UNIVERSITY OF CALIFORNIA,
IRVINE

Characterizing the Interaction of Organic Compounds with Indoor Surfaces and QM/MM Simulation of Proton Transport in the Hv1 Protein

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Chemistry

by

Saleh Riahi Samani

Dissertation Committee:
Professor Douglas Tobias, Chair
Professor Craig Martens
Professor Ioan Andricioaei

2019

TABLE OF CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to thank my PhD supervisor, Professor Douglas J. Tobias, for providing me the environment that led me to learn proper way of conducting research. Furthermore, I like to extend my sincere gratitude to my group members for their support and valuable discussions during my PhD program.

Some parts of Chapter 2 are reprint of papers published in the Journal of Physical Chemistry Letters[1] and Chemical Science[5]. The work presented in Chapter 2 is supported by the Alfred P. Sloan Foundation under grant number G-2017-9692 and G2017-9796 (MOCCIE).

CURRICULUM VITAE

Saleh Riahi Samani

EDUCATION

Doctor of Philosophy in Chemistry **2019**
University of California, Irvine *Irvine, CA*

Masters of Science in Chemistry **2014**
University of California, Irvine *St. John's, Canada*

Bachelor of Science in Chemistry **2011**
Isfahan University of Technology *Isfahan, Iran*

RESEARCH EXPERIENCE

Graduate Research Assistant **2014–2019**
University of California, Irvine *Irvine, California*

Graduate Research Assistant **2012–2014**
Memorial University of Newfoundland *St. John's, Canada*

TEACHING EXPERIENCE

Teaching Assistant **2014–2018**
University of California, Irvine *IRVINE, California*

REFEREED JOURNAL PUBLICATIONS

What Is the Driving Force behind the Adsorption of Hydrophobic Molecules on Hydrophilic Surfaces? **2019**
J. Phys. Chem. Lett.

A molecular picture of surface interactions of organic compounds on prevalent indoor surfaces: limonene adsorption on SiO_2 **2019**
Chemical Science

Why can hydrogen sulfide permeate cell membranes? **2014**
JACS

Why can hydrogen sulfide permeate cell membranes? **2014**
J. Comp. chemistry

ABSTRACT OF THE DISSERTATION

Characterizing the Interaction of Organic Compounds with Indoor Surfaces and QM/MM Simulation of Proton Transport in the Hv1 Protein

By

Saleh Riahi Samani

Doctor of Philosophy in Chemistry

University of California, Irvine, 2019

Professor Douglas Tobias, Chair

Atomic level understanding of the interactions between the prevalent organic compounds and indoor surfaces is essential to study the potential impacts of these compounds on public health. Computer simulation methods provide an accurate means to explore physiochemical processes with atomistic detail. In this study molecular simulation methods are applied to elucidate the type and strength of interactions occurring between the indoor surfaces and organic compounds. Moreover, the energetics and structural information provided by the molecular simulations exhibited in this work are used to justify the experimental observations. Furthermore, hybrid quantum mechanics/molecular mechanics simulation protocol is used to explore the proton transport in the Hv1 protein. Specifically, the role of polar side chains located at the pore region of Hv1 channel in the proton translocation is explored.

Chapter 1

Background

According to the laws/postulates of statistical mechanics,[6] a macroscopic thermodynamic property relates to the average of its microscopic counterpart according to:

$$O_{obs} = \frac{1}{\tau}\int_{t_0}^{t_0+\tau} dt\, O(\vec{p}(t), \vec{q}(t), t) = \langle O(\vec{p}(t)\vec{q}(t), t)\rangle = \int d\vec{p}\, d\vec{q}\, f(\vec{p}(t), \vec{q}(t), t)\, O(\vec{p}(t), \vec{q}(t), t) \tag{1.1}$$

where $\vec{p}(t)$ and $\vec{q}(t)$ are 3N dimensional time dependent vectors corresponding to the positions and momenta that are accessible to the system, phase space. Hamiltonian of the system provides the time evolution of $\vec{p}(t)$ and $\vec{q}(t)$:

$$\begin{cases} \dot{q}_i = \frac{\partial H}{\partial p_i} \\ \dot{p}_i = -\frac{\partial H}{\partial q_i} \end{cases}$$

$f(\vec{p}(t), \vec{q}(t), t)$ is the probability density that a given volume in the phase space is occupied. Due to the complexity of $f(\vec{p}(t), \vec{q}(t), t)$ and dimensionality of most problems, calculating thermodynamic observables directly from Equation 1.1 is not feasible. Molecular dynamics (MD) provide an approximation to the phase space average, ensemble average, through sampling the phase space by evolving the $\vec{p}(t)$ and $\vec{q}(t)$ in time. The ergodic hypothesis allows for estimating the thermodynamic properties from a trajectory obtained from MD simulation. According to the ergodic hypothesis a set of trajectories that are evolved in time based on the Hamiltonian equation of motions are representative of the probability distribution of the phase space. Therefore, the time averaged properties obtained from MD simulation are equivalent to the quantities estimated from the ensemble averages.

According to the Liouville's theorem and postulate of statistical mechanics $f(\vec{p}(t), \vec{q}(t), t) = f(\vec{p}(t_0), \vec{q}(t_0))$, i.e. $\frac{d}{dt} f = \frac{\partial}{\partial t} f = 0$. Therefore, the invariance of f during the evolution of the system suggests that $f(\vec{p}, \vec{q}) = f[H(\vec{q}, \vec{p})]$,[7] where H is the Hamiltonian of the system. It can be shown that for a system in thermal equilibrium with the surrounding the distribution function becomes:

$$f[H(\vec{q}, \vec{p})] = C \frac{e^{-\beta H(\vec{q}, \vec{p})}}{\int e^{-\beta H(\vec{q}, \vec{p})} d\vec{p}\, d\vec{q}}\,; \beta = \frac{1}{k_B T} \tag{1.2}$$

The denominator of this equation is the partition function for the canonical ensemble, $Q(N, V, T)$. A variety of thermodynamic quantities can be obtained from the partition function. For instance, Helmholtz free energy $A = -\frac{1}{\beta} \ln Q(N, V, T)$.

Considering the level of atomistic details incorporated in the simulations, three types of molecular simulations are utilized in this study:

a) classical molecular dynamics MD.

b) ab initio molecular dynamics (AIMD).

c) quantum mechanics/molecular mechanics (QM/MM).

The technical aspects of each of these methods are briefly discussed in the following sections.

1.1 Classical MD

In the classical MD simulations, nuclei are propagated in time using the Newtonian mechanics. The forces underlying the dynamics of the system are empirically parametrized such that they entail the physical property of the system under study. In the CHARMM force field, atomic motions corresponding to the bond and angle degrees of freedom are approximated by the harmonic potential. The rotational motion of two bonds with respect to each other, dihedral motion, is assumed to have trigonometric form. A partial charge is assigned to each particle and the non-bonded electrostatic interactions, between the particles separated by at least three bonds, are calculated using the Coulomb potential. London dispersion interactions are considered with the Lennard-Jones potential. Therefore, the overall functional form of the potential energy in CHARMM force field is

$$U = \sum_{\text{bonds}} K_b(r - r_0)^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_0)^2 + \sum_{\text{dihedral}} K_\chi[1 + cos(n\chi - \sigma)]$$
$$+ \sum_{\substack{\text{nonbonded} \\ \text{pairs}}} \left(\epsilon_{ij} \left[\left(\frac{R_{min,ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{min,ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{\epsilon r_{ij}} \right). \tag{1.3}$$

The first three terms in the this potential are the bonded interactions corresponding to bond and angle vibrations, and dihedral potential. The bond and angle vibrations are modeled with harmonic potential with the force constants of K_b and K_θ, that are centered at the

equilibrium bond length and bond angle of r_0 and θ_0 , respectively. The torsion of two four atoms connected by three bonds is regulated by a periodic potential with the periodicity n around the angle χ with the sift σ. The last two terms in the CHARMM force field account for the non-bonded interaction between particles, i.e. van der Waals (VDW) and electrostatic interactions. The Lennard-Jones potential (LJ) represents the with the VDW interaction. The parameters $R_{min,ij}$ and ϵ_{ij} represent the location of the minimum in the LJ potential. Lastly, Coulombic potential between two partial charges q_i and q_j form the electrostatic interaction.

Parameters such as point charges, bonds/angle force constants, and the location and depth of the Lennard-Jones function are obtained for the basic molecules, such as amino acids, so that the bulk properties of these molecules are in agreement with the experiment.

These interactions guide the positions and momenta in time while thermodynamic conditions, including temperature and pressure, are maintained by an external regulatory schemes such as thermostats and barostats . The desired structural and energetic properties are collected during the simulation and their time averages are reasonable estimates of the thermodynamics of the system.

1.2 AIMD Simulation

Despite the great success of the classical MD methods in simulating a variety of molecular events, such as biological and atmospheric processes, they are not applicable to the chemical processes where the electronic interactions dominate the overall process. To describe these events, the electronic interactions need to be incorporated during the dynamics of the system. In the AIMD simulation methods, electronic interactions are calculated from the electronic structure methods and nuclei are usually treated classically. All the bonded and non-bonded

interactions in molecules naturally emerge from the electronic interactions and the empirical force fields are no longer required. However, due to the complexity in the physics and formalism of quantum mechanics, the scale that these methods can be utilized is limited to few hundred atoms. Forces applied to the nuclei are calculated, using Born-Oppenheimer approximation, after every electronic structure calculation and nuclei positions are updated according to Newton's equations of motion. Although in principle any method under the framework of quantum mechanics can be employed to calculate the electronic interactions, methods based on density functional theory (DFT) are usually chosen in the AIMD simulations. This is due to the efficiency and accuracy of the DFT methods compared to the traditional wave function based theories. The Hohenberg-Kohn theorem is the foundation of all the DFT methods. According to this theorem,[8] two systems with different Hamiltonian operators, i. e. $H_1 - H_2 \neq \mathrm{C}$, where C is a constant, will have different electronic densities and vice versa. Therefore, a one-to-one mapping between the electron density and the energy of the system exists.

Kohn-Sham scheme provides the means to obtain the energy of the system from the electron density.[9] This is achieved by connecting the density of the interacting system to the density of a noninteracting system under certain external potential v_s. In the Kohn-Sham DFT (KS DFT), the energy of the system is defined as:

$$E[n(r)] = T[n(r)] + E_H[n(r)] + E_{XC}[n(r)] + E_{ext}(r) \tag{1.4}$$

where $n(r)$ is a unique electron density of the molecular system.

$$n(r) = 2\sum_{\alpha}^{N/2} |\phi_\alpha(r)|^2$$

and $\{\phi_\alpha\}$ construct the corresponding N-particle ground state wave function $|\Phi_0\rangle$ in the form of the Slater determinant:

$$|\Phi_0\rangle = \frac{1}{\sqrt{N!}} \begin{pmatrix} \phi_1(r_1\ \sigma_1) & ... & \phi_N(r_1\ \sigma_1) \\ . & & . \\ . & & . \\ . & & . \\ \phi_1(r_N\ \sigma_N) & ... & \phi_N(r_N\ \sigma_N) \end{pmatrix}$$

Each of the energy terms in Equation 1.4 are universal functional of electron density. However, the explicit dependence of these energies to density is only known for $T[n]$ and $E_H[n]$.

$$E^H[n] = \int d^3r \int d^3r \frac{n(r)n(r')}{|r - r'|}$$

$$T[n] = -\sum_i \int d^3r' \phi_i^*(r) \nabla^2 \phi_i(r)$$

E_{ext} is the external potential, such as electron-nuclein interaction, and E_{XC} contains all the quantum effects associated with the many electron system. The functional form of E_{XC}

can be defined through obtaining accurate energy for the simple model systems, such as uniform electron gas. Upon self consistently solving the following equation the energy of the molecular system is obtained.[10]

$$\left[-\nabla^2 + V_{ext}(r) + V_H[n](r) + V_{xc}[n](r)\right] \phi_i(r) = \epsilon_i \phi(r) \tag{1.5}$$

where $V_H[n](r) = \frac{\delta E_H[n]}{\delta n(r)}$ and $V_{XC}[n](r) = \frac{\delta E_{XC}[n]}{\delta n(r)}$

In the hybrid Gaussian and plane wave (GPW) formulation of KS DFT,[11, 12] which is primarily employed in this study, $n(r)$ is expanded in terms of Gaussian basis functions. Subsequently, plane waves are used to represent the electron density in the reciprocal space $\tilde{n}(G)$. This transformation is performed such that

$$n(r) \approx \frac{1}{\Omega} \sum_G \tilde{n}(G) e^{iG.r}$$

In this representation E_H is calculated in the reciprocal space while the $T[n]$ and $E_{XC}[n]$ are accounted for in the real space . The advantage of this scheme is that it is convenient to apply this formalism to the system under the periodic boundary conditions (PBCs) where E_H can be computed very efficiently through mapping the density on a 3D-grid and utilizing the Ewald summation methods, with the Fast Fourier Transform algorithm (FFT), to calculate the Coulombic interactions. Furthermore, GPW DFT does not suffer from the basis set superposition error as the plane wave functions are naturally delocalized.

1.3 QM/MM methods

As mentioned before, application of AIMD simulation is restricted to relatively small systems. However, the need for application of AIMD to larger systems, such as enzyme-catalyzed reactions, led to the development of hybrid Quantum Mechanics/Molecular Mechanics (QM/MM) methods.[13] In the QM/MM scheme, the full system is partitioned in two (or more) subsystems. The smaller part of the system where the chemical reaction occurs is treated with the electronic structure calculations while the rest of the system is described with the classical MD methods. Subsequently, the subsystems are coupled through the non-bonded, or bonded in the case of bonds extending across different regions, interactions.[13]

The QM/MM methods have been very successful in providing very accurate description of molecular processes that occur in enzymes and other complex macromolecules.[14] In this study the QM/MM simulation methods are employed to elucidate the proton permeation mechanism in the human proton channel (Hv1).

1.4 Free Energy Methods

In this section the basics of the free energy calculation methods that are employed in this study are briefly explained. In these methods the process under study is mapped on to a function that depends on the coordinates of the system, reaction coordinate. In order to calculate the free energy, the entire phase space along the reaction coordinate needs to be adequately sampled. Therefore, regions of space associated with barrier in free energy would be rarely visited during the free/unbiased dynamics. Variety of free energy calculation methods are devised to overcome this sampling issue. In these methods usually the system evolve under biasing restraint, constraint, or energy corresponding to umbrella sampling, blue moon ensemble, and metadynamics, respectively.

1.4.1 Umbrella Sampling

In the umbrella sampling (US),[15] simulation along the reaction coordinate is evenly divided into smaller regions, free energy windows. Subsequently, the evolution of the reaction coordinate in each window is harmonically restrained to the center of the window. Hence the system is enforced to explore the high energy regions of phase space. The effect of biasing force on the energetics and dynamics of the system can be corrected based on the relationship between the unbiased and biased probability distributions at each free energy window $\langle\ \rho(\xi)\ \rangle_i$ and $\langle\ \tilde{\rho}(\xi)\ \rangle_i$,[16] respectively,:

$$\begin{cases} \langle\ \rho(\xi)\ \rangle_i = e^{\frac{\beta}{2}(\xi-\xi_i)^2} \langle\ \tilde{\rho}(\xi)\ \rangle_i\ e^{-\beta F_i} \\ e^{-\beta F_i} = \int d\xi\ e^{\frac{-\beta}{2}(\xi-\xi_i)^2} \langle\ \rho(\xi)\ \rangle \end{cases} \tag{1.6}$$

Because of the relationship between the formulas in Equation 1.6, this Equation is solved self consistently. To obtain the free energy along the entire reaction coordinate the probabilities/free energies of all windows are combined. The weighted histogram analysis method (WHAM) achieves this through the minimization of the error associated with the binning the data from the umbrella sampling simulation.[17]

1.4.2 Blue Moon Ensemble

In the blue moon ensemble technique, a similar windowing scheme as US is devised. and free energy of a rare event is estimated by fixing the value of the reaction coordinate in each window.[18] The system is free to evolve as long as the value of reaction coordinate in the specific window remains constant, that is $\frac{d}{dt}\vec{q} \neq 0$. The values of reaction coordinate in each window are fixed using the SHAKE or RATTLE algorithms applied in the space of

generalized coordinates, $q_\alpha = \xi(\vec{q})$. Subsequently, The free energy in the Blue Moon scheme is calculated from:[19]

$$A(\xi) = A(\xi^0) + \int_{\xi^0}^{\xi} d\xi' \frac{\langle z^{\frac{-1}{2}}(r)\,[\lambda + k_B T G]\rangle_{\xi'}}{\langle z^{\frac{-1}{2}}(r)\rangle_{\xi'}}$$

where

$$\begin{cases} z = \sum_i \frac{1}{m_i} \left(\frac{\partial \xi(r)}{\partial r_i}\right)^2 \\ G = \frac{1}{z^2(r)} \sum_{i,j} \frac{1}{m_i m_j} \frac{\partial \xi(r)}{\partial r_i} . \frac{\partial^2 \xi(r)}{\partial r_i \partial r_j} . \frac{\partial \xi(r)}{\partial r_j} \end{cases} \tag{1.7}$$

and λ is the Lagrange multiplier obtained from the SHAKE algorithm. For instance, if a certain process can be described with a distance between the reactants and products, the free energy in the Blue Moon scheme at window i is reduced to the $\langle \lambda \rangle_i$.

1.4.3 Metadynamics

Many free energy calculation techniques require reducing the process under study into a single collective variable followed by biasing the dynamics and statistics of the system. The restriction in the number of simultaneous reaction coordinates and prior knowledge of reaction mechanism limit the application of these methods.

The metadynamics technique for the calculation of free energy enhances sampling of the phase space by intermittently adding a Gaussian shaped potential to the system. These added potentials frustrate the potential wells in the reaction coordinate and the system is encouraged to overcome the free energy barriers and explore the phase space. [20]

The total amount of added potential to the system at time t is:

$$V_g(\xi(\vec{q}),t) = w\sum_{t'} e^{-\frac{\left(\xi(\vec{q},t)-\xi(\vec{q},t')\right)^2}{2\delta s^2}} \tag{1.8}$$

where w is the width of the Gaussian potentials added to the system and $\xi(\vec{q},t)$ is the value of the reaction coordinate at time t.

If the width and height of the deposited potentials are comparable to the exterma on the free energy surface, at the correct frequency:

$$\lim_{t\to\infty} V_g(\xi(\vec{q}),t) \approx A(\xi) \tag{1.9}$$

Therefore, by collecting the evolution of the reaction coordinate and knowledge of the shape of the Gaussian potentials, one can calculate the free energy of the corresponding process. The power of the metadynamics scheme is in its very simplistic formalism and the possibility of sampling several reaction coordinates simultaneously.

To improve the sampling efficiency in metadynamics, a Lagrangian based extension to the original formalism was proposed. In the extended-metadynamics a mass is assigned to the collective variable and its motion is coupled to a thermostat.[21] Subsequently, through coupling the motion of the collective variable to the motion of the system further sampling of the potential surface is achieved.

Chapter 2

Characterizing the Interaction of Organic Compounds with the Indoor Surfaces

2.1 Introduction

The importance of understanding atmospheric chemistry and its implication in public health is indisputable. However, a significant portion of our daily lives is spent indoor where we are exposed to a broad range of organic compounds that are released in the air and subsequently accumulate on the accessible surfaces. For instance, the film formation of short and long length chlorinated paraffins, important components of variety of household products, on the glass surface has be reported.[22] The environmental and health risks associated with these compounds led the US[23] and European[24] Environmental Protection Agencies to limit the use of chlorinated paraffins. Ozone finds its way indoor from the outdoor atmosphere. Terpene molecules that are released indoor can be easily oxidized with ozone into potentially

more harmful compounds.[25, 26] It has been demonstrated[27] that indoor surfaces, such as glass and wall, can further enhance the reactivity of these compounds. All these processes lead to formation of volatile organic compounds (VOCs) indoor. Previous measurements have observed that the concentration of VOCs indoor is higher than outdoor.[28, 29]

Therefore, understanding the nature of interactions formed between the indoor related compounds, such as terpenes and surfaces, such as glass, is essential to explain the kinetics and mechanisms of these processes. Computer simulation methods provide us with the tools that can be employed to investigate these interactions on the atomistic scale. In this study, structures and energetics of the interactions formed between the terpineols and terpenes with the amorphous silica, rutile, and gypsum surfaces are reported. The simulation results in accordance with IR experiments shed light on the kinetics and mechanisms of film formation.[5] Combination of classical MD and AIMD simulation tools is utilized to study the surface adsorption of a variety of organic compounds. Amorphous silica structure is used to represent the glass surface. Adsorption free energy and vibrational spectra are used to characterize the strength of interaction between the adsorbent species and the silica surface. Furthermore, structural identification of conformations of organic compounds on silica and contacts formed between silica and adsorbed species is provided.

