Probing Hydrogen Bonds in Aqueous and Biological Systems: Solvation Structure, Dynamics and Spectroscopy

Garrett Dimitrios Santis

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2025

Reading Committee:

Sotiris S. Xantheas, Chair

Anne B. McCoy

Stefan Stoll

Program Authorized to Offer Degree:
Chemistry

University of Washington

Abstract

Probing Hydrogen Bonds in Aqueous and Biological Systems:
Solvation Structure, Dynamics and Spectroscopy

Garrett Dimitrios Santis

Chair of the Supervisory Committee:
Sotiris S. Xantheas
Department of Chemistry

Hydrogen bonding interactions are arguably one of the most important interactions in chemical and biological systems, determining the structure and properties of aqueous solutions, assisting the recognition of ligands by proteins, and facilitating the coupled transfer of protons and electrons. This dissertation focuses on the physical chemistry aspects governing the properties of hydrogen bonds in several systems at the molecular level.

The strength and dynamics of hydrogen bonded networks have important ramifications in protein structure and ligand docking. However, they are difficult to quantify in complex systems. There is an obvious *energetic–structural–spectral* correspondence for hydrogen bonds, relating their strength to the underlying structure of the participating covalent bonds and the ensuing infrared vibrational signatures. A physical model was developed that includes components describing the covalent DH bond of the hydrogen bond donor via a Morse potential, the Pauli repulsion, and electrostatic interactions between the constituent fragments. The model was fit to the ab initio potential energy surfaces of six archetypal hydrogen bonded dimers, namely $NH_3{-}NH_3$, $H_2O{-}H_2O$, $HF{-}HF$, $H_2O{-}NH_3$, $HF{-}H_2O$, and $HF{-}NH_3$. By applying this model, a simple linear relationship was derived for weak hydrogen bonds, viz. 8.0 kcal/mol per pm of the DH bond elongation (D refers to the hydrogen bond donor) or 4.5 kcal/mol per 100 cm^{-1} of the DH vibrational shift. In stronger hydrogen bonds, non-linear effects arise from a gradual onset of repulsion. The model accurately predicts the hydrogen bond energies in dimers and gas phase peptides from experimentally

measured red-shifts in vibrational frequencies. This relationship provided a quantitative metric for previously undetermined hydrogen bond strengths and explained the fundamental causes for red-shifts in hydrogen bond frequencies.

Hydrogen bonding interactions play a crucial role in describing the hydration and protonation sites of nicotine (NIC). Gas phase protonated NIC (NIC-H^+) can exist as two isomers that are protonated in two different sites, viz. the pyrrolidine protomer (Pyrro-H^+), which is bio-active through interactions with nicotinic acetylcholine receptors (nAChRs), and the bio-inactive pyridine protomer (Pyri-H^+). A joint experimental-theoretical approach determined that gas-phase NIC-H^+ is a mixture of the two protomers in a 6:4 (Pyri-H^+:Pyrro-H^+) ratio, whereas other less addictive NIC-like molecules (nicotinoids) largely biased the Pyri-H^+ protomer. The hydration of NIC-H^+ favored the Pyrro-H^+ protomer due to the formation of hydrogen bonded bridges between the water molecules from the one protonation site to the other. These "water bridges" provide the framework for proton transfer between the two acidic sites via the Grotthuss mechanism. Ab initio calculations of the transition states and the intrinsic reaction paths connecting the various minima explain the appearance/disappearance of peaks in the infrared (IR) spectra associated with NIC's two protonation sites upon increasing the number of water molecules solvating NIC-H^+. The calculations revealed that the proton shuffles between the two protonation sites in the presence of 4 or more water molecules via the Grotthuss mechanism at the experimental temperatures (T=130 K). This mechanism is preferred over the bi-molecular vehicle mechanism, which may become possible at room temperature (T=300 K). The solvation of the intermediate hydronium-like structures proved to be critical in lowering the reaction barriers associated with the Grotthuss mechanism and thus facilitating the proton transfer. Electronic structure calculations and a many-body decomposition of the interaction of NIC-H^+ with model binding pockets of the nAChR in the human brain suggested that the strength of interaction with these model pockets strongly correlates with the addictive character of NIC-like molecules. The difference in the binding strength were correlated to the gas phase energy difference between Pyri-H^+ and Pyrro-H^+, providing a simple molecular descriptor for the addictive character of NIC-like molecules.

