated trajectories of 1 ns in full atomistic and AdResS simulations. The velocity autocorrelation function is defined as:

$$C_{VV}(t) = \frac{1}{N} \sum_{i=1}^{N} \frac{\langle v_i(t) \cdot v_i(0) \rangle}{\langle v_i(0) \cdot v_i(0) \rangle} \tag{4.19}$$

where $\langle \cdot \rangle$ denotes the equilibrium average and $\langle v_i(t) \cdot v_i(0) \rangle$ computes the correlation between the velocities of i^{th} molecule at time '0' and 't'. In this work, the velocity auto

correlation function is calculated only for the oxygen atoms. In the same way, the dipole auto correlation function is defined as:

$$C_{\mu\mu}(t) = \frac{1}{N}\sum_{i=1}^{N}\frac{\langle \mu_i(t)\cdot\mu_i(0)\rangle}{\langle \mu_i(0)\cdot\mu_i(0)\rangle} \tag{4.20}$$

where $\langle \mu_i(t)\cdot\mu_i(0)\rangle$ computes the correlation between the dipole moment of i^{th} molecule at time 0 and t. In the current implementation of AdResS, the electrostatic interactions are calculated by short ranged reaction field method. The dipole auto correlation function results are consistent with the full atomistic simulation, also using reaction-field. The reactive flux hydrogen bond correlation [58, 59] function is defined as:

$$k(t) = -dC_{HH}/dt \tag{4.21}$$

where $C_{HH}(t)$ is the hydrogen bond autocorrelation function defined as:

$$C_{HH}(t) = \frac{\langle h(0)\cdot h(t)\rangle}{\langle h\rangle} \tag{4.22}$$

Here h is the hydrogen bond population operator for a particular pair of molecules. It is assigned a value '1', if there is a hydrogen bond between this pair, otherwise a value '0'. The criteria for considering a hydrogen bond between two water molecules is (1) inter oxygen distance is less than 0.35 nm and (2) the $O-H\ldots O$ angle is smaller than 30^o. The function $C_{HH}(t)$ is the conditional probability that a hydrogen bond between a pair of molecules is present at time 't', given that it was present at time zero. In both the full atomistic and AdResS simulations, first $C_{HH}(t)$ was calculated and then $k(t)$ was obtained by taking numerical derivative of $C_{HH}(t)$, using a time step of 0.02 ps.

4.5.2 Thermodynamic Properties

The following thermodynamic quantities are analyzed in this work:

$$Var(E) = \frac{\langle E^2\rangle - \langle E\rangle^2}{\langle E\rangle} \tag{4.23}$$

and

$$CoVar(N,E) = \frac{\langle NE\rangle - \langle N\rangle\langle E\rangle}{\langle N\rangle} \tag{4.24}$$

where $Var(E)$ is the variance in the total energy of the molecules in the atomistic subregion in AdResS and an equivalent subregion in the full-atom simulations, $CoVar(N,E)$ is the covariance between the total energy of the molecules and number of molecules which are present in the atomistic subregion in AdResS and an equivalent subregion in the full-atom simulations. The energy E consists of the sum of the kinetic energy of the molecules in the region considered, plus the energy coming from the interactions of each molecule with all the other molecules of the region considered. The interactions with the reservoir, defined in the text, "technical interactions", are not counted, for consistency with the definition of reservoir in the Bergmann-Lebowitz model. The different properties are calculated from a 2 ns long trajectory. The error in the data was calculated using "block-averaging" analysis.

4.5.3 Results-Static Properties

Figure 4.3,4.4,5.8 show the various static properties calculated in AT subregion of AdResS and in an equivalent subregion of a full-atomistic NVE simulation. The difference with the results in Ref. [111] is that the transition region is considerably smaller and the thermostat acts only on the reservoir. Figure 4.3 shows that the number density in GC-AdResS simulations agree quite well with that of full-atomistic simulations. The largest deviation is seen at the border of atomistic and transition region. This is due to the fact that we switch off the thermostat abruptly. The effect is negligible; however, there are three technical options that will allow us to reduce the effects of such a difference: (a) consider the atomistic buffer as a transition region; in accordance with the technical setup of original GC-AdResS protocol [111] (b) the convergence criteria for thermodynamic force should be made stricter (c) smoothly switching off the thermostat in the transition region. In this work, we have opted for option (a), because in any case, the deviation in the results due to this discrepancy is not more than 10%. Figure 4.4 shows the various radial distribution functions calculated in a subregion in GC-AdResS and full- atomistic NVE simulations. It can be seen that the agreement between the two set of curves is excellent, which again shows the robustness of GC-AdResS in reproducing the probability distributions of a natural grand canonical at least up to second order. Figure 5.8 reports the particle number probability distribution calculated in the subsystems of GC-AdResS and full atomistic simulations. The shape of two curves is gaussian and the curve of GC-AdResS is shifted compared to NVE system by two particles, due to deviation in the density at the border of the atomistic region. If we consider the original protocol and calculate the distribution functions in the atomistic region neglecting the region at the border; two curves overlap. Table 4.1 shows the robustness of GC-AdResS as a grand-canonical setup for the calculation of thermodynamic properties such as energy fluctuation and the covariance. Regarding the accuracy, in the worst case, the deviation is no more than 10% with results of the full-atomistic simulations, which is numerically satisfactory. However, if we apply the rigorous GC-AdResS protocol, the discrepancy reduces to 3%. We performed an additional test in order to prove that GC-AdResS satisfies a thermodynamic condition of a grand-canonical ensemble. In the thermodynamic limit, the isothermal compressibility κ_T in a grand-canonical ensemble can be related to the fluctuations in particle number distributions by [99]:

$$\rho k_B T \kappa_T = \frac{\langle N^2 \rangle - \langle N \rangle^2}{\langle N \rangle} \tag{4.25}$$

where ρ is the density of particles, k_B the Boltzmann constant and, $T = 298K$, the temperature. The test was done for a system of 20000 molecules with a reservoir of 80000 (total number of molecules 100000) at a pressure of $1atm$; in this case we obtain $\kappa_T = 44.6 \pm 1.6 \ \ 10^6 (bar^{-1})$ which should be compared with the value $45.9 \pm 1.2 \ \ 10^6 (bar^{-1})$ of the corresponding full atomistic system and with the value of about $45.25 \pm 10^6 (bar^{-1})$ of experiments [40,45] and $44.0 \pm 10^6 (bar^{-1})$ from NPT simulations of SPC/E water [40]; the overall accuracy is within 5% (in the worst case), which can be considered a satisfactory result. It must also be underlined that an effective compressibility, that is the definition of Eq.4.25 used not in the thermodynamic limit (i.e. small systems), is found to be the same in GC-AdResS and in the full atomistic simulation. We have shown that the reservoir based on the local-thermostat procedure of GC-AdResS produces grand-canonical stat-

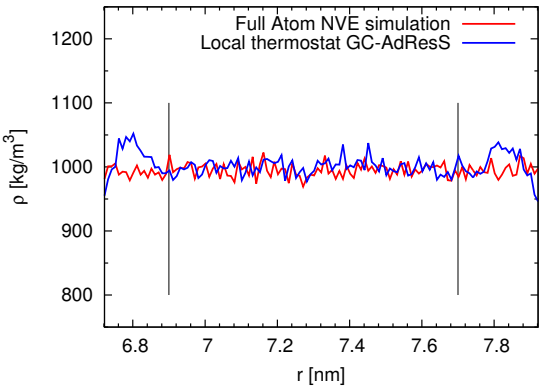

Figure 4.3: Molecular number density calculated in GC-AdResS where the thermostat is acting only in the reservoir and in a full atomistic NVE simulation. A discrepancy of about 5% is observed at the border of the AT region (vertical lines). In general, the rigorous protocol of GC-AdResS requires that this part is included in the hybrid region as a buffer of fully atomistic molecules.

istics, by calculating various static and thermodynamic quantities. We will now proceed with the calculations of equilibrium time correlation functions where the BL model for grand-ensemble provides a conceptual solidity to the numerical calculations.

Quantity	Full-Atomistic	GC-AdResS
$\frac{\langle E^2\rangle-\langle E\rangle^2}{\langle E\rangle}$	20.6 ± 0.4	19.3 ± 0.4
$\frac{\langle NE\rangle-\langle N\rangle\langle E\rangle}{\langle N\rangle}$	4.4 ± 0.2	3.9 ± 0.2

Table 4.1: Thermodynamic fluctuations calculated in atomistic subregion (EX=1.2) in GC-AdResS and full-atom simulations. There is a discrepancy of around 5-10% between the results of GC-AdResS and those of the reference full-atom simulation.

4.5.4 Results-Dynamic Properties

Figure 4.6 shows the velocity-velocity autocorrelation function, $C_{VV}(t)$ (top), (molecular) dipole-dipole autocorrelation function, $C_{\mu\mu}(t)$ (middle), reactive flux correlation function, $k(t)$ (bottom) calculated in the subsystem in GC-AdResS simulation and an equivalent subregion in full-atomistic NVE simulation. It can be seen that the agreement is remarkable. This implies that the 'ideal' reservoir in GC-AdResS is very close to the thermodynamic limit of the microscopic system. In order to further validate the concepts discussed here, we have done a systematic study by varying the size of the AT region in GC-AdResS and checking if the correlation functions converge to those obtained for the whole system in a full-atomistic NVE simulation. This principle is based on the fact that all the ensembles are equivalent in the thermodynamic limit. Figure 4.7 shows the systematic convergence of the correlation functions to the reference full atomistic results as a function of the size of

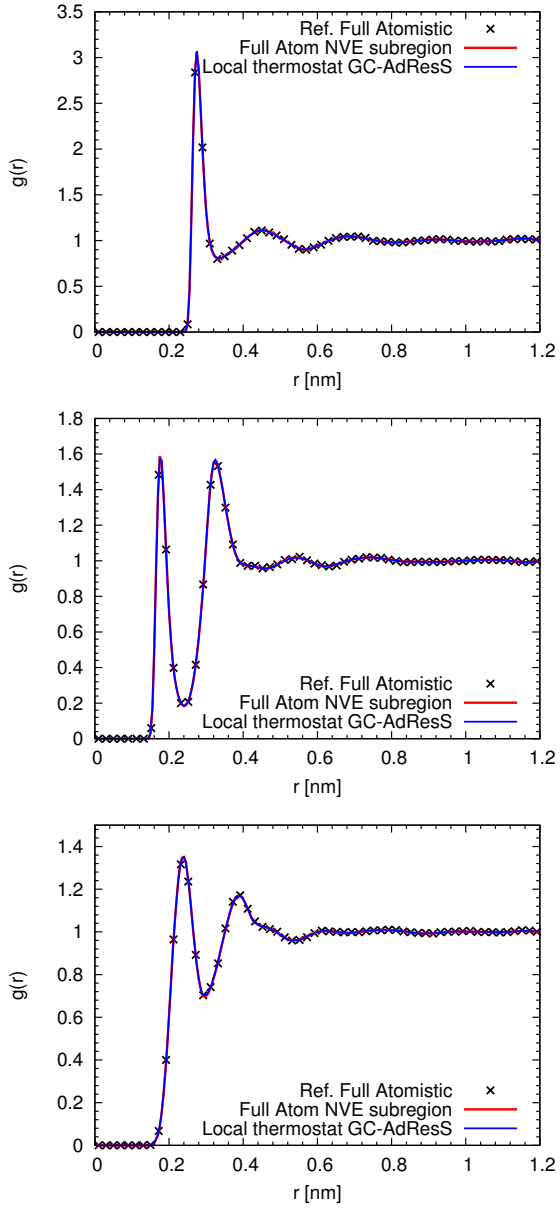

Figure 4.4: Oxygen-Oxygen (top), oxygen-hydrogen (middle) and hydrogen-hydrogen (bottom) radial distribution functions (RDF's) calculated in the subsystem (1.2 nm) of 'local-thermostat' GC-AdResS and an equivalent subregion in a fully atomistic NVE simulation. The results are compared with the same quantity calculated over the whole system in the reference full-atomistic simulations

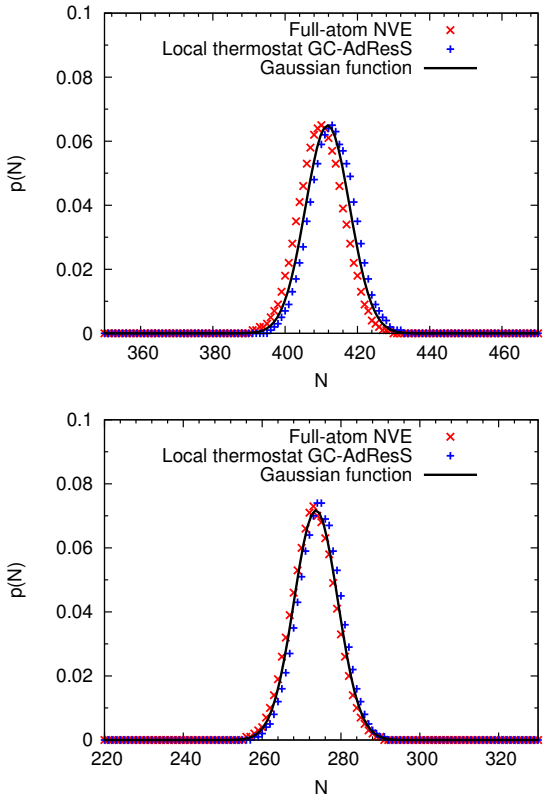

Figure 4.5: (top) Particle number probability distribution calculated in the subsystem (1.2 nm) of 'local-thermostat' GC-AdResS and an equivalent subregion in a fully atomistic NVE simulation. The shape of both curves is a Gaussian (reference black continuous curve); the curve of AdResS is shifted compared to the NVE results by only two particles. However, if we consider the additional atomistic buffer (bottom), as it should be if the principles of GC-AdResS are rigorously applied, then the two curves overlap.

Quantity	Full-Atomistic	GC-AdResS
$\frac{\langle E^2\rangle-\langle E\rangle^2}{\langle E\rangle}$	27.1 ± 0.5	26.4 ± 0.5
$\frac{\langle NE\rangle-\langle N\rangle\langle E\rangle}{\langle N\rangle}$	5.1 ± 0.2	4.9 ± 0.2

Table 4.2: Same quantities as above calculated in the region excluding the (negligible) part where the density is 5% off compared to the reference density, as discussed in Fig 4.3. The results in GC-AdResS and full-atom simulation agree now within 3%, which is numerically highly satisfactory.

the AT subsystem of AdResS. It must be noted that this procedure establishes a connection between the decay of correlation function as a function of time and the spatial locality of the process. For eg. in dense gases, the decay times are relatively large, thus if the size of explicit atomistic region is very small, then the decay time would be smaller than the real decay time, since many molecules are likely to leave the region. One way to check whether or not this method captures a certain decay process is to perform a systematic study by increasing the size of the atomistic region and observe the convergence of the functions. The connection between the decay process and spatial locality is not a limitation of the method, in fact it represents one of its main conceptual advantages since it allows one to identify the essential degrees of freedom (in space and time) required for a certain process.

4.6 Conclusions

In this part of the thesis, we have discussed Bergmann-Lebowitz model for grand ensembles in the context of statistical mechanics of open boundary systems. The model has not been employed in the development of open boundary MD techniques before, despite its conceptual solidity. We have discussed the connection between GC-AdResS and BL model and used its ideas in defining equilibrium time correlation functions in open boundary systems, where the number of particles is not constant. We have then calculated these functions for liquid water in the atomistic region in GC-AdResS, and shown that the results have a highly satisfactory agreement with the reference full-atomistic results. These results have established GC-AdResS as a conceptually robust and numerically efficient method for the simulation of open boundary systems. Therefore, one can introduce more complex molecular models in the explicit region in GC-AdResS, for eg. introducing quantum mechanics via path-integral formalism of Feynman. In this perspective, we discuss the various path-integral techniques available in literature and use the concepts developed in this to give a to combine path-integral molecular dynamics with GC-AdResS in the next chapter.

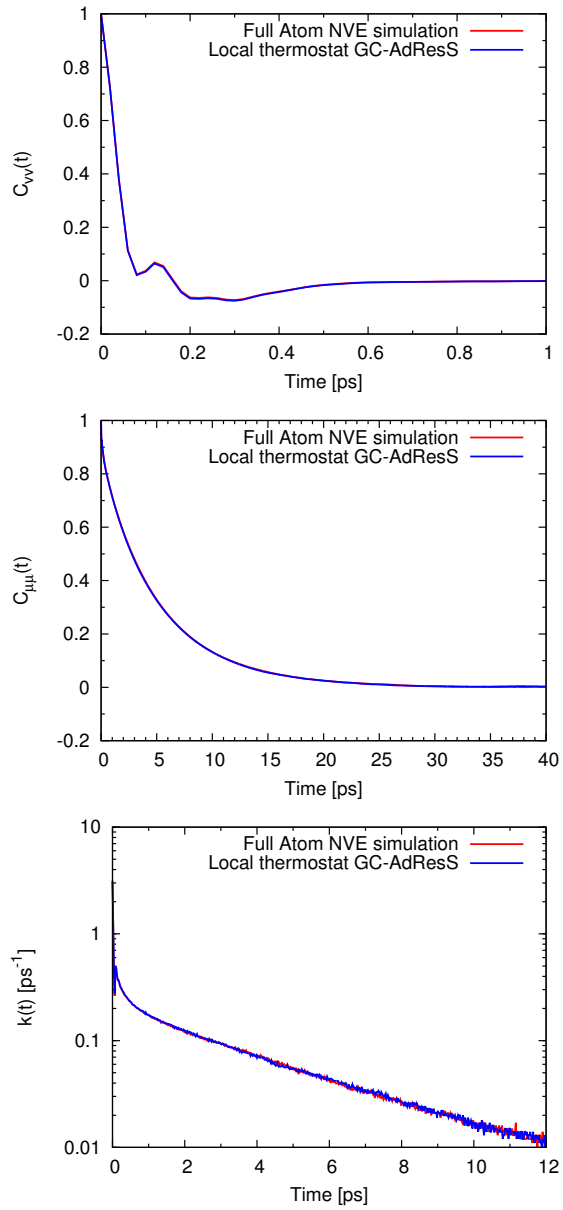

Figure 4.6: Velocity-Velocity autocorrelation function, $C_{VV}(t)$, (molecular) dipole-dipole autocorrelation function, $C_{\mu\mu}(t)$, reactive flux correlation function, $k(t)$ (semilogarithmic plot) for SPC/E water at room conditions calculated with GC-AdResS and for an equivalent subsystem in a full atomistic NVE simulation. The agreement between GC-AdResS and the full atomistic simulation is highly satisfactory.

