Quantum effects in molecular systems

by

Nabil Faruk

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Science
in
Chemistry

Waterloo, Ontario, Canada, 2014

© Nabil Faruk 2014

Author's Declaration

I hereby declare that I am the sole author or co-author of material included in this thesis: refer to the Statement of Contributions section included in the front matter. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

This thesis presents two different approaches and applications where nuclear dynamics are treated quantum mechanically in order to obtain more accurate theoretical predictions of molecular properties. In our first application, we report a first-principles prediction of the Raman shifts of parahydrogen ($p\mathrm{H}_2$) clusters of sizes $N = 4-19$ and 33, based on path integral ground-state simulations with an *ab initio* potential energy surface. The Raman shifts are calculated, using perturbation theory, as the average of the difference-potential energy surface between the potential energy surfaces for vibrationally-excited and ground-state parahydrogen monomers. The radial distribution of the clusters is used as a weight function in this average. Very good overall agreement with experiment [1] is achieved for $p(\mathrm{H}_2)_{2-8,13,33}$. A number of different pair potentials are employed for the calculation of the radial distribution functions. We find that the Raman shifts are sensitive to slight variations in the radial distribution functions. In our second application, we discuss the development of Path Integral Molecular Dynamics (PIMD) methodology, which our group has previously incorporated into the Molecular Modeling Toolkit (MMTK) [2] to account for nuclear quantum effects. This thesis is to provide a proof-of-concept for our software tools and PIMD method through the gas phase investigation of methyl beta-D-arabinofuranoside, which is a sugar residue in the cell wall of tuberculosis bacteria and is thought to provide bacterial resistance to drugs. We observe the effect of nuclear quantum sampling on

the sugar's dihedral angle distributions at different temperatures, which we then relate to nuclear magnetic resonance proton-proton coupling constants via Karplus equations. We also determine the sugar's energy convergence with path integral sampling and the energy behaviour with temperature. We find that quantum effects are non-negligible even at biological temperatures, although some challenges remain in converging our coupling constant predictions. Finally, we discuss and benchmark our extension with the Open Molecular Mechanics (OpenMM) program [3] to enable graphics processing unit-accelerated solution phase simulations for future work.

Statement of Contributions

The majority of the parahydrogen vibrational Raman shift content in this thesis has been reprinted with permission from Ref. 4 where I was the first author:

N. Faruk, M. Schmidt, H. Li, R. J. Le Roy, and P.-N. Roy, "First Principle Prediction of the Raman shifts of parahydrogen clusters", J. Chem. Phys. 141, 014310 (2014).

My co-author M. Schmidt also included an excerpt from this publication in his previously submitted thesis [5]. He was responsible for generating the radial distributions of the parahydrogen clusters in the published work and approached the content he reprinted in his thesis from the perspective of applying the molecular dynamics simulation methodology that he developed to generate those distributions. I was responsible for producing hydrogen interaction potentials, one of which was used by M. Schmidt to generate some of the distributions, and I also calculated the vibrational shifts using his distributions and the interaction potentials. I focus on that aspect in this thesis rather than the molecular dynamics simulation details.

The content on the sugar application in this thesis is still unpublished elsewhere besides my Master's proposal and is solely my work, apart from the guidance of my supervisor, P.-N. Roy.

Acknowledgements

I would first like to acknowledge my supervisor, Prof. Pierre-Nicholas Roy. I started working under him as an undergraduate research assistant four years ago and I continued working with him for this length of time because he is an excellent mentor. He is the most approachable and helpful research supervisor that I have worked with, but also pushes his students to develop independent thinking. He has helped me through all stages of this thesis, from conception of the research problems to the final edits.

I would also like to thank Prof. Robert Le Roy, who cosupervised my parahydrogen vibrational Raman shift project. He generated wavefunctions for me to use in my work and meticulously went through and suggested edits for several revisions of my J. Chem. Phys. manuscript on that work.

I extend some specific thanks to past and current graduate students and postdocs who have helped me in my research. This includes: Prof. Hui Li, who walked me through his software code for calculating vibrational shifts in clusters and helped me troubleshoot when I adapted it to pure parahydrogen clusters; Stephen Constable, who initiated me in the use of our group's Path Integral Molecular Dynamics codes and also helped me troubleshoot them; Matthew Schmidt, who collaborated with me and was a coauthor on the parahydrogen vibrational Raman shift project; Kevin Bishop, who is my neighbour in

the office and often collaborated with me in tackling bugs in our existing codebase and the development of new code; Chris Ing, who I did not interact with directly for my research, but whose Path Integral Molecular Dynamics contributions to the Molecular Modeling Toolkit software that we use has been a foundation for a great deal of the research in our group.

I give general thanks for helpful discussions and the collaborative research environment fostered by the aforementioned people and the rest of the Theoretical Chemistry group currently or formerly present at UW, including: Toby Zeng, Dmitri Iouchtchenko, Lindsay Orr, Lecheng Wang, Spencer Yim, Prateek Goel, Prof. Marcel Noojjen, Prof. Fred McCourt, and Prof. Scott Hopkins.

Dedication

This thesis is dedicated to my family.

Table of Contents

List of Tables

List of Figures

Nomenclature

$\gamma^{(0)}$	Centroid Friction
τ	$= \frac{1}{TP}$, a measure of the level of quantum sampling and closeness to the full quantum value (τ=0). Not to be confused with τ_0
τ_0	Centroid Decorrelation Time
$p\mathrm{H}_2$	parahydrogen
AHR	Adiabatic Hindered Rotor
DFT	Density Functional Theory
DMC	Diffusion Monte Carlo
GPU	Graphics Processing Unit
LePIGS	Langevin Equation Path Integral Ground State

MMTK Molecular Modeling Toolkit

NVE Microcanonical Ensemble with constant Number of particles, Volume, and Energy

NVT Canonical Ensemble, with fixed Number of particles, Volume, and Temperature

PES Potential Energy Surface

PIGS Path Integral Ground-State

PILE Path Integral Langevin Equation in Normal Modes

PIMC Path Integral Monte Carlo

PIMD Path Integral Molecular Dynamics

RPMD Ring Polymer Molecular Dynamics

TB Tuberculosis

Chapter 1

Introduction

Computer simulations hold many promises for scientific inquiry. Thousands of virtual experiments can be conducted simultaneously through simulation, depending on available computing power. Additionally, analysis of results is made easy by direct access to the microscopic details of the system under study, such as the positions of atoms at all points in time. Arguably the most important function of simulation is to act as platforms to test our theoretical understanding of the world through comparisons of our models' predictions to real-world experimental results. However, our level of theory is often ahead of what is feasible to model given computational limitations. Approximations are thus made, including commonly treating nuclei as classical point-like objects in molecular dynamics simulations, even though scientists have been developing quantum mechanics theory ever

since Max Planck presented his quantum hypothesis back in 1900[7].

This thesis explores two different approaches and applications where we are able to treat nuclear dynamics quantum mechanically in our simulations. First, the Born-Oppenheimer approximation [8] is invoked to separate the electronic and nuclear degrees of freedom. A potential energy surface can be derived by solving the electronic structure problem for fixed nuclear configurations. We then have the option of solving the nuclear dynamics problem classically using Newtonian dynamics, where the potential energy surface provides the forces acting on the nuclei, or quantum mechanically. One procedure for obtaining the quantum mechanical solution is to use basis set methods where basis states are weighted by the potential energy surface (Chapter 2 covers in part such an approach). The main problem of basis set methods is that the computational cost scales exponentially as the number of degrees of freedom increases [9, 10]. We can circumvent this problem using the Feynman path integral formulation of quantum mechanics [11] to recast the quantum nuclei in terms of a classical analogue [12]. The finite temperature version of this technique when applied to molecular dynamics simulations is called Path Integral Molecular Dynamics (PIMD) [13], and the ground state version developed and employed by our group is called Langevin Equation Path Integral Ground State (LePIGS) [14, 15]. These path integral approaches benefit from the $\mathcal{O}(N^2) - \mathcal{O}(N \log N)$ scaling [16] of established Newtonian molecular dynamics algorithms. This approach allows us to tackle much larger systems

for computing quantum ensemble average properties. Note that the eigenstate details of a basis set approach are lost when using the path integral formulation. However, it is still possible to recover ground state properties (as in Chapter 2).

One of our applications involves using LePIGS to perturbatively extend basis set calculations of a parahydrogen ($p\mathrm{H}_2$) dimer to predict vibrational frequency shifts in larger $p\mathrm{H}_2$ clusters. We thus benefit from the strengths of both the domains of basis set calculations and the path integral approach. The other application involves PIMD to study the structural and energy properties of methyl β-D-arabinofuranoside, a *Mycobacterium tuberculosis* sugar. Both applications are linked by having the same prime objective, to improve upon the approximations used in molecular modelling while remaining tractable for today's computational resources. The rest of this chapter introduces each application further.

1.1 $p\mathrm{H}_2$ cluster background

The study of $p\mathrm{H}_2$ clusters is of considerable interest because of their potential superfluid properties. $p\mathrm{H}_2$ is a molecular boson with low mass and weak intermolecular forces and thus may be expected to display superfluid properties under appropriate conditions, particularly when in the form of nano-clusters that retain liquid-like properties below the $p\mathrm{H}_2$

triple point [17]. The results of early experiments implied the presence of superfluidity in such clusters embedded in helium nanodroplets based on rovibrational Q branch behavior of dopants [18, 19, 20, 21]. The first direct measurement of the superfluid fraction was only recently conducted through comparisons of the results of Path Integral Monte Carlo (PIMC) simulations with experimental data involving pH$_2$ clusters of varying sizes, probed using infrared spectroscopy of a CO_2 chromophore [22]. Recent advances in PIMC techniques have allowed the accurate simulation of more complex asymmetric top dopants [23] and have been applied to the case of pH$_2$ clusters doped with H_2O [24] and SO_2 [25] molecules. Highly accurate Potential Energy Surfaces (PESs) that describe the molecular interactions are required for simulations of this type. The quality of a PES can be tested by using it to predict the vibrational band origin shift of a chromophore when perturbed by surrounding "solvent" cluster molecules. Li *et al.* demonstrated, for instance, that it is possible to obtain accurate shifts from theory and simulation for the case of CO_2 in He and pH$_2$ clusters of various sizes [26, 22].

An interesting question remains regarding what will occur if the chromophore is another pH$_2$, i.e., if we have a pure pH$_2$ cluster. Several PIMC studies have been carried out for pure ground state pH$_2$ clusters at finite temperature [27, 28, 29, 30, 31] and in the ground state using the Path Integral Ground-State (PIGS) formulation, [32, 33, 34] and Diffusion Monte Carlo (DMC) [35] to elucidate their structural, energetic and superfluid nature. A

summary of these studies has been presented in a review by Navarro and Guardiola [36]. However, some of these studies have shown that the choice of interaction potential used in a simulation affects the details of predicted properties [32, 37, 33, 38]. This reinforces the need to test PESs via comparison of calculated band-origin shifts with experiment. Figure 1.1 depicts the nature of the band origin shift for the fundamental vibrational energy transition of a $p\mathrm{H}_2$ chromophore in a $p\mathrm{H}_2$ cluster. Care must be taken in modelling these systems because the $p\mathrm{H}_2$ chromophore is indistinguishable from the solvent $p\mathrm{H}_2$, and bosonic exchange interactions will result in the delocalization of vibrational excitations.

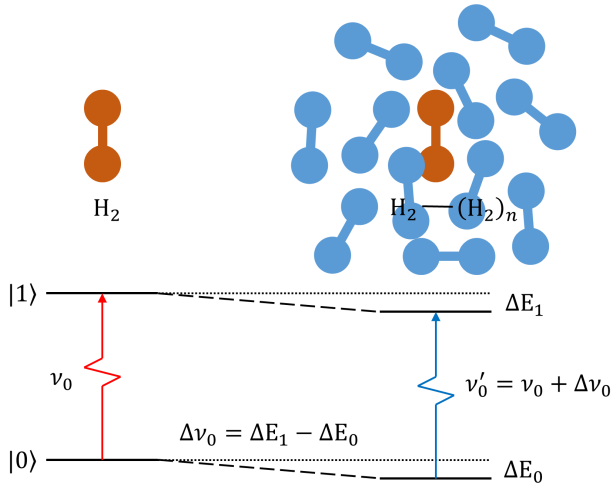

Figure 1.1: Schematic of the band origin shift of a '*para*-H_2' chromophore in a *para*-H_2 cluster.

Two recent developments stimulated this line of inquiry. Tejeda *et al.* overcame difficulties in studying liquid pure $p\mathrm{H}_2$ clusters by using cryogenic free-jet expansions to produce

clusters of varying sizes resolved in space and time, thereby allowing their clear observation with Raman spectroscopy [1]. They also developed an empirical 1-D Lennard-Jones model to fit their data that represents the difference between the PESs for the ground and the first vibrationally excited state of the chromophore. That model predicts vibrational band-origin shifts quite well when perturbatively combined with DMC calculated radial distribution functions of the $p\mathrm{H}_2$ clusters. In independent work, [39] Hinde constructed a six-dimensional *ab initio* PES for the pair interaction between H_2 molecules and used it to calculate infrared and Raman transition energies within $(\mathrm{H}_2)_2$, showing good agreement with the experiments [1].

In the present thesis, the 6-D Hinde PES [39] is used to obtain a reduced-dimension set of 1-D $p\mathrm{H}_2$ dimer PESs by solving a rovibrational Hamiltonian of the $p\mathrm{H}_2$ monomers [6]. The difference between the PES in which the $p\mathrm{H}_2$ dimer system has one quantum of vibrational excitation, and the one in which both $p\mathrm{H}_2$ monomers are in their vibrational ground states is obtained. This difference-PES is then used to predict vibrational band origin shifts in many-body $p\mathrm{H}_2$ clusters from first-principles, by averaging this difference-potential weighted by the radial distribution functions for various cluster sizes. The results are tested against the experimental observations of Tejeda *et al.* [1] and predictions generated from their empirical difference-PES.

1.2 Sugar background

PIMD and related techniques have allowed investigation within the highly quantum regime of nuclei, such as helium [40] and hydrogen clusters at low temperatures as outlined in our other application. Our group seeks to promote and explore the use of PIMD for further applications where the consideration of quantum effects might be needed for accurate results. To this end, our group was involved in developing simulation software incorporating PIMD that is accessible to use and Open Source [2, 41, 3]. We hope that these features allow other researchers to adopt it and that it inspires broader research.

This work serves as a proof-of-concept of our methods and our concern for the wider application of PIMD by demonstrating that nuclear quantum effects might even play a role in biomolecules, which are a popular subject for simulation. Biomolecule nuclei are traditionally treated classically because they are deemed too massive and the molecules active at too high temperatures to exhibit significant quantum properties. We focus on the sugar methyl β-D-arabinofuranoside found in *Mycobacterium tuberculosis*, which is responsible for Tuberculosis (TB). The sugar's conformations can be characterized in terms of dihedral angles between hydrogens bonded to the carbon backbone [42]. The involvement of these light hydrogen atoms makes its dihedral angles a prime target for evaluating the influence of quantum effects and there has been extensive classical simulation studies on

the sugar that we can compare to. Furthermore, the unique structural properties of the sugar is believed to confer drug resistance to the bacteria and thus the elucidation of sugar conformations through simulation can lead to better treatments.

An estimated 8.7 million new people experienced adverse symptoms of TB and nearly 1.4 million died from it in 2011 [43]. For perspective, it is second only to Acquired Immunodeficiency Syndrome (AIDS) for worldwide deaths due to an infectious agent [43]. Although global targets for controlling and reversing the epidemic have been reached, there is an alarming increase in multidrug resistant TB that has doubled in the number of reported cases from 2009 to 2012 in prevalent countries [43]. This has drawn attention to the development of novel treatments, where one area of focus is inhibitors of bacterial cell wall biosynthesis [44, 45]. A constituent of the mycobacterial cell wall is polysaccharides with residues of alpha and beta D-arabinofuranoside [46]. The structural formula of a monomer with the β-configuration is shown in Figure 1.2. This furanose carbohydrate with a five-membered ring is less thermodynamically stable than its six-membered ring pyranose counterpart. The image of glucose, a pyranose, in its stable chair conformation is evoked when thinking of carbohydrate conformations. In contrast, furanosides have increased ring strain and low energy barriers between states that allow them to assume many conformations [47]. This is proposed [48] to give flexibility to the polysaccharides that the arabinofuranose residues compose, which promotes efficient packing of terminal

lipid residues around the cell wall of the TB bacteria [46]. In turn, this provides a tight barrier that prevents the permeability of antibiotics and allows evasion from the host immune system [49]. Understanding the arabinofuranose conformations through simulation can thus lead to drug designs that overcome bacterial resistance by targeting the enzymes that synthesize and process the sugars [44].

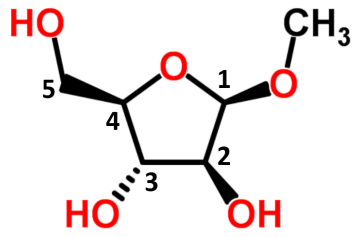

Figure 1.2: Structural formula of methyl β-D-arabinofuranoside

NMR can provide proton-proton (^{1}H-^{1}H) nuclear spin coupling constants for organic molecules that are linked to their conformations[50], giving reference points to check the accuracy of conformations predicted through simulation. In this case, we are interested in coupling constants ($^3J_{\mathrm{H,H}}$) across sequences of three bonds (fragments of H-C-C-H) around the ring and the methyl alcohol group of our sugar. These fragments are highlighted in Figure 1.3.

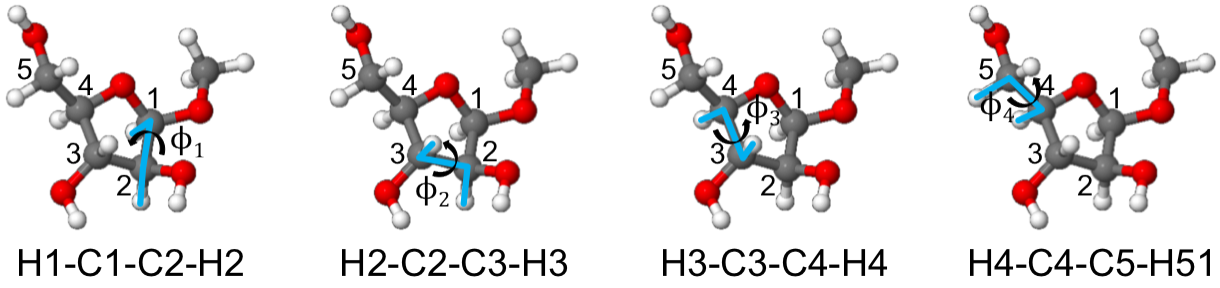

Figure 1.3: Sequences of atoms and their dihedrals (ϕ_i) of interest in methyl β-D-arabinofuranoside

These ${}^3J_{\mathrm{H,H}}$ can be related via a Karplus equation [51] to the dihedral angle that describe the relative rotation of the four atoms that comprise each sequence of three bonds [52, 53, 54, 55, 56]. There exists generalized empirical Karplus equations [52] as well as recent equations parameterized by Density Functional Theory calculations (DFT) [54, 55]. The theoretical Karplus equations have the form [54]

$$ {}^3J_{\mathrm{H,H}} = a + b\ cos\left(\phi\right) + c\ cos\left(2\phi\right) \ , \tag{1.1} $$

where ϕ is the dihedral angle and a, b, c are constants that vary between the dihedral angles.

Molecular simulations result in trajectories of molecules traversing between different configurations. Properties extracted from these trajectories can be binned and normalized in a histogram to give a probability distribution of the property over conformational space. In this case of dihedral angles as the property of interest, the average coupling constant is

calculated by[54]

$$\left\langle {}^{3}J_{\mathrm{H,H}} \right\rangle = \int_{0}^{360} {}^{3}J_{\mathrm{H,H}}\left(\phi\right) \rho\left(\phi\right) d\phi \ , \tag{1.2}$$

where $\rho\left(\phi\right)$ is the probability distribution of a dihedral angle ϕ. The calculated coupling constants can then be compared to those obtained from NMR experiments.