The network of hydrogen bonds has a significant impact on the structure and energetics of aque-

ous systems. We have established a set of descriptors to quantify the aggregation of water in the gas and the condensed phases at the molecular level. Structural descriptors such as the hydrogen bonded O-O distance (r_{OO}), O-H-O angle (θ_{OHO}), O-O-O angle (θ_{OOO}), and the modified tetrahedral order parameter (q_m, m= 2, 3, 4, 5) describe important structural changes upon aggregation. Descriptors such as the adjacency environment, hydrogen bond saturation (% HB), and number of non-short-circuited cycles describe the cooperative nature of hydrogen bonded networks. Using a previously developed database of 162,892 water cluster minima for $(H_2O)_n$, n=*3–25*, we found that r_{OO}, θ_{OHO}, and q_m correlated strongly with cluster stability. Changes in the adjacencies and cycle count provided insight into changes in the hydrogen bond network upon aggregation. This analysis has been expanded to the networks of liquid water obtained from molecular dynamic simulations at seven temperatures ranging from near the melting point (T=280 K) to near the boiling point (T=360 K). We find that changes to the adjacency environment, stemming from changes in the number of hydrogen bonds, perturb the water network and result in contracting the second solvation shell as temperature increases from 280 K to 330 K.

The above findings all together serve to describe the structural, kinetic, and spectral properties of hydrogen bonded systems in the gas and the condensed phases.

TABLE OF CONTENTS

Page

LIST OF FIGURES

GLOSSARY

NIC: nicotine

PYRRO-H^+: pyrrolidine protomer

NACHR: nicotinic acetylcholine receptor

PYRI-H^+: pyridine protomer

IR: infrared

D: hydrogen bond donor

A: hydrogen bond acceptor

ES: electrostatic

PL: polarizability

DISP: dispersion

EX: exchange repulsion

CT: charge transfer

EDA: energy decomposition analysis

MO: molecular orbital

NEDA: natural energy decomposition analysis

NBO: natural bond orbital

DEF: deformation

HF: Hartree-Fock

ALMO: absolutely localized molecular orbitals

TTM: Thole-type model

LJ: Lennard-Jones

PIP: permutationally invariant polynomials

LP: lone pair

PMF: potential of mean force

NOR: nornicotine

MP2: Møller-Plesset second-order perturbation theory

DFT: density functional theory

LDA: local density approximation

GGA: generalized gradient approximation

MGGA: meta-generalized gradient approximation

VPT2: second-order vibrational perturbation theory

PES: potential energy surface

HB: hydrogen bond

MAE: mean absolute error

MSE: mean signed error

MAXE: maximum absolute error

IMHB: intramolecular hydrogen bond

RBM: rotational bond method

QTAIM: quantum theory of atoms in molecules

QUAO: quasi-atomic orbitals

RMSD: root mean squared deviation

AE: 2-aminoethanol

BMAE: 1-amino-2-methyl-2-propanol

BFMAE: 1-amino-2,2-bis-trifluromethyl-2-ethanol

PHE: phenylalanine

ALA: alanine

LYS: lysine

AC: acetyl

DNA: deoxyribose nucleic acid

RNA: ribose nucleic acid

NTC: node total communicability

SOAP: smooth overlap of atomic positions

MC-TBP: Monte Carlo temperature basin paving

ASPL: average shortest path length

ML: machine learning

RMSE: root mean squared error

PCA: principal component analysis

PC: principal component

BE: binding energy

MD: molecular dynamics

RDF: radial distribution function

FDA: Food and Drug Administration

ACHBP: acetylcholine binding proteins

PYRI-H^+: pyridine protomer

PYRRO: pyrrolidine

ICSS: intracranial self-stimulation

ANB: anabasine

ANT: anatabine

MYO: myosmine

COT: cotinine

PDB: protein data bank

TYR: tyrosine

PCM: polarizable continuum model

3AA: 3 amino acid model

6AA: 6 amino acid model

14AA: 14 amino acid model

TRP: tryptophan

CYS: cysteine

LEU: leucine

PRO: proline

MBE: many body expansion

1B: one-body

2B: two-body

3B: three-body

OPLS-AA: optimized potentials for liquid simulations, all atoms

BSSE: basis set superposition error

CP: counterpoise

ES: electrospray

Q-MS: quadropole mass filter

QIT: quadropole ion trap

IRPD: infrared photodissociation

TOF-MS: time-of-flight mass spectrometer

CBS: complete basis set

CAS-IR: collision assisted stripping infrared

IRC: intrinsic reaction coordinate

TST: transition state theory

ACKNOWLEDGMENTS

The author wishes to acknowledge the Fujii and Ishiuchi research groups at the Institute of Science Tokyo for their long-standing collaboration on the nicotine project. Thank you for welcoming me fully in this collaboration and teaching me how to plan and interpret gas-phase spectroscopic experiments.