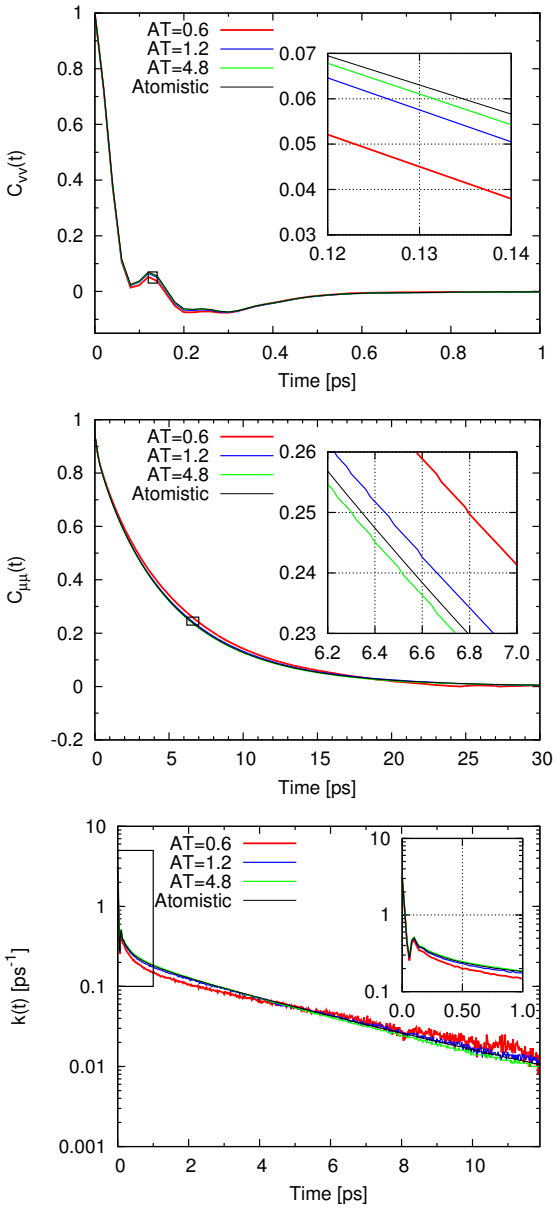

Figure 4.7: Systematic convergence of $C_{VV}(t)$, $C_{\mu\mu}(t)$ and $k(t)$ (semilogarithmic plot) of GC-AdResS to the full atomistic NVE results calculated over the whole system.

Chapter 5

Path Integral techniques within the grand canonical-like adaptive resolution (GC-AdResS) technique: Simulation of liquid water

5.1 Introduction

Atoms characterized by light nuclei such as hydrogen display considerable quantum effects, in the form of spatial delocalization. Systems composed by atoms of such kind are modelled via Feynman's path integral (PI) formalism [25]. In this approach, as discussed in Chapter 2, a classical atom is mapped onto a polymer ring with fictitious beads, and the magnitude of the quantum character is defined by the number of beads. As a consequence, the computational effort associated with path-integral simulations is massive (due to increase in number of interactions) compared to classical simulations (where the nuclei is treated as a rigid particle). The limitation of treating only small systems and short time scales may prohibit the treatment of "large-scale" problems (eg., hydration of protein in water). One of the options would be to use path integral (PI) techniques (employed in molecular dynamics) in conjunction with GC-AdResS [111, 114], where the region of relevance can be treated within the path-integral formalism, and rest of the system can be treated using simple coarse-grained models. The (open)path-integral region will exchange particles and energy with the coarse-grained reservoir, in statistical and thermodynamic equilibrium. The major difficulty in realizing a quantum/classical coupling is the different equation of motion in different regions: probabilistic for the quantum region and deterministic in the classical (or coarse-grained) region. However, the isomorphism between the path-integral partition function and the classical partition function of fictitious polymer beads transforms the problem of quantum/classical coupling into a standard coupling of different classical regions that are characterized by different number of "effective" degrees of freedom. Thus, a molecule where each atom is represented by a polymer ring, moves from the path-integral region to the coarse-grained region in a smooth way, thereby loosing (technically speaking) classical degrees of freedom. Thus, in this way, one can realize the passage from quantum to classical description and vice versa by simply employing equilibrium conditions of the standard adaptive resolution scheme. Previously PI was merged with AdResS [81–83], and the technique was used in the study of tetrahedral molecules and low-temperature parahydrogen, and the results were highly encouraging. The latest version of AdResS,

i.e. GC-AdResS, technically and conceptually more advanced, allows for the study of more complex systems. In this part of thesis, we report the technical implementation of different PI formalisms (path integral molecular dynamics (PIMD) [56,118], ring polymer molecular dynamics (RPMD) [14,62] and centroid molecular dynamics (CMD) [13,38]) into GC-AdResS, and we show its application to liquid water by calculating the structural and dynamical properties in the explicit PI subregion, and comparing the results with the reference PI results.

5.2 Path Integral Formalism and Normal Modes

In this section, we give a brief overview about path-integral formalism and the different techniques that exist in literature.

5.2.1 Background

The path-integral formalism of Feynman [25,107] in the context of dynamics of molecular systems has already been discussed in Chapter 2. Thus, in this section we will only discuss the technical aspects that are essential for its practical application in MD (see Refs. [8,16,77,107,108]). The Hamiltonian for a single particle moving under the influence of an external potential U is expressed as:

$$\mathcal{H} = \frac{\hat{p}^2}{2m} + U(\hat{x}) \tag{5.1}$$

where m_j is the mass of the j-th particle and $U(\mathbf{r}_1, ...\mathbf{r}_N)$ is the interaction potential in space. The path-integral formalism transforms the above Hamiltonian into a quantized Hamiltonian of polymer ring containing fictitious beads (each ring corresponds to an atom)(Eq. 2.106):

$$\phi(x_1...x_P) = \sum_{k=1}^{P} \left[\frac{1}{2} m\omega_P^2 (x_{k+1} - x_k)^2 + \frac{1}{P} U(x_k) \right] \tag{5.2}$$

In order to perform MD simulation, one needs momenta which can be obtained by recasting the prefactor in Eq. 2.105 by adding P-Gaussian integrals over the momentum variables $p_1, ..., p_P$:

$$\left(\frac{mP}{2\pi\beta\hbar^2} \right) = \int dp_1...dp_P \left(-\beta \sum_{i=1}^{P} \frac{p_i^2}{2m'} \right) \tag{5.3}$$

Here, $m' = \frac{mP}{(2\pi\hbar)^2}$ is an arbitrary mass parameter. Inserting Gaussian integrals in Eq. 2.105, one obtains the following partition function:

$$Z = \lim_{P\to\infty} \int dp_1...dp_P \int dx_1...dx_P \times \exp\left(-\beta \sum_{k=1}^{P} \left[\frac{p_k^2}{2m'} + \frac{1}{2} m\omega_P^2 (x_{k+1} - x_k)^2 + \frac{\beta\hbar}{P} U(x_k) \right] \right) \Bigg|_{x_1=x}^{x_{P+1}=x'} \tag{5.4}$$

There are no restrictions on the choice of mass $m^{'}$ since the prefactor does not influence equilibrium averages. The corresponding quantized Hamiltonian has the following form:

$$\mathcal{H}_P = \sum_{k=1}^{P} \left[\frac{[\mathbf{p}^{(k)}]^2}{2m^{'}} + \frac{1}{2} m\omega_P^2 (x_{k+1} - x_k)^2 + \frac{1}{P} U(x_k) \right] \tag{5.5}$$

where P is the Trotter number, $m_j^{'} = \frac{Pm}{(2\pi\hbar)^2}$ and $\mathbf{p}^i$ are fictitious mass and momentum respectively, $\omega_P = \frac{\sqrt{P}}{\beta\hbar}$ ($\beta = 1/k_B T$) and $U(\mathbf{r}_1^i, ..., \mathbf{r}_N^i)$ is the potential that acts between same bead index i of two different particles. The spatial oscillations/fluctuations of the polymer ring represent the quantum spatial delocalization of an atom. The statistical sampling of the individual bead trajectories allows for the calculation of statistical properties of the atomic/molecular system. The problem with the direct use of the Hamiltonian in Eq. 5.5 is that it leads to highly non-ergodic dynamics and inefficient sampling in the extended phase space of polymer rings, since the dynamics are characterized by a wide range of frequencies present in the system [107]. The time step in the simulation is limited by the high frequency modes, which in turn leads to a poor sampling of lower frequency modes. Thus, if the Hamiltonian in Eq. 5.5 is used directly in MD simulations, one needs to accumulate fairly long trajectories each starting from a different initial state. To circumvent the ergodicity problem, normal modes or staging transformation is preferred [8, 108]. The basic idea of the normal modes transformation is to decouple the harmonic spring term in Eq. 5.5, so that there remains a single frequency and the time step of the simulation can be chosen accordingly. The procedure is based on the transformation of the cartesian coordinates to the normal mode coordinates. In the matrix notation the second term on the R.H.S. in Eq. 5.5 can be written as follows:

$$V_{harmonic}(\mathbf{X}_I) = \frac{1}{2} M_I \omega_P^2 \mathbf{X}_I^T A \mathbf{X}_I \tag{5.6}$$

where A is the matrix that couples the coordinates of different beads. This matrix is diagonalised and the eigen vectors obtained are used to construct an orthogonal matrix O which diagonalises the above matrix. The normal mode coordinates can be written as:

$$\mathbf{X}^{'} = \frac{1}{\sqrt{P}} O\mathbf{X} \tag{5.7}$$

Now, the second term in Eq 5.5 can be formulated as:

$$V_{harmonic}(\mathbf{X}_I^{'}) = \frac{1}{2} M_I \omega_P^2 (\mathbf{X}^{'})_I^T \Omega \mathbf{X}_I^{'} \tag{5.8}$$

where $\Omega = PXAX^T$ is the diagonal matrix containing the eigen values of matrix A multiplied by P. The eigen values of the matrix can be easily calculated by hand:

$$\lambda_{2i-1} = \lambda_{2i-2} = 4P \sin^2 \left(\frac{\pi(i-1)}{P} \right) \tag{5.9}$$

Finally, the Hamiltonian in Eq. 5.5 in terms of normal mode coordinates can be written as:

$$\mathcal{H}_{normal} = \sum_{i=1}^{P} \left[\frac{p_i^2}{2m_i^{'}} + \frac{1}{2} m\omega_P^2 \lambda_i (x_i^{'})^2 + \frac{1}{P} U(x_i(\mathbf{x}^{'})) \right] \tag{5.10}$$

where $\mathbf{x}^{'}$ denotes the normal mode coordinates. The equations of motion in terms of normal mode variables can be written as:

$$\dot{x_i^{'}} = \frac{p_i}{m_i^{'}} \tag{5.11}$$

and,

$$m_i^{'}\ddot{x_i^{'}} = -m\omega_P^2\lambda_i x_i^{'} - \frac{1}{P}\frac{\partial U}{\partial x_i^{'}} \tag{5.12}$$

The different choice of mass in the above equations leads to different PI algorithms [118], which will be discussed in the next section. For N particles the quantized Hamiltonian is written as:

$$\mathcal{H}_P = \sum_{i=1}^{P}\left[\sum_{j=1}^{N}\frac{p_j^{(i)^2}}{2m_j^{'}} + \sum_{j=1}^{N}\frac{1}{2}m_j\omega_P^2(\mathbf{r}_j^{(i)} - \mathbf{r}_j^{(i+1)})^2 + \frac{1}{P}U(\mathbf{r}_1^i, ..., \mathbf{r}_N^i)\right] \tag{5.13}$$

In terms of normal mode variables, the N-particle Hamiltonian can be written as [107]:

$$\mathcal{H}_P = \sum_{i=1}^{P}\left[\sum_{j=1}^{N}\frac{p_j^{(i)^2}}{2m_j^{(i)'}} + \sum_{j=1}^{N}\frac{1}{2}m_j\lambda_j^i\omega_P^2(x^{'})_j^{(i)^2} + \frac{1}{P}U\left(\mathbf{r}_1^{(i)}(x_1^{'}), \ldots, \mathbf{r}_N^{(i)}(x_N^{'})\right)\right] \tag{5.14}$$

where $\frac{1}{P}U\left(r_1^{(i)}(x_1^{'}), \ldots, r_N^{(i)}(x_N^{'})\right)$ is the potential that acts between same indexed bead in two different particles.

5.2.2 Different path integral techniques

The different choice of fictitious masses in Eq. 5.12 leads to three different PI techniques [118]: path integral molecular dynamics (PIMD) [12, 73, 107, 108], ring polymer molecular dynamics (RPMD) [16, 31] and Centroid Molecular Dynamics (CMD) [13, 38]. From the conceptual point of view, the methods differ considerably, however, from the numerical point of view, the difference lies in the choice of masses and the use of thermostat. We will incorporate all these different PI techniques in GC-AdResS method. In this section, we briefly discuss the conceptual aspects of these methods and how they differ from the computational perspective. (From now on, PI-GC-AdResS will denote all the different PI techniques combined with GC-AdResS, while PIMD-GC-AdResS/RPMD-GC-AdResS/CMD-GC-AdResS will denote the specific PI technique combined with GC-AdResS)

Path Integral Molecular Dynamics

The standard Path Integral Molecular Dynamics (PIMD) method is an elegant way of computing equilibrium static properties, however, the real time dynamic information can not be calculated [107]. In the standard PIMD method, the masses $m_j^{(i)^{'}}$ are chosen such that all the internal modes have the same frequency. Thus, the choice of mass is:

$$\begin{aligned} m_j^{(i)^{'}} &= m_j\lambda_j^i, i = 2, \ldots, P \\ m_j^{1^{'}} &= m_j \end{aligned} \tag{5.15}$$

where m_j is the physical mass and λ_j^i are the eigenvalues obtained by the normal mode transformation. In PIMD, the expectation value of an observable A can be calculated using the following expression:

$$\langle \hat{A} \rangle = \frac{1}{Z} \lim_{P\to\infty} \left(\frac{mP}{2\pi\beta\hbar^2}\right)^{P/2} \int dx_1 ... dx_P \left(\frac{1}{P}\sum_{k=1}^{P} a(x_k)\right) e^{-\beta\phi(x_1,...,x_P)} \tag{5.16}$$

The observable can be measured by taking an average over the beads of the polymer ring and then subsequently taking an ensemble average. In path-integral calculations, one can approximate the equilibrium expectation values by using estimator functions that depend on P coordinates. If $a_P(x_1, ..., x_P)$ is defined as an appropriate estimator for $\langle \hat{A} \rangle$, then:

$$a_P(x_1, ..., x_P) = \frac{1}{P}\sum_{i=1}^{P} a(x_k) \tag{5.17}$$

For finite P, the expectation value $\langle \hat{A} \rangle$ can be approximated as an average of the estimator w.r.t. the probability distribution function $f(x_1, .., x_P)$:

$$\langle \hat{A} \rangle_P = \lim_{P\to\infty} \langle a_P(x_1, ..., x_P) \rangle_f \tag{5.18}$$

where $\langle ... \rangle_f$ denotes the average w.r.t. probability distribution function $f(x_1, ..., x_P)$. The expectation value can be written as:

$$\langle \hat{A} \rangle = \lim_{P\to\infty} \langle \hat{A} \rangle_P \tag{5.19}$$

Ring Polymer Molecular Dynamics

Craig and Manolopoulos [16] have developed Ring Polymer Molecular Dynamics (RPMD) which approximates the effects of quantum fluctuations on the dynamics. In RPMD, the choice of the masses is as follows:

$$m_j^{(i)'} = m_j \tag{5.20}$$

which means that the (real) physical mass of the bead is used. The original formulation of the RPMD method is a bit different from the formulation in Eq. 5.13. In the original formulation, the classical Hamiltonian in RPMD is written as:

$$\mathcal{H}_P = \sum_{i=1}^{P}\left[\sum_{j=1}^{N}\frac{[\mathbf{p}^{(i)}]_j^2}{2m_j} + \sum_{j=1}^{N}\frac{m_j}{2\beta_P^2\hbar^2}(\mathbf{r}_j^{(i)} - \mathbf{r}_j^{(i+1)})^2 + U(\mathbf{r}_1^i, ...\mathbf{r}_N^i)\right] \tag{5.21}$$

where $\beta_P = \beta/P$. This effectively translates to performing the simulation at P times the original temperature. Moreover, the harmonic bead-bead interaction and the potential energy terms are scaled by P relative to Eq 5.14. Ring Polymer Molecular Dynamics (RPMD) provides an approximation to the Kubo-transformed correlation functions [46,47] by using the classical ring polymer trajectories that are generated by the Hamiltonian in Eq. 5.21. The Kubo-transformed correlation function of the operators $\hat{A}$ and $\hat{B}$ is defined by:

$$K_{AB}(t) = \frac{1}{\beta Z}\int_0^\beta d\lambda \left[e^{-(\beta-\lambda)\hat{\mathcal{H}}}\hat{A}e^{-\lambda\hat{\mathcal{H}}}e^{i\hat{\mathcal{H}}t/\hbar}\hat{B}e^{-i\hat{\mathcal{H}}t/\hbar}\right] \tag{5.22}$$

where Z is the canonical partition function:

$$Z = tr\left[e^{-\beta\hat{\mathcal{H}}}\right] \tag{5.23}$$

The RPMD approximation is given by [31]:

$$\tilde{c}_{AB}(t) \approx \frac{1}{(2\pi\hbar)^{9PN} Z_P} \int\int d^P p_0 d^P r_0 e^{-\beta_P \mathcal{H}_P(p_0,r_0)} \frac{1}{N} \sum_{i=1}^{N} A_P^i(r_0) B_P^i(r_t) \tag{5.24}$$

where Z_P is the canonical partition function, and r_t indicates the time evolution of the positions at time t. The functions $A_P(r)$ and $B_P(r)$ are calculated by averaging over the positions of the beads in the polymer rings.

$$\begin{aligned} A_P(r) &= \frac{1}{P}\sum_{j=1}^{P} A(r_j) \\ B_P(r) &= \frac{1}{P}\sum_{j=1}^{P} B(r_j) \end{aligned} \tag{5.25}$$

There are no thermostats used in RPMD simulations since the beads in the polymer ring are treated as dynamical variables. Thus, the RPMD simulations are performed under NVE conditions where the starting configurations are either taken from massively thermostated PIMD simulations [77] or re-sampling of momenta from Maxwell-Boltzmann distribution is performed [32] after every few picoseconds.