Prior classical molecular dynamics work looked at coupling constants and conformer populations for monomers and oligosaccharides of both alpha and beta D-arabinofuranoside [57, 53, 54, 55, 56, 58]. Ring averaged charges were calculated for the monomers to account for their flexibility [57, 53], more advanced water models were examined for solution phase simulations [53, 54], results from DFT-derived Karplus equations were contrasted with those from empirical equations [54], and the effect of different simulation force fields and DFT basis sets were compared [55]. Although the more sophisticated procedures generally improved predictions, some coupling constants and relative magnitudes of conformer populations still differed from experiment, particularly for the beta configuration [54, 55]. Taha et al. posited that the discrepancies may arise from the charge derivation procedure, errors in fitting DFT data to Karplus equations, problems in the conformational model, or lack of polarizability in the force fields tested [53, 54, 55].

However, in this work we raise the question as to whether it was the lack of nuclear

quantum effects that led to poorer agreement with experiment. We limit our study to the gas phase of a methyl β-D-arabinofuranoside monomer for simplicity in determining optimal PIMD parameters, in part due to reduced computational cost. This means that our coupling constant predictions can not be directly compared to previous classical simulations and experiments conducted in the solution phase. Nevertheless, observing a change in the predicted dihedral angle distributions and coupling constants in the gas phase with increasing quantum sampling is enough to warrant further investigation.

1.3 Organization of the thesis

The remainder of this thesis is organized as follows: A first principle prediction of the Raman vibrational shifts of parahydrogen clusters is presented in Chapter 2. The development of a path integral methodology for the simulation of more complex molecular systems is contained in Chapter 3. In this chapter, a sugar molecule is used to highlight the challenges associated with the quantum simulation of molecular systems down to low temperatures, where quantum effects dominate. Concluding remarks and future research avenues are presented in Chapter 4.

Chapter 2

First-principles prediction of the vibrational Raman shifts in parahydrogen clusters

2.1 Theory and methods

Vibrational band-origin shifts are defined as the difference between the vibrational transition energy of the free chromophore molecule and that of the chromophore molecule in the cluster. This can also be described as the difference between the changes in the upper, $v_{\mathrm{t}} = 1$, and lower, $v_{\mathrm{t}} = 0$, vibrational energies when the free chromophore is introduced into the cluster. For the fundamental vibrational transition of a chromophore in a cluster size of N, this may be written as

$$\Delta\nu_0^{[N]} = \Delta E_{\mathrm{cluster}}^{[N]} - \Delta E_{\mathrm{free}} = E_{v_{\mathrm{t}}=1}^{[N]} - E_{v_{\mathrm{t}}=0}^{[N]} \ , \tag{2.1}$$

in which ΔE_{free} is the vibrational spacing of a free molecule, $\Delta E^{[N]}_{\text{cluster}}$ is that spacing when the molecule is in a cluster of N $p\text{H}_2$ molecules, and $E^{[N]}_{v_\text{t}}$ is the energy of the system expressed relative to the dissociation limit for N $p\text{H}_2$ molecules when the *total* number of vibrational quanta of the $p\text{H}_2$ monomers is either $v_\text{t} = 1$ or 0.

Direct calculation of these energies or energy shifts via simulation poses a challenge. Conventionally, two simulations per cluster size would be undertaken, one with the chromophore having $v_\text{t} = 0$ and the other with the chromophore having $v_\text{t} = 1$. The total energies of the simulated clusters will eventually converge to the statistical ensemble average values and their difference relative to the free chromophore in either state will give the vibrational band origin shift. However, a PIMC study of CO_2 in He clusters found very slow convergence of the statistical errors when increasing the number of simulation steps [26]. Furthermore, in this case of pure $p\text{H}_2$ clusters, an individual molecule should not be singled out and specified as the chromophore in different excitation states during simulation, due to the presence of exchange interactions. Fortunately, a perturbative approach exists to predict the shifts that converges faster by making direct use of the difference between the PESs of the ground and first vibrationally excited states of the chromophore while performing a simulation of just the ground vibrational state of the clusters [26, 1].

We begin by noting that the Hamiltonian for the $v_\text{t} = 1$, the vibrational state of the cluster is nearly identical to that in the $v_\text{t} = 0$ state, except for the small difference in their

potential energies. Here "t" indicates the total quanta of vibrational excitation among the cluster molecules, considering that the excitation can be diffuse due to exchange. Note that if clusters have a more rigid structure, [33] they could potentially be trapped in a metastable state that would prevent the vibrational excitation to be delocalized across the whole cluster. We have not investigated this effect here and therefore assume that the state of the cluster is totally symmetric upon identical particle permutations. The excited state Hamiltonian can therefore be defined approximately as,

$$\hat{H}_{v_t=1} \approx \hat{H}_{v_t=0} + \Delta\hat{V}^{[N]} \,, \tag{2.2}$$

in which $\Delta\hat{V}^{[N]} \equiv \hat{V}^{[N]}_{v_t=1} - \hat{V}^{[N]}_{v_t=0}$ is the difference between the total potential energy functions of clusters of size N when one chromophore is in its ground or its first excited state, in which N includes the chromophore in the count. First-order perturbation theory then gives the vibrational frequency shift as

$$\Delta\nu_0^{[N]} = \left\langle \Psi_0^{[N]} \left| \Delta\hat{V}^{[N]} \right| \Psi_0^{[N]} \right\rangle \,, \tag{2.3}$$

in which $|\Psi_0(R)|^2$ is the probability in one dimension of finding a solvent $p\mathrm{H}_2$ molecule at a radial distance R from the chromophore, given by the radial probability distribution function of the cluster, and $\Delta\hat{V}^{[N]}$ is given by the addition of pairwise evaluations of the

difference-potential for each $p\mathrm{H}_2$ at each position. With this knowledge, Eq. (2.3) becomes

$$\Delta\nu_0^{[N]} = (N-1)\int_0^\infty \Delta V^{1\mathrm{D}}(R)\,\rho^{[N]}(R)\,dR\,, \tag{2.4}$$

in which $(N-1)$ scales the result in response to the normalization of the radial probability distribution function $\rho^{[N]}(R)$, and $\Delta V^{1\mathrm{D}}$ is the difference-PES reduced to one dimension.

Tejeda *et al.* obtained an empirical Lennard-Jones 1-D difference-PES for pure $p\mathrm{H}_2$ clusters of the form $\Delta\mathrm{V}^{1\mathrm{D}}_{\mathrm{emp}}(R) = \alpha R^{-12} + \beta R^{-6}$, with the parameters α and β being determined by fitting their simulated shifts to their Raman spectroscopic data for the frequency shifts, using a weight distribution function $\rho^{[N]}(R)$ obtained from DMC calculations [1]. Generation of higher dimensional distribution functions is computationally prohibitive, necessitating the formulation of reduced 1-D PESs to accompany the radial probability distribution functions.

Hinde's 6-D PES must, therefore, first be reduced to effective 1-D potentials for the $v_{\mathrm{t}} = 0$ and $v_{\mathrm{t}} = 1$ vibrational states. For our *ab initio* study, Fig. 2.1 shows the degrees of freedom expressed in Jacobi coordinates for an interacting pair of $p\mathrm{H}_2$. Those are the coordinates used to represent the 6-D PES, $\mathrm{V}^{6\mathrm{D}}(R, r_1, r_2, \theta_1, \theta_2, \phi)$. In order to obtain a 1D PES, we use the Adiabatic Hindered Rotor (AHR) approach established by H. Li *et al.*, who employed it for calculating vibrational frequency shifts for CO_2 in $p\mathrm{H}_2$ clusters

[6, 22, 59]. It was shown that the AHR approach preserves some rotational details of the interacting system that a simple spherical average does not [6, 60]. In the AHR approach a five-dimensional intermolecular rovibrational Hamiltonian is diagonalized for a set of fixed R values and the locus of lowest eigenvalues define the reduced 1-D PES. In the space fixed frame, the Hamiltonian for the $(H_2)_2$ dimer is

$$\hat{H}^{\mathrm{AHR}}(R) = \hat{h}_1^{\mathrm{rovib}}(\mathbf{r_1}) + \hat{h}_2^{\mathrm{rovib}}(\mathbf{r_2}) + V^{\mathrm{6D}}(R, r_1, r_2, \theta_1, \theta_2, \phi)\,, \tag{2.5}$$

where $\hat{h}_i^{\mathrm{rovib}}$ is the kinetic rovibrational Hamiltonian operator of the ith H_2 monomer molecule. However, we directly incorporate pH$_2$ monomer rovibrational kinetic energies that were precomputed by Ref. 61 instead of specifying our own $\hat{h}_i^{\mathrm{rovib}}$.

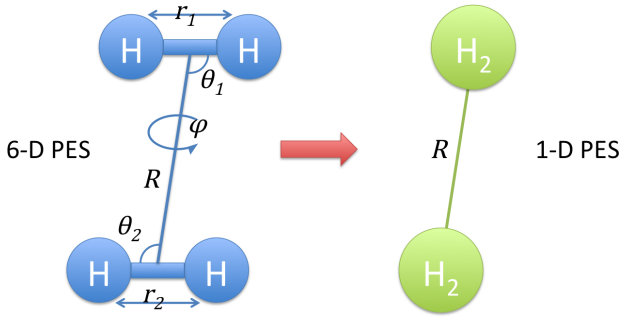

Figure 2.1: Schematic of PES reduction [6].

The pH$_2$ monomer rovibrational radial wavefunctions of Ref. 61 ($\psi_{v_i}^{j_i}(r_i)Y_{j_i m_i}(\theta_i, \phi_i) = \psi_{v_i}^{j_i}(r_i)P_{j_i}^{m_i}(\theta_1)\frac{e^{im_i\phi_i}}{\sqrt{2\pi}}$) are used as primitive basis functions for constructing the Hamiltonian matrix. Only even rotational states ($j = 0, 2, 4$) are allowed due to the singlet nature of

the nuclear spin wavefunction. For the $(p\mathrm{H}_2)_2$ dimer, we symmetrize the basis functions to account for bosonic exchange of rovibrational excitations between the two $p\mathrm{H}_2$ bosonic monomers, such that

$$\begin{aligned} z_{j_1,j_2,m_1,m_2}(r_1,r_2,\theta_1,\phi_1,\theta_2,\phi_2) = \frac{1}{\sqrt{2}}\Bigg(&\psi_{v_1}^{j_1}(r_1)P_{j_1}^{m_1}(\theta_1)\frac{e^{im_1\phi_1}}{\sqrt{2\pi}}\psi_{v_2}^{j_2}(r_2)P_{j_2}^{m_2}(\theta_2)\frac{e^{im_2\phi_2}}{\sqrt{2\pi}} \\ &+ \psi_{v_2}^{j_2}(r_1)P_{j_2}^{m_2}(\theta_1)\frac{e^{im_2\phi_1}}{\sqrt{2\pi}}\psi_{v_1}^{j_1}(r_2)P_{j_1}^{m_1}(\theta_2)\frac{e^{im_1\phi_2}}{\sqrt{2\pi}}\Bigg)\ . \end{aligned} \tag{2.6}$$

We wish to base our 1-D AHR potential on the ground state of the Hamiltonian, so we choose $m_1 + m_2 = 0$ or $m_2 = -m_1$, since otherwise there would exist additional overall rotational kinetic energy in the system. For readability we hide the arguments of z and set $m \equiv m_1$, giving

$$\begin{aligned} z_{j_1,j_2,m} = \frac{1}{\sqrt{2}}\Bigg(&\psi_{v_1}^{j_1}(r_1)P_{j_1}^{m}(\theta_1)\frac{e^{im\phi_1}}{\sqrt{2\pi}}\psi_{v_2}^{j_2}(r_2)P_{j_2}^{-m}(\theta_2)\frac{e^{-im\phi_2}}{\sqrt{2\pi}} \\ &+ \psi_{v_2}^{j_2}(r_1)P_{j_2}^{-m}(\theta_1)\frac{e^{-im\phi_1}}{\sqrt{2\pi}}\psi_{v_1}^{j_1}(r_2)P_{j_1}^{m}(\theta_2)\frac{e^{im\phi_2}}{\sqrt{2\pi}}\Bigg)\ , \end{aligned} \tag{2.7}$$

$$\begin{aligned} z_{j_1,j_2,m} = \frac{1}{\sqrt{2}}\Bigg(&\frac{1}{2\pi}\psi_{v_1}^{j_1}(r_1)\psi_{v_2}^{j_2}(r_2)P_{j_1}^{m}(\theta_1)P_{j_2}^{-m}(\theta_2)e^{im(\phi_1-\phi_2)} \\ &+ \frac{1}{2\pi}\psi_{v_2}^{j_2}(r_1)\psi_{v_1}^{j_1}(r_2)P_{j_2}^{-m}(\theta_1)P_{j_1}^{m}(\theta_2)e^{-im(\phi_1-\phi_2)}\Bigg)\ . \end{aligned} \tag{2.8}$$

We further represent our basis in terms of $\phi = \phi_1 - \phi_2$, since the potential only depends on the relative ϕ. Additionally, let $A = \frac{1}{2\pi}\psi_{v_1}^{j_1}(r_1)\psi_{v_2}^{j_2}(r_2)P_{j_1}^{m}(\theta_1)P_{j_2}^{-m}(\theta_2)$ and

$B = \frac{1}{2\pi}\psi_{v_2}^{j_2}(r_1)\psi_{v_1}^{j_1}(r_2)P_{j_2}^{-m}(\theta_1)P_{j_1}^{m}(\theta_2)$ for more readability, producing

$$z_{j_1,j_2,m} = \frac{1}{\sqrt{2}}\left(Ae^{im\phi} + Be^{-im\phi}\right) \ . \tag{2.9}$$

We then convert to the real form of the spherical harmonic functions for simplicity, using the following relationships

$$R_{j_1,j_2,m} = \begin{cases} \frac{1}{\sqrt{2}}(z_{j_1,j_2,m} + z^*_{j_1,j_2,m}) & m > 0 \\ z_{j_1,j_2,0} & m = 0 \\ \frac{i}{\sqrt{2}}(z_{j_1,j_2,m} - z^*_{j_1,j_2,m}) & m < 0 \end{cases} , \tag{2.10}$$

$$R_{j_1,j_2,m} = \begin{cases} (A+B)\cos(m\phi) & m > 0 \\ \frac{1}{\sqrt{2}}(A+B) & m = 0 \\ -(A-B)\sin(m\phi) & m < 0 \end{cases} . \tag{2.11}$$

We focus on the $m \geq 0$ cases because the symmetry of the basis functions causes cancellation of $m < 0$ matrix elements. Substituting the original component basis functions

into A and B results in

$$R_{j_1,j_2,m} = \frac{1}{\alpha}\left(\psi_{v_1}^{j_1}(r_1)\psi_{v_2}^{j_2}(r_2)P_{j_1}^{m}(\theta_1)P_{j_2}^{-m}(\theta_2)\right. \\ \left. + \psi_{v_2}^{j_2}(r_1)\psi_{v_1}^{j_1}(r_2)P_{j_2}^{-m}(\theta_1)P_{j_1}^{m}(\theta_2)\right)\cos\left(m\phi\right), m \geq 0\,, \tag{2.12}$$

where the normalization factor α is determined explicitly for each matrix element since it varies due to the presence or lack of cross terms when expanding basis functions for certain matrix elements. For the $v_{\mathrm{t}} = 1$ case, we arbitrarily set $v_1 = 1$ and $v_2 = 0$ and the basis function symmetry accounts for the exchange of the vibrational excitation between the two $p\mathrm{H}_2$ monomers. The difference-potential $\Delta V^{1\mathrm{D}}(R)$ is then simply the difference between the two 1-D PESs obtained after direct diagonalization of the Hamiltonian matrix for the $v_{\mathrm{t}} = 0$ and $v_{\mathrm{t}} = 1$ cases.

The radial probability distribution functions of the clusters used in the perturbative model were generated from Langevin equation Path Integral Ground State (LePIGS) simulations, which is a molecular dynamics transmutation of the PIMC PIGS method and offers more simple formulation, since specialized Monte Carlo moves are not required [14, 15]. In turn, the PIMC implementation of PIGS is an improvement over DMC methods because it does not suffer from the population-size bias issue affecting convergence in the latter [62]. Thus, LePIGS has two-fold advantages over the DMC method. Details of the LePIGS formulation is the subject of past theses [3, 5] and papers [14, 15] of our group members,

and will not be covered here. However, it is similar to the finite temperature PIMD formalism presented in Section 3.1 and a brief note is provided at the end of that section to point out key differences. Multiple radial probability distributions for each cluster size were generated for comparison via LePIGS using a number of different $p\mathrm{H}_2$ interaction potentials that were described in earlier studies [39, 63, 64, 65]. The Hinde [39] and Szalewicz [65] potential functions are *ab initio*, while the empirical Silvera-Goldman [64] and Buck [63] potential functions were parameterized to reproduce different types of experiments. A Jastrow-type trial wavefunction [32] is used to simulate cluster sizes of $N = 4$ and above. However, smaller clusters interacted too weakly and dissociated when utilizing this trial wave function. Instead, a direct calculation is used for $N = 2$ and a perturbative approach is used for $N = 3$. The simulation parameters were optimized using the $N = 4$ cluster, resulting in a thermodynamic β value of 1.00 K^{-1}, τ of 0.003 K^{-1}, time step of 5 fs, and 0.2 ps skipped between trajectory output to obtain decorrelated data. The simulations were run for 20 ns.

The output of the simulation trajectories contains a series of positions where the $p\mathrm{H}_2$ are located at each step. These are converted to raw radial probability distributions by calculating all $p\mathrm{H}_2$-$p\mathrm{H}_2$ pair distances. These raw radial probability distributions are directly used to calculate vibrational shifts instead of further processing them into radial probability distribution functions, $p(R)$, in order to maintain accuracy (however, the var-

ious associated $p(R)$ plots are available in the online supplementary material of Ref. 4). The value of the difference-PES is then sampled for each of those raw pair distances and the results averaged. This turns Eq. (2.4) into

$$\Delta\nu_0^{[N]} = \frac{1}{n}\sum_i^n \Delta V^{1\mathrm{D}}\left(R_i^{[N]}\right), \tag{2.13}$$

where n is the number of data points throughout a simulated trajectory for a particular cluster.

2.2 Results and discussion

Vibrational frequency shifts calculated for clusters using various radial probability distributions and two choices of the difference-potential are presented in Fig. 2.2. Results in the upper panel were obtained using the *ab initio* 1-D difference-potential derived from Hinde's 6-D PES [39] while those in the lower panel were generated using the empirical difference-potential of Ref. 1. In the upper panel, the shifts predicted with radial probability distributions obtained from the *ab initio* Hinde potential simulations (blue open circles) are in better agreement with experiment (filled circles) at small cluster sizes ($N < 13$), where shifts predicted with radial probability distributions using the Buck potential (red open squares) simulations are systemically less negative than those from the Hinde results

and from experiment. The agreement is reversed at large cluster sizes ($N > 13$), where shifts obtained from the Hinde radial probability distributions become more negative compared to the experiment, while the Buck distribution prediction at $N = 33$ agrees with the experimental result. A feature of note present in all series of shifts is the change in slope of the trend in shifts with respect to the cluster size at $N = 13$, a magic number.