2.2 Simulation Protocols

2.2.1 Classical Molecular Dynamics Simulations

The amorphous silica structure was generated applying an annealing procedure.[30] Initially, an alpha-quartz supercell composed of 11 $\times$ 11 $\times$ 8 unit cells was built. To accommodate periodic boundary conditions, bonds were introduced between the atoms located at the borders of the crystal with their bonding partners located at the opposite surface. Upon

completion of the annealing process a 24 × 52 × 50 $Å^3$ slab was selected from the annealed bulk structure. To ensure that all Si atoms located at the surface in the X direction (normal to the SiO_2 surface) satisfied tetrahedral coordination, a few oxygen atoms were added to the system. In the simulations involving the silica slab, a 40 Å thick region of vacuum was added to both sides of the slab in the X direction. The hydroxylated silica surface was generated by hydrogenating the surface oxygen atoms that had only one Si–O bond, resulting in a silanol surface density of 6.7 nm^{-2}.

All of the molecular dynamics (MD) simulations were performed using the LAMPPS package.[31] The equations of motion were integrated using the velocity–Verlet algorithm with a 1 fs time step. Electrostatic interactions were evaluated with the particle–particle particle–mesh solver[32] with a 14 Å cutoff distance for the short–ranged nonbonded interactions. The simulation temperature was maintained at 295 K using the Nosè – Hoover thermostat with a relaxation time of 100 fs. CHARMM-compatible bonded and nonbonded force field parameters optimized for the SiO_2 were employed, and the CHARMM CGenFF force field[33] parameters were used for limonene. Subsequently, a D–limonene molecule was introduced in the vacuum region adjacent to the slab and a 0.5 μs MD trajectory was generated at constant volume and a constant temperature of 295 K.

Umbrella sampling[15] was employed to calculate the potential of mean force (PMF) or free energy profile for the desorption of limonene from the silica surface. The distance of the center of mass of limonene from the surfaces was chosen as the reaction coordinate, and the desorption process was divided into 31 windows at 0.5 Å increments. A harmonic restraining potential with a force constant of 20.9 kJ $mol^{-1}Å^{-2}$ was applied in each window. The free energy profile was generated from 10 ns long biased trajectories for each PMF window using the WHAM scheme.[16] The limonene desorption enthalpy was estimated by energy minimization as a function of the separation between the limonene center of mass and the silica surface. This procedure was repeated for 100 different initial structures extracted

from the 0.5 μs MD simulation. The activation energies for the transitions from the C* up to the C* down and the C*down to C*up configurations were estimated from the relative energies of the C* up, C* down, and vertical orientations of the limonene ring with respect to the silica surface. The energies were calculated at the M06–2X/6–311++G(d,p) level[34] using structures of silica and limonene optimized at the M06–2X/6–311G(d) level. The energies reported in Table 2.1 were corrected for basis set superposition error.[35] The initial structures for these calculations were obtained from the force field-based MD simulation. The size of the silica cluster used in the electronic structure calculations, 112 atoms, was chosen such that it can fully contain the different orientations of the limonene molecule. All of the electronic energy calculations were performed using the GAUSSIAN16 package.[36]

2.2.2 AIMD Simulation

AIMD simulations were utilized to elucidate the origin of the change in the vibrational frequency of SiO_2 upon adsorption of organic compounds. AIMD simulations of systems composed of limonene, cyclohexene, benzene, and cyclohexane adsorbed on a cluster of hydroxylated SiO_2 were carried out using the CP2K program.[11] The SiO_2 cluster was obtained from a force field-based MD simulation of amorphous SiO_2. The size of the silica cluster with $Si_{23}O_{66}H_{40}$ composition was chosen such that it can fully accommodate the adsorption of the organic compounds. For each system 40 ps of AIMD simulation at 295 K with time step of 0.5 fs was performed. The BLYP–D3 exchange–correlation functional with the DZVP–MOLOPT–SR–GTH basis set[37] and the GTH pseudopotentials[38] in the QUICKSTEP module of CP2K package were employed.[12] A SCF convergence criterion of 10^{-8} (a. u.) with the orbital transformation[39] scheme was applied during the simulation. Each system was placed in an orthorhombic box of 20$\times$ 21 $\times$ 22 Å^3 under periodic boundary conditions in the Y and Z dimensions (the X direction is along the normal to the SiO_2 surface). The simulation temperature was maintained at 295 K using a

Nosè–Hoover thermostat[40] with a relaxation time of 100 fs. The binding energy of organic compounds to SiO_2 was calculated using the Gaussian 16 package.[36] Optimized structures of organic compounds and SiO_2 cluster in the isolated and complex forms were obtained with M06–2X/6–311G(d).[34] Subsequently, counterpoise corrected binding energies were estimated by M06–2X/6–311++G(d,p). Vibrational power spectra calculated from the AIMD simulation were utilized to shed light on the interaction of organic compound with the SiO_2. IR and Power spectra are computed from the autocorrelation function of velocity or dipole moment [41, 42]:

$$p(\omega) = m \int \langle\ v(0)v(t)\ \rangle_\tau\ e^{-i\omega t} dt$$

$$A(\omega) \propto \int \langle\ \dot{\mu}(0)\dot{\mu}(t)\ \rangle_\tau\ e^{-i\omega t} dt$$

where $v(t)$ and $\dot{\mu}$ are the velocity and time derivative of dipole moment of the selected atoms with mass m at time t and ω is the frequency. The brackets represent an average over atoms and time origins in the trajectory. The dipole moment of each molecule is obtained from the Wannier orbital localization scheme.[43] Wannier orbitals are obtained through a unitary transformation of the wave function in the plane wave representation such that the transformed orbitals have minimum spread. Subsequently, the dipole moment of each molecule is computed according to:

$$\mu = -2 \sum_e r_e + e \sum_N Z_N R_N$$

where the Z_N is the effective nuclear charge located at R_N and r_e is the center of the Wannier

orbital that contains 2 electrons.

2.3 Limonene Adsorption on the Silica Surface

Structural Characterization of Limonene on the SiO_2 Surface During a 0.5 μs of classical MD simulation, we observed that limonene has two predominant configurations, in which the six-membered ring is in contact with and parallel to the SiO_2 surface. The limonene molecule stays in the more stable half-chair conformation with the propenyl group in the equatorial position[44] during the entire trajectory. The two predominant configurations of the limonene molecule on the SiO_2 surface are related by a roughly 180° rotation about the long axis and can be classified by the position of the chiral carbon atom, which we label C*. In one configuration, which we refer to as "C* up", the C* atom is further from the SiO_2 surface than in the other configuration, which we refer to as "C* down". Additionally, the populations of the C* up and C* down conformations are 43% and 35%, respectively. In the C* up configuration, the limonene molecule forms a more favorable interaction with the surface, namely, closer contact between the propenyl group and the SiO_2 surface, which results in the C* up orientation being the more probable conformation.

The free energy profile for the reorientation of the limonene molecule on the surface, i.e., the transition from the C* up configuration to the C* down configuration, was calculated from the probability distribution of the angle, θ, between normal vectors in the limonene ring and the SiO_2 surface (defined in Figure 2.1) according to:

$$F(\theta) = -k_B T \ln p(\theta)$$

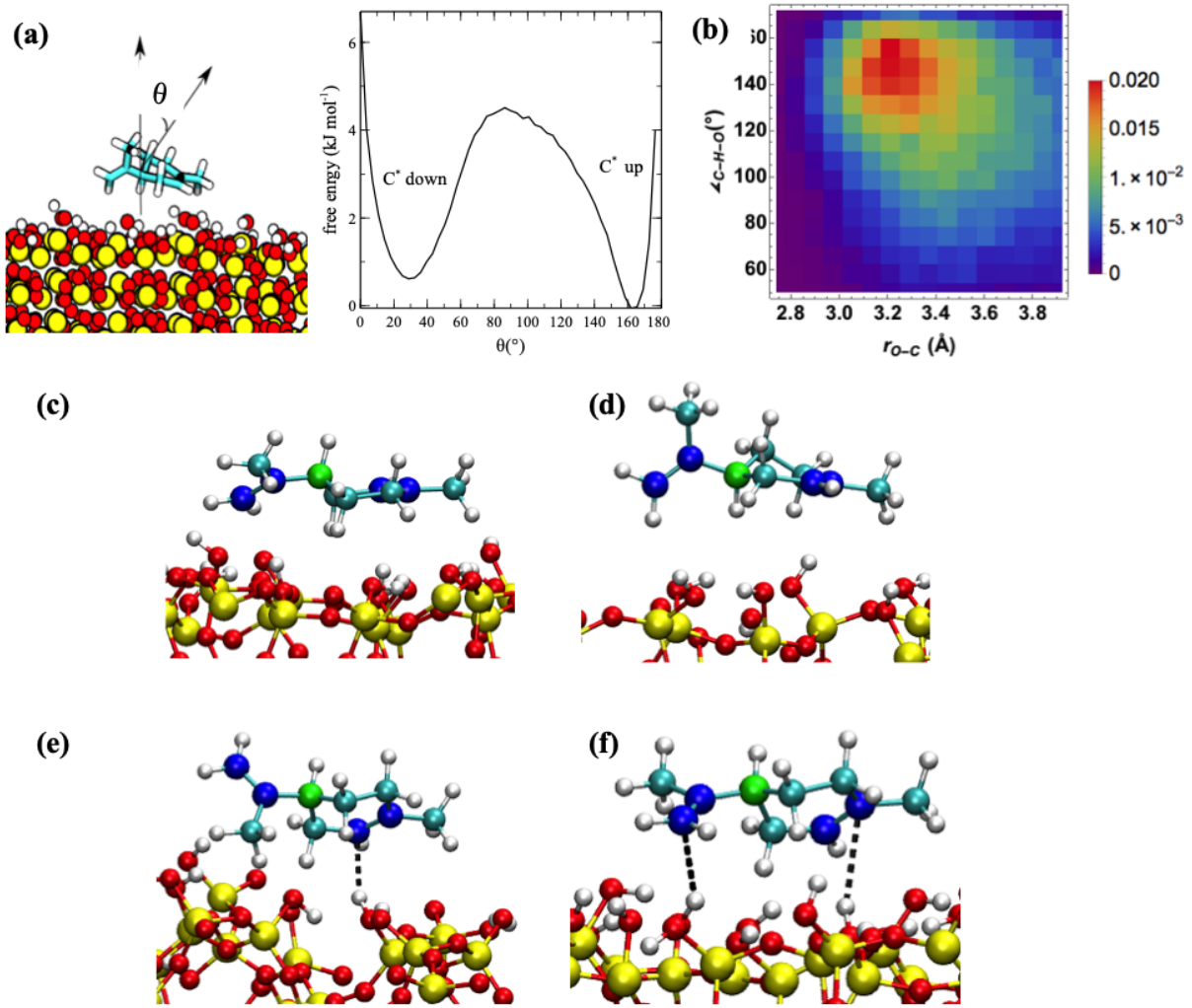


Figure 2.1: (a) Limonene reorientation free energy profile on the silica surface computed from a 0.5 μs of classical MD simulation. (b) Probability distribution of the O–C_{sp^2} distance and C–H...O angle for the nearest hydrogen atoms of the silica to the sp^2 carbon atoms. The probabilities are computed by dividing the count in each bin by the number of steps in the trajectory. (c–d) Snapshots of the limonene molecule in the C* up (c) and C* down (d) orientations. The chrial carbon atoms are colored green, the sp^2 carbon atoms blue, the sp^3 atoms cyan, the sulfur atoms yellow, the oxygen atoms red, and the hydrogen atoms white. (e–f) Snapshots of structures corresponding to the one (e) and two (f) hydrogen-bonding interactions between the limonene and silica surface. The dashed line depicts the hydrogen-bonding interaction.

limonene orientation	relative energy (kJ mol^{-1})
C* up	0.0
C* down	4.7
vertical	29.3

Table 2.1: Relative energies of limonene–silica complex for different orientations of limonene molecule on the silica surface. Energies are reported relative to the lowest energy C* up configuration.The optimized structures were obtained using M06-2X/6311-G(d) DFT calculations.

Here $p(\theta)$ is the probability distribution of θ acquired from the 0.5 μs MD simulation, k_B is Boltzmann's constant, and T is the temperature (295 K). The zero of the free energy profile was taken to be that of the most probable C* up configuration (θ=165°). According to the resulting free energy profile plotted in Figure 2a, the free energy barrier to the transition from the C* down configuration to the C* up configuration is 4.0 kJ mol^{-1}, and the barrier to the transition from the C* up configuration to the C* down configuration is 4.6 kJ mol^{-1}.

The enthalpic barrier for the transition between the C* down and C* up configurations was estimated using the relative energies of the limonene–silica complex in the C* down, C* up, and vertical configurations of limonene with respect to the silica cluster. The energies were calculated with M06–2X/6–311++G(d,p)//M06–2X/6–311G(d) electronic structure method.[34] Results are summarized in Table 2.1 after correcting for basis set superposition error (BSSE).[35] According to these calculations, the enthalpic barrier for passing from the C* down to the C* up configuration is 29.3 kJ mol^{-1}.

The radial distribution functions (RDFs) between the hydrogen atoms of the SiO_2 surface and the sp^3 and sp^2 carbon atoms of limonene are presented in Figure 2.2. The value of r_{HC} at the first peak in the RDF is the distance at which preferred interactions occur. For the sp^3 carbon atoms, the first peak in the RDF is at 3.5 Å, whereas for the sp^2 atoms it is at 2.5 Å. The latter is indicative of a π–H bonding interaction between the hydrogen atoms of the SiO_2 surface and the double bonds of the limonene molecule. The height and location of the first peak in the RDFs for the C* up and C* down configurations are the same, indicating

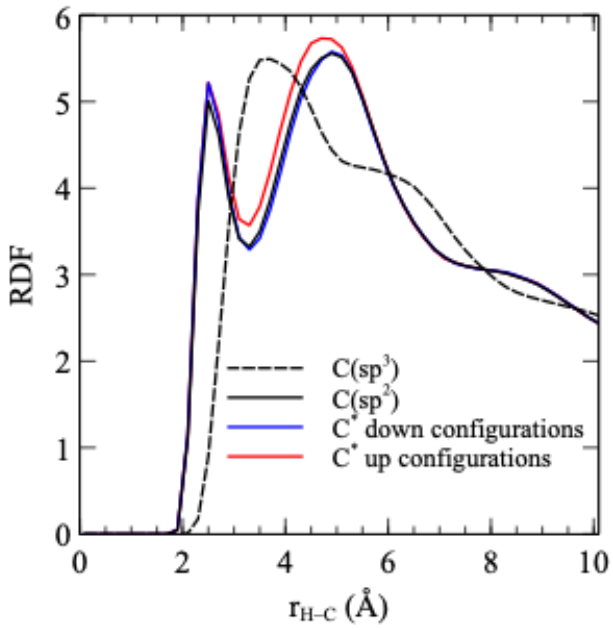


Figure 2.2: H(silica)–C radial distribution functions computed for the distances between the sp^3 and sp^2 carbon atoms of limonene and the H atoms of the silica surface calculated for the full trajectory and separately for structures corresponding to the C* up and C* down conformations.

that the strength of π–H bonding in the two configurations is similar.

The histogram of the O–C_{sp^2} distance and OH....C_{sp^2} angle plotted in Figure 2.1(b) demonstrates that the most probable O–C_{sp^2} distance and O–H...C_{sp^2} angle occur at 3.2 Å and 155°, respectively, as expected for a hydrogen-bonding interaction. Based on Figure 2.1(b), we defined the criterion for $\pi-$H bonding as the O–C_{sp^2} distance < 3.4 Å and the O–H...C_{sp^2} angle between 135° and 165°. According to this criterion, the probability of limonene engaging in one and two hydrogen bonds with the SiO_2 surface is reported in the Table 2.2. Snapshots depicting the one and two hydrogen-bonding interactions between limonene and the SiO_2 surface are depicted in Figure 2.1(e–f)

According to the results provided in Table 2.2, in the C* up configuration the probability of formation of two hydrogen-bonds between the limonene molecule and SiO_2 increases by 0.5 %, while the number of one hydrogen-bonding interactions decreases by 1.0% relative to the C* down configuration. As the propenyl group possesses more rotational flexibility

Configuration	no HB	1 HB	2 HB
C* down	0.679	0.299	0.022
C* up	0.684	0.289	0.027

Table 2.2: The probability of formation of no, one, and two hydrogen-bonding interactions between the limonene molecule and silica surface for the C* down, C*up configurations.

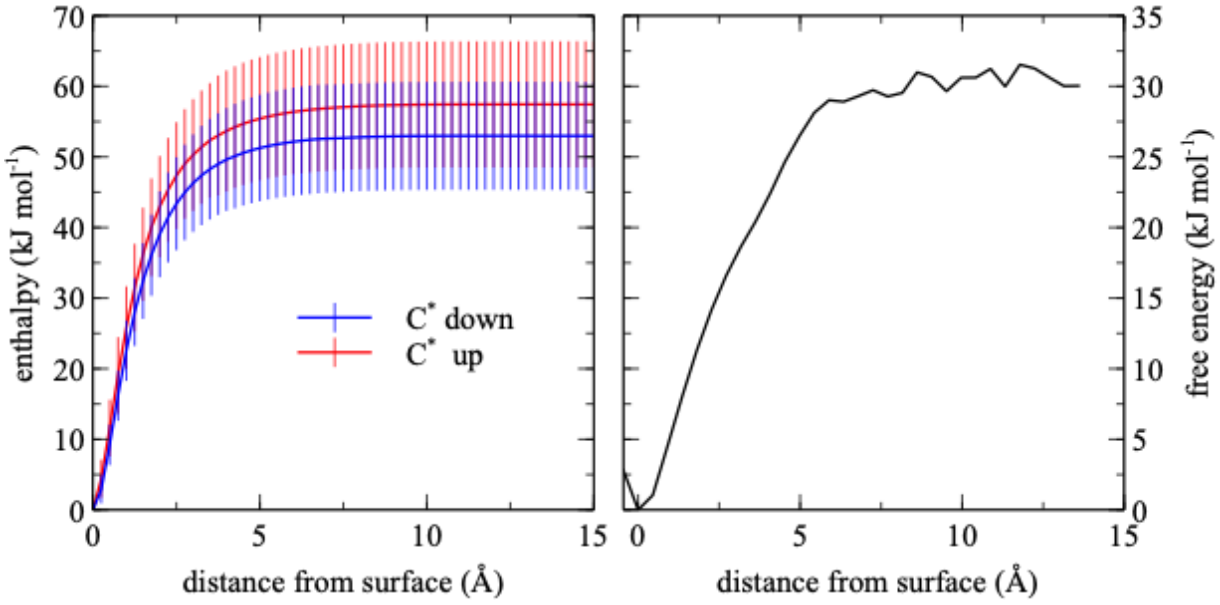


Figure 2.3: Desorption enthalpy and free energy computed from MD simulation. The error bars are based on the standard deviation of the energy computed for the 100 initial structures.

compared to the rest of molecule, the probability of the sp^2 C atoms in the propenyl group being involved in the hydrogen–bonding interaction with the surface is 8% lower than the sp^2 carbon atoms located in the ring. For instance, 58% of 1 HB structures in the C* up configuration originate from the contacts between the double bond on the ring and surface.