Centroid Molecular Dynamics

Centroid Molecular Dynamics (CMD) [13] provides an approximation to real time quantum dynamics. Centroid is a quasiclassical object that is defined as an average over all the beads in a ring polymer.

$$\begin{aligned} x_c &= \frac{1}{P}\sum_{i=1}^{P} x_i \\ p_c &= \frac{1}{P}\sum_{i=1}^{P} p_i \end{aligned} \tag{5.26}$$

The normal mode transformation is the most optimal choice for performing CMD simulations, since the centroid separates out as the first normal mode coordinate defined in Eq. 5.7. The evolution of the centroid is governed by the following equations:

$$\dot{x}_c = \frac{p_c}{m} \tag{5.27}$$

and,

$$m_c \ddot{x}_c = -\frac{\partial V_o(x_c)}{\partial x_c} \tag{5.28}$$

where m_c is the physical mass and V_o is the potential of mean force generated by the dynamics of the non-centroid modes. The rigorous CMD procedure involves sampling of the phase space pertaining to the non-centroid modes at each position of the centroid. This is computationally highly expensive, and thus one uses adiabatic decoupling to separate the fictitious motion of the non-centroid modes from the physical motion of the centroid. This version of Centroid Molecular Dynamics is called Adiabatic Centroid Molecular Dynamics (ACMD) [74, 77, 118]. The adiabatic decoupling is achieved by reducing the masses of the non-centroid modes by a factor γ^2, where $0 < \gamma^2 < 1$. The effect is that the motion of the centroid is comparatively slower compared to the the non-centroid modes, which implies that the centroid moves on the potential of mean force generated "on-the-fly" by rest of the modes. Thus, the choice of masses in CMD is:

$$\begin{aligned} m_j^{(i)'} &= \gamma^2 m_j \lambda_j^i, i = 2, \ldots, P \\ m_j^{1'} &= m_j \end{aligned} \tag{5.29}$$

where γ is the adiabaticity factor. There exists another formulation of ACMD, called partially ACMD [30,38], with the only difference that larger values of γ are used in PACMD. Due to a partial separation between the non-centroid and centroid modes, PACMD is computationally less expensive than RPMD. It was shown in Ref. [38] that the dynamical properties for liquid parahydrogen were similar in both ACMD and PACMD methods. In this work, we have implemented PACMD in GC-AdResS. From now on, we will refer to PACMD as CMD. The dynamics of the non-centroid modes in CMD is artificial and is carried out to sample $V_o(x_c)$. This requires a canonical sampling over the non-centroid modes, which is achieved by coupling the internal modes to a thermostat for rapid equilibration [77, 118]. Since the dynamics of the centroid mode is Newtonian, there are no thermostats attached to the centroid. The Kubo transformed quantum time correlation between two operators $\hat{A}$ and $\hat{B}$ is approximated by [77]:

$$C_{AB}(t) = \frac{1}{Z} \int \frac{dx_c dp_c}{2\pi\hbar} A(x_c(0)) B(x_c(t)) e^{-\beta \mathcal{H}_c} \tag{5.30}$$

where $\mathcal{H}_c = p_c^2/2m + V_o(x_c)$ is the Hamiltonian governing the evolution of centroids on the potential $V_o(x_c)$ created by the internal modes. Since in GC-AdResS, one needs to calculate the thermodynamic force before starting any production run, it becomes important to investigate the technical aspects of how to merge different PI formalisms in GC-AdResS.

5.3 Path Integral Molecular Dynamics in GC-AdResS

In this section, we discuss the computational and theoretical aspects of combining various PI schemes with GC-AdResS. The section is divided into four parts: In the first part, we report the changes that were made in force-based AdResS to incorporate different PI techniques. We also discuss why the latest version of AdResS, i.e. GC-AdResS is conceptually more robust than the earlier versions, in context of PI techniques in MD simulations. In the second part, we verify "numerically" as to why one does not require a global Hamiltonian formalism for PI-GC-AdResS simulations. In the third part, we discuss how the thermodynamic force is calculated in GC-AdResS method and in the

last part, we extend the definition of Kubo-transformed correlation functions to grand-canonical ensembles and discuss how such functions are calculated in RPMD-GC-AdresS and CMD-GC-AdResS simulations.

5.3.1 Theory

The basic idea behind AdResS can be extended to incorporate PI techniques employed in MD simulations as well. In fact, there is an exact mapping between a quantum mechanical particle and a classical ring polymer, the dynamics of polymer ring is nothing else than the dynamics of classical degrees of freedom. Thus, all the tools of standard AdResS can be used in straightforward way with just one modification [81–83].

$$F_{\alpha\beta} = w(X_\alpha)w(X_\alpha)F^{PI}_{\alpha\beta} + [1 - w(X_\alpha)w(X_\alpha)]F^{cm}_{\alpha\beta} \tag{5.31}$$

Here, $F^{PI}_{\alpha\beta}$ is the intermolecular force between beads of the rings representing the atoms of molecule α and β. Figure 5.1 shows the schematic representation of PIMD-GC-AdResS, where path-integral polymer rings are interfaced with coarse-grained spheres. In the force-based AdResS approach, one cannot express the coupling between the path-integral polymer rings and the coarse-grained particles in the Hamiltonian form. However, in the previous work involving tetrahedral test system and low-temperature parahydrogen [81–83], PIMD/AdResS was able to reproduce the structural properties obtained from the reference reference PIMD simulations. Since PIMD is based on a rigorous Hamiltonian framework, and as mentioned before, the coupling between the path-integral explicit region and the coarse-grained reservoir cannot be expressed via a rigorous Hamiltonian, the procedure in Refs. [81–83] was empirical, which could only be verified *a posteriori*. The detailed discussion on this subject has already been done in Chapter 4. Here, we will only discuss those aspects which are relevant for this particular study. The reason why the procedure in Refs. [81–83] is successful is because in terms of energetic contribution, the coupling between the path-integral polymer rings and rest of the system is negligible, under the condition that the coarse-grained region is much larger than both the explicit and hybrid regions, and the explicit region is much larger than the hybrid region. However even in the "worst case" scenario, where all the regions are comparable to each other in size, we have numerically verified that the results are satisfactory. The latest version of AdResS (GC-AdResS) gives a good justification as to why such a setup of PI-GC-AdResS is robust from a conceptual point of view. According to the model of Bergmann and Lebowitz [7,52,53], in case of grand ensemble, one does not require an explicit coupling between the path-integral subregion and rest of the system. The necessary and sufficient condition is the knowledge of distribution of molecules in the reservoir i.e. the molecular density (at first order). Infact, the interaction of the molecules of the path-integral subregion with the molecules in the reservoir does not play any relevant role as far as the conceptual development of the method is concerned, but it allows for an efficient implementation of the algorithm. Such an interaction plays a role of a "capping" potential, which ensures that two molecules entering into the PI region do not overlap in space. Thus, in GC-AdResS, the Hamiltonian to be considered is the Hamiltonian of the explicit path-integral subregion; i.e. the quantized path-integral Hamiltonian embedded in a reservoir without any perturbation term. It should be noted here that while Bergmann-Lebowitz model provides a strong conceptual structure to PI-GC-AdResS, it is not a strict requirement for justifying the existence of an

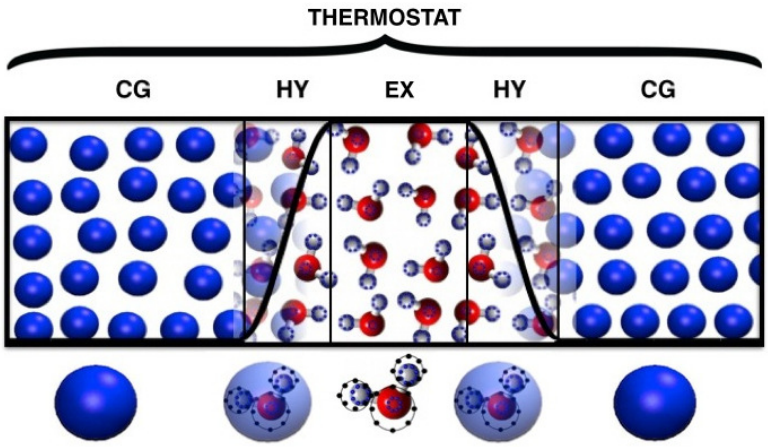

Figure 5.1: Schematic representation of the PIMD-GC-AdResS scheme; CG indicates the coarse-grained region, HY the hybrid region where the molecule changes its resolution via a space-dependent, slowly varying, function $w(x)$ and EX (or PI) is the path-integral region (that is the region of interest) that consists of path-integral polymer rings. The standard set up with the thermostat that acts globally on the whole system, used in the calculation of static properties via PIMD method.

accurate Hamiltonian in the explicit region. In the next section, we show that the interaction energy for the molecules in the explicit region is one order of magnitude larger than the interaction energy between molecules in the explicit region and rest of the system, for the systems considered in this work, which is a "worst-case" scenario.

5.3.2 Energy contribution to the coupling term

If we consider the i-th molecule in the explicit PI (EX) region (at position $\boldsymbol{r}_i$ and weighting function $w(\boldsymbol{r}_i) = 1$), then the force acting on the particle can be written down as the sum of two parts: one is the force generated by the interaction of molecule i with other molecules in the explicit region:

$$\boldsymbol{F}_{i,j} = \boldsymbol{F}_{i,j}^{PI}, \forall j \in EX \tag{5.32}$$

and the second is the force generated by the interaction of molecule i with molecules in the hybrid region (HY) and the coarse-grained region (CG).

$$\boldsymbol{F}_{i,j} = w(\boldsymbol{r}_j)\boldsymbol{F}_{i,j}^{PI} + [1 - w(\boldsymbol{r}_j)]\boldsymbol{F}_{i,j}^{CG}, \forall j \in HY + CG. \tag{5.33}$$

From Eq. 5.32, it follows:

$$\boldsymbol{F}_i = \sum_{j \neq i} \boldsymbol{F}_{i,j}^{PI} = \sum_{j \neq i} \nabla_i U_{PI}^{ij} \tag{5.34}$$

where ∇_i is the gradient w.r.t. molecule i and U_{PI}^{ij} represents the bead-bead interaction within molecules i and j. Eq. 5.33 represents the coupling force between molecules of the explicit region and rest of the system, i.e. an external force. Eq. 5.33 can also be written

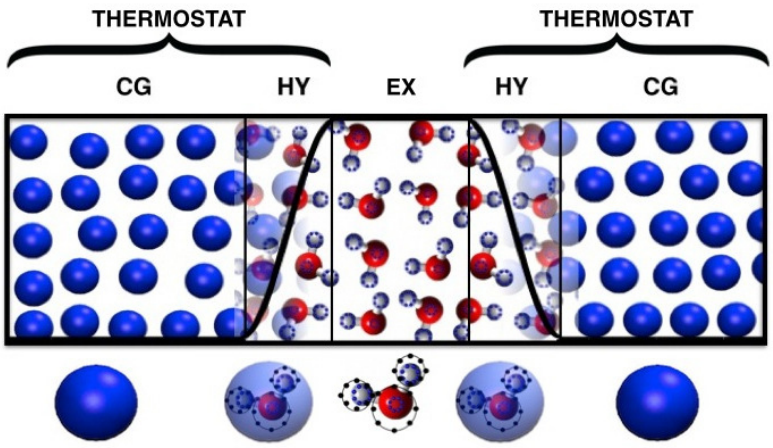

Figure 5.2: Schematic representation of the RPMD-GC-AdResS scheme; CG indicates the coarse-grained region, HY the hybrid region where the molecule changes its resolution via a space-dependent, slowly varying, function $w(x)$ and EX (or PI) is the path-integral region (that is the region of interest) that consists of path-integral polymer rings. The "local" thermostat technique employed in this work in the calculation of dynamical properties via RPMD method.

as:

$$\boldsymbol{F}_i = \sum_{j\in HY+CG} [w(\boldsymbol{r}_j)\boldsymbol{F}^{PI}_{i,j} + [1-w(\boldsymbol{r}_j)]\boldsymbol{F}^{CG}_{i,j}] = \sum_{j\in HY+CG} [w(\boldsymbol{r}_j)\nabla_i U^{ij}_{PI} + [1-w(\boldsymbol{r}_j)]\nabla_i U_{CG}]. \tag{5.35}$$

The energy of the $i-th$ molecule at a certain time t associated with force in Eq. 5.34 can be written as:

$$W_{PI-PI}(t) = \sum_{i<j} U^{ij}_{PI}; i,j \in PI \tag{5.36}$$

This corresponds to the interaction energy between molecules of the EX region. The energy of the $i-th$ molecule at a certain time t associated with force in Eq. 5.35 can be written as:

$$W^i_{PI-Res}(t) = \sum_{j\in HY+CG} [w(\boldsymbol{r}_j)U^{ij}_{PI} + [1-w(\boldsymbol{r}_j)]U^{ij}_{CG}], \tag{5.37}$$

where $Res = HY + CG$. The total energy of coupling is given by:

$$W_{PI-Res}(t) = \sum_{i\in PI} W^i_{PI-Res}(t)\,. \tag{5.38}$$

If

$$\frac{|W_{PI-PI}(t)| - |W_{PI-Res}(t)|}{|W_{PI-PI}(t)|} \approx 1; \forall t \tag{5.39}$$

then it seems reasonable to approximate the total energy of the EX region by the quantized Hamiltonian in the EX region; thus we can numerically justify the Hamiltonian formalism

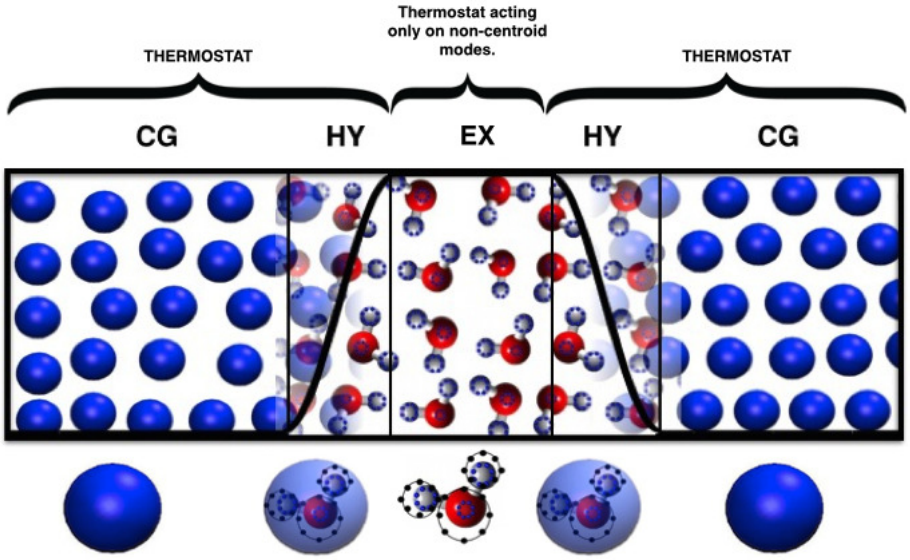

Figure 5.3: Schematic representation of the CMD-GC-AdResS scheme; CG indicates the coarse-grained region, HY the hybrid region where the molecule changes its resolution via a space-dependent, slowly varying, function $w(x)$ and EX (or PI) is the path-integral region (that is the region of interest) that consists of path-integral polymer rings. The thermostat does not act on the centroid modes in the path-integral region; this setup is used for the calculation of dynamical properties via CMD method.

in PI-GC-AdResS. Figure 5.4 shows the difference in energy is atleast an order of magnitude at every t, for two different systems considered in this work. The first system corresponds to PIMD-GC-AdResS simulation and the second system corresponds to the RPMD-GC-AdResS simulation. The CMD-GC-AdResS simulation has the same behaviour as PIMD-GC-AdResS system since they have the same thermodynamic conditions. It should be noted that we have performed simulations where the technical conditions (i.e. size of the atomistic, coarse-grained and hybrid regions) are not optimal and still the condition in Eq. 5.39 holds. The relation in Eq. 5.39 would certainly hold with the standard technical conditions as prescribed by the theory.