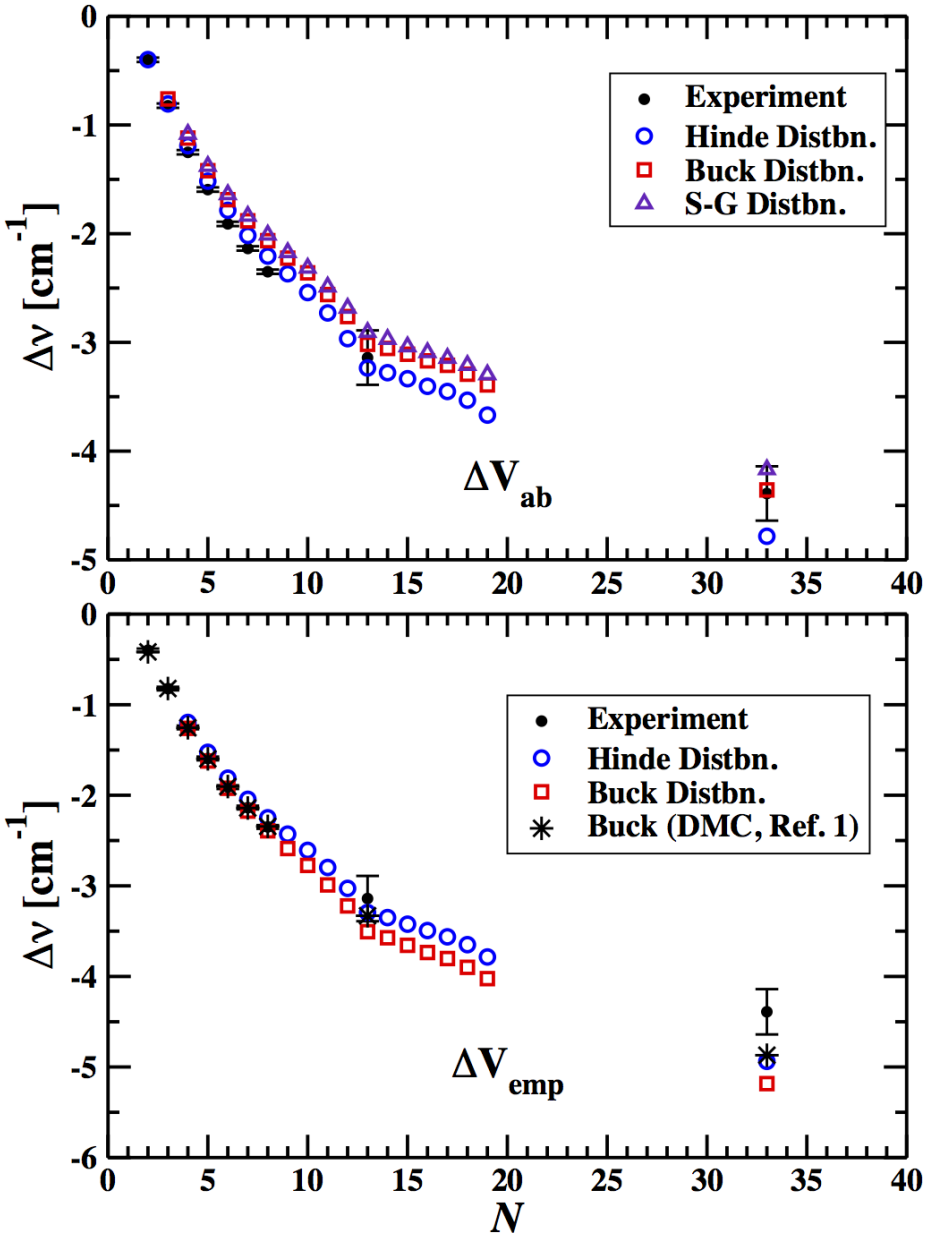

Figure 2.2: Variation with cluster size of predicted vibrational frequency shifts calculated using radial probability distributions generated from different simulation potentials (see key). Upper Panel: Results obtained using *ab initio* difference-potential. Lower Panel: Results obtained using empirical difference-potential. The error bars represent experimental uncertainty to 2σ.

The shifts obtained using distributions generated from four different potentials, together with the experimental values, are tabulated in Table 2.1. The discrepancy between the predicted vibrational frequency shifts at small and large cluster sizes from the different radial probability distributions might be explained by the origin of the potentials. The Hinde potential is an accurate *ab initio* pair potential that should perform better for small clusters where pair interactions dominate. A similar trend in the predicted shifts (not shown in Fig. 2.2) was observed for distributions generated using the *ab initio* Szalewicz potential, in that they are closer to experiment at small cluster sizes but are more negative for $N = 13$ and 33. Perhaps at larger cluster sizes many-body effects become prominent, which these *ab initio* pure pair potentials do not account for. This hypothesis may be supported by the fact that predictions generated using the Silvera-Goldman distributions that perform better than the *ab initio* potential distributions at $N = 33$, shown in the upper panel of Fig. 2.2, considering that the version of the Silvera-Goldman potential used includes an effective many-body term that was fit to solid state data [64]. However, the Buck potential distributions also appears to perform well at large cluster sizes even though it has been obtained from a fit to scattering cross section data, and scattering is an inherently two-body process, meaning the Buck potential is a pure pair potential [63]. This suggests that the good performance of the empirical potential distributions at large cluster sizes may simply be an accident. In fact, the Raman spectral peaks for the $N = 13$

and 33 clusters were unresolved as reported in Ref 1 and we can deduce a large error in the experimentally observed shifts from the broad peak widths, which might account for some of our discrepancy.

Table 2.1: $\Delta\nu_0$ (cm^{-1}) obtained from radial probability distributions generated using different potentials and the two ΔV. Std. errors are ≤ 0.005 cm^{-1}. Experimental (Exp.) are provided for comparison purposes.

		$\Delta\nu_0$, ΔV_{ab}					$\Delta\nu_0$, ΔV_{emp}			
N	Exp. [1]	Hinde	Szalewicz	Buck	S-G	Buck DMC [1]	Hinde	Szalewicz	Buck	S-G
4	-1.251	-1.19	-1.203	-1.118	-1.092	-1.255	-1.197	-1.223	-1.261	-1.180
5	-1.594	-1.517	-1.527	-1.417	-1.385	-1.597	-1.527	-1.566	-1.622	-1.515
6	-1.910	-1.785	-1.800	-1.683	-1.648	-1.904	-1.812	-1.863	-1.923	-1.809
7	-2.136	-2.016	-2.025	-1.878	-1.852	-2.141	-2.045	-2.098	-2.175	-2.043
8	-2.350	-2.206	-2.217	-2.059	-2.028	-2.344	-2.248	-2.315	-2.392	-2.257
9		-2.369	-2.384	-2.219	-2.181		-2.431	-2.493	-2.588	-2.447
10		-2.542	-2.553	-2.367	-2.335		-2.611	-2.674	-2.774	-2.628
11		-2.729	-2.750	-2.545	-2.516		-2.800	-2.887	-2.990	-2.821
12		-2.965	-2.981	-2.752	-2.702		-3.028	-3.109	-3.222	-3.036
13	-3.140	-3.236	-3.276	-3.010	-2.922	-3.330	-3.297	-3.405	-3.507	-3.270
14		-3.279	-3.289	-3.043	-3.002		-3.352	-3.449	-3.573	-3.369
15		-3.334	-3.339	-3.111	-3.064		-3.425	-3.519	-3.657	-3.452
16		-3.404	-3.404	-3.160	-3.122		-3.495	-3.595	-3.736	-3.529
17		-3.453	-3.472	-3.209	-3.175		-3.566	-3.665	-3.802	-3.606
18		-3.532	-3.550	-3.284	-3.250		-3.651	-3.753	-3.899	-3.694
19		-3.668	-3.683	-3.378	-3.333		-3.787	-3.892	-4.024	-3.793
33	-4.390	-4.784	-4.811	-4.423	-4.169	-4.870	-4.941	-5.085	-5.183	-4.794

The change in slope of all the sets of predicted shifts at $N = 13$ occurs because the first solvation shell of the cluster is completed at that point. The first solvation shell makes a larger contribution to the shift compared to outer shells because the peak of its distribution is situated closer to the minimum of the difference-potential, so there is a steeper slope in the trend of shifts as it gets filled. The lower panel in Fig. 2.2 presents shifts predicted

using the empirical Lennard-Jones difference-potential reported by Tejeda *et al.*. It also shows the predicted shifts that they calculated using DMC radial probability distributions generated from the Buck potential. Shifts using the empirical difference-potential are good for the small cluster sizes, as is to be expected, since a fit to observed shifts was used to define the empirical difference-potential. However, they all diverge from experiment in the same manner at large cluster sizes. This may be expected because the empirical difference-potential was defined by a fit to the experimental shifts for only $N = 2 - 8$ clusters that had resolved Raman spectral peaks. Interestingly, our LePIGS radial probability distributions using the Buck potential do not fully reproduce those shifts predicted by the earlier DMC distributions, despite using the same potential. This may be attributed to LePIGS distributions being less noisy than those of Ref. 1.

A comparison of the empirical 1-D difference-potential from Ref. 1 with the one we determine from the 6-D *ab initio* Hinde PES is shown in the upper panel of Fig. 2.3. There is a marked difference in the position and depth of the potentials, despite similar shift predictions for small N, with the empirical difference-potential having a much deeper well that is centered at a much shorter distance. Thus, the difference-potentials probe the radial probability distributions with different emphases, e.g. the region where the empirical difference-potential is most strongly negative is only sampled by the inner tail of the radial probability distribution functions as shown in the upper panel of Fig. 2.3 for an

$N = 8$ cluster, while the region where the empirical difference-potential turns positive is not sampled at all. In contrast, the radial probability distribution functions sample both positive and negative domains of the *ab initio* difference-potential. Additional insight into the reasons why the empirical and *ab initio* difference-potentials give similar predictions of shifts for small clusters is given by the lower panel of Fig. 2.3. It depicts plots of the integrand of Eq. 2.4 for the two difference-potentials and the $N = 8$ distribution. The shapes of the empirical and ab initio integrand curves are different due to the wells and repulsive walls of the difference-potentials matching up with different portions of the radial probability distribution function, however the net area under the curves representing the shifts is similar. The *ab initio* difference-potential is clearly the correct one and this result illustrates the difficulty of determining empirical difference-potentials using bulk-averaging.

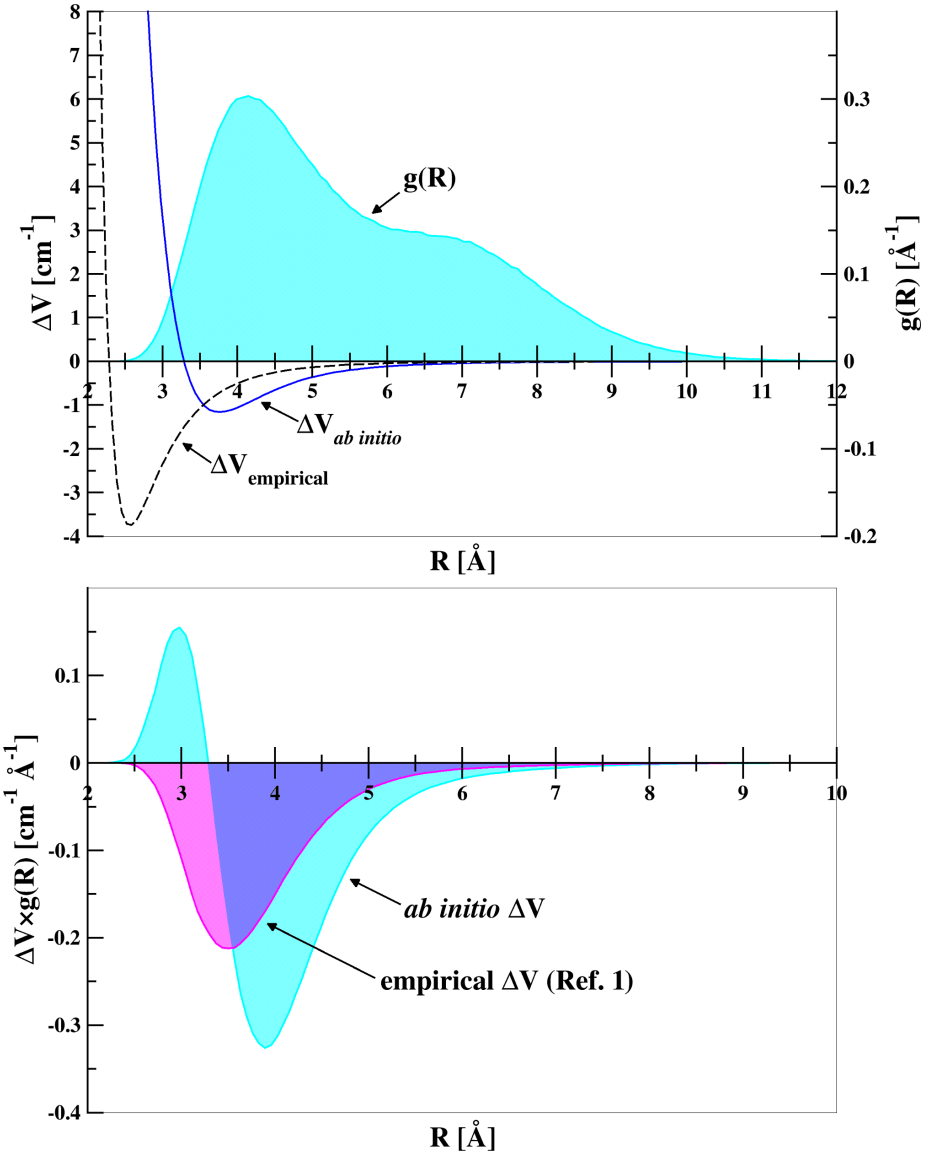

Figure 2.3: Upper Panel: Comparison of difference-potentials beside an $N = 8$ radial probability distribution function. Lower Panel: The integrand of Eq. (2.4) for the two difference-potentials when combined with the $N = 8$ radial probability distribution function.

A comparison of the four 1-D potentials used to generate the radial probability distri-

butions is presented in Fig. 2.4. The top panel compares a plot of the 1-D $v_t = 0$ potential that was generated from the 6-D *ab initio* Hinde PES with the empirical Buck potential. The bottom panel shows plots of the difference between the Hinde potential and other 1-D potentials used to generate radial probability distribution functions, together with a plot of the Hinde difference-potential ΔV_{ab} (black solid line). In the well region of the potentials, the Buck (red dashed line) and Silvera-Goldman (purple dash-dotted line) potentials are less attractive than the Hinde potential. This means the pH$_2$ are less bound to the vicinity of the potential minima, and their radial probability distributions can seep further inward or outward. This will result in a less negative shift for the Buck and Silvera-Goldman potentials relative to the Hinde potential because there is less probing of the attractive well of the difference-potential. In addition, in the region of the repulsive walls the Buck and Silvera-Goldman potentials are less repulsive than the Hinde potential, indicating they have gentler slopes. This means the pH$_2$ particles can approach closer together in those cases and this will give rise to larger values for the radial probability distributions at short distances. In this region the difference-potential is positive and so will again tend to produce less negative shifts when combined with the larger radial probability distributions of the Buck and Silvera-Goldman cases relative to the Hinde potential case. The Szalewicz potential (green dotted line) is slightly deeper than the Hinde potential in the region of the potential well, meaning the radial probability distributions are more concentrated at

the potential minimum, and more positive in the later part of the repulsive wall region, indicating a steeper slope. This leads to slightly more negative shifts relative to the Hinde potential case.

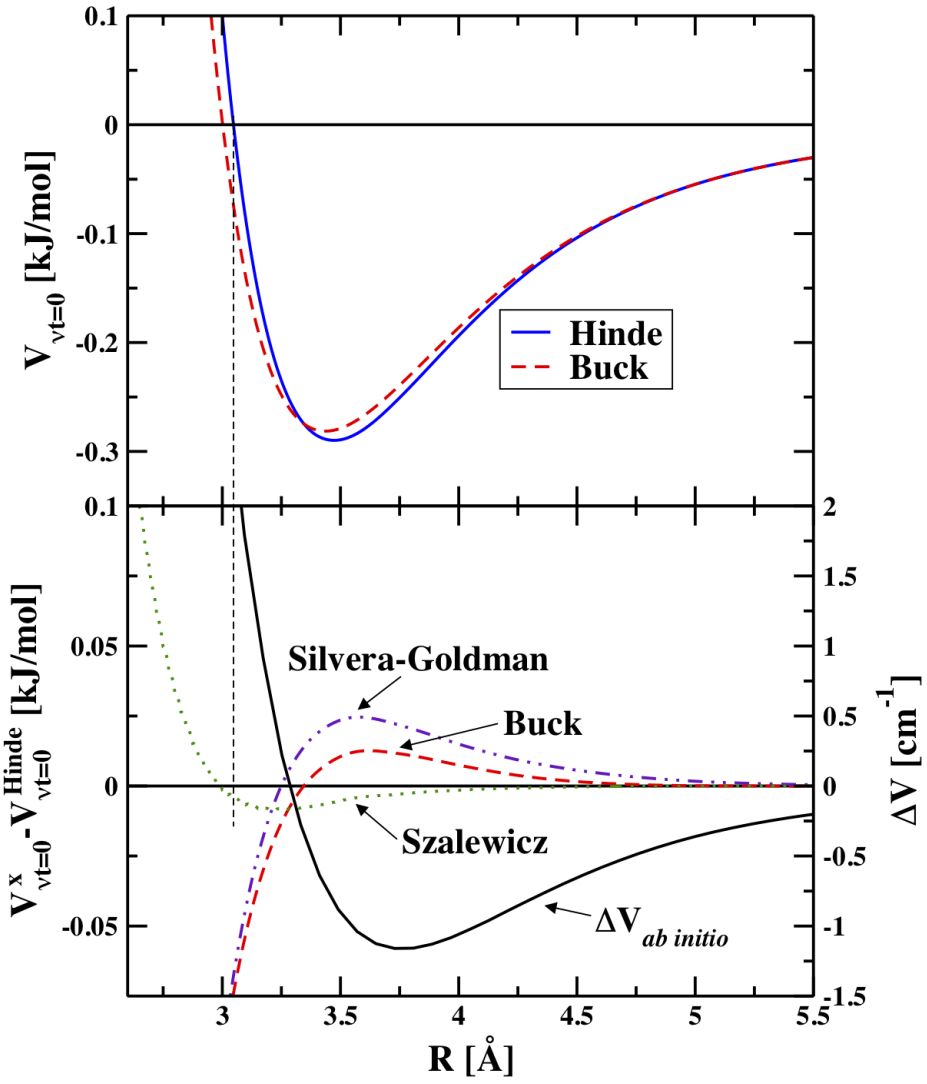

Figure 2.4: Upper Panel: Key ground vibrational state 1-D potentials used in simulation to generate radial probability distributions. Lower Panel: Differences of various 1-D potentials with the *ab initio* potential of Hinde

Simulations to obtain a dimer or trimer radial probability distribution with LePIGs using a Jastrow trial wavefunction is not possible because the clusters rapidly dissociate. Instead, for the $N = 2$ dimer the eigenvalue problem in R was solved using exact diagonalization with a Colbert-Miller DVR basis [66] after reduction of the Hinde PES for the ground and first vibrational states, with the resulting energy difference between them giving the vibrational frequency shift for the dimer. As shown in Table 2.2, the result is in good agreement experiment and with the binding energy calculations Hinde carried out himself.

Table 2.2: Direct calculation of shift for $p\mathrm{H}_2$ dimer

Method	$\Delta\nu_0$ (cm^{-1})
Observed	-0.400
Tejeda et al.: $\Delta V_{\mathrm{emp}}(R)$ [1]	-0.417
Hinde: direct *ab initio* [39]	-0.405
Our Work: $\Delta V_{\mathrm{ab}}(R)$	-0.399

To obtain the shift for the $N = 3$ trimer, we employed the usual perturbative approach using the *ab initio* difference-potential, with the trimer ground state radial pair density being obtained from exact diagonalization using the approach of Refs. 32, 67–68. The ground state interaction potential used in the evaluation was either the ground vibrational state Hinde potential or the Buck potential. The result in Table 2.3 shows good agreement with experiment for the Hinde potential case.

For comparison purposes, we have also calculated the vibrational shift based on the

Table 2.3: Perturbative calculation of shift for $p\mathrm{H}_2$ trimer

Method	$\Delta\nu_0$ (cm^{-1})
Observed	-0.822
Tejeda et al.: $\Delta V_{\mathrm{emp}}(R)$ [1]	-0.821
$\Delta V_{\mathrm{ab}}(R)$, empirical Buck $V_{v_{\mathrm{t}}=0}$	-0.748
$\Delta V_{\mathrm{ab}}(R)$, *ab initio* Hinde $V_{v_{\mathrm{t}}=0}$	-0.807

spherical averaging of the 6D Hinde hydrogen potential for the difference potential. Results are presented in the online supplementary material of Ref. 4. The AHR results generally agree better with experiments for clusters with $3 \leq N \leq 8$. For the dimer ($N = 2$) the AHR and spherically averaged results are essentially the same within errors. The spherically averaged difference potential, however, leads to a better agreement with experiment for $N = 13, 33$. Future work will focus on a detailed analysis of these differences.