Energetics of the Desorption Process From the PMF for limonene desorption from the SiO_2 surface plotted in Figure 2.3 (right panel), we estimate that the desorption free energy is 30 kJ mol^{-1}. We note that this value is a population–weighted average of both the C* down and C* up configurations. Using energy minimization, we estimated the enthalpy as a function of the distance between the limonene molecule and the SiO_2 surface, separately for the two predominant limonene configurations, obtaining a desorption enthalpy of 57 ± 8.9

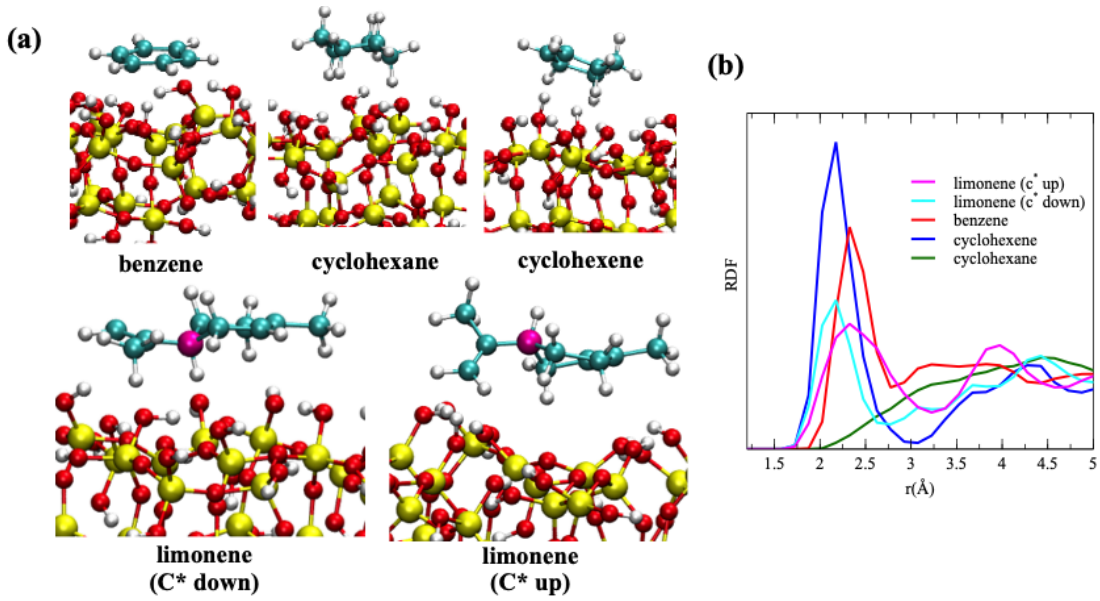


Figure 2.4: Snapshots from the AIMD simulations of the cyclic molecules adsorbed on the SiO_2 surface. The sp^2 carbon, chiral carbon, oxygen, hydrogen, and silicon atoms are colored blue, pink, red, white, and yellow, respectively. (b) Radial distribution functions calculated between the hydroxyl hydrogen atoms and sp^2 carbon atoms of limonene and cyclohexene, center of mass of benzene, and carbon atoms in cyclohexane

kJ mol^{-1} for the C^* up configuration and 53 $\pm$ 7.6 kJ mol^{-1} for the C^* down configuration (Figure 2.3). However, considering the large error bars, the difference between the adsorption enthalpies of the two configurations is not statistically significant.

2.4 Adsorption of 6-member Ring Analogues of Limonene on Silica Surface

In order to characterize the role of hydrogen-bonding and London dispersion interaction in the close contacts formed between limonene and the silica surface we further studied other cyclic compounds, including cyclohexane, cyclohexene, and benzene.

Within 40 ps of AIMD simulation, we observed that the cyclic part of all of the organic

adsorbed cyclic organic compound	location of the peak(Å)
limonene (C* up)	2.3
limonene (C* down)	2.2
benzene	2.3
cyclohexene	2.2
cyclohexane	4.4

Table 2.3: Locations of the first feak in the RDFs (Figure 2.4 b) Depicting the interactions of organic compounds with the H atoms of the SiO_2 surface. adsorbed cyclic organic compound location of the first peak (Å)

adsorbed cyclic hydrocarbon	6–311G(d)	6–311++G(d,p)
limonene (C* up)	-42.19	-46.26
limonene (C* down)	-43.89	-41.57
benzene	-28.33	-27.50
cyclohexene	-28.05	-29.15
cyclohexane	-20.04	-24.26

Table 2.4: binding energies (kJ/mol) calculated using the M06–2X method. The reported values are corrected for the basis set superposition error. The 6–311++G(d,p) binding energies were determined from single–point energy evaluation of the M06–2X/6-311G(d)-optimized structures.

compounds considered predominantly stays parallel to the SiO_2 cluster, allowing more favorable contacts and/or interactions with the SiO_2 surface as shown in Figure 2.4a. The radial distribution function (RDF) is used to assess the interaction of organic compounds with the surface. The RDF is calculated between the SiO_2 hydroxyl hydrogen atoms and sp^2 carbon atoms of limonene and cyclohexene, the center of mass of benzene, and carbon atoms of cyclohexane. The positions of the peaks in RDF correspond to the most probable distance of hydrogen atoms of hydroxyl groups near the selected carbon atoms. The location of the first peak for each system is summarized in Table 2.3. As shown in Figure 2.4b and Table 2.3, compounds with C=C bonds (limonene, benzene, and cyclohexene) localize closer to the surface hydroxyl group than the saturated cyclohexane.

The hydroxyl H atoms stay within 2.2–2.3 Å of the unsaturated carbon bonds. This close contact is associated with the $\pi-$hydrogen bonding interaction between the SiO_2 and the

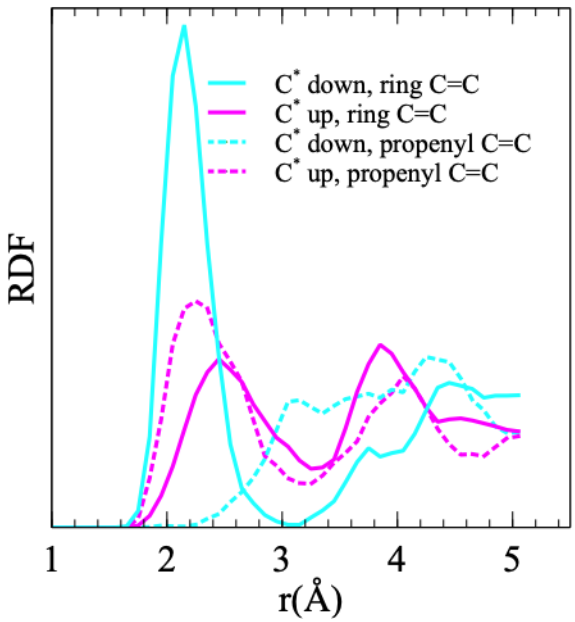


Figure 2.5: Radial distribution function calculated between the hydroxyl hydrogen atoms and sp^2 carbon atoms located at the ring and propenyl group of limonene in the C* up and C* down conformations. The RDF is calculated from AIMD simulation

unsaturated cyclic hydrocarbons. Although the first peak for the C* up configuration is 0.1 Å farther from the SiO_2 H atoms than that of the C* down configuration, the sp^2 carbon atoms located at the propenyl group form a significantly closer contact with hydroxyl groups in the C* up conformation than in the C* down conformation, as illustrated in Figure 2.5. The position of the first peak in the RDF for cyclohexane (Figure 2.4b) is indicative of a weak interaction with the SiO_2 surface, dominated by dispersion interactions. The higher RDF peak of benzene is linked to the rigidity and more delocalized electron density of benzene (Figure 2.4b) compared to the other cyclic compounds.

Binding energies between the organic molecules and SiO_2 cluster quantify the strength of the interaction between SiO_2 and the cyclic compounds. The M06–2X/6 –311G(d)[34] density functional theory method was employed to obtain the optimized structure and energy of the individual fragments (organic molecule and SiO_2 cluster) and complex (organic molecule adsorbed on the SiO_2 cluster). The binding energies, calculated as the difference between the energy of the complex and that of the individual fragments, were corrected for basis set su-

perposition error (BSSE).[35] Subsequently, single-point energy calculations were performed to calculate the binding energies with the 6–311++G(d,p) basis set using the geometries optimized with the 6–311G(d) basis set. According to these calculations (Table 2.4), and consistent with the Fourier transform infrared (FTIR) data (Figure 2.6), the interaction between limonene and SiO_2 is significantly stronger than those of the other compounds explored in this study. This is due presumably to the larger size of limonene (more dispersion interaction) and the existence of two unsaturated bonds that can participate in $\pi-$hydrogen bonding interactions. Although limonene has fewer unsaturated bonds than benzene, its more flexible structure leads to stronger interaction between the C=C moieties and hydroxyl groups, in addition to enhanced dispersion interactions, compared to benzene.

Vibrational Spectra: An additional property that can be used to characterize the strength of interaction between the adsorbate and surface is the vibrational frequencies of the silanol group. One signature of the formation of a HB interaction is the red shift in the vibrational frequencies of groups involved in the HB. In fact, experimentally IR measurements are utilized to study the surface adsorption.

Experimental IR measurments Because we are most interested here in understanding the molecular level interaction of limonene adsorbed on hydroxylated SiO_2, we focus on the O–H stretching region and the hydrogen bonding interactions. Experimental FTIR measurements reveal that the negative peak at 3742 cm^{-1} for adsorbed limonene (Figure 2.6d) is attributed to the loss of isolated surface silanol groups due to the SiO_2 surface interaction with limonene that results in the broad band near 3504 cm^{-1} that is assigned to the hydrogen bond between Si–OH groups and limonene. The intensity of the 3742 cm^{-1} peak decreases with an increase in gaseous limonene pressure and an increase in the surface coverage of limonene. These data show that limonene interacts with the SiO_2 surface through a hydrogen bonding interaction with surface hydroxyl groups. This interaction is reversible;

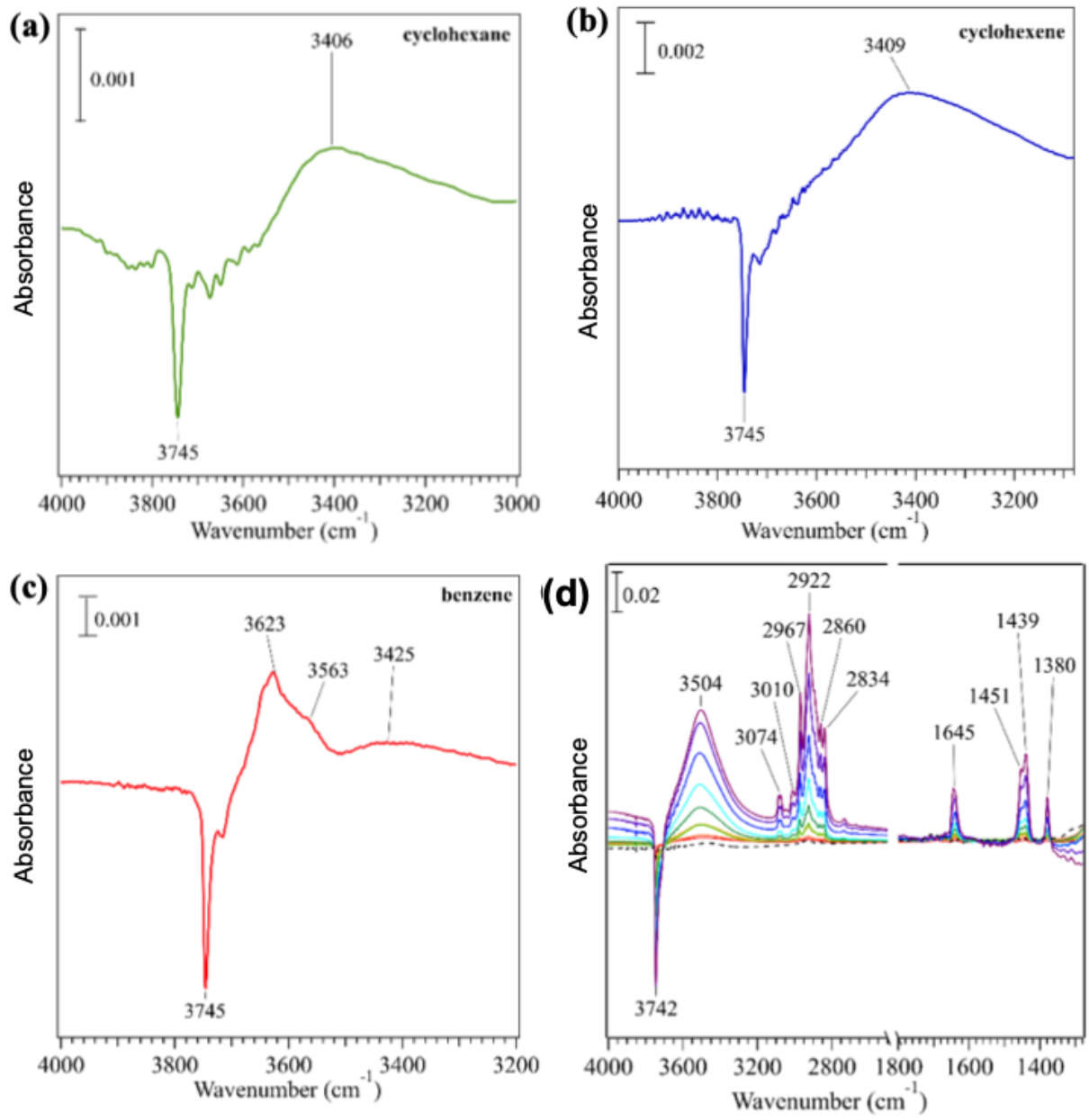


Figure 2.6: FTIR spectra of: (a) cyclohexane (3000 to 4000 cm^{-1}, 1 Torr); (b) cyclohexene (3080 to 4000 cm^{-1}, 500 mTorr); and (c) benzene (3200 to 4000 cm^{-1}, 1 Torr). (d) Absorbance spectra of limonene adsorbed on SiO_2 under dry conditions (RH $<$ 1 %) as a function of limonene pressure (1, 5, 10, 25, 50, 100, 200, 500 and 1000 mTorr) in the 1280 4000 cm^{-1} spectral regions. Note that SiO_2 is opaque below 1280 cm^{-1} due to lattice vibrations. Gas phase limonene has been subtracted from these spectra. The surface spectrum following overnight evacuation is shown as a dashed line.(reproduced by premission from Ref [1])

the intensity of absorption bands due to adsorbed limonene decreases upon evacuation, while the negative peak attributed to isolated hydroxyl groups (at 3742 cm^{-1}) reappears. Thus, limonene reversibly adsorbs on hydroxylated SiO_2.

Other cyclic compounds also form hydrogen bonds with the surface hydroxyl groups present on the SiO_2 surface as suggested by the negative peak centered at 3745 cm^{-1} in panels a–c of Figure 2.6, which is attributed to the loss of isolated surface hydroxyl groups. Adsorbed surface species appear at very low intensities for these three cyclic compounds in comparison to limonene. Figure 2.6d shows the normalized loss of the peak intensity of surface hydroxyl groups at 3745 cm^{-1} as a function of pressure for each cyclic compound. Limonene has a significantly stronger surface interaction in comparison to those of other cyclic compounds because there is a much larger loss of isolated surface hydroxyl groups in the presence of limonene than is observed for these other cyclic hydrocarbons. Among the three compounds, the loss of surface hydroxyl groups is smallest when cyclohexane is introduced, indicating that cyclohexane is weakly adsorbed on the SiO_2 surface. Benzene results in a much larger loss of surface hydroxyl groups, reaching a plateau at pressures of > 100 mTorr. Meanwhile, cyclohexene adsorption results in a larger loss of isolated surface O–H groups compared with cyclohexane, suggesting that more cyclohexene is adsorbed than cyclohexane. The interaction strength of the three cyclic compounds at pressures of <1 Torr decreases in the following order: benzene > cyclohexene > cyclohexane. At 1 Torr, cyclohexene adsorption becomes more comparable to that of benzene.

In this study, IR and power spectra are used to identify the strength of interaction. Furthermore, this will allow the simulation results to be more comparable to the experiments. As Figure 2.7 shows, the experimental and calculated vibrational spectra for limonene(g) are in good agreement with each other. Therefore, calculated spectra for the surface adsorbed organic compounds can be also compared to the experiment.

Frequency distributions amenable for qualitative comparison with the experimental FTIR

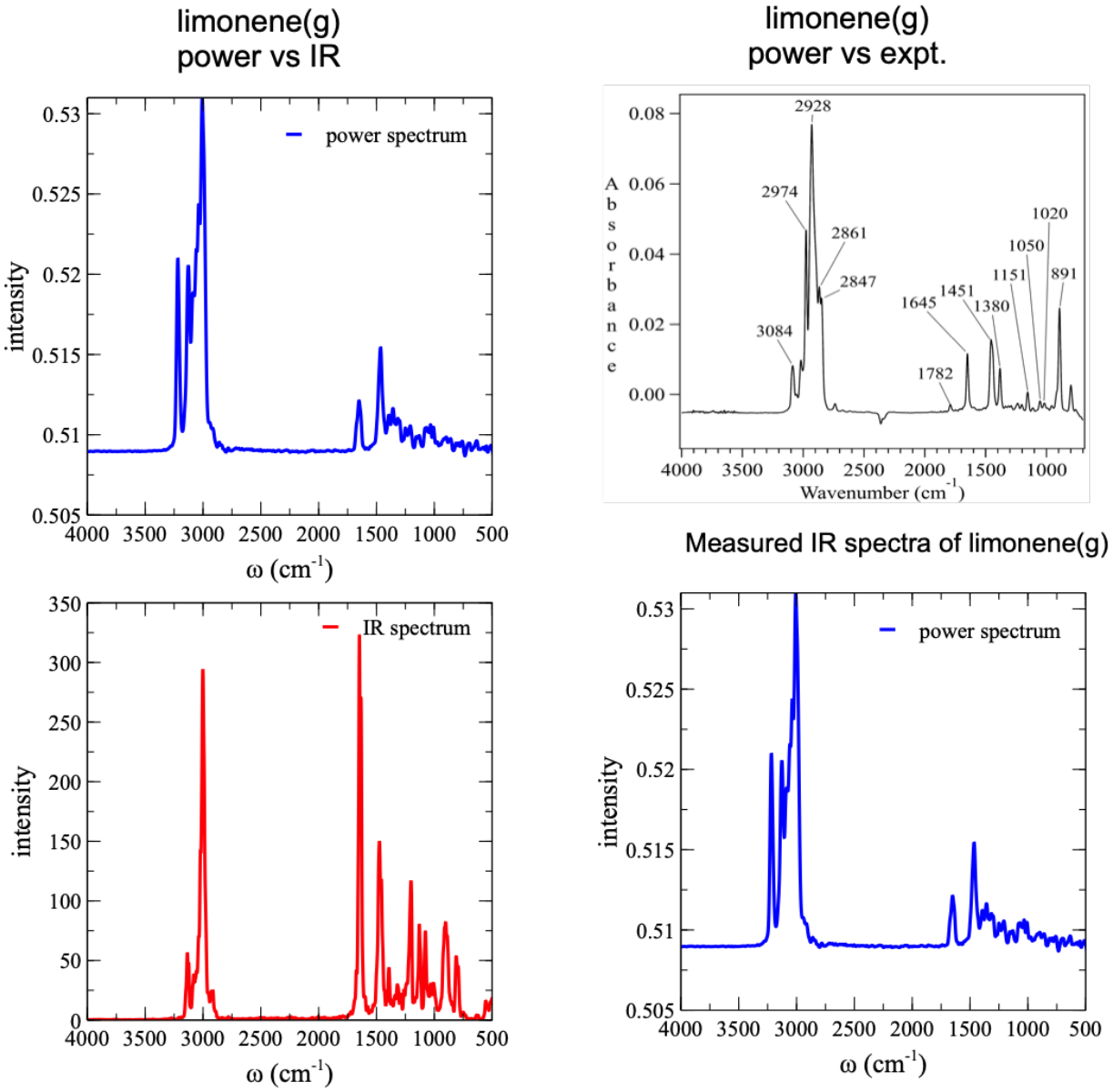


Figure 2.7: Comparison between the power and IR spectra obtained from AIMD simulation for a limonene molecule in gas phase (left panel). Comparison between the experimental IR spectra and calculated power spectra for limonene(g) (right panel). The experimental spectrum is collected in the Grassian lab at UCSD and reprinted with permission.

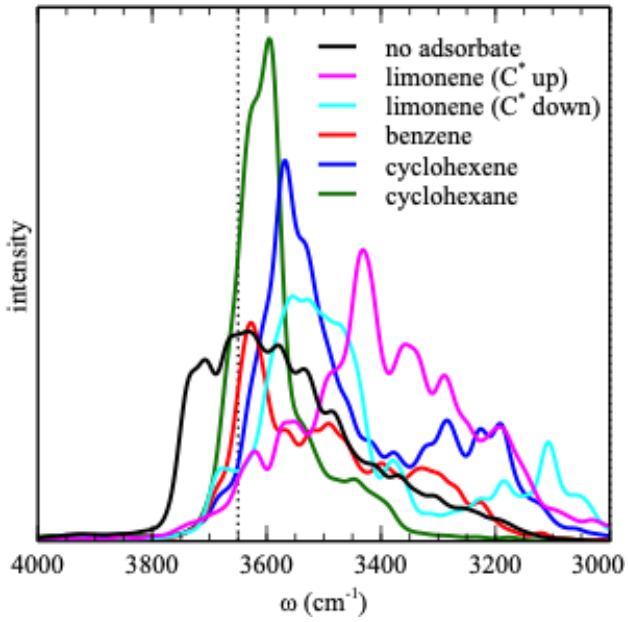


Figure 2.8: Power spectra calculated for the hydroxyl group in the vicinity of the organic compounds. The dashed line marks 3650 cm^{-1}.

spectra were calculated as Fourier transforms of the velocity autocorrelation functions of selected atoms. Subsequently, a 900 fs wide Gaussian windowing function is applied to the correlation function. The power spectra were obtained by discrete Fourier transformation of the smoothed correlation functions. The velocity autocorrelation function computed for all atoms present in each system was used to obtain the power spectra spanning the full range of frequency (Figure 2.20). In Figure 2.8 only the hydroxyl H atoms in the vicinity of sp^2 carbon atoms included in the calculation of the power spectra.