5.3.3 Calculation of the thermodynamic force in Path Integral-GC-AdResS

The concept of thermodynamic force has already been discussed in the previous chapters, here we will only discuss the technical details of how the thermodynamic force can be calculated in PI-GC-AdResS simulations. The thermodynamic force $F_{th}(x)$ can be expressed as:

$$F_{th}(x) = \frac{M}{\rho_o}\nabla P(x) \tag{5.40}$$

where M is the mass of the molecule and $P(x)$ is the pressure which characterizes different resolutions (for the initial configuration). We can approximate $P(x)$ in terms of linear interpolation of the molecular number density:

$$P(x) = P_{atom} + \frac{M}{\rho_o \kappa}\left[\rho_o - \rho(x)\right] \tag{5.41}$$

where ρ_o is the target density, κ is the compressibility and $\rho(x)$ is the density generated if there is no thermodynamic force used in the simulation. The thermodynamic force is then calculated using an iterative procedure:

$$F^{th}_{k+1}(x) = F^{th}_{k}(x) - \frac{M_\alpha}{[\rho_o]^2\kappa}\nabla\rho_k(x) \tag{5.42}$$

After each iteration, we obtain a density profile $\rho(x)$ which is different from the $\rho(x)$ obtained from the previous iteration. The process converges once the density is equal to the target density ρ_o. Once the equilibration steps are completed (i.e. calculation of the thermodynamic force), one can start the production run using $F_{th}(x)$ obtained after the final iteration. The calculation of thermodynamic force in PI-GC-AdResS is based on the same theoretical concepts introduced in Refs. [27, 114] for classical systems, i.e. balance of grand-potential for interfaced open boundary systems:

$$\left[P_{quantum} + \rho_o \int_\Delta F_{th}(r)dr\right] V = P_{CG}V \tag{5.43}$$

where ρ_o is the target density of the reference path-integral system. As for the classical case, $P(x)$ can be written as:

$$P(x) = P_{quantum} + \frac{M_a}{\rho_o \kappa}\left[\rho_o - \rho(x)\right] \tag{5.44}$$

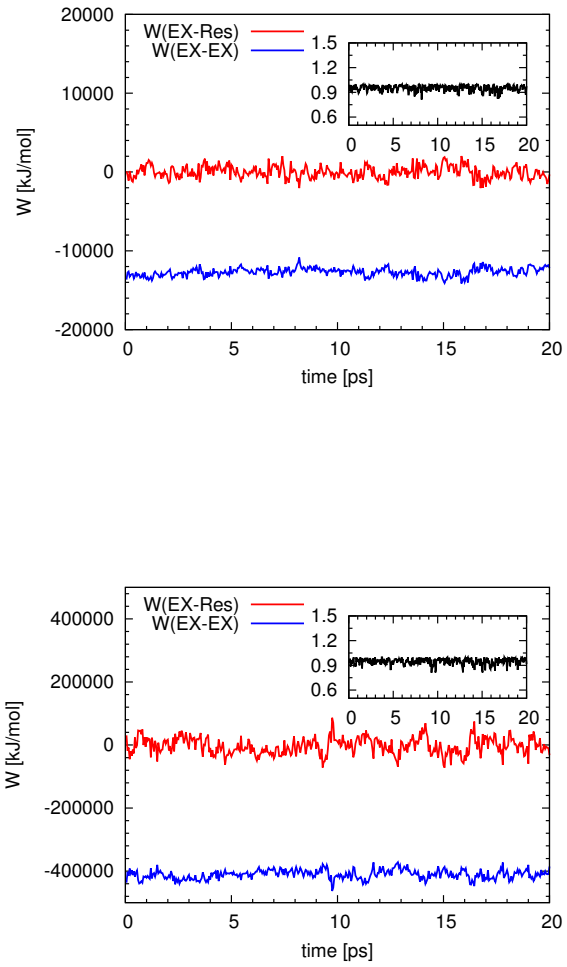

Figure 5.4: Comparison of $W_{PI-PI}(t)$ with $W_{PI-Res}(t)$. **Inset**: The relative amount of the interaction between the PI region and the rest of the system along the trajectory: $\frac{|W_{PI-PI}(t)|-|W_{PI-Res}(t)|}{|W_{PI-PI}(t)|}$; the contribution is, at most 10%. Calculations are done with PIMD-GC-AdResS simulation (top) and RPMD-GC-AdResS (bottom).

While this approach is efficient for classical system; in case of path integral, it is highly inefficient to run a PI-GC-AdResS simulation to calculate the thermodynamic force, before the actual production run, as path-integral simulations are inherently expensive. In this work, we have proposed an empirical strategy to obtain $F_{th}(x)$, which requires minimal computational effort. In case of PIMD-GC-AdResS and CMD-GC-AdResS methods where the simulation temperature is just the physical (standard) temperature, the calculation of thermodynamic force is straightforward. Since thermodynamic force calculation is based on the balance of pressure in the explicit PI and coarse-grained region, and since the thermodynamic conditions (state point) are same for the classical and path-integral systems, we expect that $F_{th}(x)$ calculated for classical systems would provide thermodynamic equilibrium for path-integral systems as well (or be very close, at least). For verification, we have calculated $F_{th}(x)$ for liquid water by considering systems with different number of beads per polymer ring ($P = 1, 4, 6, 8$ and 10, where $P = 1$ represents the classical limit). It can be seen from Figure 5.5 that the thermodynamic force remains essentially the same in all the systems. Thus, in PIMD-GC-AdResS and CMD-GC-AdResS approaches, once can use the thermodynamic force calculated from classical systems, if nuclear quantum effects on the pressure of the system are small. We use this thermodynamic force in production runs in PIMD-GC-AdResS simulations of liquid water (using $P = 32$), and find that the density in the transition region is similar to the reference density. In case of RPMD, the situation is a bit more complicated, since the simulation temperature is artificial (and the effective thermodynamic state point) and thus changes as the number of beads is changed. In this case, there will not be any option other than running a RPMD-GC-AdResS simulation with $P = 32$ and calculate the thermodynamic force before the actual production run. In this work, we have avoided such an expensive calculation by calculating $F_{th}(x)$ for systems with different number of beads ($P = 1, 4, 6, 8$ and 10) and extrapolating thermodynamic force for $P = 32$, using space dependent factors calculated from thermodynamic force for smaller values of P (Figure 5.6). We use this thermodynamic force in production runs in RPMD-GC-AdResS simulations of liquid water and find that the density of water molecules in the transition region differs at worst by 3% relative to the target density; this is a satisfactory confirmation of the validity of this procedure.

5.3.4 Equilibrium time correlation functions: Theoretical and Computational aspects

In this section, we discuss how equilibrium time correlation functions are calculated in RPMD-GC-AdResS and CMD-GC-AdResS. We also discuss the role of thermostat in GC-AdResS for the calculation of dynamical properties.

Time correlation functions in Ring Polymer Molecular Dynamics-GC-AdResS

The RPMD approximation to the Kubo-transformed correlation function of the operators $\hat{A}$ and $\hat{B}$ is given by Eq. 5.24. The equation needs to be rewritten within the formalism of grand-canonical ensemble, if one wants to calculate the correlation functions using RPMD within GC-AdResS. In grand-canonical ensemble, the Kubo-transformed correlation func-

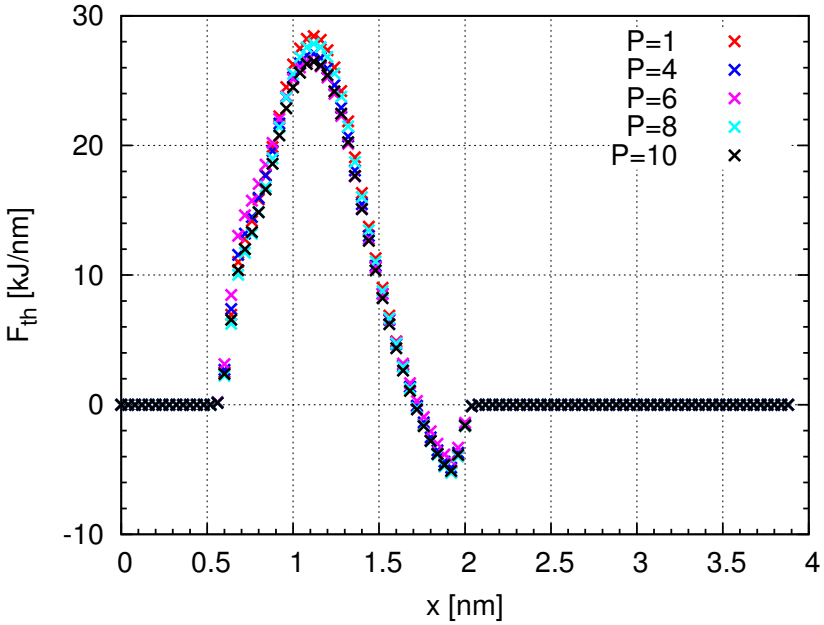

Figure 5.5: Thermodynamic force calculated in PIMD-GC-AdResS and CMD-GC-AdResS simulation for systems with different number of ring polymer beads per particle.

tion of the operators $\hat{A}$ and $\hat{B}$ is defined by:

$$\tilde{c}_{AB}(t) \approx \frac{1}{Z_P^{GC}} \sum_N \frac{1}{(2\pi\hbar)^{9PN} N!} \int\int d^P p_0(N) d^P r_0(N) e^{-\beta_P [\mathcal{H}_P(N)(p_0(N), r_0(N)) - \mu N]} \times \frac{1}{N'} \sum_{i=1}^{N'} A_P^i(r_0(N)) B_P^i(r_t(r_0(N))) \tag{5.45}$$

where μ is the chemical potential and N' is the number of molecules that stay within the path-integral subregion for the time duration considered, i.e. number of molecules at time '0' that remain correlated at time 't'; $Z_P^{GC} = \sum_N e^{\beta\mu N} Z_P$ is the grand-canonical partition function and $\mathcal{H}_P(N)$ is the Hamiltonian of the path integral subregion and N is the instantaneous number of molecules. It should be noted that in GC-AdResS, μ is calculated on-the-fly during the equilibration phase, and thus, *a priori* knowledge of μ is not required. We have used the same numerical procedure as in Chapter 4 where equilibrium time correlations were calculated in open subsystems using classical MD. This procedure is based on the description of the reservoir in Bergmann-Lebowitz model [7, 52, 53], which states that as the molecule enters into the reservoir, it looses its microscopic identity, and thus, its corresponding correlations. Therefore, if a molecule at time t_o enters into the reservoir at time t, then the contribution of this molecule outside the time duration $[t_o,\ t + t_o]$ to the correlation functions shall be neglected. In Chapter 4, it was shown that such a principle is physically consistent on the basis of results of classical MD. Since RPMD simulations are performed in NVE conditions, we use the "local-thermostat" procedure [113] in GC-AdResS so as to keep the dynamics of the beads Newtonian in the PI region. In this procedure, we use the thermostat only in the hybrid and coarse-grained

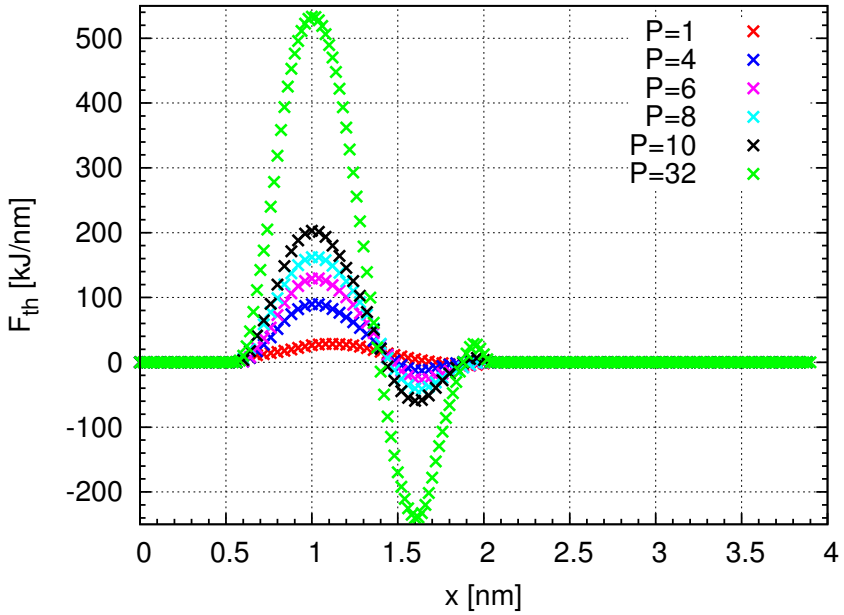

Figure 5.6: Thermodynamic force calculated in RPMD-GC-AdResS for systems with different number of ring polymer beads per particle. The force is calculated for different number of polymer ring beads. The thermodynamic force for $P = 32$ is extrapolated by using space-dependent scaling factors calculated using thermodynamic force for $P = 1, 4, 6, 8$ and 10.

regions, while no thermostats are used in the explicit PI region. This makes sure that molecules which are present in the explicit PI subregion are not subjected to the dynamical perturbation caused by the thermostat. Figure 5.2 shows a schematic representation of the RPMD-GC-AdResS system.

Time correlation functions in Centroid Molecular Dynamics-GC-AdResS

The direct extension of Eq. 5.30 to the case of an open boundary system/grand-canonical-like ensemble is:

$$C_{AB}(t) = \frac{1}{Q_{GC}} \sum_{N} \frac{1}{N!} \int \frac{dx_c^N dp_c^N}{2\pi\hbar} A(\mathbf{x}_c^N(0)) B(\mathbf{x}_c^{N'}(t)) e^{-\beta[\mathcal{H}_c^N(x_c^N, p_c^N) - \mu N]} \tag{5.46}$$

where Q_{GC} is the grand-canonical partition function, μ is the chemical potential, N is the number of path centroids (which is now a variable number in the system) and $\mathcal{H}_c^N$ is the Hamiltonian governing the evolution of centroids with N, instantaneous number of path centroids. The number of centroids N' at time $t = t' > 0$ is likely to be different from the number of centroids N at time $t = 0$. We use the correspondence between the Bergmann and Lebowitz model [7, 52, 53] and GC-AdResS to interpret the quantity $B(x_c^N(t))$ in this context. For the calculation of equilibrium time correlation functions, we have used the same procedure as described in the previous section. As discussed in Section 5.2.2, the centroids are not coupled to any thermostat. In the context of AdResS, this would simply

translate to having a thermostat in the coarse-grained and hybrid regions, while in explicit region, no thermostats are attached to the centroid mode, while non-centroid modes move under the action of the thermostat. Figure 5.3 shows the CMD-GC-AdResS system and the application of the thermostat in different regions.

5.4 Numerical Results

In this section, we report the technical details and the comparison between results obtained from reference PI and PI-GC-AdResS simulations. The section is divided into four parts, the first part reports the technical setup used to calculate the structural and dynamical properties in PI simulations; the second part reports issues regarding employment of a Langevin Thermostat in PIMD simulations; and the third part report results of static (i.e molecular number density across the system, radial distribution functions and the particle number probability distribution) and dynamic (equilibrium time correlation functions) properties calculated using PIMD/RPMD/CMD GC-AdResS simulations. It must be noted that all the properties are calculated only within the explicit PI subregion in PI-GC-AdResS simulations and in an equivalent subregion in reference PI simulations. The subsystem of a large reference PI simulation box is a natural grand-canonical ensemble; thus, if the subsystem of AdResS reproduces the results of reference PI subsystem, then we can state that the PI subregion in AdResS samples the phase space according to grand-canonical distribution.

5.4.1 Technical Setup

All path integral simulations are performed by home-modified GROMACS [93] and the thermodynamic force in GC-AdResS simulations is calculated using VOTCA [98] The quantum model for liquid water used in this work is q-SPC/FW [72]. It was shown that the thermodynamic and dynamical properties calculated using this water model agree quite well with the experiment data. The number of water molecules in system are 1320, and the box dimensions are $5.8 \times 2.6 \times 2.6\ nm^3$, corresponding to a density 990 kgm^{-3}. In AdResS simulations, the resolution of the molecules changes along x-axis, as depicted in Figure 5.1. The transition region, which has dimensions $2.6 \times 2.6 \times 2.6\ nm^3$ is fixed in all the simulations. In the coarse-grained region, we use a generic WCA potential given by Eq. 3.60. The values of σ and ϵ are taken to be 0.30 nm and 0.65 kJ/mol respectively. We use 32 ring polymer beads in all the simulations, which is sufficient to obtain converged results for both static and dynamical properties. Reaction field method is used to compute the electrostatic properties with dielectric constant for water equal to 80. These technical details are common to all the different PI techniques (PIMD, RPMD, CMD) studied in this work. In the following sections, we report the technical details that are specific to each of these techniques.

Path Integral Molecular Dynamics

We use three different sizes of explicit region ($PI = 0.5, 1.2$ and 2.4) in GC-AdResS simulation to demonstrate the computational and conceptual robustness of the method. The smallest size represents the limiting case of statistically relevant number of molecules in

the path-integral subregion while the largest size represents the limiting case of the size of the reservoir which may not fulfil the condition of being statistically large enough. All the static properties are computed from 250 ps long trajectories. The simulations are performed at 298 K. The time step used in PIMD simulations is 0.1 fs.