2.3 Conclusions

In conclusion, we have described the development and application of *ab initio* reduced PESs for the prediction of Raman spectral shifts of pure para-hydrogen clusters. The results confirm that the 6-D Hinde PES is a high quality pair PES, with its reduced 1-D surfaces predicting shifts in good agreement with experiment for small cluster sizes including the $(p\mathrm{H}_2)_2$ dimer. However, for a cluster size of $N = 33$, the shift predicted using radial probability distributions generated from the ground state reduced Hinde potential was

more negative than the experimental one. In that domain, radial probability distributions generated from empirical potentials performed better, through accident in the case of the Buck potential or possibly by accounting for many-body effects that may become non-negligible after the first solvation shell in the case of the Silvera-Goldman potential. By combining radial probability distributions generated from these empirical potentials and the *ab initio* difference-potential, we are able to extend the prediction of shifts to large $p\mathrm{H}_2$ cluster sizes that have not yet been observed, such as those we have shown between $N = 13-33$. If this were done with the empirical Lennard-Jones difference-potential, those predicted shifts would diverge from experiment, as is expected at the large cluster sizes that it was not parameterized for. However, it is clear that our *ab initio* difference-potential is closer to physical reality. Future work will involve the assessment of the importance of many-body effects that may account for the difference between our predictions and the result of experiment for $N = 33$. In that regard, the work of Hinde [69] on three-body effects on the energy of the H_2 trimer will be of interest. One other possibility worth exploring is an incorrect assignment of the experimental vibrational shift for $N = 33$ as suggested in Ref. 37.

Chapter 3

Conformation and energy analysis of sugars using path integral molecular dynamics

3.1 Theory

Molecular dynamics can reproduce statistical ensemble averages of properties, such as as our $\langle {}^3J_{\mathrm{H,H}} \rangle$ of interest, through the ergodic hypothesis. The hypothesis states that a member of an ensemble allowed to sufficiently evolve in time will visit all the accessible states occupied by the entire ensemble, such that the time average of a property across the trajectory of a single simulated copy of the system will eventually converge to the ensemble

average. This is expressed formally as

$$\langle A \rangle = \lim_{T \to \infty} \frac{1}{T} \sum_{t=0}^{t=T} A(p(t), q(t)) \ , \tag{3.1}$$

where q is the particle's position and p is its momentum. We particularly want to replicate the canonical ensemble where the number of particles, volume, and temperature are fixed (NVT), in order to mirror the conditions of experiments.

The time evolution of the system is given by numerically integrating Hamilton's Equations of Motion,

$$\frac{\mathrm{d}p(t)}{\mathrm{d}t} = -\frac{\partial H(p,q)}{\partial q} \ , \tag{3.2}$$

$$\frac{\mathrm{d}q(t)}{\mathrm{d}t} = \frac{\partial H(p,q)}{\partial p} \ , \tag{3.3}$$

through molecular dynamics, where $H(p, q)$ is the classical mechanics Hamiltonian that describes the total energy of the system. For the quantum nuclei case, the path integral formulation allows us to transform the quantum canonical partition function into one that describes a classical analogue system with an effective Hamiltonian [12] that we can subject to the same equations of motion.

Following the notes of Tuckerman [70] and Refs. 41 and 3, we start with the quantum

canonical partition function in the position representation for one particle,

$$Z = \int dq \, \langle q | \, e^{-\beta \hat{H}} \, | q \rangle = \int dq \, \langle q | \, e^{-\beta (\hat{K} + \hat{V})} \, | q \rangle \ , \tag{3.4}$$

where $\hat{K}$ and $\hat{V}$ are kinetic and potential energy operators, respectively. It is not possible to solve this expression for Z directly because $\hat{K}$ and $\hat{V}$ do not generally commute in the quantum case. The foundations for an approximate approach to the solution is provided by Trotter factorization [71] to separate the exponential into P slices,

$$Z = \lim_{P \to \infty} \int \mathrm{d}q \langle q | \left(e^{-\frac{\beta}{2P} \hat{V}} e^{-\frac{\beta}{P} \hat{K}} e^{-\frac{\beta}{2P} \hat{V}} \right)^P | q \rangle = \lim_{P \to \infty} \int \mathrm{d}q \langle q | \hat{\Omega}^P | q \rangle \ . \tag{3.5}$$

Insertion of the identity operator $I = \int \mathrm{d}q |q\rangle\langle q|$ between each of the P applications of the Ω operator in the power allows for further break down into a cyclic form given by

$$Z = \int dq_1 \ldots dq_P \, \langle q_1 | \, \hat{\Omega} \, | q_2 \rangle \ldots \langle q_P | \, \hat{\Omega} \, | q_1 \rangle \ . \tag{3.6}$$

The potential energy part of the $\hat{\Omega}$ operator simply evaluates to the potential energy at each q_i coordinate. Looking at the individual integrands in the expression for Z, this results

in

$$\begin{aligned}\langle q_i|\,\Omega\,|q_{i+1}\rangle &= \langle q_i|\,e^{-\frac{\beta}{2P}\hat{V}}e^{-\frac{\beta}{P}\hat{K}}e^{-\frac{\beta}{2P}\hat{V}}\,|q_{i+1}\rangle \\ &= e^{-\frac{\beta}{2P}V(q_i)}\,\langle q_i|\,e^{-\frac{\beta}{P}\hat{K}}\,|q_{i+1}\rangle\,e^{-\frac{\beta}{2P}V(q_{i+1})}\ .\end{aligned} \tag{3.7}$$

The remaining matrix element containing the kinetic energy operator can be evaluated by applying a complete set of momentum eigenstates:

$$\langle q_i|\,e^{-\frac{\beta}{P}\hat{K}}\,|q_{i+1}\rangle = \left(\frac{mP}{2\pi\beta\hbar^2}\right)^{\frac{1}{2}} e^{-\frac{mP}{2\beta\hbar^2}(q_{i+1}-q_i)^2}\ . \tag{3.8}$$

Substituting these results back into the expression for Z gives the discretized path integral form of the quantum mechanical canonical partition function,

$$\begin{aligned}Z = \lim_{P\to\infty}\left(\frac{mP}{2\pi\beta\hbar^2}\right)^{\frac{P}{2}}\int dq_1\ldots dq_P \\ \times\exp\left\{-\beta\sum_{i=1}^{P}\left[\frac{mP}{2\beta^2\hbar^2}\left(q_{i+1}-q_i\right)^2+\frac{1}{P}V\left(q_i\right)\right]\right\}_{q_{P+1}=q_1}\ .\end{aligned} \tag{3.9}$$

In molecular dynamics, the P slices in the partition function can be represented by identical classical particles or *beads* that compose a kind of cyclical polymer to stand in for a quantum atomic nucleus. The exponential for the partition function now contains an effective potential with a term for the potential energy at each bead that is governed

by interactions with equivalent beads of neighbouring atoms, i.e. what would normally be given through the force field used in molecular dynamics, but now distributed over the beads, and a term resembling a spring potential energy between two neighhbouring beads of the same atom that gives rise to the polymer analogy,

$$V_{\text{eff}}(q_1, \ldots, q_p) = -\sum_{i=1}^{P} \left[\frac{m\omega_P^2}{2} \left(q_{i+1} - q_i\right)^2 + \frac{1}{P} V\left(q_i\right) \right]_{q_{P+1}=q_1} , \tag{3.10}$$

where $\omega_P = \sqrt{P}/\beta\hbar$. We can introduce fictitious classical momenta to enable molecular dynamics sampling of states from the partition function [13] by transforming the partition function prefactor into a set of P Gaussian integrals,

$$Z = \lim_{P\to\infty} \int dp_1 \ldots dp_P \int dq_1 \ldots dq_P \times \exp\left[-\beta \sum_{i=1}^{P} \frac{p_i^2}{2Pm_i} + V_{\text{eff}}(q_1, \ldots, q_p) \right]_{q_{P+1}=q_1} , \tag{3.11}$$

where we arbitrarily set $m_i = m$, i.e. the full mass of the atom. Observing within the partition function exponential then gives the fictitious effective classical Hamiltonian as

$$H_P(p,q) = \sum_{i=1}^{P} \left[\frac{p_i^2}{2Pm} + \frac{m\omega_P^2}{2} \left(q_{i+1} - q_i\right)^2 + \frac{1}{P} V\left(q_i\right) \right]_{q_{P+1}=q_1} , \tag{3.12}$$

$$H_P(p,q) = \sum_{i=1}^{P} \left[H_{P,\text{free}}(p,q) + \frac{1}{P} V\left(q_i\right) \right]_{q_{P+1}=q_1} , \tag{3.13}$$

where $H_{P,\text{free}}$ is the free ring polymer component. A beneficial aspect of the path integral approach in addition to this classical isomorphism is that while the exact quantum situation is sampled as $P \to \infty$, a limit can be specified on the number of beads used in simulations, making the simulations computationally tractable. Note that $P = 1$ corresponds to a classical simulation.

However, problems exists when attempting to integrate through the equations of motion with the effective Hamiltonian in Cartesian space [72]. A primary concern is that increasing P leads to stiffening of the harmonic spring term, which begins to dominate interactions and results in non-ergodicity. A solution to this issue is to transform the position and momenta of the discrete cyclic path into normal mode coordinates via a Fourier expansion in order to uncouple the spring term into P independent harmonic oscillators [73],

$$\tilde{p}_i^{(k)} \leftarrow \sum_{j=1}^{P} p_i^{(j)} C_{jk}, \;\; \tilde{q}_i^{(k)} \leftarrow \sum_{j=1}^{P} q_i^{(j)} C_{jk} \;, \tag{3.14}$$

where the coefficients for even P are

$$C_{jk} = \begin{cases} \sqrt{1/P} & k = 0 \\ \sqrt{2/P}\cos(2\pi jk/P) & 1 \leq k \leq P/2 - 1 \\ \sqrt{1/P}(-1)^j & k = P/2 \\ \sqrt{2/P}\sin(2\pi jk/P) & P/2 + 1 \leq k \leq P - 1 \end{cases} . \tag{3.15}$$

The free ring polymer part of the Hamiltonian for N particles in the normal mode representation is then

$$\tilde{H}_{P,\text{free}}(\tilde{p}, \tilde{q}) = \sum_{i=1}^{N} \sum_{k=0}^{P-1} \left(\frac{[\tilde{p}_i^{(k)}]^2}{2Pm_i} + \frac{m_i \omega_k^2}{2} [\tilde{q}_i^{(k)}]^2 \right) , \tag{3.16}$$

where $\omega_k = 2\omega_P \sin^2(k\pi/P)$. New normal mode equations of motions can be derived from this Hamiltonian to propagate the free ring polymer component in a separate step from evaluating the potential energy component in Cartesian coordinates, thus avoiding the influence of a problematic harmonic difference term.

To maintain constant temperature and sample the canonical ensemble, we employ a thermostat based on a white-noise Langevin equation [74] formulated in the path integral normal mode representation (PILE) [73]. The PILE thermostat stochastically models the

interaction between the system of interest and a heat bath by modifying the momenta of system particles according to

$$\tilde{p}_i^{(k)} \leftarrow c_1^{(k)} \tilde{p}_i^{(k)} + \sqrt{\frac{m_i P}{\beta}} c_2^{(k)} \xi_i^{(k)} \ , \tag{3.17}$$

where ξ is a random number sampled from a Gaussian distribution for each degree of freedom and path integral normal mode to account for high-energy collisions between the particle and the bath particles, and the coefficients are

$$c_1^{(k)} = e^{\Delta t/2 \gamma^{(k)}} \ , \tag{3.18}$$

$$c_2^{(k)} = \sqrt{1 - [c_1^{(k)}]^2} \ . \tag{3.19}$$

The friction parameter γ emulates damping of a system particle's motion by heat bath particles. For nonzeroth order modes, it is analytically chosen to minimize the autocorrelation time of $\langle H_{P,\text{free}}(0) H_{P,\text{free}}(t) \rangle$ and so allows for the fastest sampling of statistically distinct free ring polymer energy states. These free ring polymer friction coefficients should also allow for similar performance with the interacting ring polymer since the high frequency internal modes are sufficiently decoupled from the influence of the potential. The optimal

friction values are found to be

$$\gamma^{(k)} = \begin{cases} 1/\tau_0 & k = 0 \\ \\ 2\omega_k & k > 0 \end{cases}, \tag{3.20}$$

where the centroid friction $\gamma^{(0)}$ is determined numerically by extracting the average centroid decorrelation time τ_0 from a series of autocorrelation functions $\langle A(0)A(t)\rangle$ of a property of interest sampled from a Ring Polymer Molecular Dynamics (RPMD) format [75]. In RPMD, several microcanonical simulations with constant system energy (NVE) are undertaken with initial configurations obtained from an NVT simulation in order to obtain dynamical properties (whereas PIMD gives equilibrium properties). Autocorrelation functions are then built from each of the NVE simulations. For the sugar application, we denote centroid dihedral angles as the dihedral angles calculated from the centroid positions of the beads representing the sugar's nuclei and we determine a centroid friction from autocorrelation functions of these centroid dihedral angles.

The normal mode transformation and the PILE thermostat are incorporated into the popular Velocity Verlet algorithm [76, 77] used in molecular dynamics to provide the fol-

lowing scheme for numerically integrating through the equations of motion

$$\tilde{p}_i^{(k)}(t) \leftarrow \sum_{j=1}^{P} p_i^{(j)}(t) C_{jk} \ , \tag{1}$$

$$\tilde{p}_i^{(k)}(t) \leftarrow c_1^{(k)} \tilde{p}_i^{(k)}(t) + \sqrt{\frac{m_i P}{\beta}} c_2^{(k)} \xi_i^{(k)} \ , \tag{2}$$

$$p_i^{(j)}(t) \leftarrow \sum_{j=1}^{P-1} C_{jk} \tilde{p}_i^{k}(t) \ , \tag{3}$$

$$p_i^{(j)}(t + \Delta t/2) \leftarrow p_i^{(j)}(t) + \frac{\Delta t}{2} \nabla V(q_i^{(j)}(t)) \ , \tag{4}$$

$$\tilde{p}_i^{(k)}(t + \Delta t/2) \leftarrow \sum_{j=1}^{P} p_i^{(j)}(t + \Delta t/2) C_{jk} \ , \tag{5}$$

$$\tilde{q}_i^{(k)}(t) \leftarrow \sum_{j=1}^{P} q_i^{(j)}(t) C_{jk} \ , \tag{6}$$

$$\begin{pmatrix} \tilde{p}_i^{(k)}(t + \Delta t/2) \\ \tilde{q}_i^{(k)}(t + \Delta t) \end{pmatrix} \leftarrow \begin{pmatrix} \cos(\omega_k \Delta t) & -m\omega_k \sin(\omega_k \Delta t) \\ (m\omega_k)^{-1} \sin(\omega_k \Delta t) & \cos(\omega_k \Delta t) \end{pmatrix} \begin{pmatrix} \tilde{p}_i^{(k)}(t + \Delta t/2) \\ \tilde{q}_i^{(k)}(t) \end{pmatrix} \ , \tag{7}$$

$$p_i^{(j)}(t + \Delta t/2) \leftarrow \sum_{j=1}^{P-1} C_{jk} \tilde{p}_i^{k}(t + \Delta t/2) \ , \tag{8}$$

$$q_i^{(j)}(t + \Delta t) \leftarrow \sum_{j=1}^{P-1} C_{jk} \tilde{q}_i^{k}(t + \Delta t) \ , \tag{9}$$

$$p_i^{(j)}(t+\Delta t) \leftarrow p_i^{(j)}(t+\Delta t/2) + \frac{\Delta t}{2}\nabla V(q_i^{(j)}(t+\Delta t)) \ , \tag{10}$$

$$\tilde{p}_i^{(k)}(t+\Delta t) \leftarrow \sum_{j=1}^{P} p_i^{(j)}(t+\Delta t) C_{jk} \ , \tag{11}$$

$$\tilde{p}_i^{(k)}(t+\Delta t) \leftarrow c_1^{(k)}\tilde{p}_i^{(k)}(t+\Delta t) + \sqrt{\frac{m_i P}{\beta}} c_2^{(k)} \xi_i^{(k)} \ , \tag{12}$$

$$p_i^{(j)}(t+\Delta t) \leftarrow \sum_{j=1}^{P-1} C_{jk}\tilde{p}_i^{k}(t+\Delta t) \ , \tag{13}$$

where Δt is the time step between subsequent states in the trajectory, ∇V is the gradient of the potential, or force, acting on the particles, and where Step 7 contains the normal mode equations of motion.

In addition to dihedral angles and proton spin coupling constants, we are also interested in observing the convergence of the internal energy of our sugar system as function of path integral beads as another property to gauge the "quantumness" of the system. Given that

$$\langle \epsilon \rangle = -\left(\frac{\partial \ln Z}{\partial \beta}\right)_{NVT} = \frac{1}{Z}\left(\frac{mP}{2\pi\beta\hbar^2}\right)^{\frac{P}{2}} \int dq_1 \ldots dq_P \epsilon_{\mathrm{T}} e^{-\beta V_{\mathrm{eff}}} \ , \tag{3.21}$$

a primitive or Barker estimator of the internal energy is obtained as [78]

$$\epsilon_{\mathrm{T}} = \frac{3NP}{2\beta} - \frac{Pm}{2\hbar^2\beta^2}\sum_{j=1}^{P}(q_j - q_{j+1})^2 + \frac{1}{P}\sum_{j=1}^{P}V(q_j)\ . \tag{3.22}$$

However, the mean square fluctuations of this estimator increase $\propto \frac{P}{2\beta^2}$ due to the spring term. The virial theorem can be used to recast the spring energy in a form with better variance, giving the centroid virial estimator [40]

$$\epsilon_{\mathrm{CV}} = \frac{3N}{2\beta} + \frac{1}{2P}\sum_{j=1}^{P}(q_j - \bar{q})\cdot V'(q_j) + \frac{1}{P}\sum_{j=1}^{P}V(q_j)\ , \tag{3.23}$$

where $\bar{q} = \frac{1}{P}\sum_{j=1}^{P} q_j$ is the path centroid.

We compare the temperature dependence of both quantum and classical internal energies with analytic quantum and classical harmonic oscillator models based on the sugar's vibrational frequencies. Biomolecules are often treated using a harmonic approximation via normal mode analysis of the system's Hessian [79, 80], and our comparison allows us to establish whether this approach will be sufficient to describe the sugar's energy. The energy of a classical harmonic oscillator is given by

$$\langle\epsilon_{\mathrm{ClHO}}\rangle = V_{\min} + \frac{3}{\beta} + \sum_i \frac{1}{\beta} \tag{3.24}$$

where i is the vibrational mode and V_{min} is the potential energy minimum of the simulated system. V_{min} is required to match up the zero of the harmonic model with the contribution from a multitude of nonharmonic interactions in a simulated system. The energy of a quantum harmonic oscillator on the other hand is

$$\langle \epsilon_{\text{QHO}} \rangle = V_{\text{min}} + \sum_i h\nu_i \left(\frac{1}{e^{\beta h\nu_i} - 1} + \frac{1}{2} \right) + \frac{3}{\beta} \,, \tag{3.25}$$

where ν_i is the frequency of vibrational mode i.

Finally, a brief note on how the LePIGS method utilized in Chapter 2 differs from PIMD. Path Integral Ground State theory leads to an open path of beads representing the quantum ground state nuclei instead of a closed ring. The end beads of the open path are acted upon by trial wavefunctions that help to propagate the system to the true ground state distribution [14, 15]. Sampling from the PIGS partition function via molecular dynamics using LePIGS is also similar to the PILE PIMD integration scheme, except that a cosine transform into the normal mode coordinates is required instead of a Fourier transform [14, 15].

3.2 Methods

The Molecular Modelling Toolkit (MMTK) [81] software is central to running and analyzing our simulations. It consists of libraries for the Python scripting language that allow for easy object-oriented setup and analysis of simulations by scientific users. Scientists are able to use programming semantics related to molecular modelling terminology familiar to them. However, performance critical aspects of the tasks, such as integration, are executed by passing information to fast C language code under the hood. MMTK comes prepackaged with many useful algorithms for molecular modelling tasks including solvation, support for arbitrary force fields, and of special interest to us, PIMD, which was codeveloped for the software by C. Ing while he was a member of our research group [41, 2].