Note that only the positions and not the intensities of the peaks in the power spectra are comparable to the experimental data. Consistent with the experimental data, the calculated power spectra presented in Figure 2.8 for the H atoms of the hydroxylated SiO_2 surface display a red shift in the hydroxyl frequency upon the addition of the cyclic organic compounds. Figure 2.20 shows the power spectra for all atoms over the complete vibrational frequency range. The shifts in the vibrational frequency in the presence of the organic compounds versus bare SiO_2 are attributed to the hydrogen bonding interaction between the

SiO_2 surface and sp^2 carbon atoms of the organic compounds. The trend in the shifts in O–H stretching frequency is in agreement with the binding energies (Table 2.4) and peak locations in the radial distribution functions (Figure 2.4b). As is evident in Figure 2.8, the C^* up conformation of the limonene molecule induces the largest frequency shift, followed by the C^* down conformation of limonene.

Interestingly, cyclohexane also alters the O–H vibrations by 50 cm^{-1}, which might be a signature of hydrogen bonding interactions, due to the intermolecular charge transfer between the C–H and O–H groups.[45, 46] However, the role of dispersion forces in these types of interactions cannot be neglected. The first peak of the RDF calculated between the SiO_2 O and the H atoms of cyclohexane occurs at 2.85 Å, which is significantly greater than the peak locations for the other compounds listed in Table 2.3. Thus, we do not characterize the interaction between cyclohexane and the SiO_2 surface as a hydrogen bonding interaction.

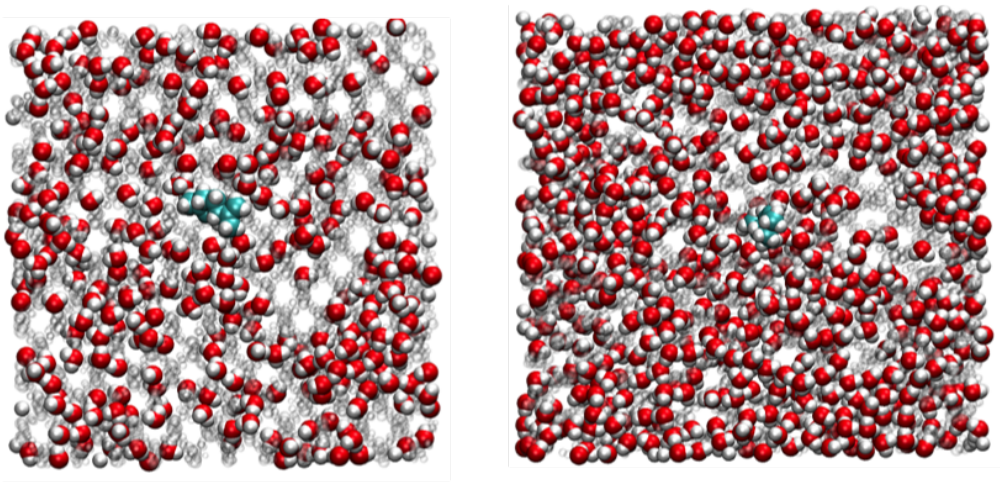


Figure 2.9: A snapshot of 50% and 100% covered surface used to compute the free energy profile in the presence of humidity. TIP3P model was used for the water molecules.

2.5 Limonene Adsorption on Silica in the Presence of Humidity

We explored the free energy of adsorption of limonene on the silica surface in the presence of humidity. Two water coverage states corresponding to 50% and 100% surface coverage, i.e. 159 and 318 water molecules, respectively, were added to the silica surface (Figure 2.9). Lower OH surface density for silica was considered, compared to the results presented for the dry silica slab, so that the surface is a better representative of the glass surface (silanol suface density=4.7 OH/nm^2). TIP3P water model[2] was used to model the water molecules. Upon equilibration a limonene molecule was introduced to the system. Subsequently, Umbrella Sampling simulations, with the similar protocols to the dry silica, were carried out. First, it should be noted that the desorption free energy of limonene from the dry silica surface has decreased by 4 kJ/mol upon decreasing the silanol surface density from 6.5 to 4.7 OH/nm^2 (as it is exhibited in Figure 2.10). Furthermore, the free energy in the presence of humidity is approximately similar to the dry surface. Lastly, the locations of minima of the free energy profile of the wet surface have shifted towards higher value in the presence of water. This

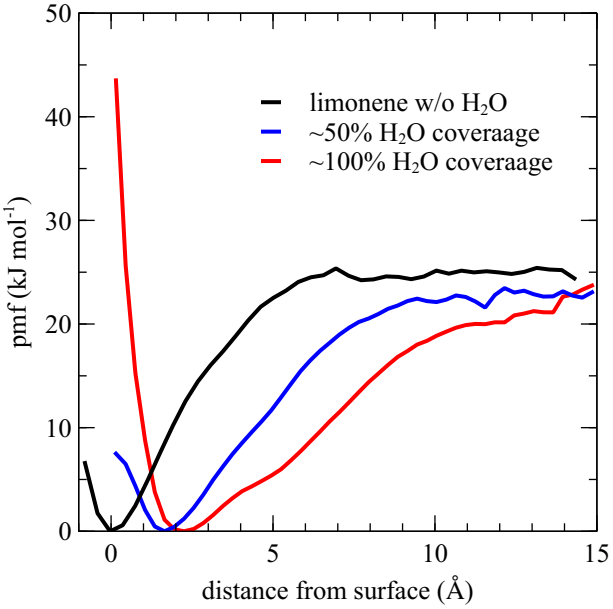


Figure 2.10: The free energy profile of desorption of limonene from wet silica surface. Two different water coverage are considered. A fully covered surface and half exposed surface, corresponding to 318 and 159 h2O molecules, respectively. Water molecules are represented by the TIP3P water model.[2]

indicates that limonene interaction with the silica surface is not sufficient to part the water molecules that are adsorbed on the surface.

2.6 Adsorption of Limonene Isomers on Silica Surface

As presented before, limonene participates in hydrogen bonding interaction with the silanol groups on the silica surface through its unsaturated bonds. To establish how the separation between the C=C groups affect the H-bonding interaction with silica surface, we investigated three isomers of limonene, including terpinolene, α-terpinene, and γ-terpinene. In addition to their similarity to limonene, these compounds are also the main scent component of a variety of cleaning products. Using classical MD simulations we estimated the adsorption free energy and surface mass accommodation coefficient in a similar manner as described for limonene. The CGenFF force field[33] parameters were adopted for these compounds.

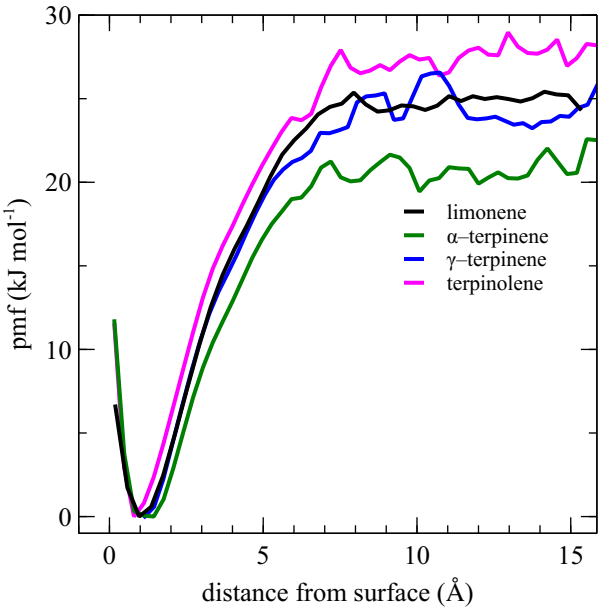


Figure 2.11: The free energy profiles calculated for three isomers of limonene. Umbrella sampling calculations were performed using the hydoxylated silica slab with the dimension and force field parameters as described before. The silanol surface density of 4.7 OH/nm^2 was used to prepare the surface. The CGenFF parameters were assigned to the organic compounds.

Similar to limonene, the surface mass accommodation coefficient for these compounds is close to one. The adsorption free energy profiles are presented in Figure 2.11. The adsorption free energy of these four isomers varies between 22–28 kJ/mol with α-terpinene and terpinolene being the weakest and strongest of the four, respectively.

Furthermore, The O–C$_{sp^2}$ and O–H...C$_{sp^2}$ distance-angle distribution for these molecules are calculated from the AIMD simulations (Figure 2.12). According to this analysis, both C=C moieties in limonene have high propensity to form close contacts with surface. Whereas, in terpinolene one of the C=C groups (C=C$_2$) is predominantly interacting with the surface and the probability of the contacts formed by this group is higher than the contacts formed in limonene. The most probable O–C distance for compounds other than γ-terpinene is located at 2.9–3.0 Å. This distance for γ-terpinene is 0.1 Å shorter than other isomers.

The power spectra calculated for the O–H groups on silica located in the vicinity of each

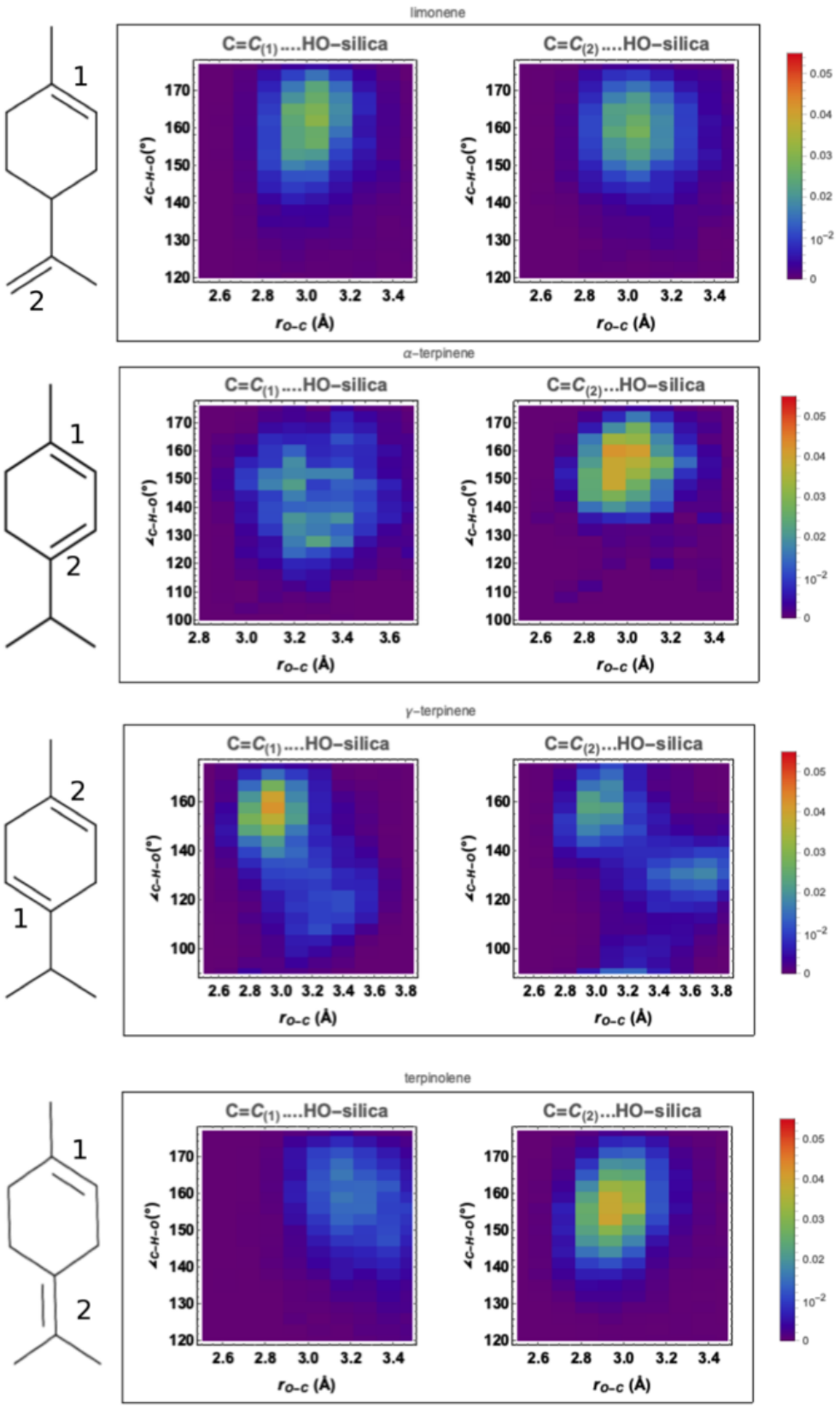


Figure 2.12: The HB donor–acceptor distance angle distribution calculated for limonene, α-terpinene, γ-terpinene, and terpinolene. The analysis was performed on the trajectories obtained from 40 ps of AIMD simulation.

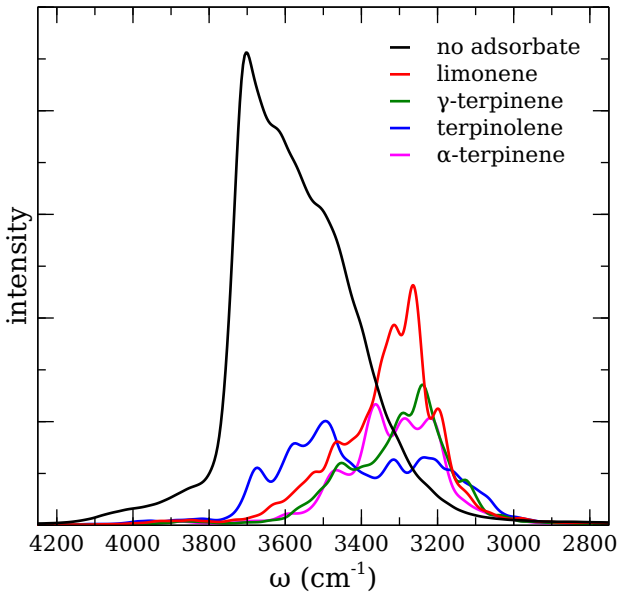


Figure 2.13: The power spectra of silica O–H groups in the vicinity of the organic compounds (limonene isomers).

compound is presented in Figure 2.13. As is evident in this Figure, all the isomers significantly alter the O–H vibrations upon HB formation. The vibrational frequency of H atoms in HB with terpenes span 3200–3400 cm^{-1} frequency range. The extent of this effect is similar for all of these compounds. The IR spectra for the entire silica cluster and terpene molecules are presented in Figure 2.21 , 2.22. IR spectra are obtained using the maximally localized Wannier orbitals and the TRAVIS code[3] is used to generate these spectra.

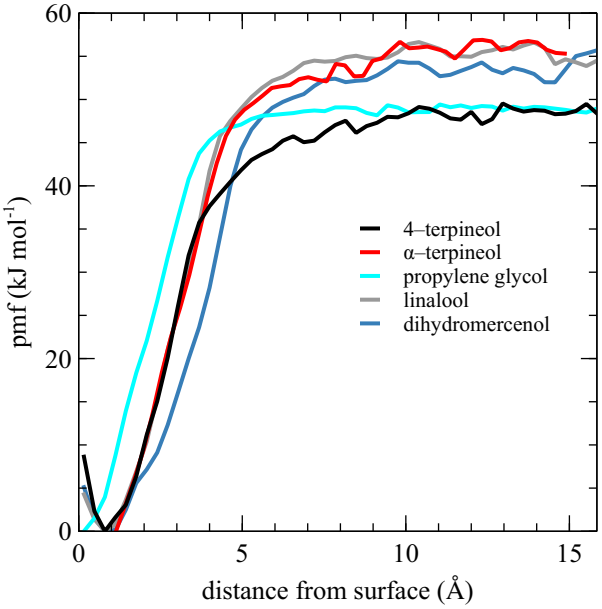


Figure 2.14: Desorption free energy of 4-terpineol, α-terpineol, linalool, propylene glycol, and dihydromercenol. The umbrella sampling calculations were performed using the hydoxylated silica slab with the dimension and force field parameters as described before. The silanol surface density of 4.7 OH/nm^2 was used to prepare the surface. The CGenFF parameters were assigned to the organic compounds

2.7 Adsorption of Hydrophilic Terpenoids on Silica

We also explored the the interaction of hydrophilic terpenes, e.g. terpenoids, that are prevalent in household and cosmetic products. To complete the analysis of prevalent organic compounds with the silica surface, we included two terpineol isomers, i.e. 4-terpineol and α-terpineol, as well as linalool, dihydromercenol, and propylene glycol. These compounds are extensively used in perfumes and deodorants. Therefore, it is important to explore their interaction with the silica surface.

The simulation of surface mass accommodation from 100 scattering trajectories (classical MD), suggest that similar to limonene and other terpenes, all the hydroxylated compounds have propensity to accumulate on the silica surface. The surface mass accommodation coefficient is near one for all of these compounds. The free energy of adsorption profiles are computed using a similar strategy as described for limonene. The force field parameters

for these molecules were adopted from CGenFF force field.[33] The desorption free energy profiles are exhibited in Figure 2.14. As these results suggest, organic compounds with OH groups form stronger interactions with the silica surface compared to ordinary terpenes. The desorption free energies of these compounds are approximately twice those of the terpene molecules (Figure 2.11). In this class of molecules, larger compounds, e.g. linalool and dihydromyrcenol, form stronger interactions with the silica surface. Although propylene glycol is smallest in this group, it actively participates in two HB with silica through its two OH groups.

The HB donor–acceptor distance-angle distribution and power spectra calculations provide more insight into the interaction of these compounds with silica. As Figure 2.15 exhibits, in the case α-terpineol and 4-terpineol, HB formation through donating their hydroxyl H to the oxygen atoms on silica is the most probable form of HB interaction. Moreover, the donor–acceptor distance where these contacts occur (2.5–2.6 Å) are significantly closer than the other possible HB forms, i.e. H acceptor site at O and C=C. The HB formation beween O–H....C=C in 4-terpineol is similar to terpenes presented before. Whereas, α-terpineol does not participates in HB interaction through the C=C group with surface.

The power spectra of the silanol groups near terpineol compounds are presented in Figure 2.16. The peak locations in the vibrational spectra of silanol groups interacting with 4-terpineol and α-terpineol are similar to each other (3300 cm^{-1}). Compared to other terpene compounds (Figure 2.13), these spectra are broader and have considerable intensity at $\omega <$ 3200 cm^{-1}. This is indicative of stronger HB formation due between the OH groups of these compounds and silica compared to the hydrophobic terpenes.

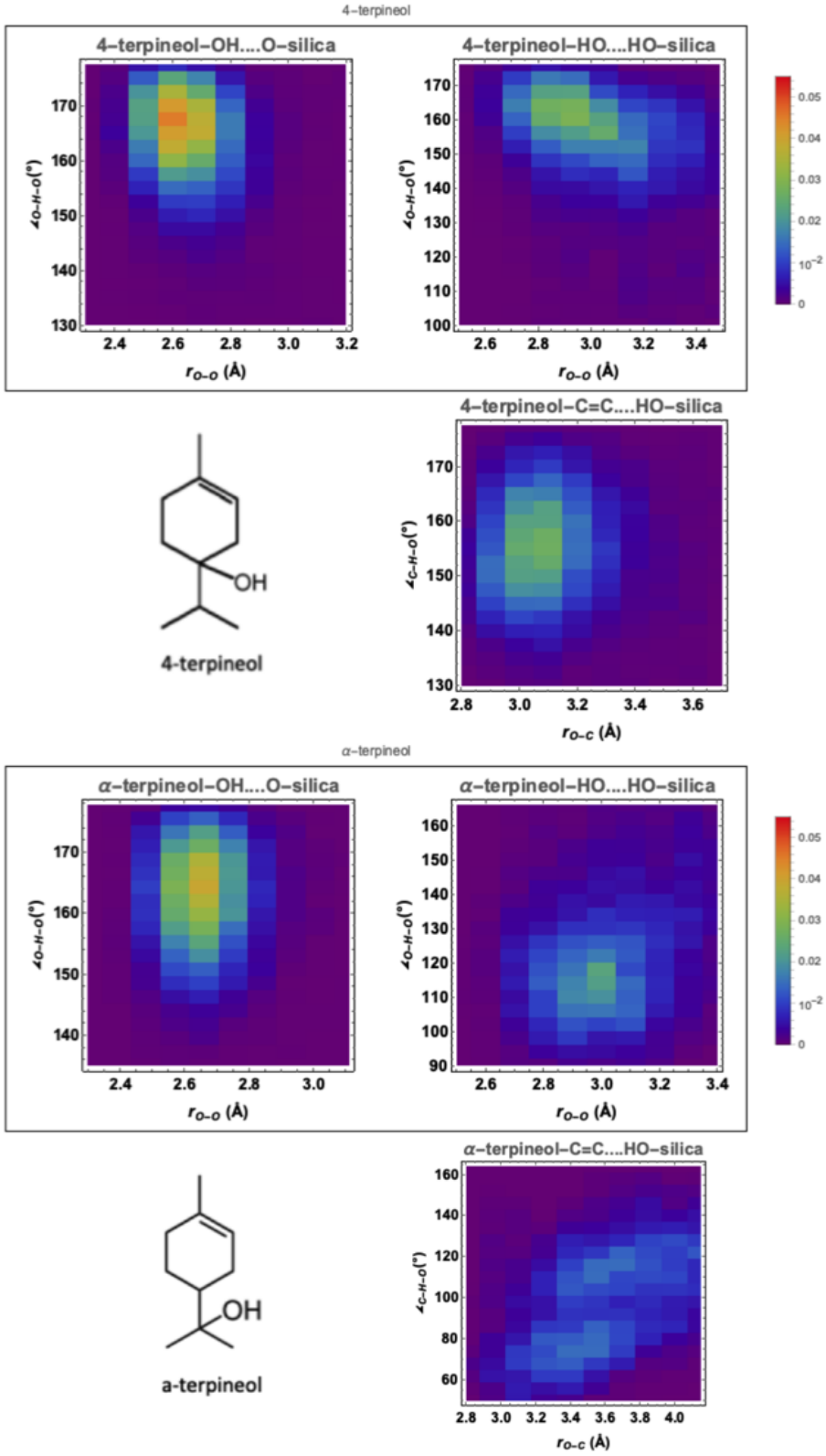


Figure 2.15: HB donor-acceptor distance-angle distribution calculated for α-terpineol and 4-terpineol from 40 ps of AIMD simulation. These compounds have 3 possible HB sites; C=C bond, and HB acceptor from O atom and HB donor from hydroxyl H atom.