Ring Polymer Molecular Dynamics

We perform RPMD simulations with starting configurations taken every 8 ps along PIMD trajectories. A total of 25 trajectories each of length 25 ps are generated. For the first 5 ps, we keep the thermostat switched on, in order to adjust the velocities as masses are different in PIMD and RPMD methods. We have followed the procedure proposed in Ref. [77]. After this initial equilibration run, the thermostat is switched off, and the NVE trajectories generated are used to compute various time correlation functions. We use the same strategy for RPMD-GC-AdResS simulations; the only difference is that the initial configurations are now taken every 8 ps from PIMD-GC-AdResS trajectories. For the first 5 ps, the thermostat acts in the explicit as well as the hybrid and coarse-grained regions. After the short equilibration run, the thermostat is switched off in the explicit region, while the hybrid and coarse-grained region are kept under the action of the thermostat for the production run. The RPMD simulations are performed at temperature $298 \times 32 = 9536\ K$ A time step of 0.1 fs has been used in all RPMD simulations. We calculate the velocity auto correlation functions and first and second order orientational correlation functions. The explicit formulae for RPMD approximation to Kubo-transformed correlation functions are given in Ref. [65]. We also calculate the diffusion constant and the l^{th} order relaxation times. The diffusion coefficient is obtained from the time integral of the velocity auto-correlation function:

$$D = \frac{1}{3}\int_0^\infty C_{vv}(t)dt \tag{5.47}$$

The l^{th} order relaxation times are obtained from the time integrals of the corresponding orientational correlation functions:

$$\tau_l^n = \int_0^\infty C_l^n dt \tag{5.48}$$

The time correlation functions are calculated in the explicit region in the last 20 ps, i.e. excluding the equilibration run. The correlation functions are calculated up to 10 ps by averaging over the 25 trajectories. Since the orientational correlation functions are calculated for 10 ps, an exponential tail was fitted to the correlation functions for computing the integral.

Centroid Molecular Dynamics

An adiabaticity parameter of $\gamma^2 = P^{-(P+1)/(P-1)}$ [30] is used, and a time step of 0.1 fs is found to be sufficient for this adiabaticity parameter. The starting configurations are taken every 8 ps along a PIMD trajectory. We initiate 25 ps long CMD trajectories from the saved configurations. All the modes are coupled to a thermostat for first 5 ps. After this initial warm up run, we decouple the centroid mode from the thermostat while non-centroid modes move under the action of thermostat. In the CMD-GC-AdResS simulations, this translates to having a thermostat coupled to the whole system for the first 5 ps, following

which the thermostat is coupled to all the modes in coarse-grained and hybrid regions and only non-centroid modes in the explicit path-integral region. The time correlation functions are calculated in the explicit region in the last 20 ps. The functions are calculated up to 10 ps by averaging over the 25 trajectories. The explicit formulae for the correlation functions can be found in in Ref. [72].

5.4.2 Thermostat issue in Path Integral Molecular Dynamics simulations

Massive thermostating is required in the PIMD simulations since the forces arising due to the high frequencies in the polymer ring and the forces due to the potential $U(x)$ are weakly coupled [107]. Tuckerman et al. [108] coupled each normal mode variable to separate Nose-Hoover chains, thereby ensuring proper ergodic sampling of the phase space. Manolopoulos et. al [14] developed specific Langevin equations for thermostat, that are tuned to sample all the internal modes of the ring polymer quite efficiently. However, in this work we choose the standard Langevin equations of thermostat with time scale 0.1 ps, which is strong enough for sampling the phase space effectively, though it may not be be the most efficient choice from computational point of view. The reason is that in the initial stage of validating PIMD-GC-AdResS, we need to show that the properties obtained in the reference PIMD simulations are reproduced exactly in AdResS. Since we use the same thermostat conditions in both the simulations, there should not be any discrepancy arising due to the thermostat. However, we have verified that the static properties calculated in our reference PIMD simulation agree very well with those available in literature.

5.4.3 Results

Path Integral Molecular Dynamics

We compare molecular number density, quantum (bead-bead) radial distribution functions and the particle number probability distribution in reference PIMD and PIMD-GC-AdResS simulations. Figure 5.7 shows the molecular density calculated over the whole box. It can be seen that the agreement is highly satisfactory and that the fluctuations are of the same order in PIMD-GC-AdResS simulations as in reference PIMD simulations. This was an important test to verify that the subregion samples the correct grand-canonical distribution following the mathematical formulation in Ref. [111]. For further verification, we have calculated the particle number probability distribution functions in the PI subregion in GC-AdResS and in an equivalent subregion in reference PIMD simulations, which are shown in Figure 5.8. It can be seen that GC-AdResS and reference PIMD distributions agree quite well with each other and the shape of both the curves is Gaussian, as one would expect. This again highlights the fact that AdResS samples the phase space of the subregion in a correct way. However, it must be noted that there is a systematic shift in the probability distribution functions by two-three particles. One explanation is that in PIMD-GC-AdResS, the density is not exactly flat across the box, since we have used an empirical approach to calculate the thermodynamic force instead of the standard one, as explained in Section 5.3.3. Such a shift seems to correlate with the width of the distribution, i.e. extension of the PI region. An explanation of the effect is that in AdResS for larger PI regions the frequency in space of small density fluctuations is larger than for smaller PI regions. The discrepancy is, however, numerically negligible which does not

affect the quantum (or bead-bead) radial distribution functions. The radial distribution function (g(r)) represents two-body correlation function which is one order higher than the molecular density. It differs considerably when quantum models for liquid water are used [32, 72]. We have calculated the quantum g(r)'s in the PI subregion in GC-AdResS and in an equivalent subregion in reference PIMD simulations. Figure 5.9 to Figure 5.11 show that that GC-AdResS is able to exactly reproduce the results obtained from reference PIMD simulations.

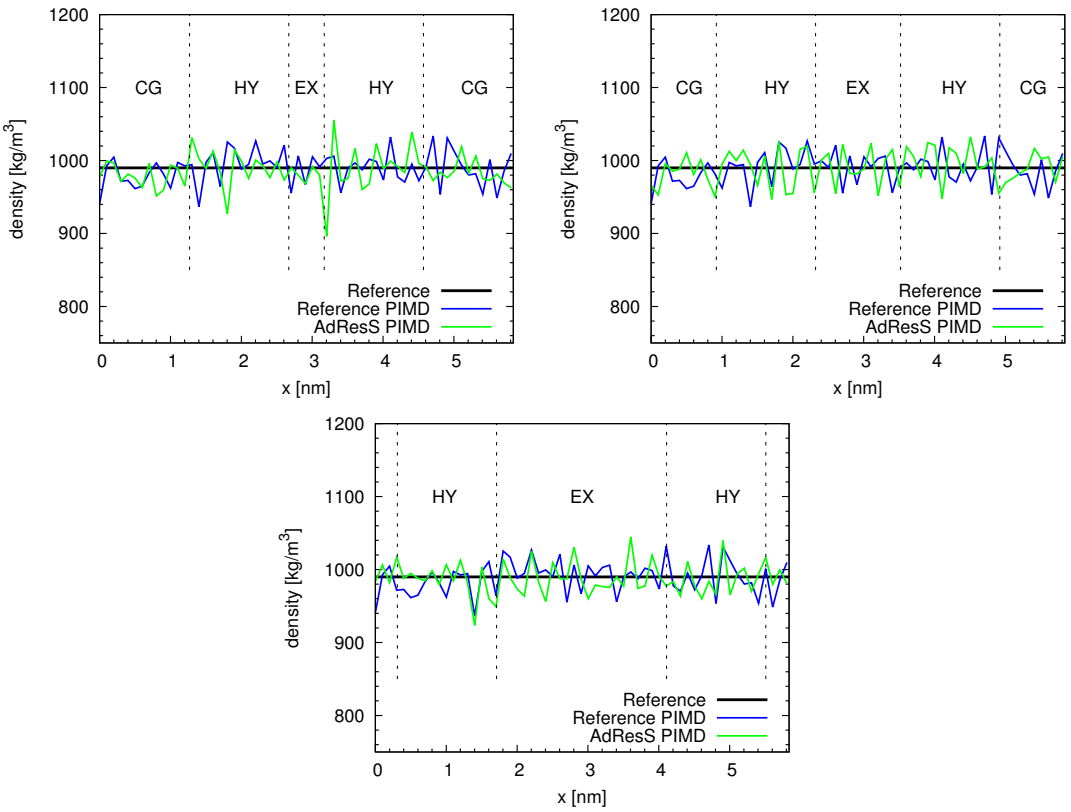

Figure 5.7: Molecular number density calculated with PIMD-GC-AdResS for different size of quantum subregion. Results are compared with the density obtained in a reference path integral simulation.

Ring Polymer Molecular Dynamics

We calculate the static properties (molecular number density, radial distribution functions) using RPMD and compare them with the properties calculated in RPMD-GC-AdResS simulations for the most relevant case, $PI = 1.2\,nm$. It can be seen from figure 5.12 that the density in the hybrid region differs by around 3-5%, as compared to the reference density. As discussed in Section 5.3.3, this discrepancy occurs due to the empirical calculation of the thermodynamic force which does not ensure a perfect flat density across the box. However, the error is negligible and it does not influence the quantum radial distributions (shown in Figure 5.14) at all. Figure 5.13 shows the number of molecules that remain

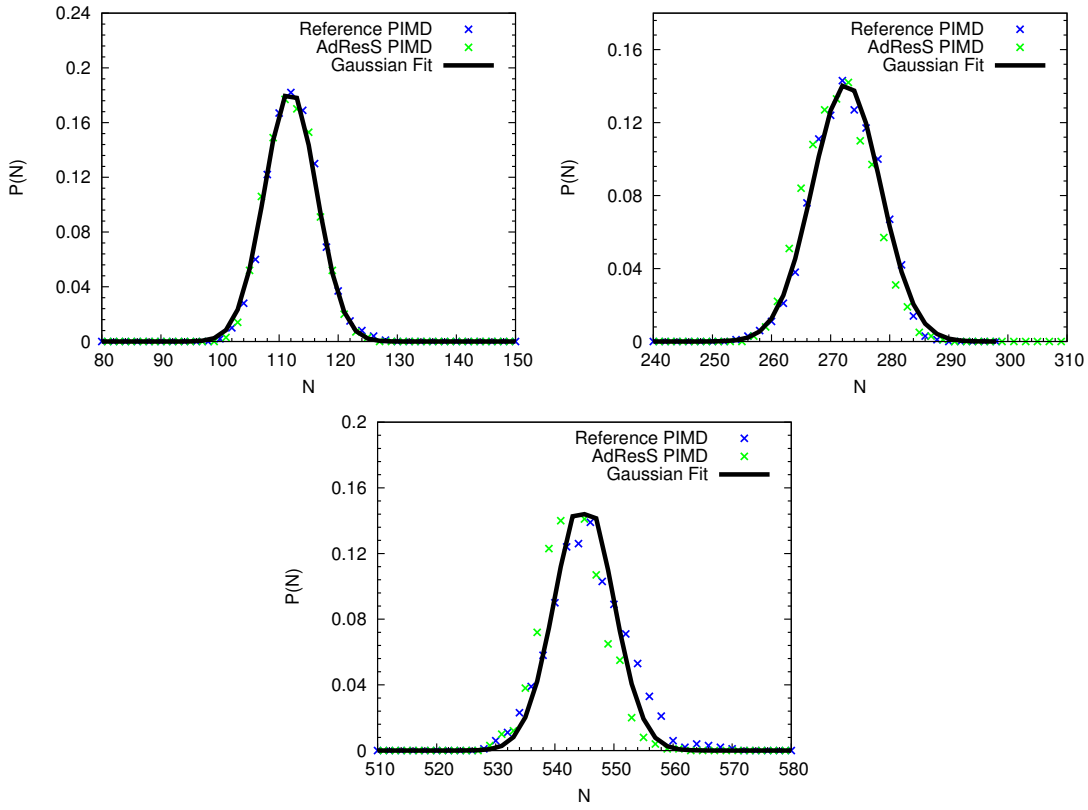

Figure 5.8: Particle number probability distribution functions calculated in the explicit PI subregion in PIMD-GC-AdResS compared with the functions calculated in an equivalent subregion in reference PIMD simulation, for different size of quantum subregion. The shape of both curves is a Gaussian (reference black continuous curve) in all the three different simulations.

within the explicit path-integral region 'ν' as a function of time. This is calculated as following: we label all the molecules in the trajectory at time '0' and calculate how many of those labeled molecules are present at time 't'. This is an important quantity that describes the movement of the molecules in and out of the explicit path-integral region. This quantity is calculated in RPMD-GC-AdResS and reference RPMD simulations. It can be seen that the two curves overlap. This result confirms, once again, that GC-AdResS subregion has the same average dynamical behaviour as the reference RPMD subregion. This shows that RPMD-GC-AdResS is able to sample basic properties of a grand-canonical ensemble. We also calculate the velocity auto-correlation and the first and second order orientational correlation (dipole moment axis is chosen as the inertial axis of the molecule) functions [65] in RPMD-GC-AdResS and reference RPMD simulations. From a physical point of view, it must be clarified that the functions calculated in subregion are local in space and time if compared to the functions calculated over the whole box. Figure 5.15 show the three correlation functions calculated in the explicit region and Table 5.2 reports the local diffusion constant (D ($\text{Å}^2 ps^{-1}$)) and l^{th} order relaxation times (τ_1^{μ} (ps) and τ_2^{μ} (ps)). It can be seen that the local diffusion constant is same in RPMD-GC-AdResS and

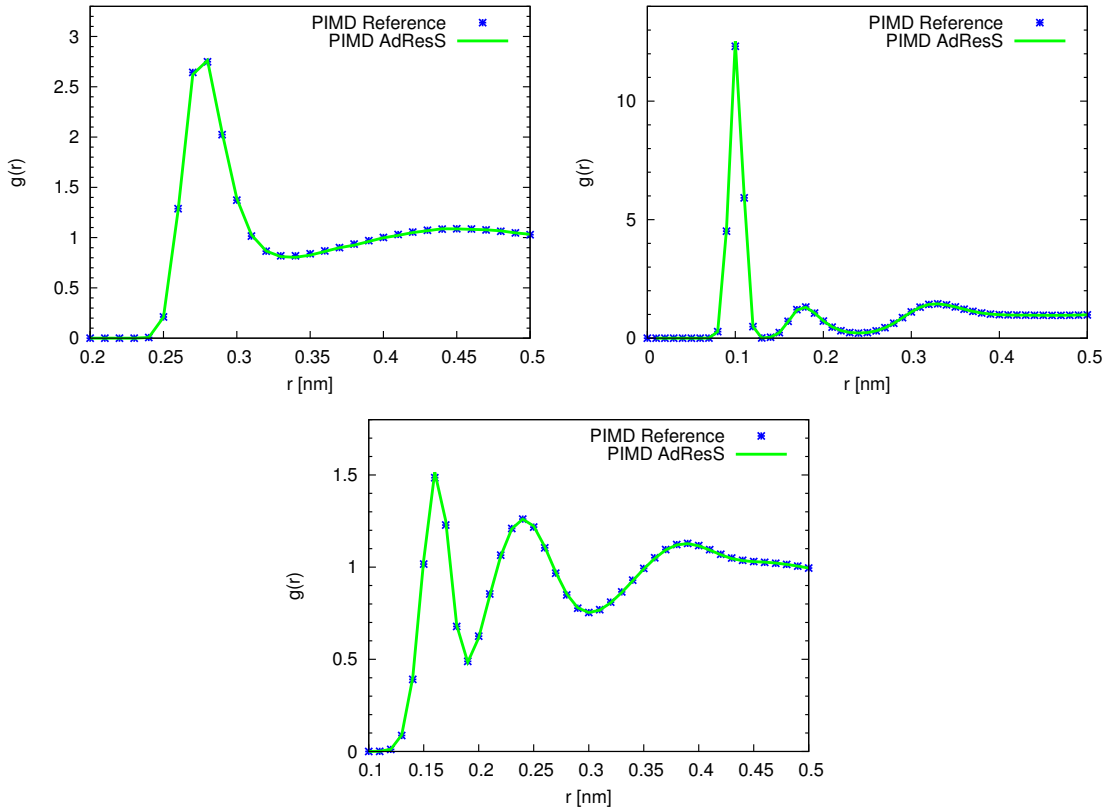

Figure 5.9: From left to right: quantum (bead-bead) oxygen-oxygen, oxygen-hydrogen and hydrogen-hydrogen radial distribution functions calculated with PIMD-GC-AdResS compared with the results obtained for an equivalent subsystem ($PI = 0.5nm$) in a reference path integral simulation.

reference RPMD simulations, while there is a discrepancy of 10% in the 1^{st} order relaxation time. The difference is not significant, however it must be reported. Thus, GC-AdResS can be considered a robust computational method to calculate quantum-based static and dynamic properties of liquid water, and as a result, for systems where water plays a major role.