Simulations take place in an *infinite universe* where the sugar is free to roam without interaction with other particles and emulates gas phase behaviour. GLYCAM-Web Carbohydrate Builder was used to generate the topology and initial configuration of the methyl β-D-arabinofuranoside sugar, including partial charges on the atoms [82]. The AMBER 99 [83] force field is used to evaluate forces with carbohydrate parameter modification GLYCAM06 [84] to tailor the force field parameters to our sugar application.

Integration of the equations of motion achieves a speed up by using as large of a time step as possible while maintaining energy stability of the system. The primitive and

centroid virial estimators can be compared to determine a maximum time step. Both estimators should provide similar energy values and where they diverge in relation to the size of the time step indicates instability introduced by too large of a step because particles may jump too deeply into the repulsive wall region of their potential [2]. NVT simulations with different time steps were run at 200 K with 10 fs of equilibration, 300 ps of production and a centroid friction of 3.45×10^{-2} ps^{-1} from preliminary decorrelation analysis of the sugar's exocylic dihedral angle. A safe time step of 0.5 fs that was below the divergence point of the estimators was selected for use in subsequent simulations.

To determine the optimal centroid friction for our sugar system, we first ran several simulations with different temperatures and path integral beads following the RPMD scheme outlined earlier. We wished to establish whether those parameters have an effect on the centroid friction. For each value of the parameters, 10 ps of equilibration and 300 ps of production are carried out in NVT simulation, and no steps skipped between outputs to the trajectory files. The centroid friction used in these initial NVT simulations are essentially arbitrary, only set so that the thermostat samples the canonical ensemble, even if the sampling may be inefficient. However, we alternatively tried $0.01/\Delta t$ applied to all atoms or $m_i \times 0.01$ $\frac{\text{mol}}{\text{ps}\cdot\text{g}}$, where m_i is the molar mass of atom i in g/mol, applied separately on each atom as preliminary sources of centroid frictions (based on Müser's rule [85] for 1st method and K. Hinsen's MMTK example file [81] for 2nd method). The maximum

dihedral angle decorrelation time of the production NVT trajectory is determined from the endocyclic centroid dihedrals calculated from the centroid positions of the sugar nuclei throughout the trajectory, and this decorrelation time is used to select statistically distinct initial configurations from the production NVT trajectory. 100 samples of NVE simulation are then carried out from an equivalent number of these initial configurations for 5 ps each. Dihedral angle decorrelation times are calculated for the centroids of each NVE trajectory and averaged. The inverse of this average decorrelation time gives the centroid friction, which we apply to all atoms in subsequent simulations.

A similar format was used to determine the skip steps required for general NVT simulation of the sugar to provide statistically distinct states in simulation trajectories. Except in this case, the sampling simulations are also NVT simulations and the objective centroid friction established previously is supplied to both the preliminary NVT simulation and the sampling simulations. The simulation lengths are the same as the previous centroid friction determination scheme, while the number of samples is 50 and the skip steps is chosen to be the overall average of the endocyclic centroid dihedral angle decorrelation time of the sample trajectories plus two standard deviations. We add the standard deviations in order to ensure that our skip steps are large enough to provide decorrelated states for all cases. The temperature and number of path integral beads are varied as before for these simulations. In a separate study, the centroid friction is also varied several magnitudes around

the value determined previously in order to confirm whether it is optimum and observe the thermostat's robustness. An optimum centroid friction should give a minimum NVT decorrelation time and thus fewer steps will need to be skipped such that more usable data is obtained with a shorter simulation length.

The convergence properties of dihedral angle distributions with path integral beads at different temperatures was examined using NVT simulation after the centroid friction and skip steps were established. The simulations involved 200 ps of equilibration and 20 ns of production. Visual inspection of the resulting distribution plots was used to establish convergence. Coupling constants were then calculated for each distribution using the Karplus equations of Ref. 55.

Energies converge at a different rate compared to structural properties, so a separate convergence study of the quantum energy across a range of temperatures was undertaken using the centroid virial estimator. The Trotter factorization error is $\mathcal{O}\tau^2$, where $\tau = \frac{1}{TP}$, so we expect the energy convergence to be quadratic for small *tau*. For each temperature, several simulations of varying beads were run for 100 ps of production if $T \leq 405$ K or 1 ns of production if $T \geq 405$ K to obtain energy data points throughout this quadratic convergence region. A fitting function was then used to extrapolate the energies to the $\tau = 0$ full quantum value. We attempted to collect data over a similar range of τ for all temperatures. We noticed an unexpected divergence of the energies after $\sim \tau = 4{\times}10^{-5}$

K^{-1} and discuss this in Section 3.3. Decreasing the time step to 0.25 fs fixed the issue and we began using this new time step for subsequent path integral simulations. After developing familiarity with the behaviour and analysis of the energy convergence data, extrapolated quantum energies and classical energies were calculated for a large range of temperatures and compared to the analytic energies from the quantum and classical harmonic oscillator models.

3.3 Results and discussion

The primitive and centroid virial estimates of the average internal energy at 200 K for different simulation time steps are shown in Fig. 3.1. The results of the two estimators begin to diverge after $\Delta t = 0.8$ fs, where the system becomes unstable. A much smaller time step of $\Delta t = 0.5$ fs was thus selected for a margin of safety, since we also simulate higher temperatures where the faster motion of particles increases the likelihood of instability with too large of a time step.

Some representative centroid dihedral angle correlation functions and their exponential fitting functions calculated from samples in the RPMD centroid friction determination scheme are displayed in Fig. 3.2. Fig. 3.3 contains the average decorrelation times of the NVE samples for each condition of temperature and number of beads and for each endo-

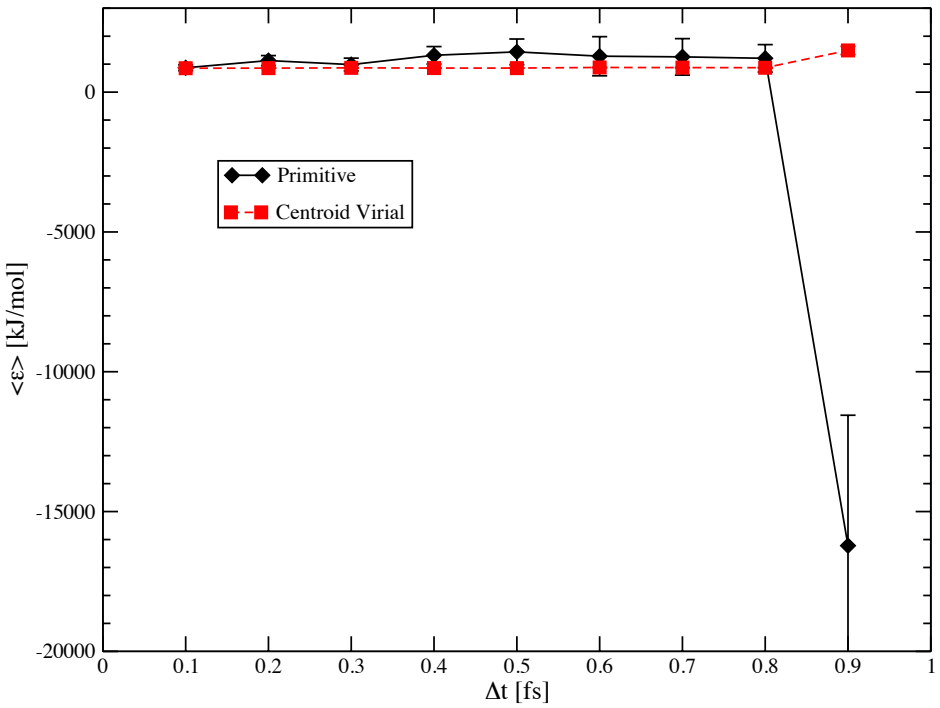

Figure 3.1: Finding a safe Δt based on the divergence of average primitive and centroid virial energy estimates

cyclic dihedral angle. However, subsequent NVT decorrelation analysis for determining skip steps, which will be discussed shortly, indicates that the change in NVE decorrelation times with temperature is small and will not have appreciable effect on the NVT decorrelation times. Therefore, we move ahead with an overall average NVE decorrelation time for all conditions and dihedral angles of $\tau_0^{\mathrm{NVE}} \sim 0.12$ ps, which gives a centroid friction of $\gamma^{(0)} = 8.33$ ps^{-1} that we use as a standard for subsequent NVT simulations.

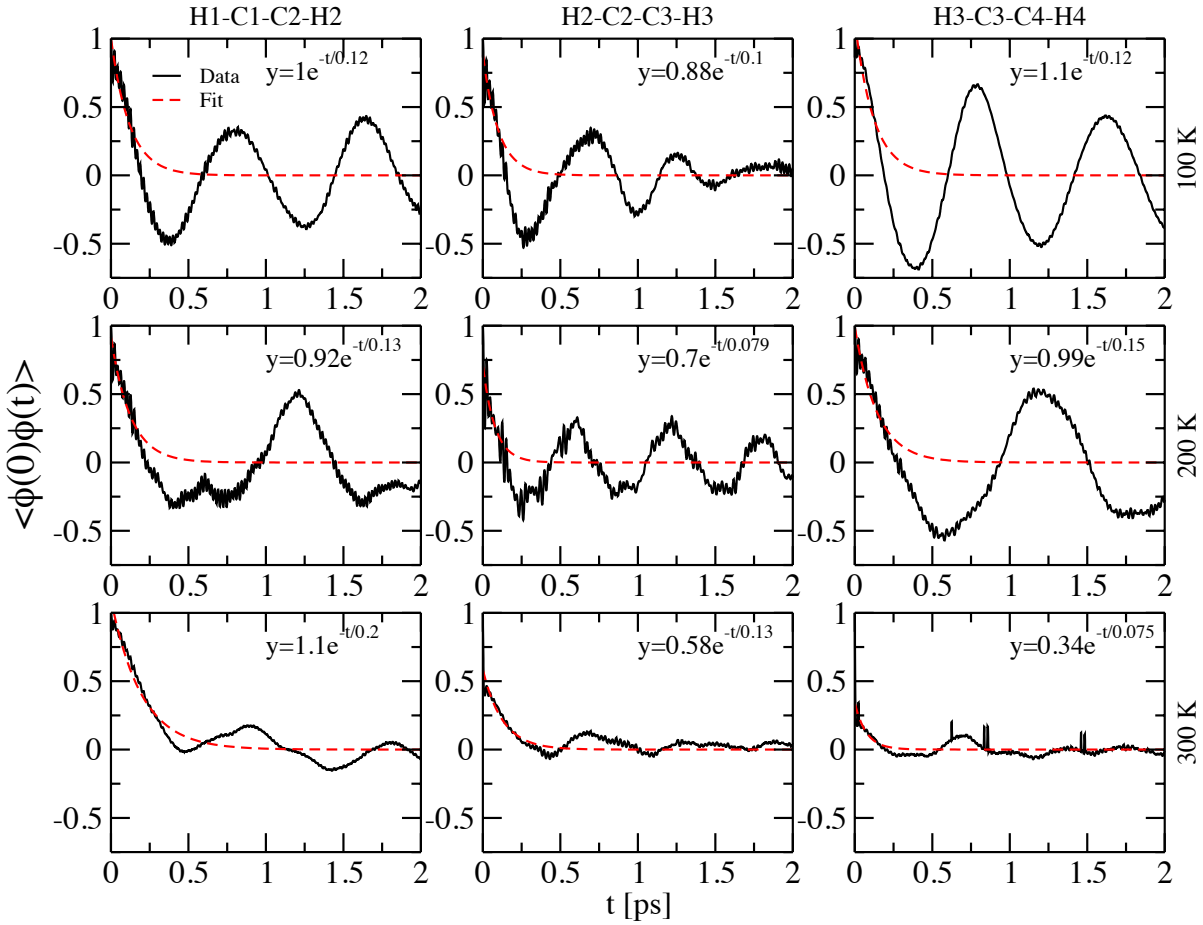

Figure 3.2: Some representative centroid dihedral angle autocorrelation functions from the NVE sampling runs

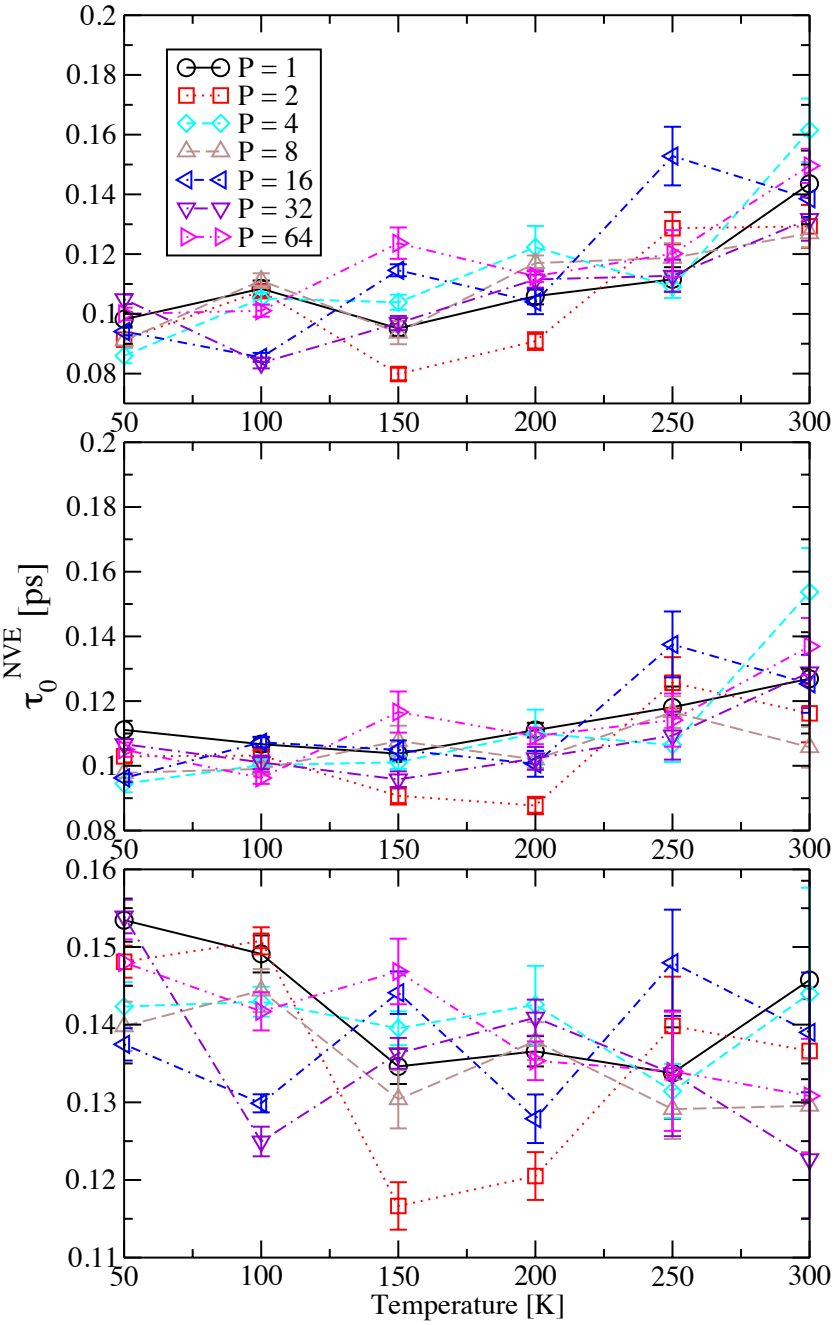

Figure 3.3: Average NVE decorrelation times for endocyclic dihedral angles

The average decorrelation times of NVT simulations using this centroid friction for a range of temperatures, number of beads, and for each endocyclic centroid dihedral angle are similarly shown in Fig 3.4. The overall average $\tau_0^{\mathrm{NVT}} + 2\sigma_{\tau_0} = 0.385$ ps, giving a skip step of ~ 0.5 ps, where we have rounded up for convenience. Additionally, plots of the H1-C1-C2-H2 decorrelation times from simulations using centroid frictions that vary several orders of magnitude about the centroid friction previously established through RPMD decorrelation analysis, which we now denote $\gamma^{(0)}_{\mathrm{RPMD}}$, are compiled in Fig. 3.5. The relatively flat trend of low decorrelation times to the left of $\gamma^{(0)}_{\mathrm{RPMD}} = 8.33$ fs^{-1} figure reveals that system states are sampled almost equally efficiently with centroid frictions $\leq \gamma^{(0)}_{\mathrm{RPMD}}$. However, with centroid frictions $> \gamma^{(0)}_{\mathrm{RPMD}}$ the decorrelation times increase and so larger skip steps would be required, indicating there is a one-sided robustness of the normal mode Langevin thermostat towards lower centroid frictions for this sugar system. This picture of the thermostat's robustness is not general because changing the centroid friction in either direction did not have a significant effect on τ_0^{NVT} for ground state *para*-H_2 simulations, [15] whereas changing the centroid friction in either direction from the analytically predicted optimal value for water dimer simulations using umbrella sampling increased τ_0^{NVT} [86].