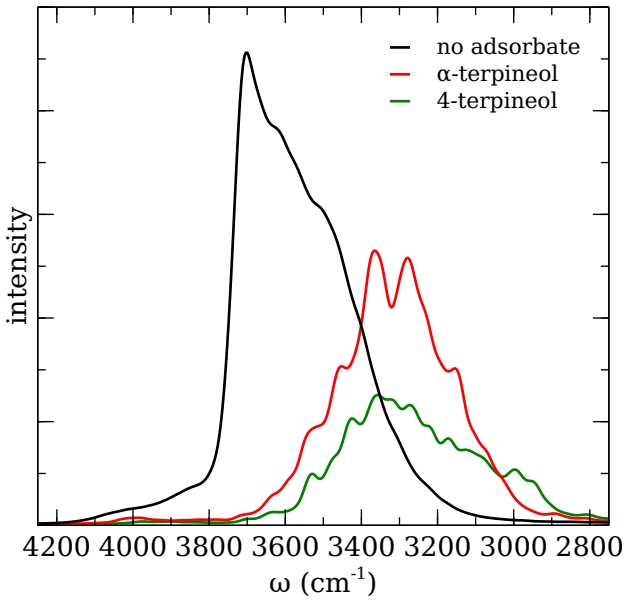


Figure 2.16: Power spectra calculated for the silanol hydrogen atoms of the silica cluster located in the proximity of the terpineol compounds and the OH groups in these compounds during 40 ps of AIMD simulation. The power spectra calculations were performed using the TRAVIS code.[3]

The analysis of interactions formed between possible HB sites on propylene glycol, linalool, and dihydromyrcenol and silanol groups (Figures 2.17 and 2.18) suggests that in these molecules HB formation from C=C and OH sites are equally probable. They involve both HB donors and HB acceptors from the hydroxyl groups. The donor-acceptor distance at which these HB interactions are formed is 2.6 Å similar to terpineol compounds. Linalool and dihydromyrcenol also form significant HB interactions at their C=C sites. Similar to the terpenes, in both of these compoundss the most probable C=C....O distance is located at 2.9–3.0 Å.

The vibrational frequency of silanol OH groups located near these compounds display broader redshifts compared to the hydrophobic compounds (Figure 2.19). This is in agreement with the closer contacts formed between the HB sites of these compounds and silanol groups compared to limonene isomers. The IR spectra for the entire silica cluster and hydrophilic compounds are exhibited in Figure 2.23 and 2.24. IR spectra are obtained using the maxi-

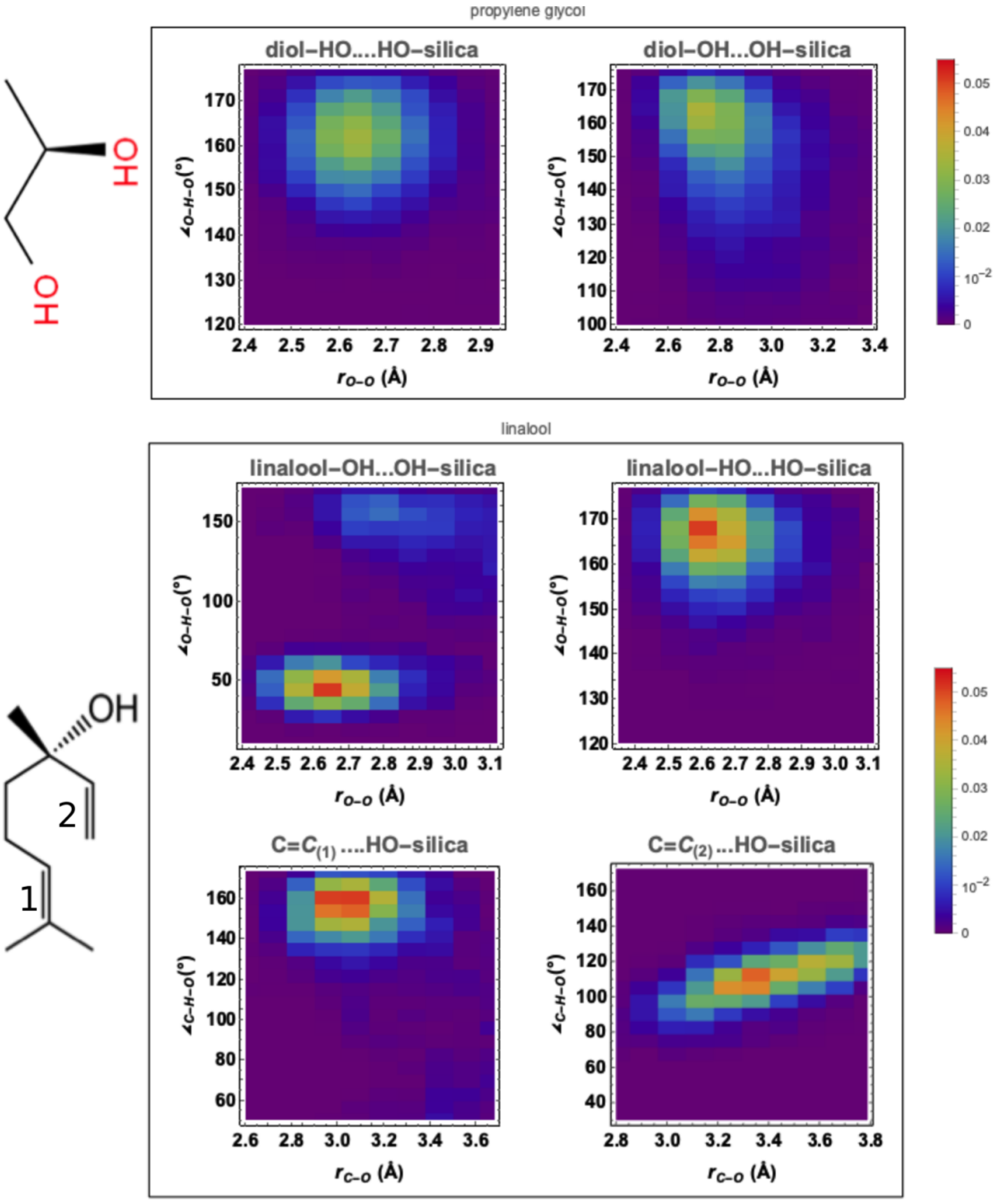


Figure 2.17: Density plots of the distance-angle distribution of the HB donor-acceptor in propylene glycol and linalool obtained from 40 ps of AIMD simulation.

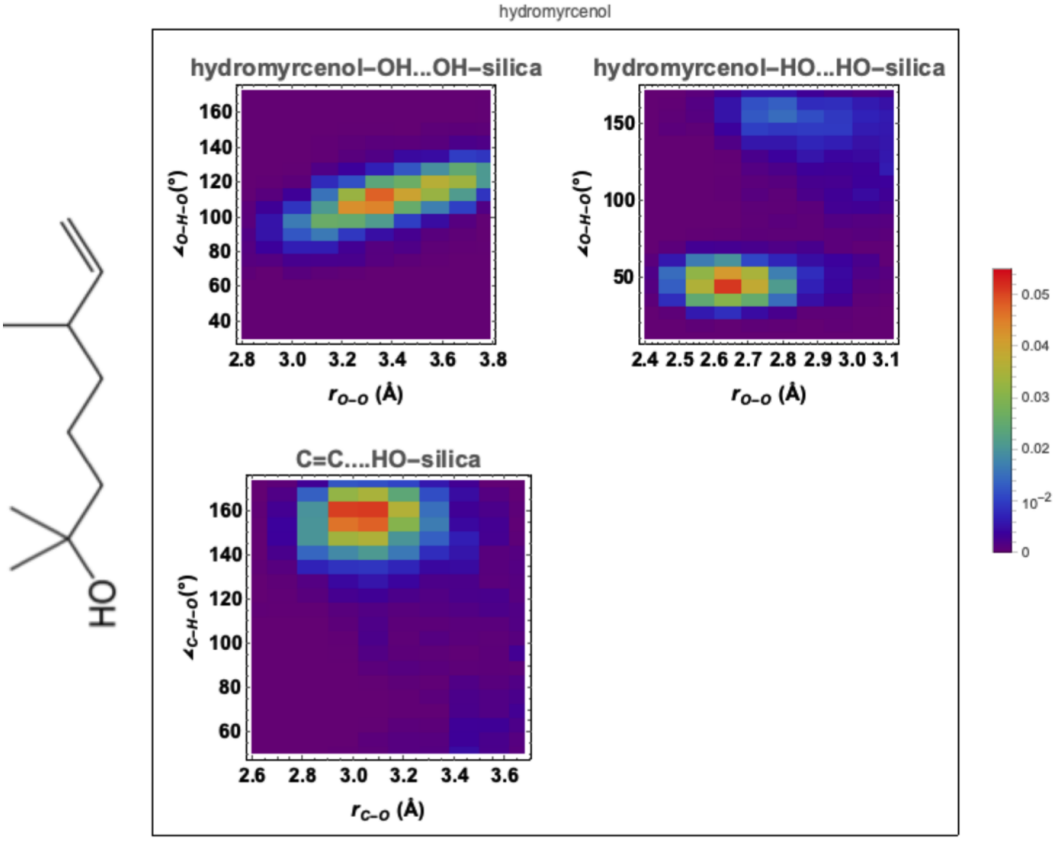


Figure 2.18: Density plots of the distance-angle distribution of the HB donor-acceptors in dihydromyrcenol obtained from 40 ps of AIMD simulation.

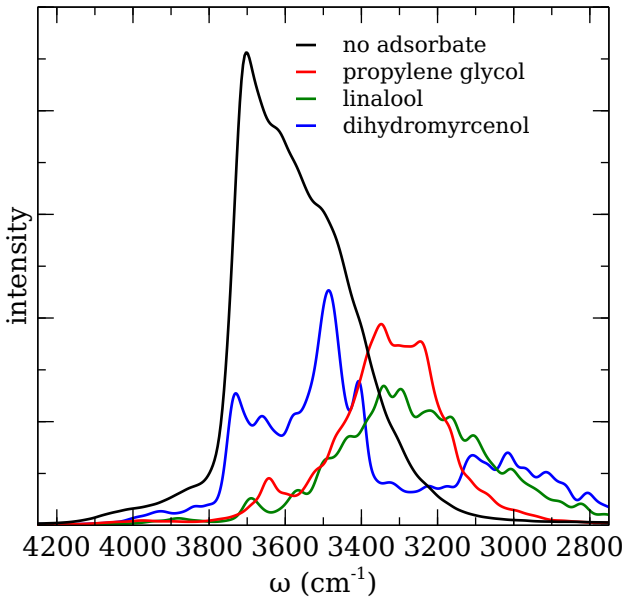


Figure 2.19: The power spectra calculated for the OH groups of silanol near adsorbents and OH groups in the adsorbents calculated from 40 ps of AIMD simulation using TRAVIS program.[3]

mally localized Wannier orbitals. And the TRAVIS code[3] is used to generate these spectra.

2.8 Conclusions

Using combination of classical MD and AIMD simulations we investigated the interaction between the prevalent terpenes and terpenoids with the hydroxylated silica surface. Our results suggest that even hydrophobic terpenes, such as limonene and terpinolene, form strong intermolecular interactions with the silica surface. Based on the distribution of C=C...O distance and C...H–O angle we identify these intermolecular interactions as π-hydrogen bonding interaction. Three limonene isomers involve in similar HB interactions with silica and they significantly alter silanol O–H virational frequency. Hydrophilic terpenoids, such as α-terpineol and linalool, participate in stronger intermolecular interaction with the silica surface compared to limonene isomers and their desorption free energies are approximately twice thos of the limonene isomers. Furtheremore, the hydroxyl groups in these compounds

form closer HB interactions with silanol groups compared to the C=C moieties. The stronger HB interaction of these compounds with silica is evident in the broad redshifted vibrational spectra of the silanol groups in the vicinity of these terpenoids.

2.9 Supplementary Information

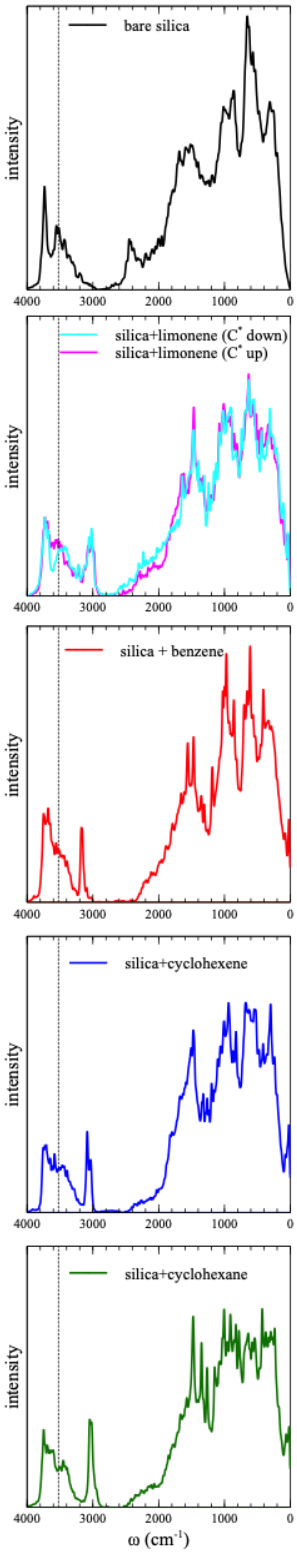


Figure 2.20: Power spectra calculated for the complete range of frequency for all the atoms in each system. Dashed line marks the 3500 cm^{-1} vibration.

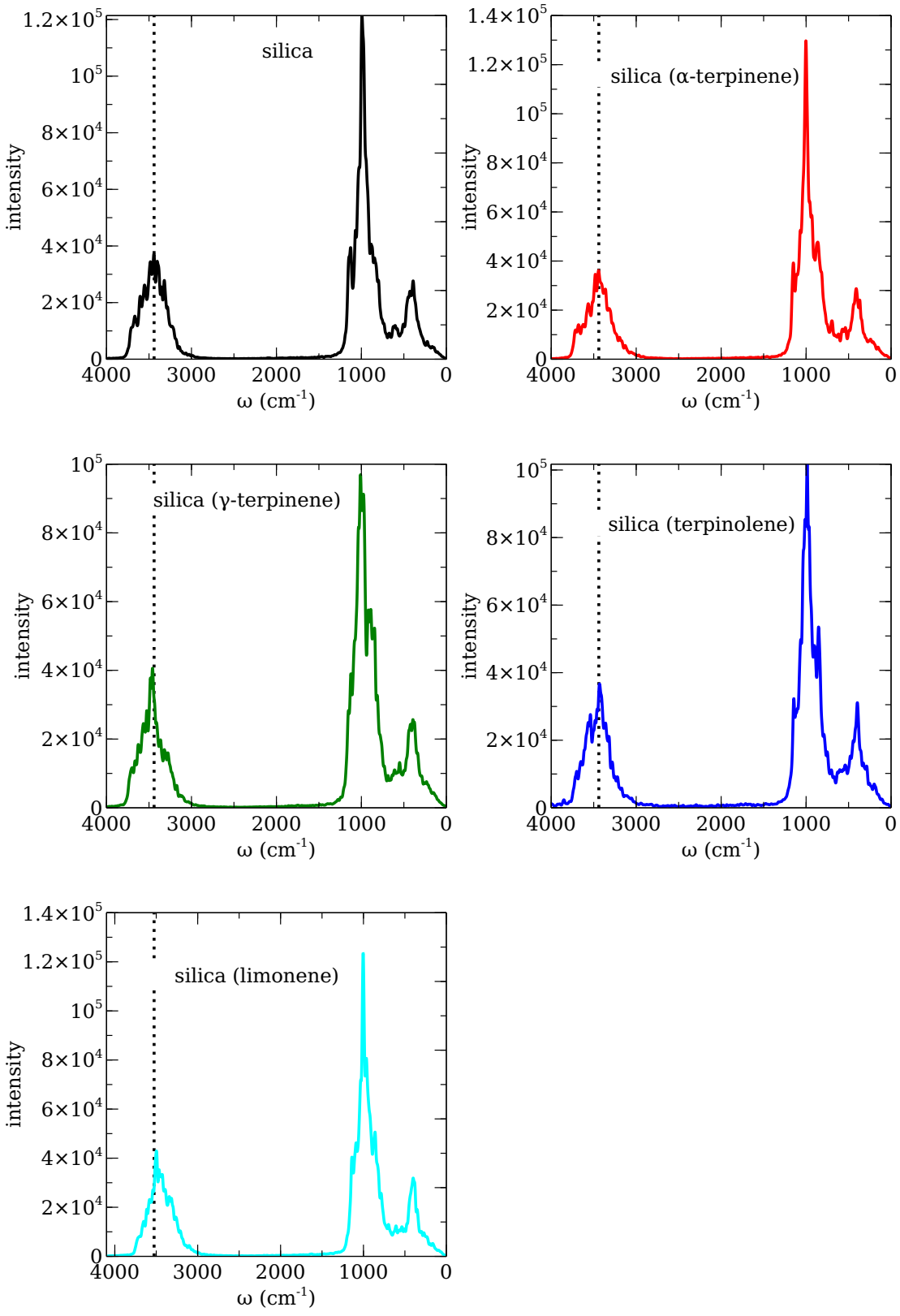


Figure 2.21: IR spectra of the silica cluster exposed to the terpene molecules. The identity of the adsorbate molecule in each panel is provided in parenthesis.