Centroid Molecular Dynamics

We calculate the static properties (molecular number density, radial distribution functions) using CMD and compare them with the properties calculated in CMD-GC-AdResS simulations where the size of the explicit region is $PI = 1.2\ nm$. Figure 5.16 shows the centroid density for liquid water. The agreement between the reference results and AdResS results is highly satisfactory. We also calculate the number of molecules that stay within the explicit path-integral region 'ν' as a function of time, as described in the previous section. We calculate this property in the explicit region in CMD-GC-AdResS and an equivalent subregion in reference CMD simulations. The agreement is highly satisfactory. This result indirectly shows the Grand Canonical-like character of GC-AdResS. Figure 5.18 shows the

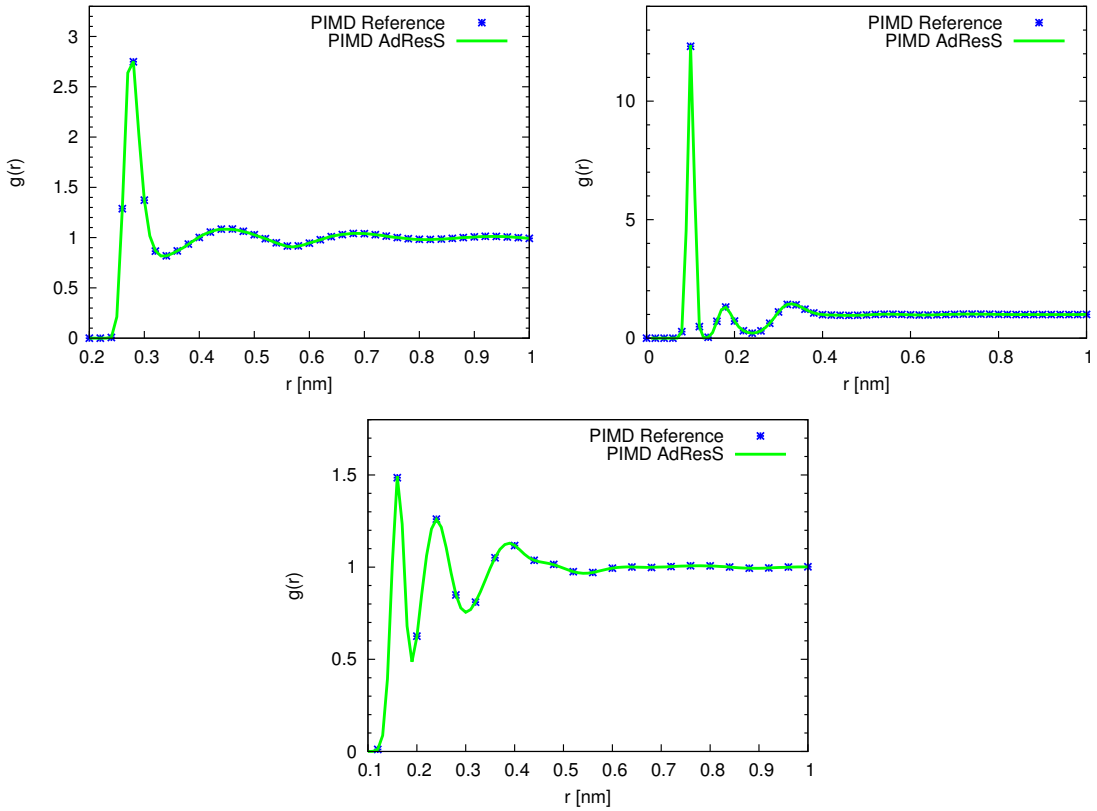

Figure 5.10: From left to right: quantum (bead-bead) oxygen-oxygen, oxygen-hydrogen and hydrogen-hydrogen radial distribution functions calculated with PIMD-GC-AdResS compared with the results obtained for an equivalent subsystem ($PI = 1.2nm$) in a reference path integral simulation.

centroid RDF's calculated in the explicit region in CMD-GC-AdResS and an equivalent region in reference CMD simulations. It should be noted that centroid RDF's are not same as the quantum (bead-bead) RDF's and a deconvolution procedure [9] is used to convert centroid RDF's to the actual quantum RDF's. However, it is an important numerical quantity to show that the explicit path-integral region in AdResS reproduces detailed structural properties in a proper way. It can be seen that the curves overlap which shows the robustness of the method in reproducing the two-body distribution functions. Figure 5.19 shows the velocity auto-correlation function, first and second order orientational correlation functions (dipole moment axis is chosen as the inertial axis of the molecule) [72] calculated in the explicit region in CMD-GC-AdResS and an equivalent region in reference CMD simulations and Table 5.2 reports the the local diffusion constant (D ($\text{Å}^2 ps^{-1}$)) and l^{th} order relaxation times (τ_1^{μ} (ps) and τ_2^{μ} (ps)). It can be seen that the local diffusion constant is same in both CMD-GC-AdResS and reference CMD results, while there is some discrepancy in the 1^{st} order relaxation time, as observed with RPMD simulations. We also

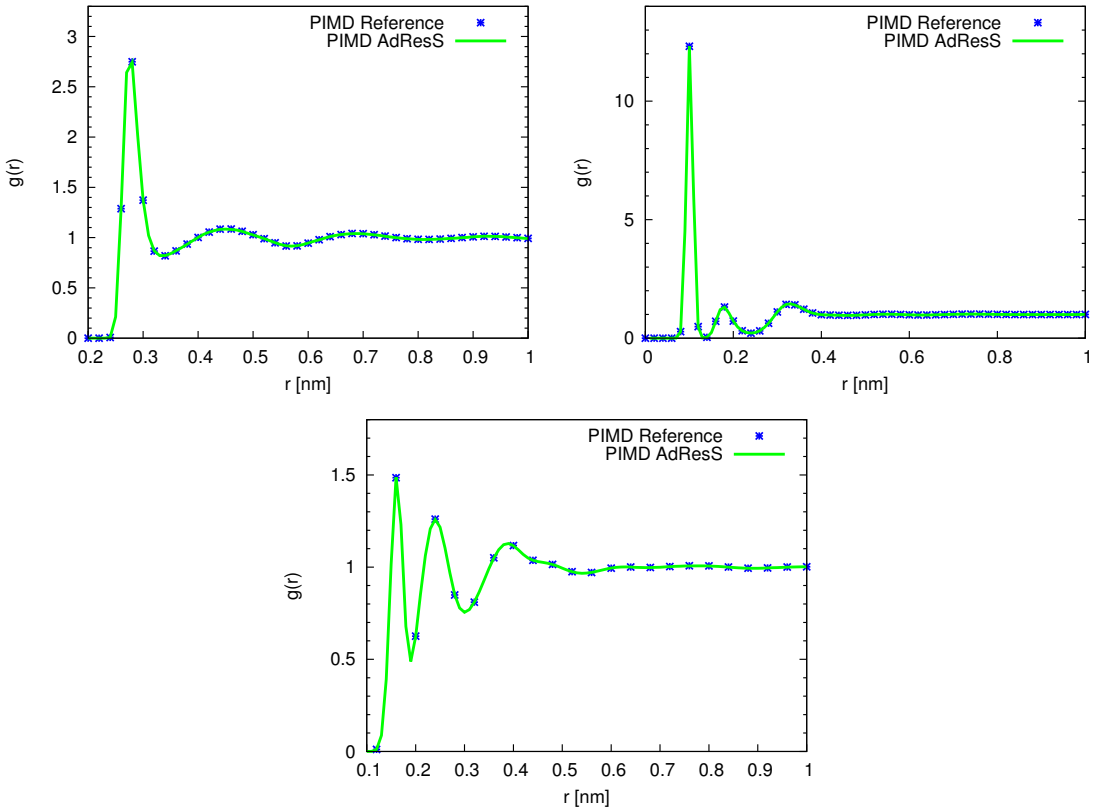

Figure 5.11: From left to right: quantum (bead-bead) oxygen-oxygen, oxygen-hydrogen and hydrogen-hydrogen radial distribution functions calculated with PIMD-GC-AdResS compared with the results obtained for an equivalent subsystem ($PI = 2.4nm$) in a reference path integral simulation.

calculate the infrared absorption coefficient $\alpha(\omega)$ using the following relation [72]:

$$\alpha(\omega) = \left[\frac{4\pi^2\omega}{3V\hbar cn(\omega)}\right](1 - e^{-\beta\hbar\omega})\frac{1}{2\pi} \times \int_{-\infty}^{+\infty} e^{-i\omega t}\langle M(0)M(t)\rangle dt, \tag{5.49}$$

where $\langle M(0)M(t)\rangle$ is the total dipole moment auto-correlation function, c is the speed of light, V is the volume of the box and $n(\omega)$ is the refractive index of the system at frequency ω. Figure 5.20 shows the infrared spectrum calculated in the path-integral subregion in CMD-GC-AdResS and reference CMD simulations. The agreement is remarkable, and strongly supports the numerical and conceptual solidity of the method since the spectrum is a quantity of primary importance also from an experimental point of view.

In general, it should be pointed out that the current PI-GC-AdResS simulations are not performed under optimal conditions, i.e. a very large reservoir and (ideally) a relatively small hybrid region. The computational set up employed in this work represents a "worst case scenario" that tests the technical frontiers of the method; it is natural to expect that when theoretical conditions are fully met then the level of accuracy can only rise (as shown in Chapter 4). However, already under non optimal conditions the results are highly satisfactory.

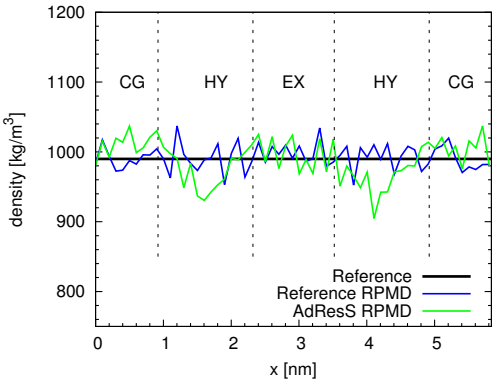

Figure 5.12: Molecular number density calculated with RPMD-GC-AdResS for different size of quantum subregion. Results are compared with the density obtained in a reference path integral simulation.

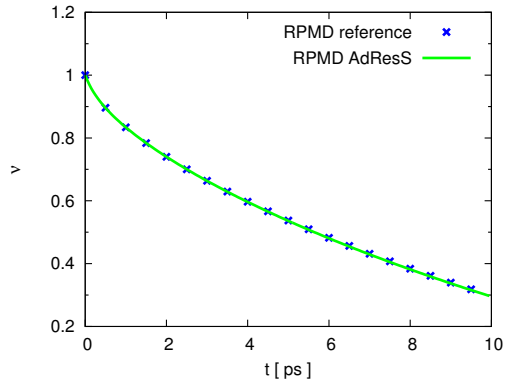

Figure 5.13: Number of molecules that remain within the explicit path-integral region ν as a function of time. This quantity is calculated in reference RPMD and RPMD-GC-AdResS simulations.

Table 5.1: Local diffusion constant and l^{th} order relaxation times for liquid water calculated in explicit region of RPMD-GC-AdResS and an equivalent subregion in the reference RPMD simulations.

Parameter	Reference RPMD	AdResS RPMD
D (Å$^2 ps^{-1}$)	0.30	0.30
τ_1^μ (ps)	3.9	3.6
τ_2^μ (ps)	1.2	1.1

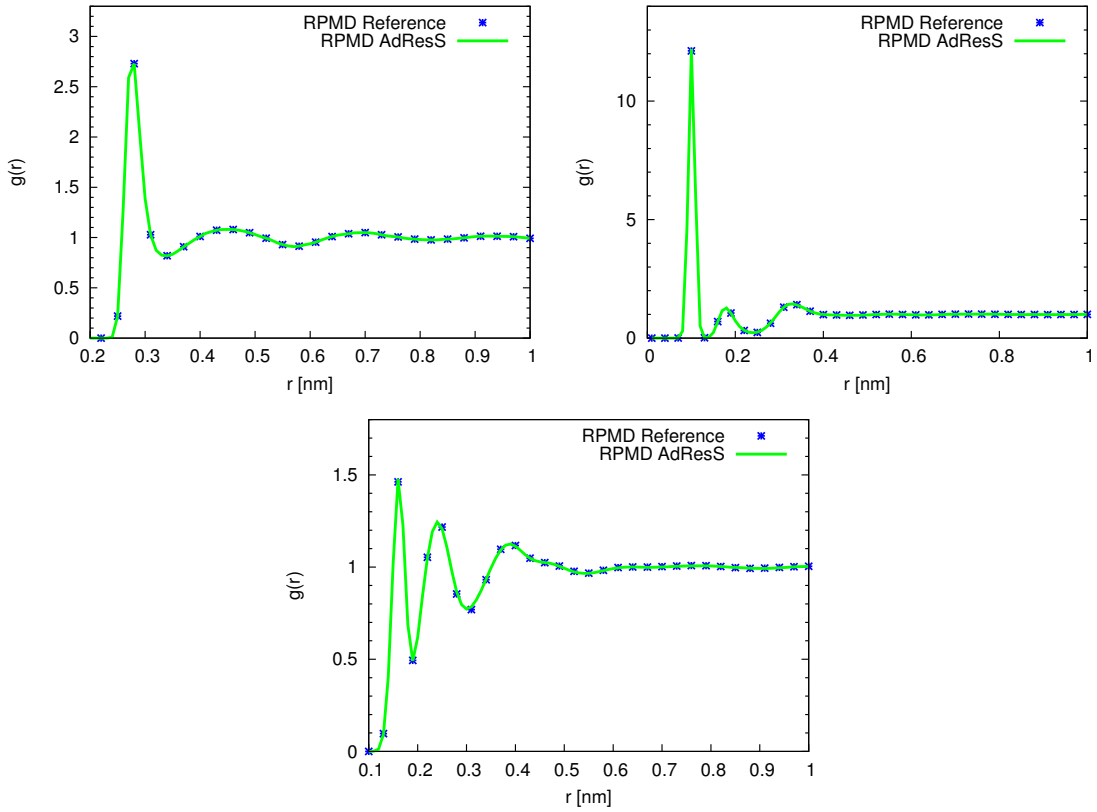

Figure 5.14: From left to right: quantum (bead-bead) oxygen-oxygen, oxygen-hydrogen and hydrogen-hydrogen radial distribution functions calculated with RPMD-GC-AdResS compared with the results obtained for an equivalent subsystem ($PI = 1.2nm$) in a reference RPMD simulation.

Table 5.2: Local diffusion constant and l^{th} order relaxation times for liquid water calculated in explicit region of CMD-GC-AdResS and an equivalent subregion in the reference CMD simulations.

Parameter	Reference CMD	AdResS CMD
D ($\text{Å}^2 ps^{-1}$)	0.32	0.32
τ_1^{μ} (ps)	4.0	3.7
τ_2^{μ} (ps)	1.3	1.2

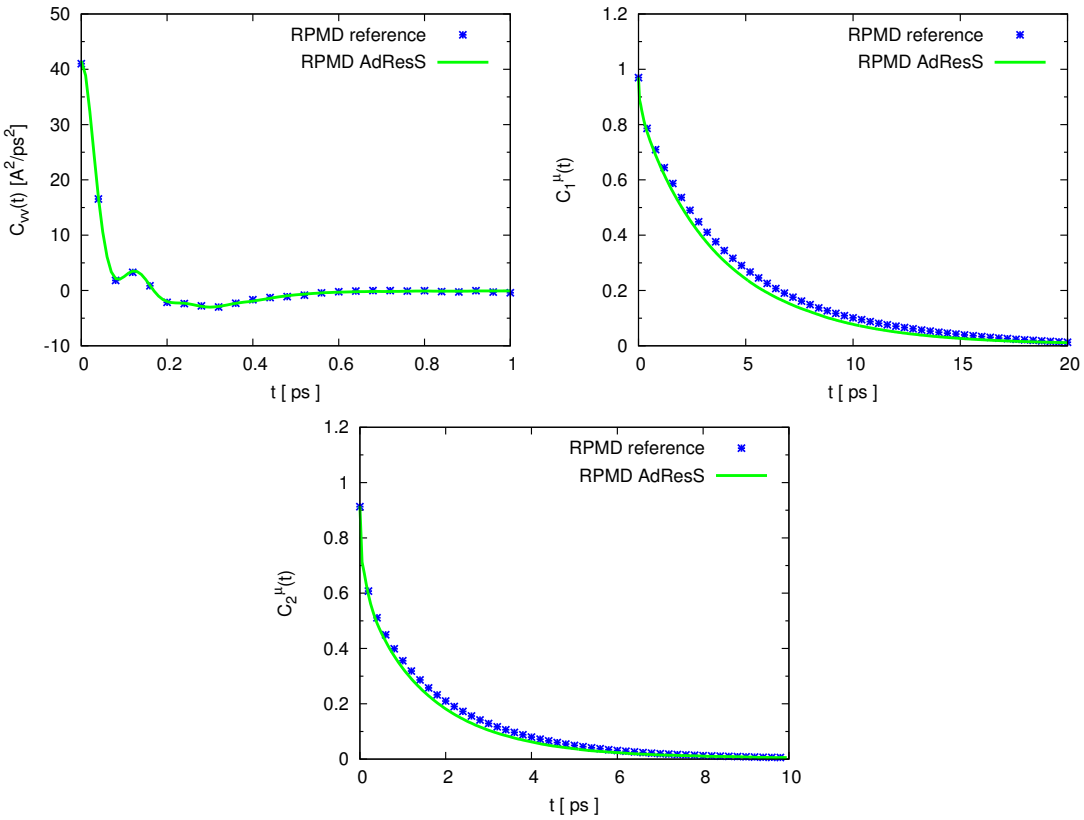

Figure 5.15: Kubo-transformed equilibrium time correlation functions calculated in explicit PI region of RPMD-GC-AdResS and an equivalent subregion in reference RPMD simulation. For the first order correlation function, an exponential tail has been fitted beyond 10 ps.

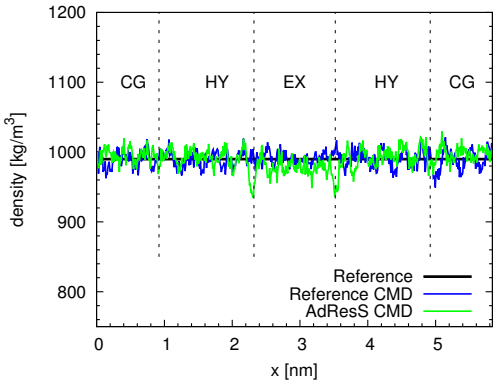

Figure 5.16: Centroid density in the explicit path-integral region in reference CMD and CMD-GC-AdResS simulations for liquid water.

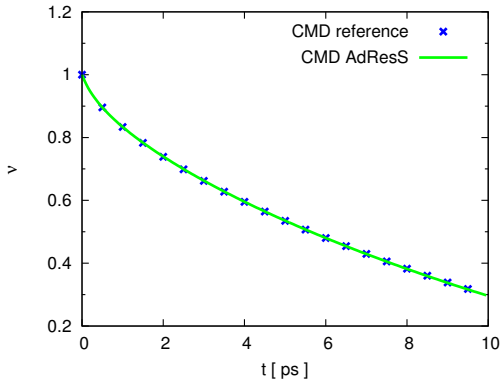

Figure 5.17: Number of molecules that remain within the explicit path-integral region ν as a function of time. This quantity is calculated in reference CMD and CMD-GC-AdResS simulations.