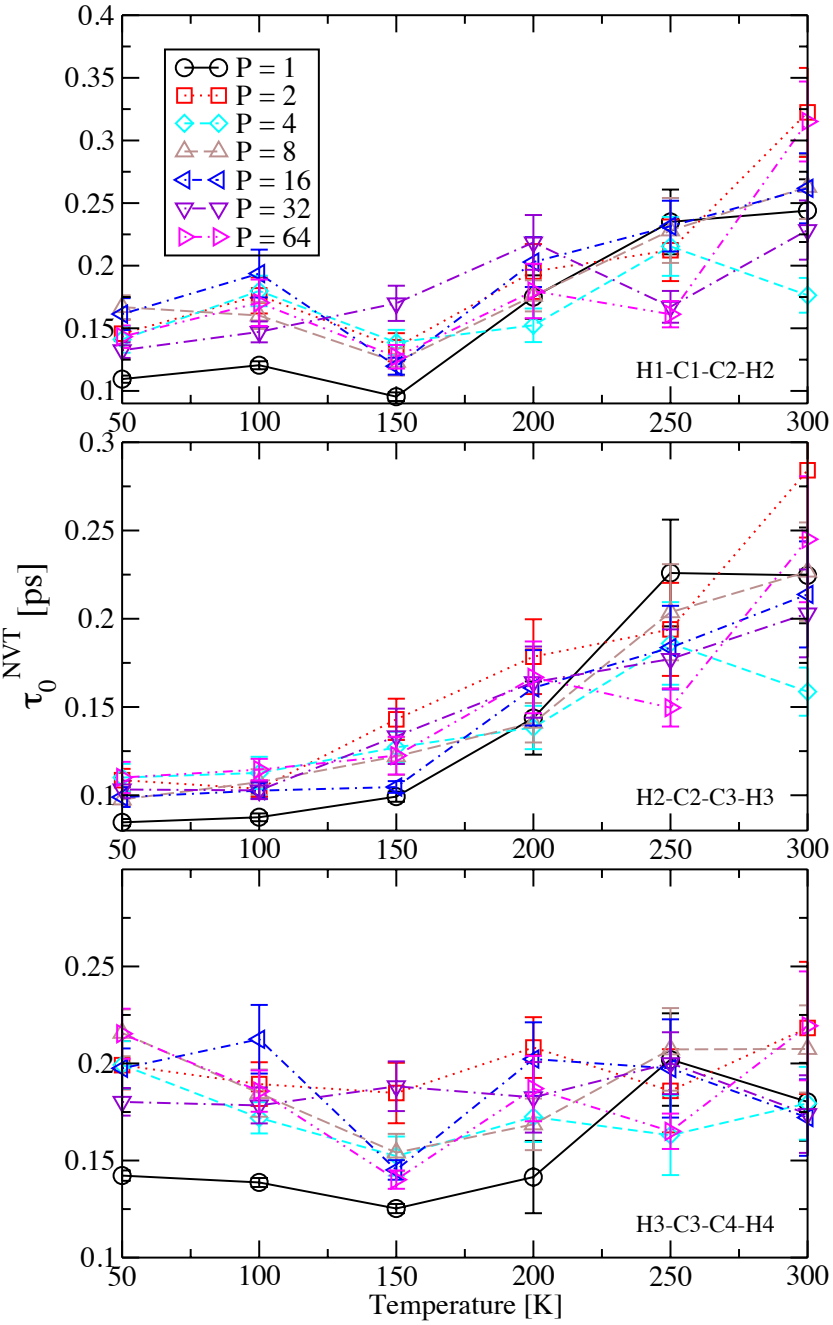

Figure 3.4: Average NVT decorrelation times for endocyclic dihedral angles

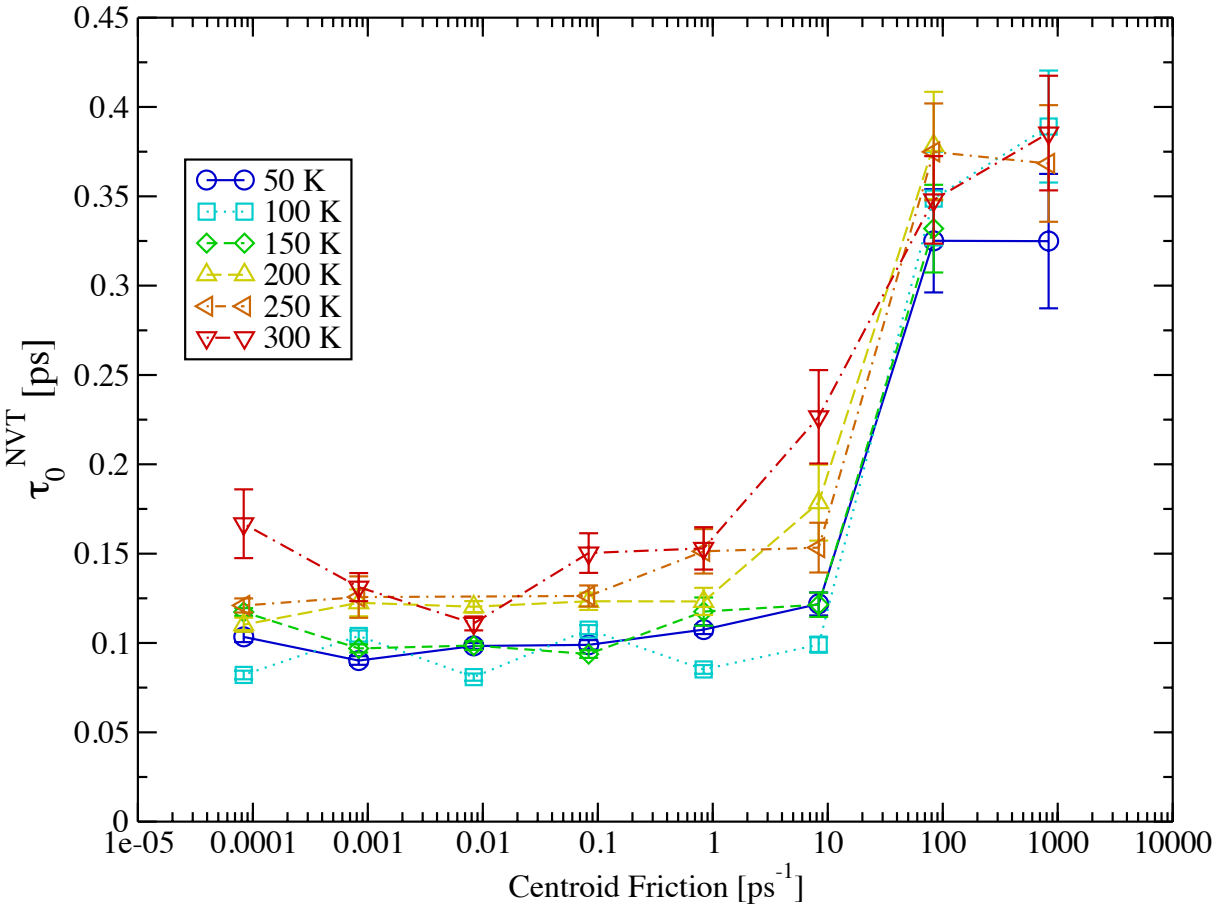

Figure 3.5: Checking the optimization of skip steps via τ_0^{NVT} by varying the centroid friction for H1-C1-C2-H2

One concern with the preceding results is that our algorithm to fit an exponential decay to the autocorrelation functions did not account for large amplitude oscillations over long timescales. In Fig. 3.2, these oscillations in the centroid dihedral angle correlations are visible at 200 K and more pronounced at 300 K, whereas they are sufficiently damped by $\sim$

0.5 ps at 100 K. Longer decorrelation times would be obtained for the lower temperatures if the algorithm fit to the envelope of the oscillations. Molecules have less kinetic energy at lower temperatures such that transitioning into different configurations will be hindered, which makes longer decorrelation times expected. This issue of fitting the oscillations will be addressed in the future.

With simulation parameters established, we can now look at physical properties of the sugar itself. The convergence of endocylic dihedral angle distributions with number of beads at 200 K is depicted in Fig. 3.6. As expected, increasing quantum sampling results in decreasing distribution peak heights and broadens their widths. The distributions for each dihedral angle are also bimodal, with one mode clearly favoured over the other in each case. The distributions have roughly converged with $P = 32$ beads or $\tau = 1.56{\times}10^{-4}$ K^{-1}. This τ is maintained for PIMD simulations at 100 K ($P = 21$) and 300 K ($P = 64$) and a compilation of classical dihedral angle distributions and distributions at $\tau = 1.56{\times}10^{-4}$ K^{-1} are shown in Fig. 3.7 for all three temperatures. At 100 K, there is starker contrast between the tall and narrow classical distributions and the shorter and spread out quantum distributions. This result is in line with well known quantum theory whereby quantum tunneling dominates at low temperatures. At 300 K, which is near biological temperatures, the quantum distributions are still noticeably less concentrated compared to the classical distributions, an important finding that may have implications for drug

design. A mimic of the sugar may potentially require a similar conformational flexibility entailed by the quantum spreading to act as an effective inhibitor of the sugar's enzymes.

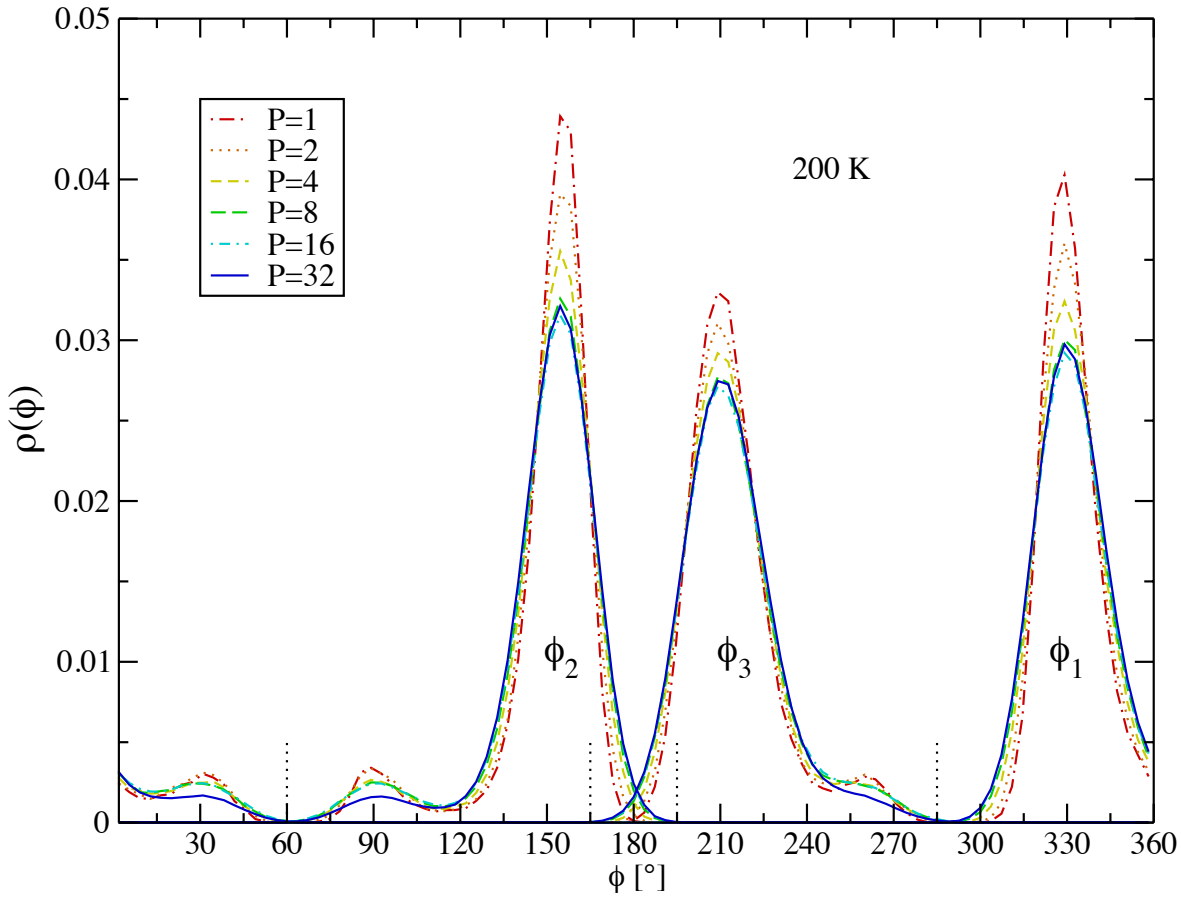

Figure 3.6: Path integral bead convergence of endocyclic dihedral angle distributions. The dotted vertical lines are guides to deliniate the different distributions

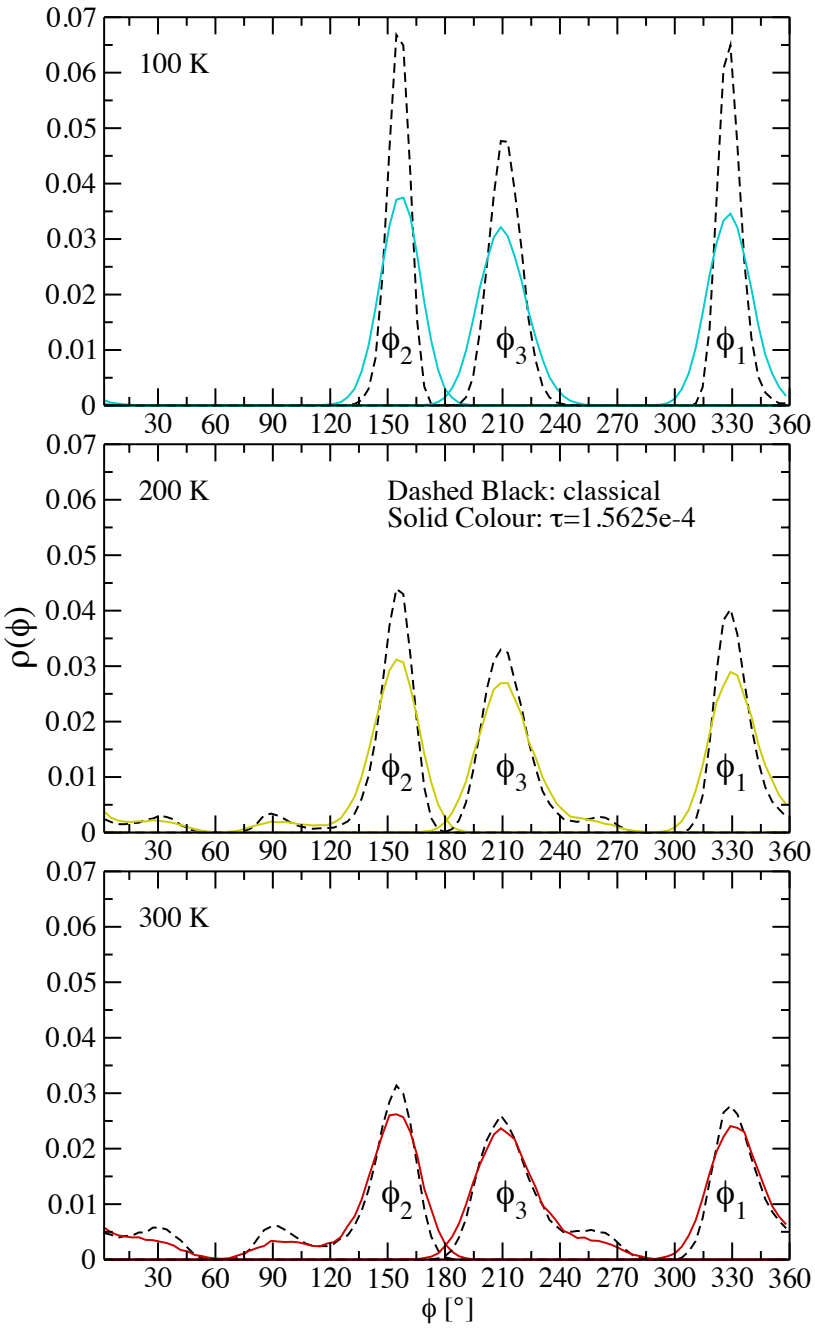

Figure 3.7: Temperature dependence of endocyclic dihedral angle distributions

A more quantitative look at quantum effects is provided by $^3J_{\mathrm{H,H}}$ coupling constant predictions. The τ dependence of the coupling constant for each endocyclic dihedral angle at 100 K, 200 K, and 300 K is displayed in Fig. 3.8. There appears to be a trend of decreasing predicted coupling constant values with increasing τ for the first few points starting at high τ (looking from the left hand side of the graphs). The first point on the left hand side of each graph is the classical prediction. But the results become noisy at low τ, shown by the scattered points on the graphs. Upon further inspection, we noticed that the more extremely misbehaving points at low τ corresponded to dihedral angle distributions where the relative populations of the two dihedral modes did not follow the same trend as the distributions for other τ. This suggests there may be ergodicity problems in our simulations whereby the sugar becomes trapped more in one dihedral mode over the other relative to the expected trend and that 20 ns of production simulation is not enough to reach equilibrium distributions. In future work we will increase the simulation length to improve sampling of the system's phase space or employ umbrella sampling to force better sampling [56]. It is important to note that these coupling constant results cannot be compared to previous classical simulation predictions and experimental NMR observations reported in Ref. 55 because those are based on the solution phase. We plan to carry out solution phase simulations in the future while taking advantage of Graphics Processing Unit (GPU) acceleration to distribute the computational load of the many water molecules

required for the solution phase. We discuss our software and hardware set up for these GPU-accelerated simulations and preliminary benchmarking in Section 4.2.

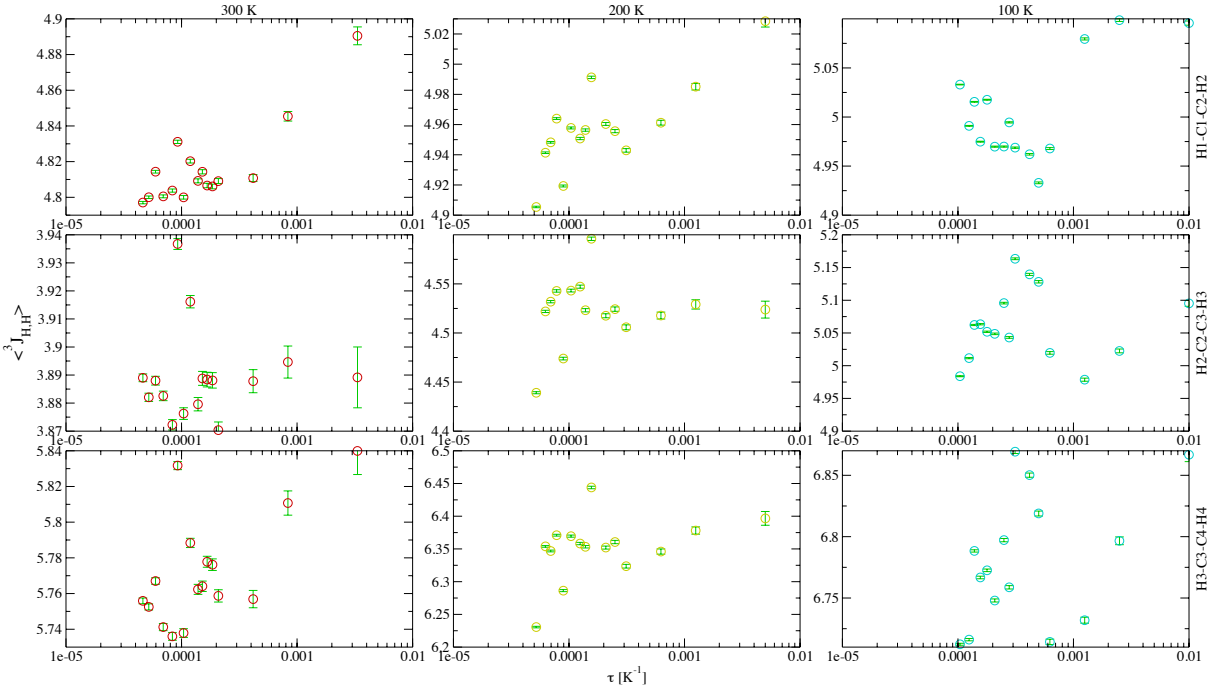

Figure 3.8: τ convergence of endocyclic coupling constants

After completing the investigation into the sugar's structural properties, we moved on to its energy properties. Our first attempt in establishing energy convergence with respect to τ using the centroid viral estimator for different temperatures is shown in Fig. 3.9. The convergence appears quadratic up until about $\tau \sim 4{\times}10^{-5}$ K^{-1}, after which there is a divergence in the energy. The simulation time step Δt was the first suspect in causing this

discrepancy because it is a common culprit in energy instability for classical simulations. We then calculated the energy for a series of simulations with varying Δt and τ and found that indeed a lower Δt is required for lower τ. We will attempt to uncover the source of this energy instability with time step size in future work. A time step of $\Delta t = 0.25$ fs was found to be sufficient to maintain energy stability for our energy convergence studies that required going down to $\tau \sim 3.5{\times}10^{-5}$ K^{-1} for convergence. This time step is used for subsequent PIMD simulations.