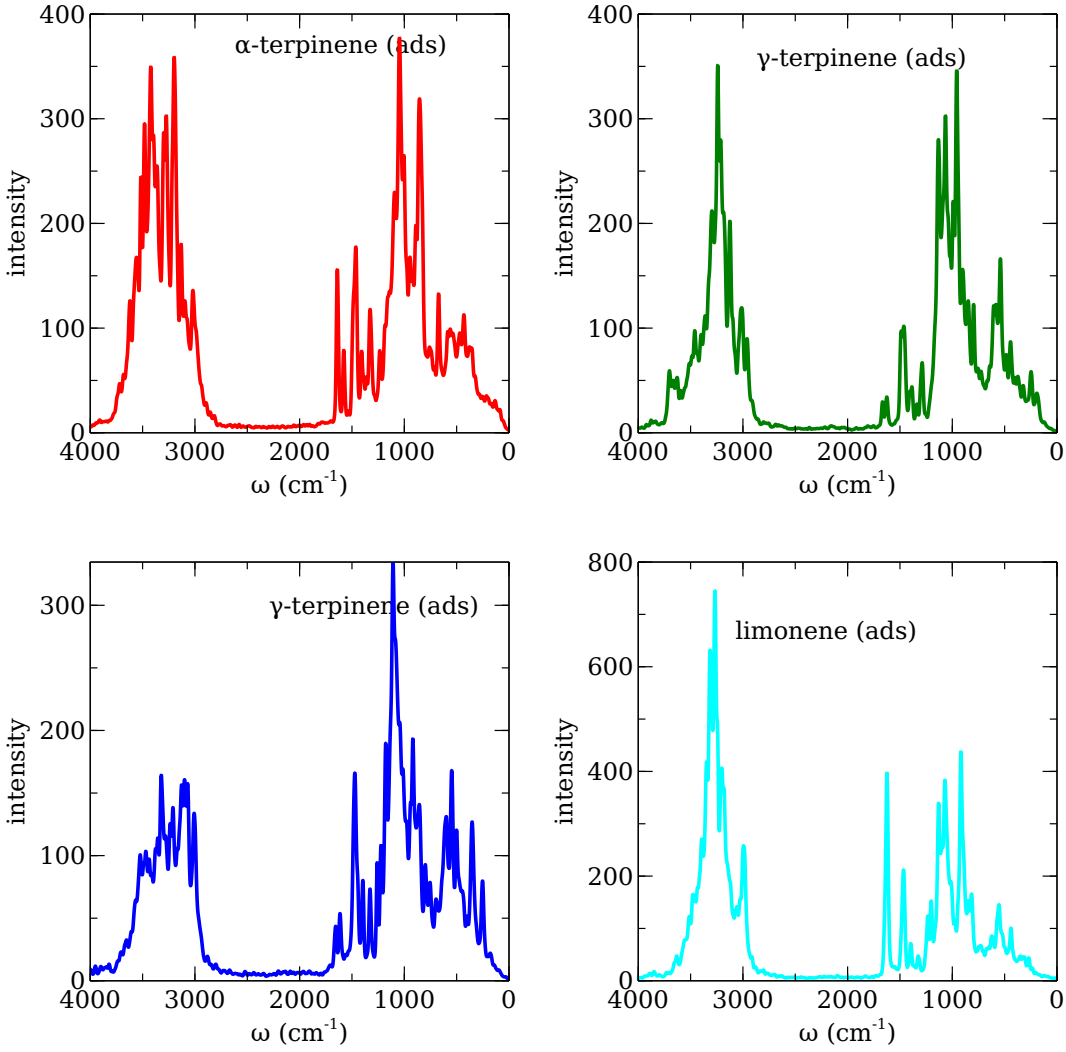


Figure 2.22: IR sectra of terpene molecules adsorbed on the silica cluster.

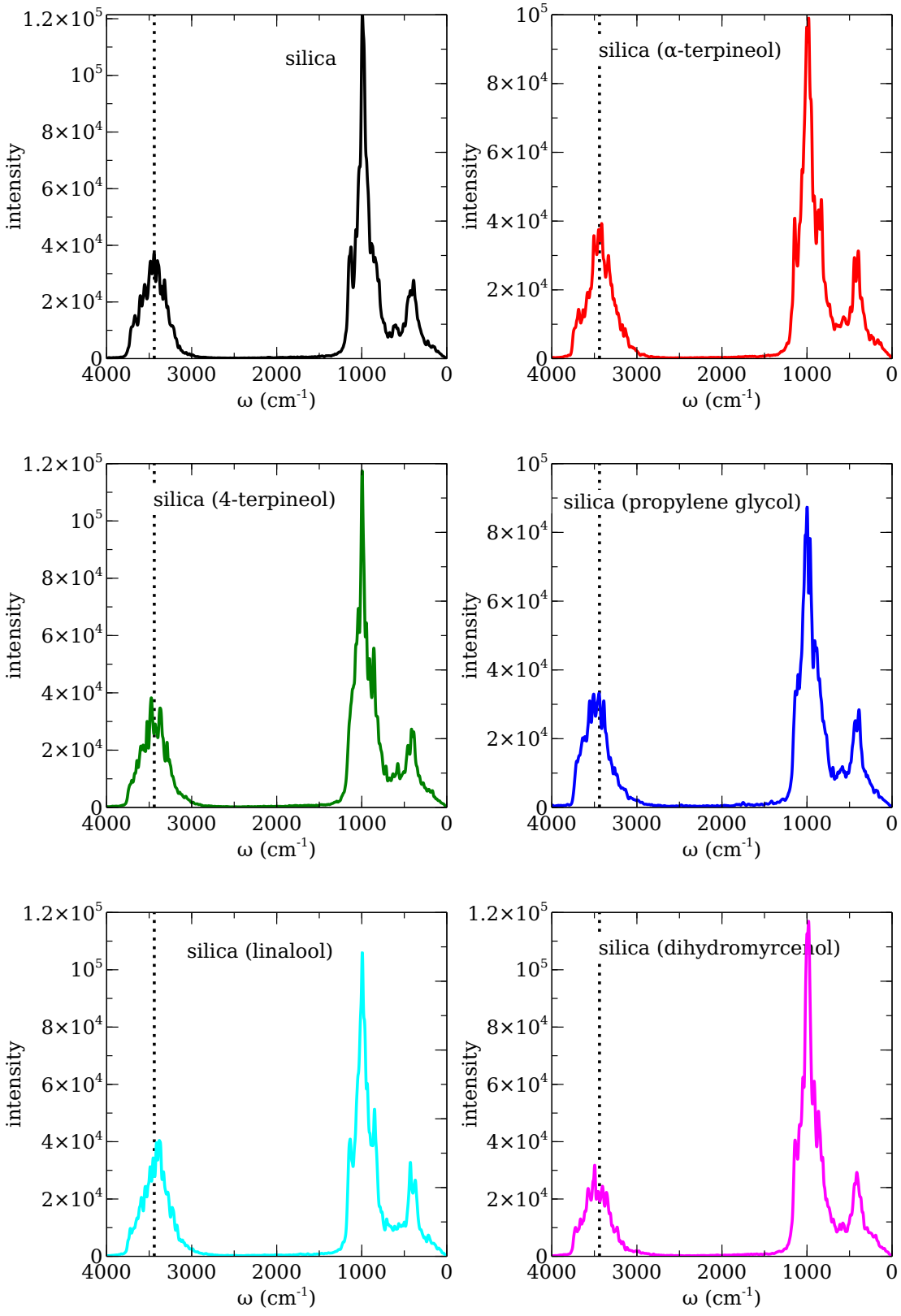


Figure 2.23: IR spectra of the silica cluster interacting with hydrophilic compounds. The identity of the adsorbate molecule in each panel is provided in parenthesis.

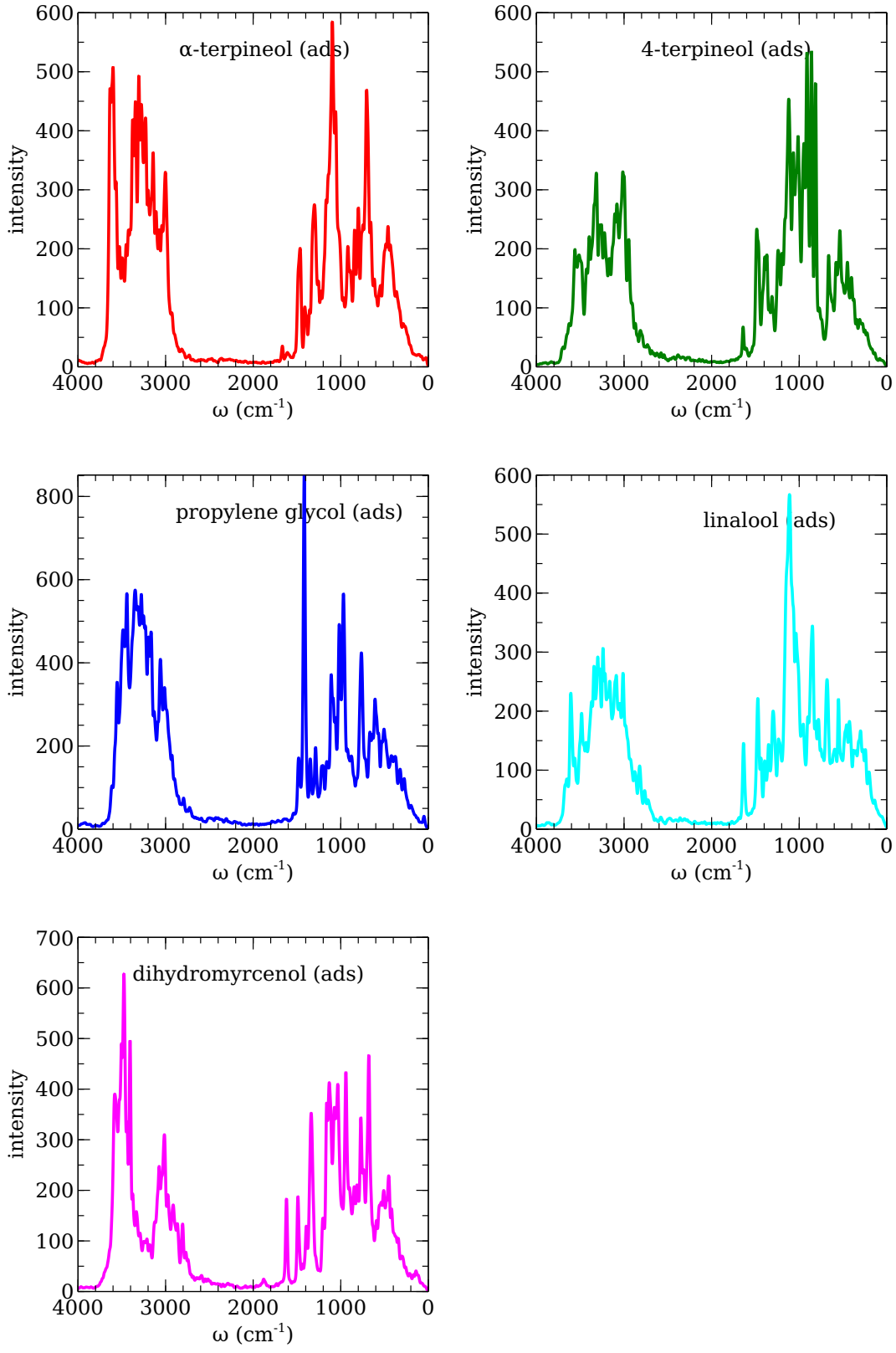


Figure 2.24: IR spectra of hydrophilic compounds adsorbed on the silica cluster.

Chapter 3

QM/MM Silumation of H^+ Permeation in Hv1

3.1 Introduction

H^+ ions are essential to variety of cellular activities. They are either product or reactant of several important biological processes, such as ATP generation and NADPH function.[47, 48] Furthermore, pH regulates processes that are vital to cell life, such as cell growth[49] and neurotransmitters.[50] Nevertheless, H^+ cannot readily pass through the cell membranes. Proton channels and transporters are the active pathways that move H^+ ions in and out of the cell. Human proton channel (Hv1) is an ion channel that translocates H^+ ions out of the cell. Similar to other voltage-dependent ion channels, Hv1 possesses a voltage sensing domain (VSD)[51] that is rich in basic amino acids, namely three arginines R205, R208, and R211. This motif induces a conformational change in the structure of Hv1 in response to the changes in transmembrane voltage which results in the channel opening. Hv1 is composed of four helices (S1–S4) and Hv1 dimers are connected through the S4 and transport

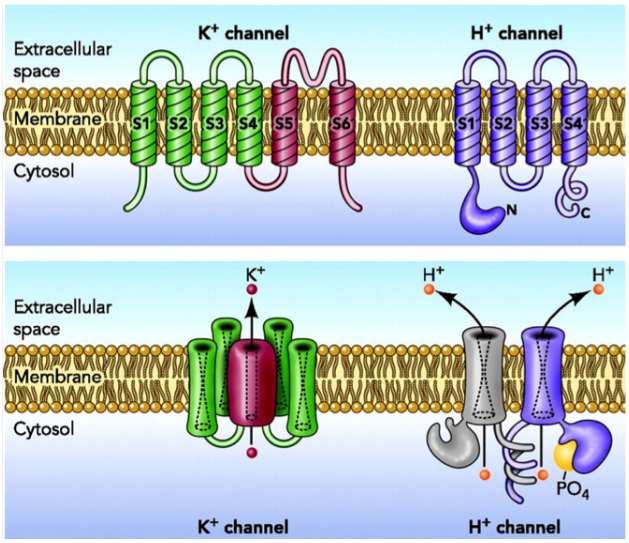


Figure 3.1: Comparison between H^+ and K^+ channels. Reprinted with permission from Ref. [4]

H^+ cooperatively.[52] Unlike most of other voltage-sensitive ion channels, every Hv1 unit contains a pore that transports H^+ ions. Moreover, the super-selectivity of Hv1 to protons is unique.[53] This becomes more impressive considering that cellular concentrations of other cations, such as Na^+, are several orders of magnitude higher than H^+.[54] The extraordinary selectivity of Hv1 to H^+ ions is associated to the D112–R211 selectivity filter.[53] However, the exact mechanism of proton permeation is not clear. The structure for the open state of the Hv1 channel is unknown and there is not a consensus whether the proton permeates through the water wire, i.e. Grotthuss mechanism, or acidic/basic side chains also participate in proton permeation.[55, 56]

In this study a QM/MM simulation method is employed to shed light on the role of titratable side chains located in the pore region in the proton permeation. The atomistic model of Geragotelis et al.[57] for the open state of Hv1 is used as the initial structure for the QM/MM simulations presented in this Chapter.

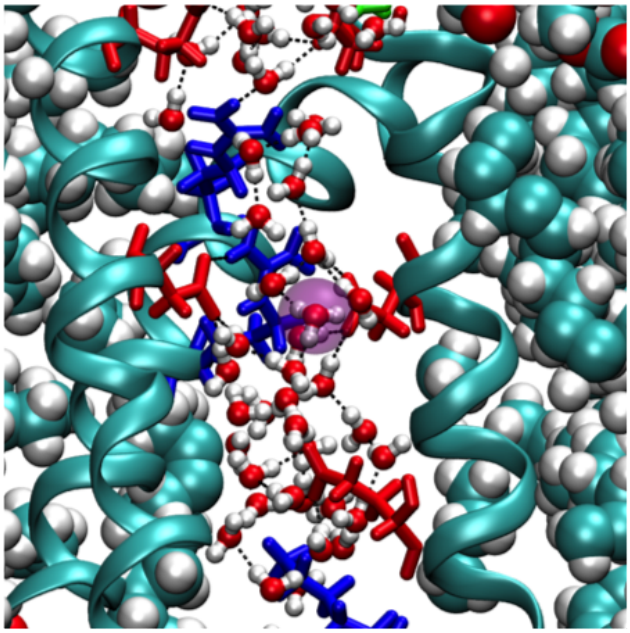
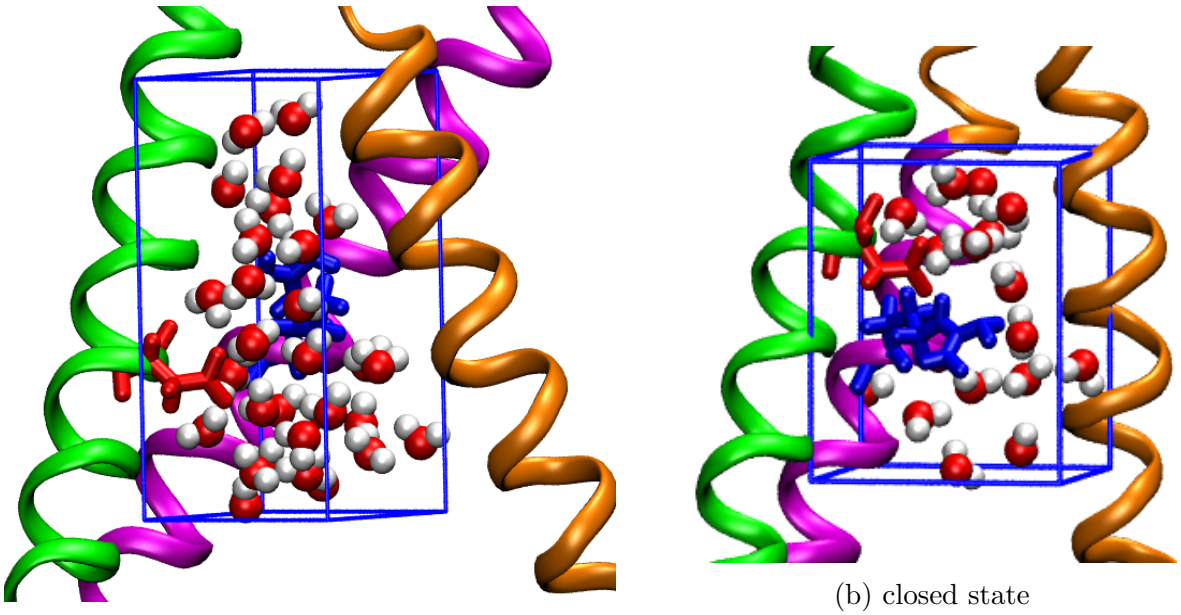


Figure 3.2: A snapshot of the HB network near the SF of the Hv1 channel.

3.2 Deprotonation of D112 in Hv1

Since the role of D112 in the selectivity of Hv1 is demonstrated and it resides in the pore region, it is a plausible assumption that the D112 side chain transitions between the protonated and deprotonated states during the proton translocation. The free energy profile for the deprotonation reaction of D112 side chain can characterize the role of D112 in the proton translocation.

We utilized a QM/MM simulation protocol to obtain the free energy of deprotonation of D112 in the open and closed conformations of the Hv1 channel. The QM region contains the D112 and R211 side chains along with the water molecules in the vicinity of these residues (Figure 3.3). All the QM/MM simulations of Hv1 are carried out using the QUICKSTEP module of the CP2K package.[12] The D3-dispersion[58] corrected BLYP functional, with the TZV2P-MOLOPT-GTH basis set,[37] for the valence electrons, and GTH pseudopotentials, for the core electrons, with the orbital transformation scheme and the SCF convergence of 5.0×10^{-8}, are used. The plane wave cutoff is 300 Ry. The IMOMM scheme[59] is applied



(a) open state

(b) closed state

Figure 3.3: A snapshot of the QM region of the open and closed state of Hv1 protein. D112 and R211/R205 side chains + water molecules in the vicinity of these side chains are included in the QM region.

to link the MM and QM, the C_α–C_β atoms of the side chains, for the residues included in the QM system. Periodic boundary conditions (PBCs) are applied to both QM and MM regions. The electrostatic interactions between the QM and MM parts are included with the Gaussian expansion of electrostatic potential (GEEP) technique with the QM decoupling scheme.[60, 61] The vdW interactions between the QM and MM particles are included with the CHARMM36 Lennard-Jones parameters. The MM region is composed of 174 dipalmitoyl-phosphatidylcholine (POPC) lipids that form the membrane bilayer (88 in the upper leaflet and 86 in the lower leaflet). 10788 water molecules are added to solvate the system. 34 Na^+ and 36 Cl^- ions are evenly added to either side of the membrane bilayer to to form the ionic concentration and neutralize the entire system. The MM region is modeled with the CHARMM36 force fields for proteins [62] and lipids,[63] and the TIP3P model[2] is used for for water. The nuclei are propagated with a 0.5 fs time step. A Nosè-Hoover thermostat[40] with 0.1 ps relaxation time is employed to maintain the temperature of the system near 300 K.

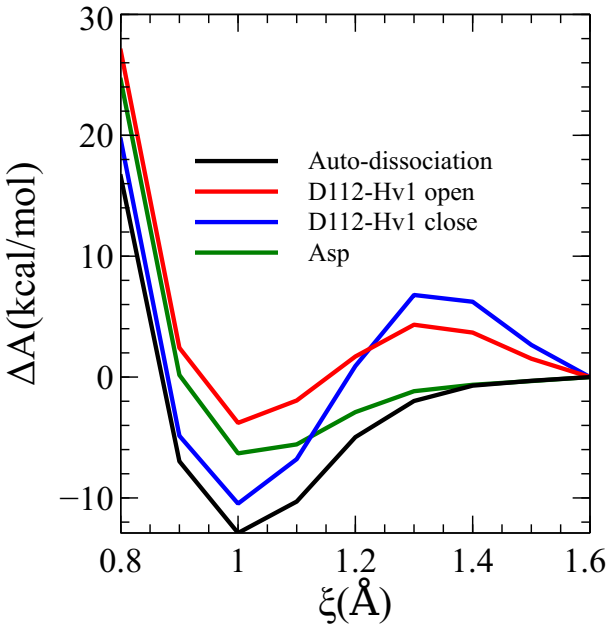


Figure 3.4: Derotonation free energy surfaces calculated using distance as reaction coordinate under the Blue Moon Ensemble scheme.