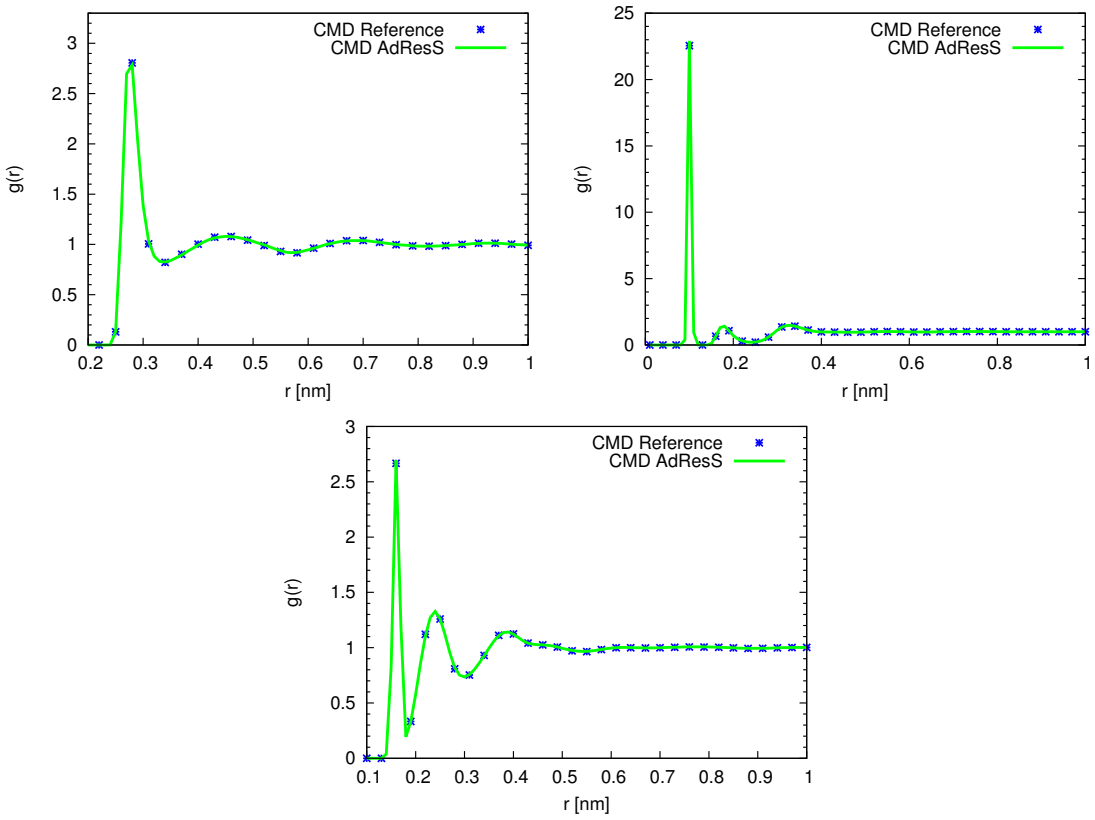

Figure 5.18: From left to right: centroid oxygen-oxygen, oxygen-hydrogen and hydrogen-hydrogen radial distribution functions calculated with CMD-GC-AdResS compared with the results obtained for an equivalent subsystem ($PI = 1.2nm$) in a reference CMD simulation.

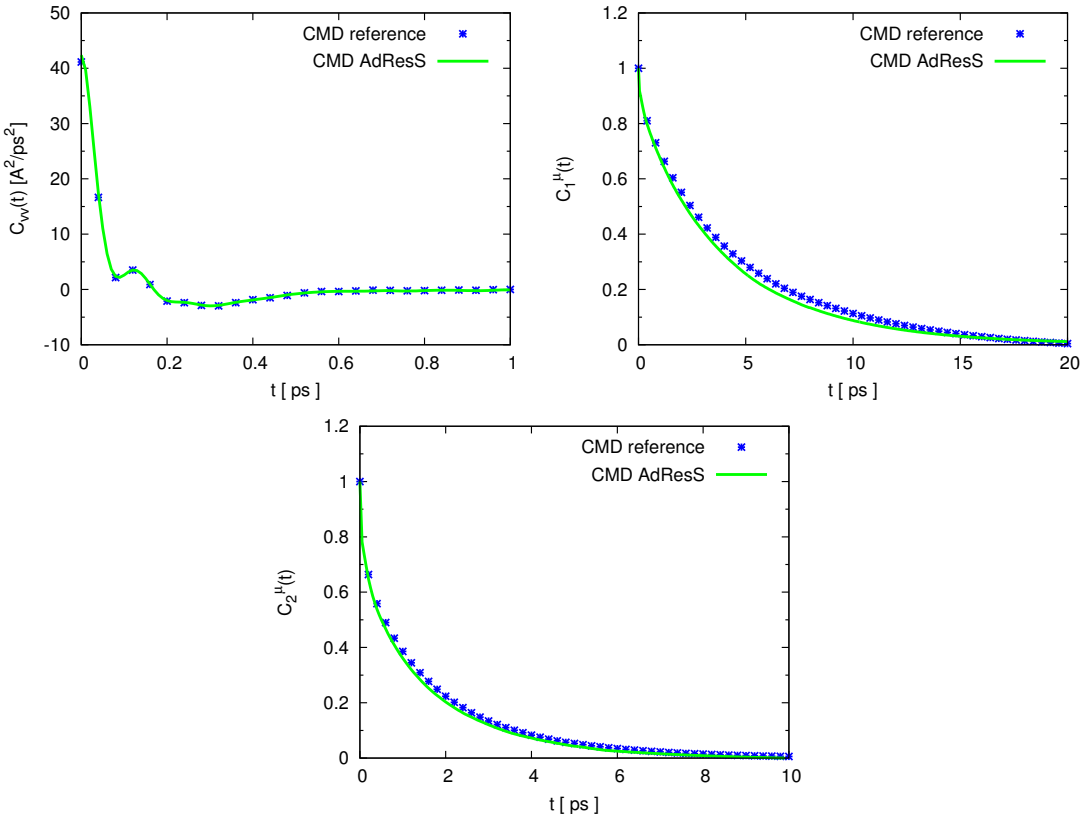

Figure 5.19: From left to right: centroid (bead-bead) velocity auto-correlation function, first order orientational correlation function and second order orientational correlation function calculated with CMD-GC-AdResS compared with the results obtained for an equivalent subsystem ($EX = 1.2nm$) in a reference CMD simulation. For the first order correlation function, an exponential tail has been fitted beyond 10 ps.

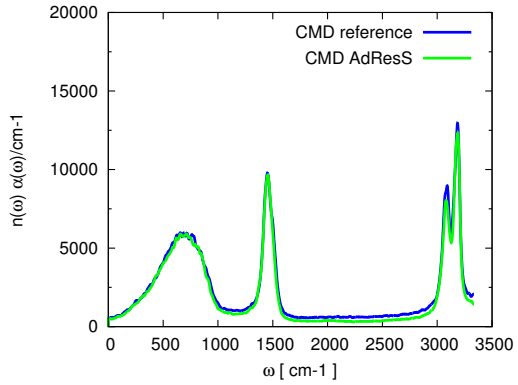

Figure 5.20: Infrared spectrum for liquid water at 298K calculated in explicit region of CMD-GC-AdResS and an equivalent subregion in the reference CMD simulations.

5.5 Conclusions

We have implemented different PI schemes in the grand-canonical version of AdResS, i.e. GC-AdResS, where the (open) region of interest treated at PI resolution is in contact with a generic reservoir of energy and particles. We have reported the theoretical and computational aspects of combining the three methods. We have performed simulations of liquid water using all the different formalisms of PI. The results regarding the static and dynamic properties are highly satisfactory and qualify PI-GC-AdResS as a robust method for "large-scale" PI simulations which are currently prohibitive.

Chapter 6

Summary and Outlook

6.1 Summary

In this thesis, grand-canonical adaptive resolution method (GC-AdResS) was extended to incorporate different path-integral (PI) techniques, which are used to model systems where nuclear quantum effects have considerable role. Prior to such an implementation, it was validated that GC-AdResS is a robust method for the calculation of equilibrium structural and thermodynamic properties for classical MD systems. In Chapter 3, it was shown that in GC-AdResS, the excess chemical potential is automatically calculated during the equilibration phase, i.e. the calculation of thermodynamic force. The excess chemical potential was calculated for a wide variety of liquids and mixtures, and the results agree quite well with the values obtained using thermodynamic integration. It was shown that there is no need for specific coarse-grained models and one can obtain very accurate results with a generic Weeks-Chandler-Andersen (WCA) potential in the coarse-grained region. In Chapter 4, the correspondence between the Bergmann-Lebowitz (BL) model for grand-ensemble and GC-AdResS was made. The BL model not only gave a theoretical justification to the simulation of open boundary systems, but it also provided a theoretical definition of equilibrium time correlation functions in open boundary systems. The simulations of liquid water were performed using GC-AdResS and equilibrium time correlation functions were calculated in the explicit PI region. The results were compared with the results from reference classical MD simulations. It was shown that GC-AdresS simulations perfectly reproduces the results of the reference full-atomistic simulations. Finally, in Chapter 5, three different path-integral (PI) techniques employed in MD simulations were discussed. The technical aspects of the implementation of these techniques in GC-AdResS were reported in detail. Since the PI formalism maps an atom onto a polymer ring containing classical fictitious beads, the concepts developed in Chapters 3 and 4 for classical MD were applied straightforwardly. The static and dynamical properties of liquid water were calculated in PI-GC-AdResS simulations and reference PI simulations. It was shown that the results of PI-GC-AdResS simulations are fully consistent with results from reference PI simulations.

6.2 Outlook

Since the PI techniques have been implemented successfully in GC-AdResS, one can now apply PI-GC-AdResS methods to solvation of molecules in water, considering in the region

of high resolution, the nuclear quantum effects and surround the solvation region with a generic reservoir as that constructed in GC-AdResS. The delocalization of hydrogen atom may affect the hydrogen bonding network in the solvation shells. The hydrogen bonding network may become more flexible which may lead to different solvation properties and different conformational states of the solvated molecules. Thus, one can investigate whether or not the inclusion of nuclear quantum effects play an important role in the solvation process. If one finds considerable differences with the classical MD results, then it becomes necessary to consider the relevance of nuclear quantum effects in different solvation processes. Furthermore, it may stimulate development of quantum force fields. The PI-GC-AdResS setup can be employed as a tool of analysis by systematically varying the size of the high resolution region and checking whether the given property changes when compared to the results of reference PI simulations. This approach allows for an unambiguous identification of essential (quantum) degrees of freedom required for a given property. The reservoir in GC-AdResS is very generic and it acts as a thermodynamic bath, thus all the quantum degrees of freedom are exclusively present in the quantum region. For classical systems, this kind of approach has already been used to investigate the locality/non-locality of the hydrogen bonding network for water around a series of fullerene molecules [49]. One can perform the same study using PI formalism in the high-resolution region. This should add the effect of quantum description to the classical locality/non-locality given by classical GC-AdResS.

Chapter 7

Zusammenfassung

Mit der von Richard Feynman entwickelte Pfadintegral (PI)-Formulierung können Kern-Quanteneffekte, die durch die Delokalisierung leichter Atomkerne auftreten, mit den Methoden der Molekulardynamik (MD) behandelt werden. Allerdings ist PIMD aufgrund des hohen Rechenaufwands auf kleine Systeme und kurze Zeitskalen beschränkt. In dieser Dissertation wird eine Erweiterung der kürzlich entwickelten "Grand Canonical AdResS"-Methode (GC-AdResS) durch verschiedene PIMD-Algorithmen behandelt. Hierbei wird der PI-Formalismus auf eine relativ kleine Region angewandt, die in eine grosses Reservoir eingebettet ist, in dem die gleichen Moleküle in einem "coarse-grained"-Model behandelt werden. Um die strukturellen und dynamischen Eigenschaften in dem Quantengleichgewicht zu berechnen, müssen zunächst diese Eigenschaften für offener Systeme definiert werden. Hierzu zeigen wir die Genauigkeit der GC-AdResS in Verbindung mit klassischen MD, indem wir das chemische Überschusspotential berechnen. Wir zeigen die Entsprechung zwischen dem Bergmann-Lebowitz (BL) Modell für offener Systeme und GC-AdResS, und verwenden die Prinzipien des BL-Modells, um die Gleichgewichts-Zeitkorrelationfunktionen für offener Systeme zu definieren. Diese berechnen wir für flüssiges Wasserd und zu zeigen, dass GC-AdresS die gleichen Ergebnisse wie MD bei vollständig atomistischer Auflösung ergibt. In der Pfadintegral-Formulierung wird jedes Atom als Ring von "fiktiven" klassischen Partikeln repräsentiert. Somit können wir die zuvor entwickelten theoretischen Konzepte anwenden, um die Quanteneigenschaften von Wasser zu untersuchen. Hierzu kombinieren wir GC-AdResS mit drei verschiedene PI-Methoden: PIMD, "Ring-polymer" MD (RPMD) und "Centroid"-MD (CMD). Wir zeigen, dass die Ergebnisse aller drei Methoden mit denen einer reinen PI-Simulation konsistent sind.

Bibliography

[1] M. Allen and D. Tildesley, *Computer Simulation of Liquids*, Oxford Science Publications, Oxford, 1989.

[2] H. Andersen, *Molecular dynamics simulations at constant pressure and/or temperature*, The Journal of Chemical Physics, 72 (1980), pp. 2384–2393.

[3] C. Bennett, *Efficient estimation of free energy differences from monte carlo data*, Journal of Computational Physics, 22 (1976), pp. 245–268.

[4] H. Berendsen, J. Postma, W. F. van Gunsteren, and J. Hermans, *Interaction models for water in relation to protein hydration*, Intermolecular Forces, Reidel, Dordrecht, 1981, p. 331.

[5] H. J. C. Berendsen, *Computer Simulation in Materials Science: Interatomic Potentials, Simulation Techniques and Applications*, Springer Netherlands, Dordrecht, 1991, ch. Transport Properties Computed by Linear Response through Weak Coupling to a Bath, pp. 139–155.

[6] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, *The missing term in effective pair potentials*, The Journal of Physical Chemistry, 91 (1987), pp. 6269–6271.

[7] P. Bergmann and J. Lebowitz, *New approach to nonequilibrium processes*, Phys. Rev., 99 (1955), pp. 578–587.

[8] B. Berne and D. Thirumalai, *On the simulation of quantum systems: Path integral methods*, Annual Review of Physical Chemistry, 37 (1986), pp. 401–424.

[9] N. Blinov and P. Roy, *Connection between the observable and centroid structural properties of a quantum fluid:âĂĆapplication to liquid para-hydrogen*, The Journal of Chemical Physics, 120 (2004), pp. 3759–3764.

[10] T. Cagin and B. M. Pettitt, *Grand molecular dynamics: A method for open systems*, Molecular Simulation, 6 (1991), pp. 5–26.

[11] ——, *Molecular dynamics with a variable number of molecules*, Molecular Physics, 72 (1991), pp. 169–175.

[12] D. Callaway and A. Rahman, *Microcanonical ensemble formulation of lattice gauge theory*, Phys. Rev. Lett., 49 (1982), pp. 613–616.

[13] J. Cao and G. Voth, *A new perspective on quantum time correlation functions*, The Journal of Chemical Physics, 99 (1993), pp. 10070–10073.

[14] M. Ceriotti, M. Parrinello, T. Markland, and D. Manolopoulos, *Efficient stochastic thermostatting of path integral molecular dynamics*, The Journal of Chemical Physics, 133 (2010).

[15] M. C. Colombo, L. Guidoni, A. Laio, A. Magistrato, P. Maurer, S. Piana, U. Röhrig, K. Spiegel, M. Sulpizi, J. VandeVondele, M. Zumstein, and U. Röthlisberger, *Hybrid qm/mm car-parrinello simulations of catalytic and enzymatic reactions*, CHIMIA International Journal for Chemistry, 56 (2002-01-01T00:00:00), pp. 13–19.

[16] I. Craig and D. Manolopoulos, *Quantum statistics and classical mechanics: Real time correlation functions from ring polymer molecular dynamics*, The Journal of Chemical Physics, 121 (2004), pp. 3368–3373.

[17] L. Delle Site, *Some fundamental problems for an energy-conserving adaptive-resolution molecular dynamics scheme*, Phys. Rev. E, 76 (2007), p. 047701.

[18] ——, *What is a multiscale problem in molecular dynamics?*, Entropy, 16 (2014), p. 23.

[19] ——, *Formulation of liouville's theorem for grand ensemble molecular simulations*, Phys. Rev. E, 93 (2016), p. 022130.

[20] B. Dünweg, *Computer Simulations of Surfaces and Interfaces*, Springer Netherlands, Dordrecht, 2003, ch. Langevin Methods, pp. 77–92.

[21] G. Emch and G. Sewell, *Nonequilibrium statistical mechanics of open systems*, Journal of Mathematical Physics, 9 (1968), pp. 946–958.

[22] B. Ensing, S. Nielsen, P. Moore, M. Klein, and M. Parrinello, *Energy conservation in adaptive hybrid atomistic/coarse-grain molecular dynamics*, Journal of Chemical Theory and Computation, 3 (2007), pp. 1100–1105. PMID: 26627429.

[23] F. Ercolessi and J. B. Adams, *Interatomic potentials from first-principles calculations: The force-matching method*, EPL (Europhysics Letters), 26 (1994), p. 583.

[24] H. Eslami and F. Muller-Plathe, *Molecular dynamics simulation in the grand canonical ensemble*, Journal of Computational Chemistry, 28 (2007), pp. 1763–1773.