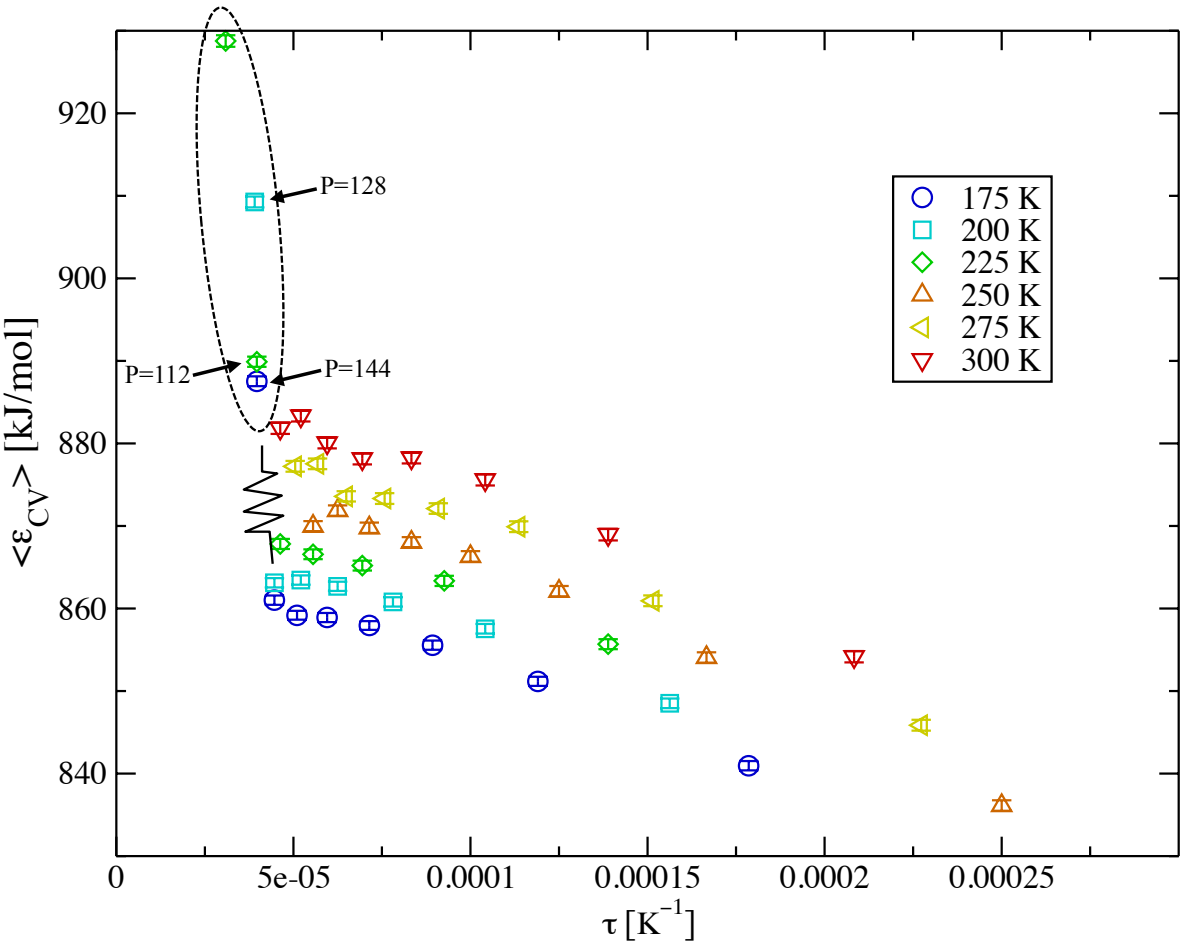

Figure 3.9: τ convergence of energies with Δt too large (0.5 fs)

New plots of the energy convergence with the updated time step are displayed in Fig. 3.10. The quadratic behaviour of convergence is consistent in these plots. This study provided an idea of the number of data points and the range of τ required for good fits in order to extrapolate to $\tau = 0$ K^{-1} full quantum energies. These parameters were roughly

maintained across temperatures in the follow up large scale study of the temperature dependence of the quantum energy. Plots of the full quantum and classical energies are shown for 50 K - 1300 K in Fig. 3.11 alongside analytic quantum and harmonic oscillator energies. As an aside, each of the data points in the quantum case (there are many more in between but not displayed) needed ~ 12 simulations of varying τ in a manner similar to Fig. 3.10. Such a large scale undertaking was only possible through our supercomputing cluster that we call `nlogn`. There is a visible systematic difference between the simulated energies and their harmonic oscillator counterparts that grows with temperature, even at low temperature as seen in the inset of Fig. 3.11. These differences represent the contribution of nonharmonic degrees of freedom and are > 12 kJ/mol (2.87 kcal/mol) by 300 K. However, the energy difference between the simulations and harmonic oscillator models are within chemical accuracy of ~ 1 kcal/mol at low temperatures, so treating the sugar harmonically through normal mode analysis is sufficient to reproduce experimental results for the energy at these temperatures and much cheaper computationally than PIMD. In the future, we plan to use Monte Carlo sampling from a multivariate Gaussian distribution to harmonically explore the sugar's configuration space and obtain dihedral angle distributions, which might also be more efficient than PIMD. We additionally plan to calculate the ground state quantum energy using LePIGS and compare that to the ground state quantum harmonic oscillator energy.

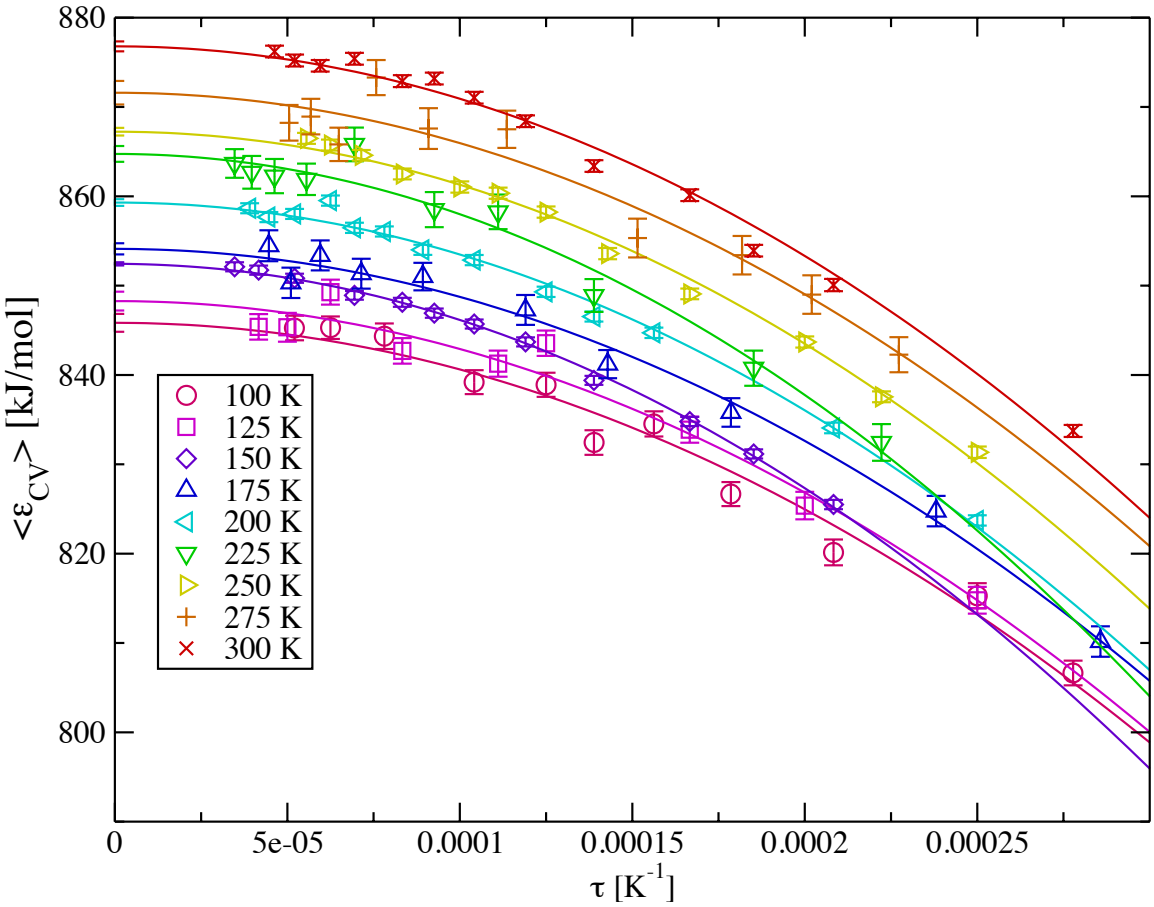

Figure 3.10: τ convergence of energies with lower Δt (0.25 fs) for stable simulations at small τ

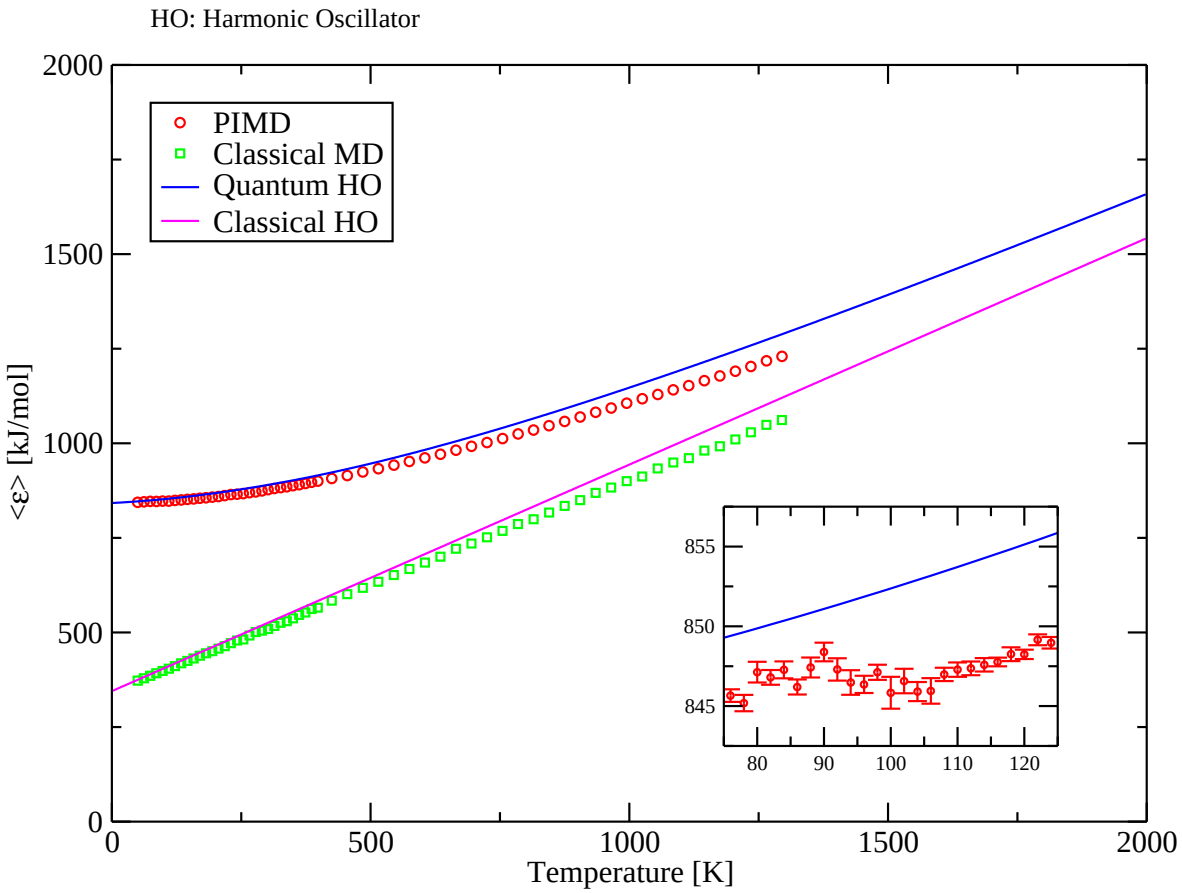

Figure 3.11: Temperature dependence of the average sugar internal energy. The PIMD results are based on Centroid Virial estimates extrapolated to τ=0

3.4 Conclusions

We have optimized PIMD simulation parameters for the gas phase study of the sugar beta D-arabinofuranoside and confirmed that significant nuclear quantum effects exist for its

structural and energy properties, even at biological temperatures. Our results indicate that systematic optimization of simulation parameters is necessary to promote efficiency and avoid system instability. Dihedral angle distributions visibly became more delocalized as quantum sampling was increased by increasing the number of path integral beads. This corresponded to a visible change in ^{1}H-^{1}H nuclear spin coupling constants as compared to classical predictions, however the low τ results were noisy and so we could not fit to a converged value. The predicted quantum internal energies were also much larger than the classical energy predictions and the quantum internal energy also had a temperature dependence that followed the contour of an analytic quantum harmonic oscillator model that is non linear at low temperature, whereas the classical internal energy had a temperature dependence that followed the linear contour of an analytic classical harmonic oscillator model. For temperatures < 300 K, the harmonic oscillator models, which are within chemical accuracy in this domain, appear better suited for predicting energies due to their much lower computational cost compared to their MD counterparts. We will extend to solution phase simulation in future work so that we are able to compare our quantum results to prior classical and experimental values.

Chapter 4

Conclusions and future work

4.1 Concluding statements

Recall that our prime objective has been to improve the accuracy of molecular modeling while remaining amenable to current computational resources. This thesis investigated two different approaches and applications where quantum nuclear dynamics was included in order to achieve this goal.

In Chapter 2, a basis set method was used to reduce a recent high quality 6D *ab initio* hydrogen pair potential into a set of 1D pair potentials to enable feasible LePIGS simulations of $p\mathrm{H}_2$ clusters. The difference of the reduced potentials were then perturbatively combined with the resulting radial distributions of the simulated clusters in order to predict Raman vibrational shifts. The results based on radial distributions generated from our ground vibrational state reduced potential were superior to that of other popular hydrogen

pair potentials, at least for small $p\mathrm{H}_2$ cluster sizes where reliable experimental shift values exist. Also, we believe that our difference potential is closer to physical reality compared to the empirical difference potential of Tejeda et al. and extensible to larger cluster sizes. Thus, for this application it appears that prime objective has been accomplished. However, there is still some doubt about the accuracy of our approach at larger cluster sizes. Future work will investigate many-body effects and possible misassignment of the large cluster experimental Raman peaks. Future applications include predicting vibrational shifts of orthodeuterium and paratritium, the larger bosonic isomers of hydrogen, as well as solid parahydrogen.

In Chapter 3, we sought to validate the broader use of PIMD by applying it to study the structural and energetic properties of methyl β-D-arabinofuranoside, a sugar previously examined through classical MD. We employed a systematic approach in establishing simulation parameters, the importance of which is underlined by identifying energy instability at low τ or high number of beads if the time step Δt is too large. Increasing the amount of quantum sampling visibly impacted structural properties including dihedral angle distributions and ^{1}H-^{1}H nuclear spin coupling constants. The dihedral angle distributions in particular became noticeably more delocalized due to quantum tunneling. However, the convergence of the coupling constants with τ is still unresolved due to fluctuations at low τ. In future work we will attempt longer simulations or umbrella sampling in order overcome

any ergodicity problems that might be the cause of the coupling constant fluctuations. It is too early to determine if we achieved our prime objective of increasing simulation accuracy with regards to the coupling constants due to this fluctuation problem and the fact that we can not compare to experiment until we undertake solution phase simulations. However, our internal energy results support a quantum picture for the sugar, considering that the temperature dependence of predicted quantum internal energy estimates followed the contour of a quantum harmonic oscillator model. These quantum internal energy predictions also behaved significantly differently from classical internal energy predictions, particularly at low temperatures. Therefore we can conclude that it is important to treat the sugar's nuclear dynamics quantum mechanically, especially at low temperatures where quantum effects dominate, and that we should investigate further. However, PIMD appears to be excessive for this task, considering that a quantum harmonic oscillator model can reproduce quantum energies within chemical accuracy and with much lower computational cost.

4.2 Future directions: solution phase sugar simulations

PIMD is more computationally demanding than traditional MD because the system grows by a factor roughly proportional to the number of beads. Additionally, for our sugar

application a long simulation time is required to get converged distributions of exocyclic conformations [53, 56] and possibly even endocylic conformations as discussed earlier with coupling constant fluctuation problem. It is therefore in our interest to optimize the simulations as much as possible, both in terms of simulation parameters and also the software and hardware infrastructure to carry them out. One time-saving measure is to adopt the paradigm of *parallel computing*, which involves splitting up tasks into *threads* for multiple physical processing cores such that they can be computed simultaneously. A schematic of parallel computing is shown in Fig. 4.1. It is only practical for tasks that do not require much communication between each other because the extra effort and downtime during sending and receiving messages will create performance bottlenecks [87]. MD lends itself to parallel computing because the forces on each particle and their translations at each time step during integration can be calculated independently of other particles, only memory of the other particles' configurational state from the end of the previous time step is required, such that separate threads can carry out calculations for different particles. Afterwards, the threads collectively update the memory of the state of the system with their results, for use by each thread for the subsequent time step calculations [88]. With this technique, we get closer to $\mathcal{O}(N)$ computational efficiency.

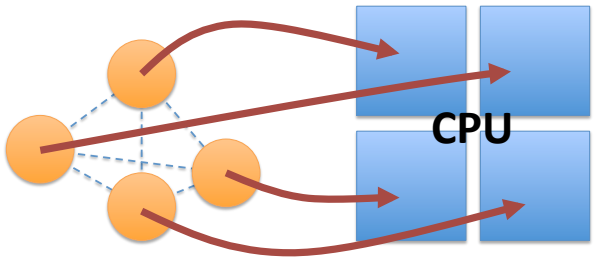

Figure 4.1: Schematic depicting parallel computing. Note that this is an oversimplification where one particle is assigned to one processing core.

Parallel computing using Central Processing Units (CPUs), the main computer chips found in typical desktop machines, is already well established. CPUs of today in average home desktops usually have 4 cores and CPUs for servers usually have up to 16 or 32 cores. Multiple CPUs are linked over a high-speed network to create a supercomputing grid or cluster with potentially thousands of cores accessible to demanding scientific applications such as MD. However, the history of CPU use has lead to designs that maximize the performance of single threads for serial processing and only a fraction of a CPU chip's area is devoted to raw computation, the rest being used as support mechanisms to help handle irregular loads [87, 89]. This makes them useful for the varied tasks we subject them to at home and in the office, but inefficient for the repeated set of instructions on large data sets used in scientific computing [87, 89].

A new entrant to parallel computing that has been quickly gaining popularity over

the last decade as an answer to these shortcomings is general-purpose Graphics Processing Unit (GPGPU) computing. Current performance grade GPUs have hundreds of small cores on a single chip with a large area focused on raw computation, which is an architecture influenced by their roots of having to process many regular graphics generation tasks simultaneously for real-time display [88, 87]. They gained widespread consumer adoption to work alongside CPUs in computers and video game consoles for their role in driving video game graphics, and so have the benefit of economies of scale in their manufacturing. Thus, they offer the advantage of a small footprint both in terms of size and cost compared to CPU parallel computing solutions. A schematic showing the overall structural difference between CPUs and GPUs is shown in Fig. 4.2. GPU manufacturers and related consortiums have opened up the GPUs to general computation beyond graphics by making the cores programmable and providing Application Programming Interfaces (APIs) that allow easier programming. A major platform for GPGPU computing is Nvidia Corporation's Compute Unified Device Architecture (CUDA) that includes a C-like API [90].

MMTK can generate multiple threads for parallel force field evaluation. However, its built-in code is designed to run on CPUs and may suffer from the limitations outlined in the proceeding section. To address this, S. Constable, another graduate of our group, removed the MMTK C language internals related to integration on the CPU and replaced them with a series of custom codes [3] to instead pass data about the system and simulation parameters

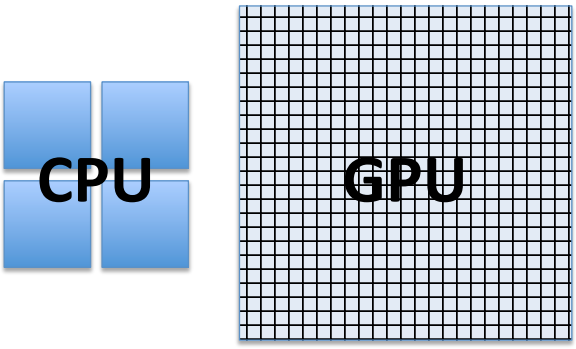

Figure 4.2: Schematic comparing CPUs and GPUs. CPUs have fewer, larger, general purpose cores, whereas GPUs have many, smaller, raw computation cores

to the Open Molecular Mechanics (OpenMM) program using the OpenMM C API [91]. In turn, OpenMM provides an interface to a general-purpose GPU computing API like CUDA, which it calls a *Platform*, that is geared specifically towards the acceleration of molecular modeling tasks via parallel execution on GPU cores. Particularly, OpenMM's `LangevinIntegrator` and `RPMDIntegrator` classes of algorithms are used for classical MD integration and PIMD integration, respectively, with both applying the Langevin equation to sample the canonical ensemble. After a specified number of integration steps, data about the new state of the system is brought back up from the GPU memory to the C level of code running on the CPU via OpenMM queries of the `Platform`, where an MMTK algorithm for trajectory output reports the data for recording into a trajectory file and makes them available at the surface Python level for further action by the user. OpenMM also has a *Reference* platform that runs the integration on the CPU instead for benchmarking the

performance of the GPU-based platforms. A schematic of this software setup is contained in Fig. 4.3.