The distance between the proton and a carboxylate oxygen atom in the D112 side chain was used to describe the reaction. The free energy calculations under the Blue Moon Ensemble were performed for the water auto-dissociation, Asp(aq), and D112 side chain in the two stats of Hv1 channel. The deprotonation reaction is divided into 15 windows spanning 0.8–1.6 Å. As Figure 3.4 and Table 3.1 show, the deprotonation free energy profile of the D112 side chain in the closed state has larger barrier (twice) compared to the open state and Asp(aq). According to these free energy profiles, an aspartic acid in solution has a much smaller barrier to dissociate than the aspartic acide side chain in the Hv1 protein. This can be due to the limited hydration in the Hv1 channel compared to the bulk solution and the presence of a positively charge arginine residue near D112 that resists the formation of hydronium ion. These free energy profiles can be utilized to estimate the relative pK_a of various side chains in water and protein environment.[64, 65]

system	deprotonation barrier (kcal/mol)	deprotonation free energy (kcal/mol)
auto-dissociation (water)	13	13
Asp(aq)	6	6
D112 (open)	8	4
D112 (closed)	16	10

Table 3.1: The deprotonation barrier and free energy calculated based on the distance between a carbolxyl oxygen and hydrogen from QM/MM simulation in the Blue Moon ensemble scheme.

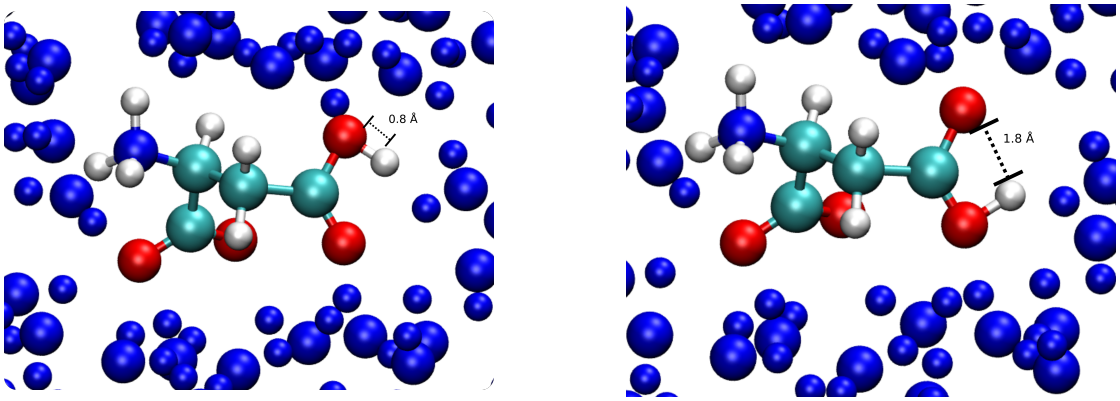
$$\frac{K_d(Asp)}{K_d(H_2O)} = \frac{\int_0^{R_c} exp\left(-\beta w_{H_2O}(r)\right)\ r^2\ dr}{\int_0^{R_c} exp\left(-\beta w_{Asp}(r)\right)\ r^2\ dr}$$

where R_c is the cutoff distance (≈ 1.4 Å) used to integrate the free energy profile. Moreover, this result suggests that the D112 side chain in the closed state of Hv1 has larger pKa than in the open state. Therefore, one can argue that D112 behaves as a proton trap in the closed state.

Although these results seem promising, there is a problem with using a distance as the reaction coordinate. As illustrated in Figure 3.5, the separated proton can bind to the other carboxylate oxygen in the aspartate side chain and the side chain thus remains protonated. Therefore, using a single distance is insufficient to describe the deprotonation reaction involving carboxylate moieties.[64]

To overcome this issue the coordination number (C.N.) is introduced to model the deprotonation reaction. The C.N. reaction coordinate is defined by:

$$\xi_{C.N.}(r) = \frac{1}{N_D} \sum_{D}^{N_D} \sum_{H}^{N_H} \frac{1 - \left(\frac{r_{DH}}{1.6}\right)^6}{1 - \left(\frac{r_{DH}}{1.6}\right)^{12}}$$



(a) ξ=0.8 (Å)

(b) ξ=1.8 Å

Figure 3.5: A representation of the problem with using single distance as reaction coordinate. In both states the aspartic acid side chain remains protonated. The water molecules in the coodination shell of aspartic acid are represented in blue.

where the summation is over all of the hydrogen atoms in the QM region that can possibly form covalent bonds with O atoms, and r_{DH} is the distance between the corresponding O and H atoms. As such, C.N. = 1 and C.N.= 0 correspond to a protonated and deprotonated side chain, respectively.

We employed the extended Lagrangian version of the metadynamics scheme[21] to compute the deprotonation free energy based on the C.N. The free energy profiles after 50 ps of QM/MM simulation are shown in Fig. 3.6. Modeling the deprotonation reaction with the C.N. yields same trends as using the distance to represent the deprotonation reaction, namely, larger free energy barrier for deprotonation of D112 in the closed state compared to the open state of Hv1 and aqueous aspartic acid. However, the free energy values are different from the results reported in Table 3.1.

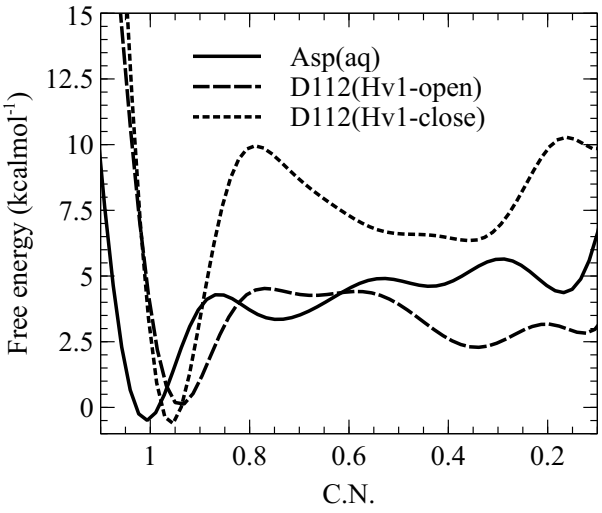


Figure 3.6: Deprotonation free energy calculated based on the coordination number using extended metadynamics technique. The free energies are obtained during 50 ps QM/MM simulation at 300 K.

3.3 Free Energy Profile of Proton Transport through Hv1

Although the deprotonation free energy of D112 is illuminating, it is highly localize to the SF region and it does not indicate which species are proton carriers near that region. To find the proton carriers across the entire channel and the free energy associated with the proton translocation, a reaction coordinate is required that is not biased to a specific species and identifies the proton carrier as the proton moves across the entire channel. The Center of excess charge (CEC) collective variable achieves both of these objectives.[66] The CEC is defined as:

$$\xi(z) = \sum_{i}^{N_H} z_i - \sum_{j}^{N_X} w_j z_j - \sum_{ij}^{N_H N_X} f(r_{ij})(z_i - z_j) + w_{pair}(m_k - m_l)(z_l - z_k),$$

$$f(r_{ij}) = \frac{1}{1 - exp[(r_{ji} - r_{sw})/d_{sw}]};\ m_j = \frac{\sum_i^{N_H} f(r_{ij})^{16}}{\sum_i^{N_H} f(r_{ij})^{15}}$$

Where N_X is the number of O and N atoms that can accept/donate proton and N_H is the total number of hydrogen atoms in the QM region participating in the proton transport, such as H atoms in water molecules and acidic/basic side–chains, and w_j is the number of hydrogen atoms that are in the chemical bond with the O and N atoms included in the sum. The r_{sw} and d_{sw} parameters are chosen such that $f(r_{ij}) = 1$ for H atoms in the covalent bond with X_j and $f(r_{ij}) = 0$ otherwise. It has been shown that $r_{sw} = 1.25$ Å and $d_{sw} = 0.04$ Å satisfy the aforementioned condition. The last term in this equation is used for the cases where several donor atoms share a proton and this term locates the shared proton in the middle of these species. The key advantage of using CEC as reaction coordinate is that it eliminates any presumption about side chain or water molecule that the H^+ ion tends to bind to. Voth et al. used the CEC to investigate proton permeation in the influenza A M2 proton channel.[67]

3.3.1 Proton Trasport between a Pair of Acetic Acid Molecules

To verify the proper implementation of the CEC collective variable in CP2K we performed the simulation of proton transport between two acetic acid molecules that are separated by 10 Å and solvated in a water sphere. Furthermore, this simulation might be helpful to explain the free energy profiles obtained for proton transport in the Hv1 channel.

Figure 3.8 shows the free energy profile associated with the transport of a proton between two acetic acid molecules that are separated by 10 Å. The acetic acid pairs are in a water sphere, as illustrated by Figure 3.7, and the simulations are performed using the tight-bonding DFT

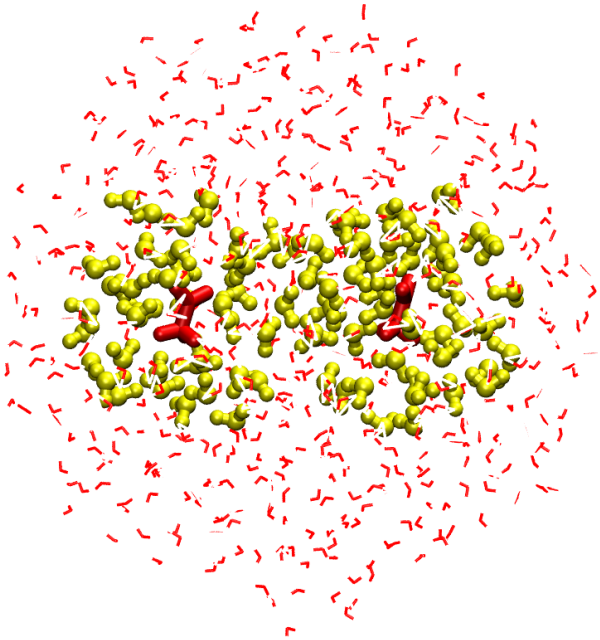


Figure 3.7: A snapshot of the system containing a pair of acetic acid molecules separated by 10 Å and surrounded by 500 water molecules. The QM water molecules are represented in yellow, whereas, the water molecules in the MM region are shown by the red lines.

method (DFTB).[68, 69] The interactions beyond 18 Å is included implicitly through the Generalized Solvent Boundary Potential (GSBP).[70] Acetic acid molecules along with a cylinder of water molecules, with 7 Å radius, are included in the DFTB calculations. The free energy profile is calculated using the umbrella sampling scheme where the location of the CEC collective variable is harmonically restrained with the spring constant of 40 kcal/mol/Å^2. The proton transfer reaction was divided into 40 windows separated by 0.25 Å increments in ξ. Each PMF widows was sampled for 20 ps. The free energy profile is shown in the Figure 3.8.. In this system, acetic acid molecules are located at 0 and 15 Å. The free energy was obtained using the WHAM scheme.[16] As it can be seen in Figure 3.8, the deprotonation of each acetic acid molecule is followed by an increase in the free energy, while the H^+ translocation through the water wire, $7< \xi <11$ Å, occurs readily.

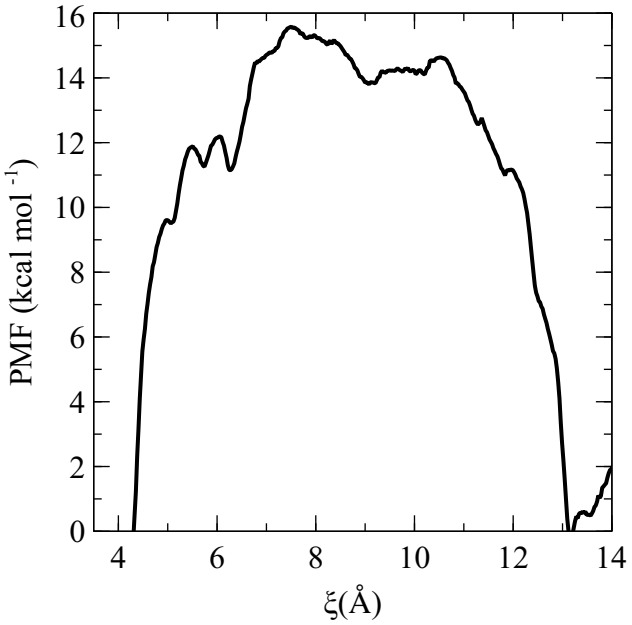


Figure 3.8: The free energy profile for transporting a proton between two acetic acid molecules located at 10 Å from each other. The center-of-excess charge is used to model the proton transport reaction. The PMF is obtained using US method with 0.25 Å spacing between the adjacent windows.

3.3.2 Proton Transport Across the Hv1 Channel

We utilized an umbrella sampling protocol and QM/MM scheme to estimate free energy of H^+ transport in the pore region of Hv1 protein using the CEC as reaction coordinate. The QM region comprises between 200 and 300 atoms, including all acidic, basic, and polar neutral amino acid side chains facing the Hv1 putative permeation pathway (i.e. the interior of the VSD), and all their solvating waters (see Figure 3.9). Bulk water regions, each spanning five Å along the transmembrane direction, on both the intracellular and extracellular side are also included as a suitable global reference for the PMF. The total span of the QM region is 55 Å along the transmembrane direction. To make the QM/MM simulations tractable, for each conformational state of the VSD, the QM region is partitioned into five overlapping subregions as depicted in Figure 3.9. An independent set of US calculations along the collective variable is performed on each one of these subregions. This partitioning scheme allows joining the partial PMF profiles, obtained from each subsystem while maintaining

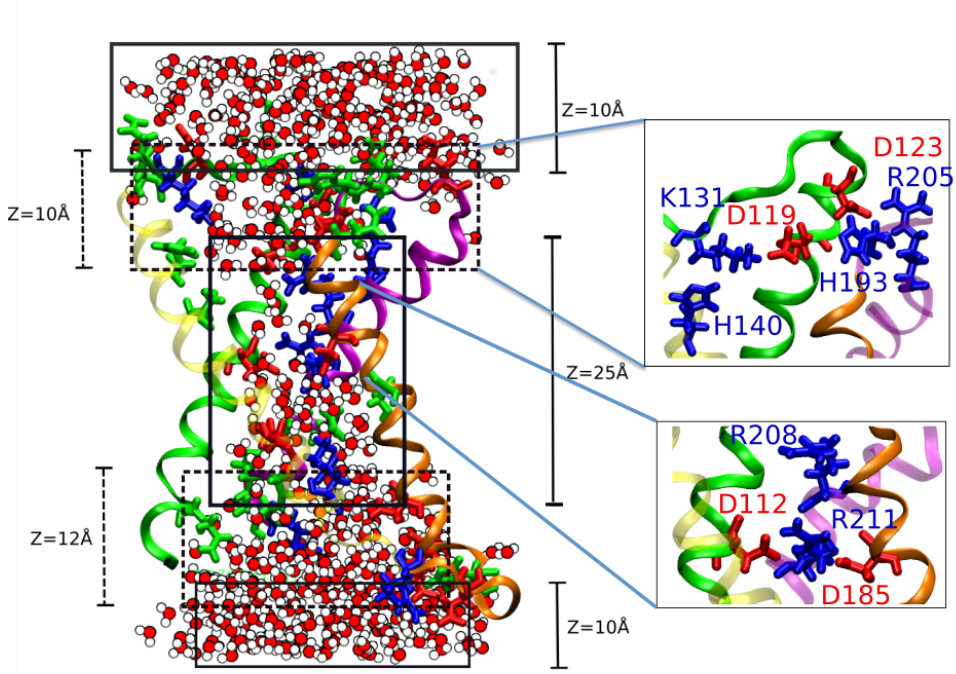


Figure 3.9: Specification of the QM region in the channel open state. All the atoms included in the QM region are shown as licorice (amino acid residue side chains) and filled spheres (waters). Amino acid residue side chains are colored by type (basic, blue; acidic, red; polar noncharged, green). Water molecules are colored by atom (O, red; H, white). Each box (either solid or dashed lines) represents one QM subregion that will be used in a set of US simulations. These regions are devised to overlap by about 2-3 Å, which will provide a way to connect the end points of the neighboring regions. The corresponding extent of each region along the transmembrane direction is explicitly indicated. The insets identify amino acid residue side chains located on the channel extracellular surface (top inset) and the interior of the channel (bottom inset) that may play a role in proton conduction

a reasonable size for the QM region. These regions are devised to overlap by about 2–3 Å, which will provide a way to connect the end points of subregions, with reference to a proton entering/leaving the channel from/to the bulk aqueous medium. All the charged amino acid side chains facing the interior of the channel are included in the QM region. In each US window, a hydronium ion is introduced to the system at the corresponding position along the transmembrane direction. The coordinates of the excess proton are monitored by the CEC collective variable. The CEC is restrained with a harmonic force constant of 40 kcal/mol/Å^2. The CEC reaction coordinate was sampled at 0.25 Å spacing between consecutive US windows. Each US window was run for at least 15 ps of QM/MM simulation. All the simulations are performed using CP2K package with a similar setup for the DFT calculations and QM–MM electrostatic coupling as described in Section 3.2.

The free energy profiles of proton transport across the entire Hv1 channel for both states of the channel, i.e. open and closed, are presented in Figure 3.10. The main outcome of these free energy profiles is that ionizable side chains other than D112 play important role in the proton transport. For instance, in both the open and closed states the E119 and D185 side chains become protonated as the proton nears them. Subsequently, moving the proton away from these side chains is associated with a free energy penalty for deprotonating the carboxylate groups of these side chains. Furthermore, the free energy profile is flat for the regions that lack a polar side chain. This result suggests that the proton is transported through a combination of Grotthuss mechanism and binding to the acidic side chains. The recent work by Delemotte et al. suggests three binding sites for the proton in Hv1, namely D174/E153, D112/D185, and E119/D123.[71] Our free energy calculations also show the importance of the E153, D185, and E119 side chains in the proton permeation. As pointed out by Warshel and co-workers, the rate limiting process in the H^+ transportation in Hv1 is not the transport though the water wire.[72] In contrast to their finding, that electrostatic interactions between the positively charge side chains and H^+ ion is the dominant factor in proton transport, we find that formation of chemical bond between the H^+ and acidic side

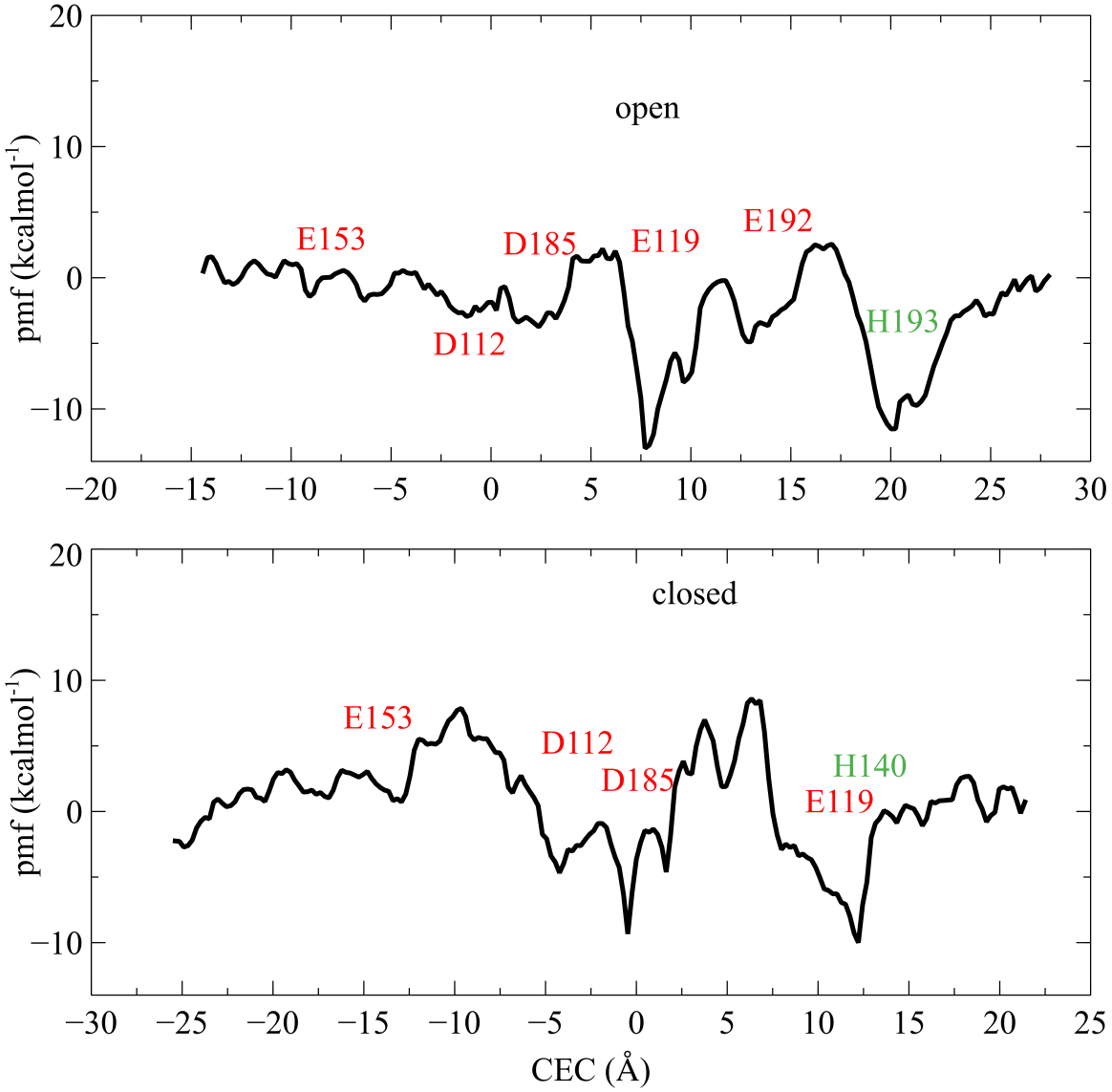


Figure 3.10: The free energy profile for translocation of a proton across the Hv1 channel. The free energy calculations were performed using a QM/MM scheme. The center-of-excess charge was employed as the reaction coordinate. The free energy profile covers intercellular, VSD, and exteracellular regions of the Hv1 protein in both the open and closed states. The labels on each graph represent an approximate location for the titratable side chains.

chains plays the most influential role in Hv1 function.