[25] R. Feynman and A. Hibbs, *Quantum Mechanics and Path Integrals*, McGraw-Hill, Inc., 1965.

[26] D. Frenkel and B. Smit, *Understanding Molecular Simulation, Second Edition: From Algorithms to Applications*, Academic Press, San Diego, 2001.

[27] 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*, Phys. Rev. Lett., 108 (2012), p. 170602.

[28] D. Geerke, C. Oostenbrink, N. van der Vegt, and W. van Gunsteren, *An effective force field for molecular dynamics simulations of dimethyl sulfoxide and dimethyl sulfoxide water mixtures*, The Journal of Physical Chemistry B, 108 (2004), pp. 1436–1445.

[29] D. Geerke and W. van Gunsteren, *Force field evaluation for biomolecular simulation: Free enthalpies of solvation of polar and apolar compounds in various solvents*, ChemPhysChem, 7 (2006), pp. 671–678.

[30] S. Habershon, G. Fanourgakis, and D. Manolopoulos, *Comparison of path integral molecular dynamics methods for the infrared absorption spectrum of liquid water*, The Journal of Chemical Physics, 129 (2008).

[31] S. Habershon, D. Manolopoulos, T. Markland, and T. Miller, *Ring-polymer molecular dynamics: Quantum effects in chemical dynamics from classical trajectories in an extended phase space*, Annual Review of Physical Chemistry, 64 (2013), pp. 387–413. PMID: 23298242.

[32] S. Habershon, T. Markland, and D. Manolopoulos, *Competing quantum effects in the dynamics of a flexible water model*, The Journal of Chemical Physics, 131 (2009).

[33] L. Hand and J. Finch, *Analytical Mechanics*, Cambridge University Press, Cambridge, 1998.

[34] A. Heyden, , and D. Truhlar, *Conservative algorithm for an adaptive change of resolution in mixed atomistic/coarse-grained multiscale simulations*, Journal of Chemical Theory and Computation, 4 (2008), pp. 217–221. PMID: 26620653.

[35] T. Hill, *Statistical Mechanics: Principles and Selected Applications*, Dover Publications, New York, 1987.

[36] H.Kleinert, *Path Integrals in Quantum Mechanics, Statistics, Polymer Physics and Financial markets*, World Scientific Publishing Co., 2009.

[37] H. Hoang and G. Galliero, *Grand canonical-like molecular dynamics simulations: application to anisotropic mass diffusion in a nanoporous medium*, The Journal of chemical physics, 136 (2012), p. 184702.

[38] T. Hone, P. Rossky, and G. Voth, *A comparative study of imaginary time path integral based methods for quantum dynamics*, The Journal of Chemical Physics, 124 (2006).

[39] W. Hoover, *Canonical dynamics: Equilibrium phase-space distributions*, Phys. Rev. A, 31 (1985), pp. 1695–1697.

[40] C. Huang, K. T. Wikfeldt, T. Tokushima, D. Nordlund, Y. Harada, U. Bergmann, M. Niebuhr, T. M. Weiss, Y. Horikawa, M. Leetmaa, M. P. Ljungberg, O. Takahashi, A. Lenz, L. OjamÃĎe, A. P. Lyubartsev, S. Shin, L. G. M. Pettersson, and A. Nilsson, *The inhomogeneous structure*

of water at ambient conditions, Proceedings of the National Academy of Sciences, 106 (2009), pp. 15214–15218.

[41] J. Ji, T. Cagin, and B. M. Pettitt, *Dynamic simulations of water at constant chemical potential*, The Journal of Chemical Physics, 96 (1992), pp. 1333–1342.

[42] J. E. Jones, *On the determination of molecular fields. i. from the variation of the viscosity of a gas with temperature*, Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 106 (1924), pp. 441–462.

[43] ——, *On the determination of molecular fields. ii. from the equation of state of a gas*, Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 106 (1924), pp. 463–477.

[44] C. Junghans and S. Poblete, *A reference implementation of the adaptive resolution scheme in espresso*, Computer Physics Communications, 181 (2010), pp. 1449 – 1454.

[45] G. Kell, *Isothermal compressibility of liquid water at 1 atm.*, Journal of Chemical and Engineering Data, 15 (1970), pp. 119–122.

[46] R. Kubo, *Statistical-mechanical theory of irreversible processes. i. general theory and simple applications to magnetic and conduction problems*, Journal of the Physical Society of Japan, 12 (1957), pp. 570–586.

[47] R. Kubo, M. Toda, and N. Hashitsume, *Statistical Physics II: Nonequilibrium Statistical Mechanics*, Springer, New York, 1985.

[48] T. Kuznetsova and B. Kvamme, *Grand canonical molecular dynamics for tip4p water systems*, Molecular Physics, 97 (1999), pp. 423–431.

[49] B. Lambeth, C. Junghans, K. Kremer, C. Clementi, and L. Delle Site, *Communication: On the locality of hydrogen bond networks at hydrophobic interfaces*, The Journal of Chemical Physics, 133 (2010).

[50] O. Lanford, *Statistical Mechanics and Mathematical Problems*, vol. 20 of Lecture Notes in Physics, Springer-Verlag, Berlin, 1973.

[51] A. R. Leach, *Molecular Modelling: Principles and Applications*, Pearson Education EMA, New York, 2001.

[52] J. Lebowitz and P. Bergmann, *Irreversible gibbsian ensembles*, Annals of Physics, 1 (1957), pp. 1 – 23.

[53] J. Lebowitz and A. Shimony, *Statistical mechanics of open systems*, Phys. Rev., 128 (1962), pp. 1945–1958.

[54] M. Lee and N. van der Vegt, *A new force field for atomistic simulations of aqueous tertiary butanol solutions*, The Journal of Chemical Physics, 122 (2005), p. 114509.

[55] B. Leimkuhler, E. Noorizadeh, and F. Theil, *A gentle stochastic thermostat for molecular dynamics*, Journal of Statistical Physics, 135 (2009), pp. 261–277.

[56] J. Lobaugh and G. Voth, *A quantum model for water: Equilibrium and dynamical properties*, The Journal of Chemical Physics, 106 (1997), pp. 2400–2410.

[57] L.Onsager, *Electric moments of molecules in liquids*, Journal of the American Chemical Society, 58 (1936), pp. 1486–1493.

[58] A. Luzar and D. Chandler, *Effect of environment on hydrogen bond dynamics in liquid water*, Phys. Rev. Lett., 76 (1996), pp. 928–931.

[59] ——, *Hydrogen-bond kinetics in liquid water*, Nature, 379 (1996), pp. 55–57.

[60] G. Lynch and B. Pettitt, *Grand canonical ensemble molecular dynamics simulations: Reformulation of extended system dynamics approaches*, The Journal of Chemical Physics, 107 (1997), pp. 8594–8610.

[61] A. Lyubartsev and A. Laaksonen, *Calculation of effective interaction potentials from radial distribution functions: A reverse monte carlo approach*, Phys. Rev. E, 52 (1995), pp. 3730–3737.

[62] T. Markland and D. Manolopoulos, *An efficient ring polymer contraction scheme for imaginary time path integral simulations*, The Journal of Chemical Physics, 129 (2008).

[63] 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, 10 (2015).

[64] S. Matysiak, C. Clementi, M. Praprotnik, K. Kremer, and L. Delle Site, *Modeling diffusive dynamics in adaptive resolution simulation of liquid water*, The Journal of Chemical Physics, 128 (2008).

[65] T. Miller and D. Manolopoulos, *Quantum diffusion in liquid water from ring polymer molecular dynamics*, The Journal of Chemical Physics, 123 (2005).

[66] D. Mobley, J. Chodera, and K. Dill, *On the use of orientational restraints and symmetry corrections in alchemical free energy calculations*, The Journal of Chemical Physics, 125 (2006), pp. –.

[67] D. Mukherji, N. 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, 8 (2012), pp. 375–379. PMID: 26596589.

[68] T. Murtola, A. Bunker, I. Vattulainen, M. Deserno, and M. Karttunen, *Multiscale modeling of emergent materials: biological and soft matter*, Phys. Chem. Chem. Phys., 11 (2009), pp. 1869–1892.

[69] S. Nielsen, C. Lopez, G. Srinivas, and M. Klein, *Coarse grain models and the computer simulation of soft materials*, Journal of Physics: Condensed Matter, 16 (2004), p. R481.

[70] W. G. Noid, J. Chu, G. S. Ayton, V. Krishna, S. Izvekov, G. A. Voth, A. Das, and H. C. Andersen, *The multiscale coarse-graining method. i. a rigorous bridge between atomistic and coarse-grained models*, The Journal of Chemical Physics, 128 (2008).

[71] S. Nose, *A molecular dynamics method for simulations in the canonical ensemble*, Molecular Physics, 52 (1984), pp. 255–268.

[72] F. Paesani, W. Zhang, D. Case, T. Cheatham, and G. Voth, *An accurate and simple quantum model for liquid water*, The Journal of Chemical Physics, 125 (2006).

[73] M. Parrinello and A. Rahman, *Study of an f center in molten kcl*, The Journal of Chemical Physics, 80 (1984), pp. 860–867.

[74] M. Pavese, S. Jang, and G. Voth, *Centroid molecular dynamics: A quantum dynamics method suitable for the parallel computer*, Parallel Computing, 26 (2000), pp. 1025 – 1041.

[75] W. Peier, *Time evolution of open systems*, Physica, 57 (1972), pp. 565 – 584.

[76] C. Perego, M. Salvalaglio, and M. Parrinello, *Molecular dynamics simulations of solutions at constant chemical potential*, The Journal of Chemical Physics, 142 (2015).

[77] A. Perez, M. Tuckerman, and M. Muser, *A comparative study of the centroid and ring-polymer molecular dynamics methods for approximating quantum time correlation functions from path integrals*, The Journal of Chemical Physics, 130 (2009).

[78] C. Peter and K. Kremer, *Multiscale simulation of soft matter systems - from the atomistic to the coarse-grained level and back*, Soft Matter, 5 (2009), pp. 4357–4366.

[79] J. R. Pliego Jr and J. M. Riveros, *Gibbs energy of solvation of organic ions in aqueous and dimethyl sulfoxide solutions*, Phys. Chem. Chem. Phys., 4 (2002), pp. 1622–1627.

[80] S. Poblete, M. Praprotnik, K. Kremer, and L. Delle Site, *Coupling different levels of resolution in molecular simulations*, The Journal of Chemical Physics, 132 (2010), pp. –.

[81] A. B. Poma and L. Delle Site, *Classical to path-integral adaptive resolution in molecular simulation: Towards a smooth quantum-classical coupling*, Phys. Rev. Lett., 104 (2010), p. 250201.

[82] ——, *Adaptive resolution simulation of liquid para-hydrogen: testing the robustness of the quantum-classical adaptive coupling*, Phys. Chem. Chem. Phys., 13 (2011), pp. 10510–10519.

[83] R. Potestio and L. Delle Site, *Quantum locality and equilibrium properties in low-temperature parahydrogen: A multiscale simulation study*, The Journal of Chemical Physics, 136 (2012).

[84] R. Potestio, S. Fritsch, P. Espanol, R. Delgado-Buscalioni, K. Kremer, R. Everaers, and D. Donadio, *Hamiltonian adaptive resolution simulation for molecular liquids*, Phys. Rev. Lett., 110 (2013), p. 108301.

[85] M. Praprotnik, L. Delle Site, and K. Kremer, *Adaptive resolution molecular-dynamics simulation: Changing the degrees of freedom on the fly*, The Journal of Chemical Physics, 123 (2005).

[86] ——, *Adaptive resolution molecular-dynamics simulation: Changing the degrees of freedom on the fly*, The Journal of Chemical Physics, 123 (2005).

[87] ——, *Adaptive resolution scheme for efficient hybrid atomistic-mesoscale molecular dynamics simulations of dense liquids*, Phys. Rev. E, 73 (2006), p. 066701.

[88] ——, *A macromolecule in a solvent: Adaptive resolution molecular dynamics simulation*, The Journal of Chemical Physics, 126 (2007).

[89] ——, *Multiscale simulation of soft matter: From scale bridging to adaptive resolution*, Annual Review of Physical Chemistry, 59 (2008), pp. 545–571. PMID: 18062769.

[90] M. Praprotnik, K. Kremer, and L. Delle Site, *Adaptive molecular resolution via a continuous change of the phase space dimensionality*, Phys. Rev. E, 75 (2007), p. 017701.

[91] ——, *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, 40 (2007), p. F281.

[92] M. Praprotnik, S. Matysiak, L. Delle Site, K. Kremer, and C. Clementi, *Adaptive resolution simulation of liquid water*, Journal of Physics: Condensed Matter, 19 (2007), p. 292201.

[93] S. Pronk, S. Pall, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. Shirts, J. Smith, P. Kasson, D. van der Spoel, B. Hess, and E. Lindahl, *Gromacs 4.5: a high-throughput and highly parallel open source molecular simulation toolkit*, Bioinformatics, 29 (2013), pp. 845–854.

[94] H. Qian, *Phosphorylation energy hypothesis: Open chemical systems and their biological functions*, Annual Review of Physical Chemistry, 58 (2007), pp. 113–142. PMID: 17059360.

[95] D. Rapaport, *The Art of Molecular Dynamics Simulation*, Cambridge University Press, Cambridge, 2004.

[96] D. Reith, M. Pötz, and F. Müller-Plathe, *Deriving effective mesoscale potentials from atomistic simulations*, Journal of Computational Chemistry, 24 (2003), pp. 1624–1636.

[97] F. Reynaud, *Application of the grand canonical ensemble in classical statistical physics to the study of orderâĂŞdisorder transitions in binary substitutional alloys*, physica status solidi (a), 84 (1984), pp. 393–400.

[98] V. Rühle, C. Junghans, A. Lukyanov, K. Kremer, and D. Andrienko, *Versatile object-oriented toolkit for coarse-graining applications*, Journal of Chemical Theory and Computation, 5 (2009), pp. 3211–3223. PMID: 26602505.

[99] J. Salacuse, *Particle fluctuations within sub-regions of an -particle, three-dimensional fluid: Finite-size effects and compressibility*, Physica A: Statistical Mechanics and its Applications, 387 (2008), pp. 3073 – 3083.

[100] T. Schlick, *Molecular Modeling and Simulation: An Interdisciplinary Guide*, Springer, New York, 2010.

[101] F. Schwabl, *Statistical Mechanics*, Springer, Berlin, 2002.

[102] B. Smit, *Grand canonical monte carlo simulations of chain molecules: adsorption isotherms of alkanes in zeolites*, Molecular Physics, 85 (1995), pp. 153–172.

[103] L. J. Smith, H. J. C. Berendsen, and W. F. van Gunsteren, *Computer simulation of urea-water mixtures: A test of force field parameters for use in biomolecular simulation*, The Journal of Physical Chemistry B, 108 (2004), pp. 1065–1071.

[104] W. Swope, H. Andersen, P. H. Berens, and K. Wilson, *A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters*, The Journal of Chemical Physics, 76 (1982), pp. 637–649.

[105] I. Tironi and W. V. Gunsteren, *A molecular dynamics simulation study of chloroform*, Molecular Physics, 83 (1994), pp. 381–403.

[106] I. Tironi, R. Sperb, P. Smith, and W. van Gunsteren, *A generalized reaction field method for molecular dynamics simulations*, The Journal of Chemical Physics, 102 (1995), pp. 5451–5459.

[107] M. Tuckerman, *Statistical Mechanics: Theory and Molecular Simulation*, Oxford University Press, New York, 2010.

[108] M. Tuckerman, B. Berne, G. Martyna, and M. Klein, *Efficient molecular dynamics and hybrid monte carlo algorithms for path integrals*, The Journal of Chemical Physics, 99 (1993), pp. 2796–2808.

[109] L. Verlet, *Computer "experiments" on classical fluids. i. thermodynamical properties of lennard-jones molecules*, Phys. Rev., 159 (1967), pp. 98–103.

[110] R. Walser, A. Mark, W. van Gunsteren, M. Lauterbach, and G. Wipff, *The effect of force-field parameters on properties of liquids: Parametrization of a simple three-site model for methanol*, The Journal of Chemical Physics, 112 (2000), pp. 10450–10459.

[111] H. Wang, C. Hartmann, C. Schütte, and L. Delle Site, *Grand-canonical-like molecular-dynamics simulations by using an adaptive-resolution technique*, Phys. Rev. X, 3 (2013), p. 011018.

[112] H. Wang, C. Junghans, and K. Kremer, *Comparative atomistic and coarse-grained study of water: What do we lose by coarse-graining?*, The European Physical Journal E, 28 (2009), pp. 221–229.

[113] H. Wang, C. Schütte, G. Ciccotti, and L. Delle Site, *Exploring the conformational dynamics of alanine dipeptide in solution subjected to an external electric field: A nonequilibrium molecular dynamics simulation*, Journal of Chemical Theory and Computation, 10 (2014), pp. 1376–1386. PMID: 26580357.

[114] 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, 8 (2012), pp. 2878–2887.

[115] S. Weerasinghe and B. M. Pettitt, *Ideal chemical potential contribution in molecular dynamics simulations of the grand canonical ensemble*, Molecular Physics, 82 (1994), pp. 897–912.

[116] B. Widom, *Some topics in the theory of fluids*, The Journal of Chemical Physics, 39 (1963), pp. 2808–2812.

[117] ——, *Some topics in the theory of fluids*, The Journal of Chemical Physics, 39 (1963), pp. 2808–2812.

[118] A. Witt, S. Ivanov, M. Shiga, H. Forbert, and D. Marx, *On the applicability of centroid and ring polymer path integral molecular dynamics for vibrational spectroscopy*, The Journal of Chemical Physics, 130 (2009).

[119] L. Woodcock, *Isothermal molecular dynamics calculations for liquid salts*, Chemical Physics Letters, 10 (1971), pp. 257 – 261.