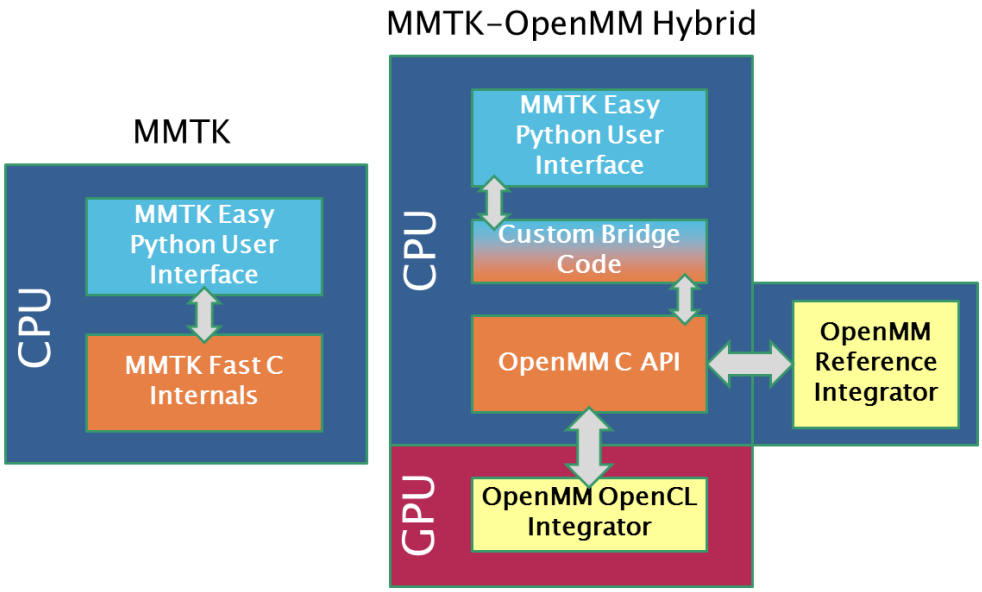

Figure 4.3: Schematic of our software setup

This MMTK-OpenMM hybrid program was developed rather than using OpenMM as standalone software because S. Constable evaluated that OpenMM does not have the level of user friendliness and features for setting up and analyzing a system like MMTK [3]. For example, this work requires the popular composite force field AMBER [83], which is not directly supported by OpenMM. Thus, our custom code that bridges between MMTK and OpenMM extracts individual components of the AMBER force field, which is originally set with MMTK, and feeds them into OpenMM, which builds the different forces separately

in its own representation that is suitable for GPU computation. With our hybrid program this tedious process is hidden from the user and all they need to do is set the force field in the MMTK Python front end with a simple command. User friendliness is important to us because we want wide consumption of our tools and our perspective on the importance of quantum sampling through PIMD by the scientific community. However, the nascent status of the MMTK-OpenMM hybrid program entails the presence of software bugs that we have to clear before general release of the program. Several 'bug fixes' have already been committed to enable future solution phase work with the sugar. These include resolving issues in the treatment of periodic boundary conditions and removing double processing of vibrational force constants when transferring AMBER forcefield components.

Solution phase simulations will take place in an *orthorhombic periodic universe* with the sugar surrounded by water molecules. A periodic universe enables computationally feasible simulation of bulk systems by breaking them into a microscopic unit cell that is virtually repeated in all directions. The unit cell in an orthorhombic universe is a box. Motions of the particles in the central, explicit cell are implicitly mimicked by motions of the virtual particles in the virtual copies of the unit cell. When a particle crosses the boundary of the explicit cell it means that a virtual copy should enter the cell from the opposite boundary. However, in practice locations of virtual copies are not stored in memory but generated for temporary use in calculations by inferring from the location of the explicit particles and

size of the unit cell. Force calculations take interactions with virtual particles into account from their inferred locations and use the *nearest image convention* to select for interaction with only the nearest copy of a particle and a *cutoff* distance to limit the extent of the interactions to those particles that have significant influence. In this way, a small finite number of particles can be used to reproduce bulk properties. A schematic of the nearest image convention is displayed in Fig. 4.4.

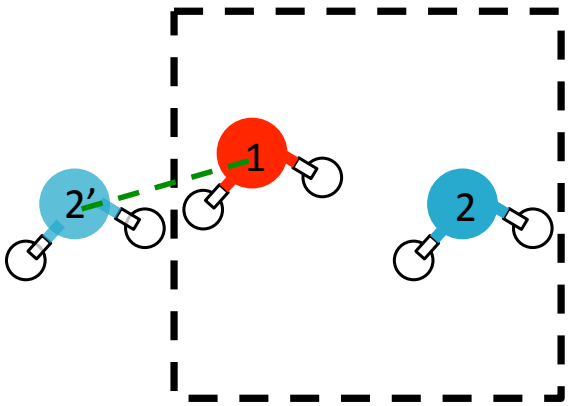

Figure 4.4: Schematic of the nearest image convention. Particle 1 only interacts with Particle 2', the virtual copy of Particle 2, because it is closer

Benefits of GPU acceleration are only seen if the system has enough particles such that parallel evaluation of energy and forces outweighs the overhead cost of transferring information between the CPU and GPU and setting up the system on the GPU. Additionally, we expect larger systems to make more use of the greater number of cores on the GPU. To investigate this system size effect on GPU acceleration, we ran several short simulations of

30 ps with different sized periodic boxes of water molecules filled to a density of 0.9970479 g/cm^3 using both the original pure MMTK software and the MMTK-OpenMM hybrid software. The results of this benchmarking are shown in Fig. 4.5 and confirm that performance gains increase with system size. For large box size of (2.5 Å)^3 containing 520 water molecules, which corresponds to the size used in Ref. 55, the GPU simulation is $\sim$100$\times$ faster than the single-threaded pure MMTK CPU simulation. These results reassure us that the investment in developing the hybrid program was a sound decision.

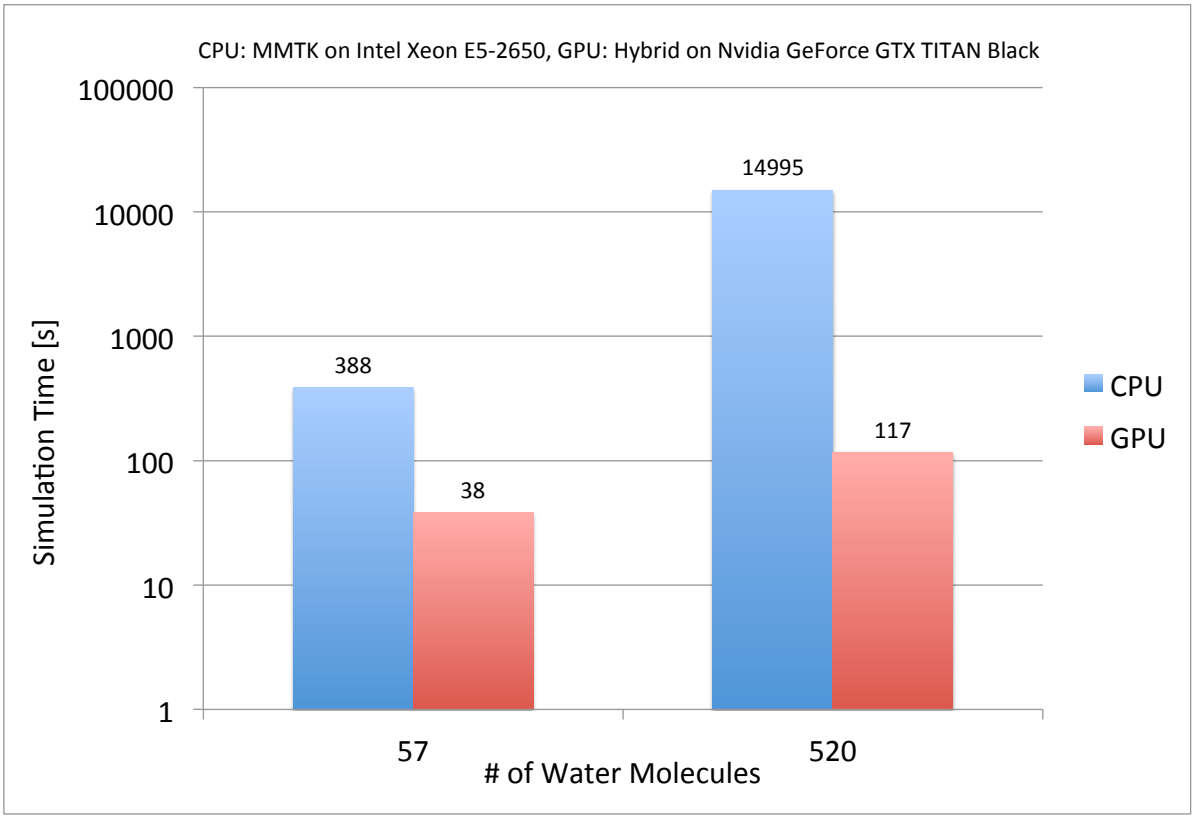

Figure 4.5: CPU and GPU benchmarking results for periodic boxes of water molecules

References

[1] G. Tejeda, J. M. Fernández, S. Montero, D. Blume, and J. P. Toennies, Phys. Rev. Lett. **92**, 223401 (2004).

[2] C. Ing, K. Hinsen, J. Yang, T. Zeng, H. Li, and P.-N. Roy, J. Chem. Phys. **136**, 224309 (2012).

[3] S. Constable, Master's thesis, Department of Chemistry, University of Waterloo, 2012.

[4] N. Faruk, M. Schmidt, H. Li, R. J. Le Roy, and P.-N. Roy, The Journal of Chemical Physics **141**, (2014).

[5] M. Constable, Master's thesis, Department of Chemistry, University of Waterloo, 2014.

[6] H. Li, P.-N. Roy, and R. J. Le Roy, J. Chem. Phys. **133**, 104305 (2010).

[7] R. Steiner, Rep. Prog. Phys. **76**, 016101 (2013), WOS:000312593700002.

[8] M. Born and R. Oppenheimer, Annalen der Physik **389**, 457 (1927).

[9] J. C. Light and T. Carrington, *Discrete-Variable Representations and their Utilization*, pages 263–310, John Wiley & Sons, Inc., 2007.

[10] B. P. Lanyon et al., Nat Chem **2**, 106 (2010).

[11] R. P. Feynman, Rev. Mod. Phys. **20**, 367 (1948).

[12] D. Chandler and P. G. Wolynes, J. Chem. Phys. **74**, 4078 (1981).

[13] M. Parrinello and A. Rahman, J. Chem. Phys. **80**, 860 (1984).

[14] S. Constable, M. Schmidt, C. Ing, T. Zeng, and P.-N. Roy, J. Chem. Phys. A **117**, 7461 (2013).

[15] M. Schmidt, S. Constable, C. Ing, and P.-N. Roy, The Journal of Chemical Physics **140**, 234101 (2014).

[16] M. J. Harvey and G. De Fabritiis, J. Chem. Theory Comput. **5**, 2371 (2009).

[17] T. Zeng and P.-N. Roy, Rep. Prog. Phys. **77**, 046601 (2014).

[18] S. Grebenev, B. Sartakov, J. P. Toennies, and . A. F. Vilesov, Science **289**, 1532 (2000).

[19] S. Grebenev, B. Sartakov, J. P. Toennies, and A. Vilesov, Phys. Rev. Lett. **89**, 225301 (2002).

[20] S. Grebenev, B. G. Sartakov, J. P. Toennies, and A. F. Vilesov, Europhys. Lett. **83**, 66008 (2008).

[21] S. Grebenev, B. G. Sartakov, J. P. Toennies, and A. F. Vilesov, J. Chem. Phys. **132**, 064501 (2010).

[22] H. Li, R. J. Le Roy, P.-N. Roy, and A. R. W. McKellar, Phys. Rev. Lett. **105**, 133401 (2010).

[23] E. G. Noya, C. Vega, and C. McBride, J. Chem. Phys. **134**, (2011).

[24] T. Zeng, H. Li, and P.-N. Roy, J. Phys. Chem. Lett. **4**, 18 (2013).

[25] T. Zeng, G. Guillon, J. T. Cantin, and P.-N. Roy, J. Phys. Chem. Lett. **4**, 2391 (2013).

[26] H. Li, N. Blinov, P.-N. Roy, and R. J. Le Roy, J. Chem. Phys. **130**, 144305 (2009).

[27] P. Sindzingre, D. M. Ceperley, and M. L. Klein, Phys. Rev. Lett. **67**, 1871 (1991).

[28] M. C. Gordillo and D. M. Ceperley, Phys. Rev. B **65**, 174527 (2002).

[29] F. Mezzacapo and M. Boninsegni, Phys. Rev. Lett. **97**, 045301 (2006).

[30] F. Mezzacapo and M. Boninsegni, Phys. Rev. A **75**, 033201 (2007).

[31] F. Mezzacapo and M. Boninsegni, Phys. Rev. Lett. **100**, 145301 (2008).

[32] J. E. Cuervo and P.-N. Roy, J. Chem. Phys. **125**, 124314 (2006).

[33] J. E. Cuervo and P.-N. Roy, J. Chem. Phys. **128**, 224509 (2008).

[34] J. E. Cuervo and P.-N. Roy, J. Chem. Phys. **131**, 114302 (2009).

[35] R. Guardiola and J. Navarro, Phys. Rev. A **74**, 025201 (2006).

[36] J. Navarro and R. Guardiola, Int. J. Quantum Chem. **111**, 463 (2011).

[37] S. A. Khairallah, M. B. Sevryuk, D. M. Ceperley, and J. P. Toennies, Phys. Rev. Lett. **98**, 183401 (2007).

[38] F. Mezzacapo and M. Boninsegni, J. Phys.: Conf. Ser. **150**, 032059 (2009).

[39] R. J. Hinde, J. Chem. Phys. **128**, 154308 (2008).

[40] D. M. Ceperley, Rev. Mod. Phys. **67**, 279 (1995).

[41] C. Ing, Master's thesis, Department of Physics, University of Waterloo, 2011.

[42] C. Altona and M. Sundaralingam, J. Am. Chem. Soc. **94**, 8205 (1972).

[43] H. M. Dias et al., Global tuberculosis report, Technical report, World Health Organization, 2012.

[44] T. L. Lowary, Mini Rev Med Chem **3**, 689 (2003).

[45] Y. L. Janin, Bioorgan. Med. Chem. **15**, 2479 (2007).

[46] P. J. Brennan and H. Nikaido, Annual Review of Biochemistry **64**, 29 (1995).

[47] J. B. Houseknecht, T. L. Lowary, and C. M. Hadad, J. Phys. Chem. A **107**, 5763 (2003).

[48] N. D. Connell and H. Nikaido, *Tuberculosis: Pathogenesis, Protection and Control*, pages 333–352, American Society for Microbiology, 1994.

[49] P. Draper and M. Daffé, *Tuberculosis and the Tubercle Bacillus*, pages 261–273, American Society for Microbiology, 1994.

[50] F. A. A. M. De Leeuw and C. Altona, J. Comput. Chem. **4**, 428 (1983).

[51] M. Karplus, The Journal of Chemical Physics **30**, 11 (1959).

[52] C. Altona, R. Francke, R. de Haan, J. H. Ippel, G. J. Daalmans, A. J. A. W. Hoekzema, and J. van Wijk, Magn. Reson. Chem. **32**, 670 (1994).

[53] H. A. Taha, N. Castillo, P.-N. Roy, and T. L. Lowary, J. Chem. Theory Comput. **5**, 430 (2009).

[54] H. A. Taha, N. Castillo, D. N. Sears, R. E. Wasylishen, T. L. Lowary, and P.-N. Roy, J. Chem. Theory Comput. **6**, 212 (2010).

[55] H. A. Taha, P.-N. Roy, and T. L. Lowary, J. Chem. Theory Comput. **7**, 420 (2011).

[56] S. M. Islam, M. R. Richards, H. A. Taha, S. C. Byrns, T. L. Lowary, and P.-N. Roy, J. Chem. Theory Comput. **7**, 2989 (2011).

[57] M. Seo, N. Castillo, R. Ganzynkowicz, C. R. Daniels, R. J. Woods, T. L. Lowary, and P.-N. Roy, J. Chem. Theory Comput. **4**, 184 (2008).

[58] S. M. Islam and P.-N. Roy, J. Chem. Theory Comput. **8**, 2412 (2012).

[59] P. L. Raston, W. Jäger, H. Li, R. J. Le Roy, and P.-N. Roy, Phys. Rev. Lett. **108**, 253402 (2012).

[60] T. Zeng, H. Li, R. J. Le Roy, and P.-N. Roy, J. Chem. Phys. **135**, 094304 (2011).

[61] C. Schwartz and R. J. L. Roy, J. Mol. Spectrosc. **121**, 420 (1987).

[62] M. Boninsegni and S. Moroni, Phys. Rev. E **86**, 056712 (2012).

[63] U. Buck, F. Huisken, A. Kohlhase, D. Otten, and J. Schaefer, J. Chem. Phys. **78**, 4439 (1983).

[64] I. F. Silvera and V. V. Goldman, J. Chem. Phys. **69**, 4209 (1978).

[65] K. Patkowski, W. Cencek, P. Jankowski, K. Szalewicz, J. B. Mehl, G. Garberoglio, and A. H. Harvey, J. Chem. Phys. **129**, 094304 (2008).

[66] D. T. Colbert and W. H. Miller, J. Chem. Phys. **96**, 1982 (1992).

[67] P.-N. Roy, J. Chem. Phys. **119**, 5437 (2003).

[68] M. P. Nightingale and P.-N. Roy, J. Phys. Chem. A **110**, 5391 (2006).

[69] R. J. Hinde, Chem. Phys. Lett. **460**, 141 (2008).

[70] M. E. Tuckerman, Path integration via molecular dynamics, in *Quantum Simulations of Complex Many-Body Systems: From Theory to Algorithms*, volume 10, pages 269–298 (John von Neumann Institute for Computing, 2002).

[71] H. F. Trotter, On the product of semi-groups of operators, in *P. Am. Math. Soc.*, volume 10, pages 545–551 (American Mathematical Society, 1959).

[72] R. W. Hall and B. J. Berne, The Journal of Chemical Physics **81**, 3641 (1984).

[73] M. Ceriotti, M. Parrinello, T. E. Markland, and D. E. Manolopoulos, The Journal of Chemical Physics **133**, 124104 (2010).

[74] G. Bussi and M. Parrinello, Phys. Rev. E **75**, 056707 (2007).

[75] I. R. Craig and D. E. Manolopoulos, The Journal of Chemical Physics **121**, 3368 (2004).

[76] L. Verlet, Phys. Rev. **159**, 98 (1967).

[77] W. C. Swope, H. C. Andersen, P. H. Berens, and K. R. Wilson, J. Chem. Phys. **76**, 637 (1982).

[78] J. A. Barker, J. Chem. Phys. **70**, 2914 (1979).

[79] M. Levitt, C. Sander, and P. S. Stern, Journal of Molecular Biology **181**, 423 (1985).

[80] D. A. Case, Current Opinion in Structural Biology **4**, 285 (1994).

[81] K. Hinsen, J. Comput. Chem. **21**, 79 (2000).

[82] Woods Group. (2005-2014) *GLYCAM Web*. Complex Carbohydrate Research Center, University of Georgia, Athens, GA. (http://glycam.org/).

[83] D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang, and R. J. Woods, J. Comput. Chem. **26**, 1668 (2005).

[84] K. N. Kirschner, A. B. Yongye, S. M. Tschampel, J. González-Outeiriño, C. R. Daniels, B. L. Foley, and R. J. Woods, J. Comput. Chem. **29**, 622 (2008).

[85] M. H. Müser, J. Chem. Phys. **114**, 6364 (2001).

[86] K. P. Bishop and P.-N. Roy, Quantum mechanical free energy profiles of the water dimer, (Unpublished).

[87] P. N. Glaskowsky, Nvidia's fermi: The first complete gpu computing architecture, Technical report, NVIDIA, 2009.

[88] J. Owens, M. Houston, D. Luebke, S. Green, J. Stone, and J. Phillips, P. IEEE **96**, 879 (2008).

[89] M. S. Friedrichs, P. Eastman, V. Vaidyanathan, M. Houston, S. Legrand, A. L. Beberg, D. L. Ensign, C. M. Bruns, and V. S. Pande, J. Comput. Chem. **30**, 864 (2009).

[90] J. Nickolls, I. Buck, M. Garland, and K. Skadron, Queue **6**, 4053 (2008).

[91] P. Eastman and V. Pande, Comput. Sci. Eng. **12**, 34 (2010).