Despite the interesting outcome of the free energy profile obtained from CEC, we should note that quantitative deduction from these results is not valid. A possible source of uncertainty in these calculations is ambiguity in the ionization state of the titratble side chains. In the starting structures that are employed in this study all the acidic side chains in Hv1, i.e. aspartate and glutamate amino acids, are in the deprotonated state. This might not be a reasonable choice as the results of free energy calculations indicate some of these side chains might be in the neutral form. Therefore, we suggest to use the results of mutagenesis studies to assign the ionization state of the acidic amino acids. For instance, it has been shown that mutations in the D185 and E153 residues do not affect proton cunduction.[73] Therefore, using the neutral form of these side chains might be a better starting point. One other source of error in the QM/MM simulations presented here is the over-polarization of the plane waves in the QM region due to the dipole moments of the TIP3P water molecules in the MM region.[74] Although the smearing of the MM charges implemented in the GEEP module of CP2K is designed to avoid the electron spill-out,[75, 76] this scheme seems to be inadequate in the regions where the QM region is completely surrounded by the TIP3P water molecules. We observed that the QM region at the intercellular and exteracellular parts accumulate near the QM walls after a few picoseconds. This problem could not be avoided even with increasing the Lennard-Jones radii of the QM–MM interaction to prevent MM water molecules approaching the QM region. To avoid this problem we recommend the use of a multilayer QM/MM approach, such as embedding methods,[77] where a buffer region is added between the QM and MM boundaries and the particles are allowed to exchange between the QM and MM regions.

3.4 Conclusions

We applied a QM/MM simulation method to study the proton permeation mechanism in the Hv1 protein. Specifically, we investigated the role of polar side chains residing in the pore region of the channel. Our results indicate that the D112 side chain has more tendency to remain in the protonated form in the closed state of the channel compared to the open state. Furthermore, using the center of excess charge collective variable we explored proton translocation across the entire channel. The free energy profile obtained with the CEC suggests that the acidic side chains located in the pore region of Hv1 have a tendency to bind a proton when the hydronium ion approaches them. Consequently, the regions where these side chains are located correspond to wells in the free energy profile. Finally, we provide suggestions on how to improve the free energy calculations using the Adaptive Buffered Force QM/MM scheme to alleviate the anomalies caused by the superfluous interaction between the QM and MM particles near the boundaries of these two regions.

Bibliography

[1] Yuan Fang and Vicki H. Riahi Saleh McDonald Andrew T. Shrestha Mona, Tobias Douglas J. Grassian. What is the driving force behind the adsorption of hydrophobic molecules on hydrophilic surfaces? *The Journal of Physical Chemistry Letters*, 10(3):468–473, Feb 2019.

[2] William L. Jorgensen, Jayaraman Chandrasekhar, Jeffry D. Madura, Roger W. Impey, and Michael L. Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2):926–935, 1983.

[3] Martin Brehm and Barbara Kirchner. TRAVIS - a free analyzer and visualizer for Monte Carlo and molecular dynamics trajectories. *Journal of Chemical Information and Modeling*, 51(8):2007–2023, 2011. PMID: 21761915.

[4] Thomas E. DeCoursey. The voltage-gated proton channel: A riddle, wrapped in a mystery, inside an enigma. *Biochemistry*, 54(21):3250–3268, 2015. PMID: 25964989.

[5] Yuan Fang, Pascale SJ Lakey, Saleh Riahi, Andrew T McDonald, Mona Shrestha, Douglas J Tobias, Manabu Shiraiwa, and Vicki H Grassian. A molecular picture of surface interactions of organic compounds on prevalent indoor surfaces: limonene adsorption on sio 2. *Chemical Science*, 10(10):2906–2914, 2019.

[6] Terrell L. Hill. *Statistical Mechanics: Principles and Selected Applications Book by Terrell L. Hill*. Dover Publications, 1987.

[7] M.E. Tuckerman. *Statistical Mechanics: Theory and Molecular Simulation*. Oxford graduate texts. Oxford University Press, 2011.

[8] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys. Rev.*, 136:B864–B871, Nov 1964.

[9] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140:A1133–A1138, Nov 1965.

[10] R.M. Dreizler and E.K.U. Gross. *Density Functional Theory: An Approach to the Quantum Many-Body Problem*. Springer Berlin Heidelberg, 2012.

[11] Gerald Lippert, Jürg Hutter, and Michele Parrinello. The gaussian and augmented-plane-wave density functional method for ab initio molecular dynamics simulations. *Theoretical Chemistry Accounts*, 103(2):124–140, Dec 1999.

[12] Joost VandeVondele, Matthias Krack, Fawzi Mohamed, Michele Parrinello, Thomas Chassaing, and Jürg Hutter. Quickstep: Fast and accurate density functional calculations using a mixed gaussian and plane waves approach. *Computer Physics Communications*, 167(2):103 – 128, 2005.

[13] A. Warshel and M. Levitt. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *Journal of Molecular Biology*, 103(2):227 – 249, 1976.

[14] Hans Martin Senn and Walter Thiel. QM/MM methods for biomolecular systems. *Angewandte Chemie International Edition*, 48(7):1198–1229, 2009.

[15] G.M. Torrie and J.P. Valleau. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics*, 23(2):187 – 199, 1977.

[16] Benoit Roux. The calculation of the potential of mean force using computer simulations. *Computer Physics Communications*, 91(1):275 – 282, 1995.

[17] Shankar Kumar, John M Rosenberg, Djamal Bouzida, Robert H Swendsen, and Peter A Kollman. The weighted histogram analysis method for free-energy calculations on biomolecules. i. the method. *Journal of computational chemistry*, 13(8):1011–1021, 1992.

[18] E.A. Carter, Giovanni Ciccotti, James T. Hynes, and Raymond Kapral. Constrained reaction coordinate dynamics for the simulation of rare events. *Chemical Physics Letters*, 156(5):472 – 477, 1989.

[19] Michiel Sprik and Giovanni Ciccotti. Free energy from constrained molecular dynamics. *The Journal of Chemical Physics*, 109(18):7737–7744, 1998.

[20] Alessandro Laio and Michele Parrinello. Escaping free-energy minima. *Proceedings of the National Academy of Sciences of the USA*, 99(20):12562–12566, 2002.

[21] Marcella Iannuzzi, Alessandro Laio, and Michele Parrinello. Efficient exploration of reactive potential energy surfaces using car-parrinello molecular dynamics. *Phys. Rev. Lett.*, 90:238302, Jun 2003.

[22] Wei Gao, Jing Wu, Yawei Wang, and Guibin Jiang. Distribution and congener profiles of short-chain chlorinated paraffins in indoor/outdoor glass window surface films and their film-air partitioning in beijing, china. *Chemosphere*, 144:1327 – 1333, 2016.

[23] William W Nazaroff. Indoor air chemistry: Cleaning agents, ozone and toxic air contaminants (final report). *California Air Resources Board*, 144, 2006.

[24] Commission of the European Communities. *Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy.* Office for Official Publications of the European Communities, 2000.

[25] Charles J. Weschler and Helen C. Shields. Indoor ozone/terpene reactions as a source of indoor particles. *Atmospheric Environment*, 33(15):2301 – 2312, 1999.

[26] Tsung-Hung Li, Barbara J Turpin, Helen C Shields, and Charles J Weschler. Indoor hydrogen peroxide derived from ozone/d-limonene reactions. *Environmental Science & Technology*, 36(15):3295–3302, 2002.

[27] Jerald L. Schnoor. Environmental science & technology presents the 2014 excellence in review awards. *Environmental Science & Technology*, 48(22):13023–13024, Nov 2014.

[28] Ramina Alwarda, Shouming Zhou, and Jonathan PD Abbatt. Heterogeneous oxidation of indoor surfaces by gas-phase hydroxyl radicals. *Indoor Air*, 28(5):655–664, 2018.

[29] Charles J. Weschler and William W. Nazaroff. Semivolatile organic compounds in indoor environments. *Atmospheric Environment*, 42(40):9018 – 9040, 2008.

[30] Daniel J. Cole, Mike C. Payne, Gábor Csányi, S. Mark Spearing, and Lucio Colombi Ciacchi. Development of a classical force field for the oxidized Si surface: Application to hydrophilic wafer bonding. *The Journal of Chemical Physics*, 127(20):204704, 2007.

[31] Steve Plimpton. Fast parallel algorithms for short-range molecular dynamics. *Journal of Computational Physics*, 117(1):1 – 19, 1995.

[32] Rolf E. Isele-Holder, Wayne Mitchell, and Ahmed E. Ismail. Development and application of a particle-particle particle-mesh ewald method for dispersion interactions. *The Journal of Chemical Physics*, 137(17):174107, 2012.

[33] K. Vanommeslaeghe, E. Prabhu Raman, and A. D. MacKerell. Automation of the CHARMM general force field (CGenFF) ii: Assignment of bonded parameters and partial atomic charges. *Journal of Chemical Information and Modeling*, 52(12):3155–3168, Dec 2012.

[34] Yan Zhao and Donald G. Truhlar. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theoretical Chemistry Accounts*, 120(1):215–241, May 2008.

[35] S.F. Boys and F. Bernardi. The calculation of small molecular interactions by the differences of separate total energies. some procedures with reduced errors. *Molecular Physics*, 19(4):553–566, 1970.

[36] M. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, and H. Nakatsuji. Gaussian16 revision b.01, 2016. 2018.

[37] Joost VandeVondele and Juerg Hutter. Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases. *The Journal of Chemical Physics*, 127(11):114105, 2007.

[38] S Goedecker, M Teter, and Jürg Hutter. Separable dual-space gaussian pseudopotentials. *Physical Review B*, 54(3):1703, 1996.

[39] Joost VandeVondele and Jürg Hutter. An efficient orbital transformation method for electronic structure calculations. *The Journal of Chemical Physics*, 118(10):4365–4369, 2003.

[40] Denis J Evans and Brad Lee Holian. The Nose–Hoover thermostat. *The Journal of Chemical Physics*, 83(8):4069–4074, 1985.

[41] Martin Thomas, Martin Brehm, Reinhold Fligg, Peter Vöhringer, and Barbara Kirchner. Computing vibrational spectra from ab initio molecular dynamics. *Physical Chemistry Chemical Physics*, 15(18):6608–6622, 2013.

[42] Katharina Wendler, Martin Brehm, Friedrich Malberg, Barbara Kirchner, and Luigi Delle Site. Short time dynamics of ionic liquids in aimd-based power spectra. *Journal of Chemical Theory and Computation*, 8(5):1570–1579, 2012.

[43] Nicola Marzari, Arash A. Mostofi, Jonathan R. Yates, Ivo Souza, and David Vanderbilt. Maximally localized wannier functions: Theory and applications. *Rev. Mod. Phys.*, 84:1419–1475, Oct 2012.

[44] Frederick R. Jensen and C. Hackett Bushweller. Conformational preferences and interconversion barriers in cyclohexene and derivatives. *Journal of the American Chemical Society*, 91(21):5774–5782, Oct 1969.

[45] Solveig Gaarn Olesen and Steen Hammerum. Hydrogen bonding to alkanes: Computational evidence. *The Journal of Physical Chemistry A*, 113(27):7940–7944, 2009. PMID: 19522536.

[46] Alexander N. Isaev. Intermolecular charge transfer as evidence for unusual O-H...C_{sp^3} hydrogen bond. *Computational and Theoretical Chemistry*, 1090:180 – 192, 2016.

[47] K Shiozuka, K Tani, S Mizushima, and H Tokuda. The proton motive force lowers the level of ATP required for the in vitro translocation of a secretory protein in escherichia coli. *Journal of Biological Chemistry*, 265(31):18843–7, 1990.

[48] L M Henderson, J B Chappell, and O T G Jones. The superoxide-generating NADPH oxidase of human neutrophils is electrogenic and associated with an H^+ channel. *Biochemical Journal*, 246(2):325–329, 1987.

[49] Reinhard Dechant, Shady Saad, Alfredo J. Ibáñez, and Matthias Peter. Cytosolic pH regulates cell growth through distinct GTPases, Arf1 and Gtr1, to promote Ras/PKA and TORC1 activity. *Molecular Cell*, 55(3):409 – 421, 2014.

[50] Z.-L. Chen and R.-Q. Huang. Extracellular pH modulates GABAergic neurotransmission in rat hypothalamus. *Neuroscience*, 271:64 – 76, 2014.

[51] Liang Hong, Medha M. Pathak, Iris H. Kim, Dennis Ta, and Francesco Tombola. Voltage-sensing domain of voltage-gated proton channel Hv1 shares mechanism of block with pore domains. *Neuron*, 77(2):274 – 287, 2013.

[52] Francesco Tombola, Maximilian H. Ulbrich, Susy C. Kohout, and Ehud Y. Isacoff. The opening of the two pores of the Hv1 voltage-gated proton channel is tuned by cooperativity. *Nature Structural &Amp; Molecular Biology*, 17:44, Dec 2009.

[53] Todor Dudev, Boris Musset, Deri Morgan, Vladimir V. Cherny, Susan M. E. Smith, Karine Mazmanian, Thomas E. DeCoursey, and Carmay Lim. Selectivity mechanism of the voltage-gated proton channel, HV1. *Scientific Reports*, 5:10320 EP, May 2015.

[54] Thomas E. DeCoursey. The voltage-gated proton channel: A riddle, wrapped in a mystery, inside an enigma. *Biochemistry*, 54(21):3250–3268, 2015. PMID: 25964989.

[55] Bennett Ashley L. and Ramsey Ian Scott. Crosstalk opposing view: proton transfer in Hv1 utilizes a water wire, and does not require transient protonation of a conserved aspartate in the S1 transmembrane helix. *The Journal of Physiology*, 595(22):6797–6799, 2017.

[56] Mona L. Wood, Eric V. Schow, J. Alfredo Freites, Stephen H. White, Francesco Tombola, and Douglas J. Tobias. Water wires in atomistic models of the Hv1 proton channel. *Biochimica et Biophysica Acta BBA - Biomembranes*, 1818(2):286 – 293, 2012.

[57] Geragotelis Andrew, Wood Mona, L. Goeddeke Hendrik, Riahi Saleh, Hollingsworth Scott A., Freites J. Alfredo, Tombola Francesco, and Tobias Douglas J. Multi-microsecond molecular dynamics simulations of the HV1 proton channel. *Biophysical Journal*, 110(3):282a–283a, Feb 2016.

[58] Stephan Ehrlich, Jonas Moellmann, Werner Reckien, Thomas Bredow, and Stefan Grimme. System-dependent dispersion coefficients for the DFT-D3 treatment of adsorption processes on ionic surfaces. *ChemPhysChem*, 12(17):3414–3420, 2011.

[59] Feliu Maseras and Keiji Morokuma. IMOMM: A new integrated ab initio + molecular mechanics geometry optimization scheme of equilibrium structures and transition states. *Journal of Computational Chemistry*, 16(9):1170–1179, 1995.

[60] Laino Teodoro, Mohamed Fawzi, Laio Alessandro, and Parrinello Michele. An efficient real space multigrid QM/MM electrostatic coupling. *Journal of Chemical Theory and Computation*, 1(6):1176–1184, 2005. PMID: 26631661.

[61] P. E. Blöchl. Electrostatic decoupling of periodic images of plane-wave-expanded densities and derived atomic point charges. *The Journal of Chemical Physics*, 103(17):7422–7428, 1995.

[62] Jing Huang and Alexander D. MacKerell Jr. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *Journal of Computational Chemistry*, 34(25):2135–2145, 2013.

[63] R. W. Pastor and A. D. MacKerell. Development of the CHARMM force field for lipids. *The Journal of Physical Chemistry Letters*, 2(13):1526–1532, 2011. PMID: 21760975.

[64] Ivaylo Ivanov, Bin Chen, Simone Raugei, and Michael L. Klein. Relative pKa values from first-principles molecular dynamics: The case of histidine deprotonation. *The Journal of Physical Chemistry B*, 110(12):6365–6371, 2006. PMID: 16553455.

[65] Arindam Bankura, Michael L. Klein, and Vincenzo Carnevale. Proton affinity of the histidine-tryptophan cluster motif from the influenza a virus from ab initio molecular dynamics. *Chemical Physics*, 422:156 – 164, 2013. A tribute to Dr. Robin Hochstasser.

[66] P. H. König, N. Ghosh, M. Hoffmann, M. Elstner, E. Tajkhorshid, Th. Frauenheim, and Q. Cui. Toward theoretical analyis of long-range proton transfer kinetics in biomolecular pumps. *The Journal of Physical Chemistry A*, 110(2):548–563, 2006. PMID: 16405327.

[67] Ruibin Liang, Hui Li, Jessica M. J. Swanson, and Gregory A. Voth. Multiscale simulation reveals a multifaceted mechanism of proton permeation through the influenza A M2 proton channel. *Proceedings of the National Academy of Sciences*, 111(26):9396–9401, 2014.

[68] M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, Th. Frauenheim, S. Suhai, and G. Seifert. Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys. Rev. B*, 58:7260–7268, Sep 1998.

[69] Qiang Cui, Marcus Elstner, Efthimios Kaxiras, Thomas Frauenheim, and Martin Karplus. A QM/MM implementation of the self-consistent charge density functional tight binding SCC-DFTB method. *The Journal of Physical Chemistry B*, 105(2):569–585, 2001.

[70] Wonpil Im, Simon Bernèche, and Benoit Roux. Generalized solvent boundary potential for computer simulations. *The Journal of Chemical Physics*, 114(7):2924–2937, 2001.

[71] Siri C. van Keulen, Gianti, Eleonora, Vincenzo Carnevale, Michael L. Klein, Rothlisberger, Ursula, and Delemotte Lucie. Does proton conduction in the voltage-gated H^+ channel hHv1 involve Grotthuss-like hopping via acidic residues? *The Journal of Physical Chemistry B*, 121(15):3340–3351, 2017. PMID: 27801578.

[72] Myungjin Lee, Chen Bai, Mikolaj Feliks, Raphael Alhadeff, and Arieh Warshel. On the control of the proton current in the voltage-gated proton channel Hv1. *Proceedings of the National Academy of Sciences*, 115(41):10321–10326, 2018.

[73] Thomas E. DeCoursey, Deri Morgan, Boris Musset, and Vladimir V. Cherny. Insights into the structure and function of HV1 from a meta-analysis of mutation studies. *The Journal of General Physiology*, 148(2):97–118, 2016.

[74] Nadia Rega, Srinivasan S. Iyengar, Gregory A. Voth, H. Bernhard Schlegel, Thom Vreven, and Michael J. Frisch. Hybrid ab-initio/empirical molecular dynamics: Combining the ONIOM scheme with the atom-centered density matrix propagation ADMP approach. *The Journal of Physical Chemistry B*, 108(13):4210–4220, 2004.

[75] Teodoro Laino, Fawzi Mohamed, Alessandro Laio, and Michele Parrinello. An efficient linear-scaling electrostatic coupling for treating periodic boundary conditions in QM/MM simulations. *Journal of Chemical Theory and Computation*, 2(5):1370–1378, 2006. PMID: 26626844.

[76] Teodoro Laino, Fawzi Mohamed, Alessandro Laio, and Michele Parrinello. An efficient real space multigrid QM/MM electrostatic coupling. *Journal of Chemical Theory and Computation*, 1(6):1176–1184, 2005. PMID: 26631661.

[77] Letif Mones, Andrew Jones, Andreas W. Götz, Teodoro Laino, Ross C. Walker, Ben Leimkuhler, Gábor Csányi, and Noam Bernstein. The adaptive buffered force QM/MM method in the CP2K and AMBER software packages. *Journal of Computational Chemistry*, 36(9):633–648, 2015.