Thank you to Anne McCoy and her research group for being strong mentors during my tenure at UW. You provided a research home in Seattle, especially during the COVID-19 pandemic, that has elevated my research and science communication.

Thank you all past and present members of the Xantheas research group for making a supportive research environment these past five years. A special thanks to Kristina Herman for mentorship and guidance on our many adventures, Max Hoffman for challenging me and helping me explain my science fully, and Madisen Hallsten for expanding my early work with a targeted perspective of biological interplay.

Finally, a special thank you to Sotiris Xantheas. You have provided tremendous mentorship and plentiful opportunities these past five years that have elevated my ability to tackle scientific problems.

DEDICATION

to Jack. Thank you for keeping me grounded through the end of this degree.

Chapter 1

INTRODUCTION: DESCRIPTION AND APPLICATIONS OF HYDROGEN BONDING

The hydrogen bond is arguably one of the most important non-covalent interactions in chemical and biological systems. [1–3] Hydrogen bonds exist between an electron-poor hydrogen (traditionally covalently bonded to oxygen, nitrogen, or fluorine atoms) and a Lewis base (traditionally the lone pair on an oxygen, nitrogen, or fluorine atom). [3] These interactions are prevalent in many important chemical systems, guiding the bulk behavior of water, [4, 5] determining the secondary structure of proteins, [6–11] pairing nucleotides in DNA, [12–15] forming self-healing polymers, [16–18] contributing to the stability and catalytic properties in metal-organic frameworks, [19, 20] and diversifying the interactions in ionic liquids. [21, 22] The important structural properties and functionality in the aforementioned chemical systems can be attributed to the properties of their constituent hydrogen bonds. For that reason, understanding the nature of the hydrogen bonds in these systems is important for understanding their macroscopic physical properties and for developing the next-generation materials that exploit the properties of hydrogen bonds.

In this thesis, we will explore the physical chemistry aspects of hydrogen bonds. The interactions in hydrogen bonds will be dissected to attribute the correlations between hydrogen bond properties to physical concepts (chapters 3-5). Subsequently, we will explore the cooperative properties of hydrogen bond networks as they relate to water (chapters 6-7). Finally, the effects of hydrogen bonds on molecular structure will be investigated in a series of alkaloids, where experimentation can determine the protonation state of each drug and theory explains the thermodynamic and kinetic roots for changing structure (chapters 8-10). First, we will describe the hydrogen bond, from both an energetic and a structural perspective. Subsequently, we will explore the quintessential examples of hydrogen bonding, providing broader context to the specific applications addressed throughout this thesis.

1.1 Physical Interactions in Hydrogen Bonds

The hydrogen bond is a unique interaction with properties that lie between classical and quantum descriptions. To fully understand the hydrogen bond, one must consider how each atom contributes to the overall structure and interactions of the hydrogen bonded system. Hydrogen bonds consist of a hydrogen bond donor (D) and hydrogen bond acceptor (A). Historically, the hydrogen bond donor was defined as the hydrogen atom bound to either oxygen, nitrogen, or fluorine, and the hydrogen bond acceptors were restricted to the lone pairs on oxygen, nitrogen, and fluorine. [3] A more updated description of the hydrogen bond expands this definition to the interactions between any electron poor hydrogen atom bound to an electronegative atom and a strong Lewis base.

The hydrogen bond donor and acceptor are molecular functional groups, consisting of atoms and their subatomic particles. Protons and neutrons sit in the nucleus. Nuclei are heavy, fairly localized, and highly positive in charge. Electrons surround the nuclei in a delocalized fashion creating a sea of negative charge interacting with the positive nuclei. Electronic structure, the quantum mechanical nature of electrons in molecules, governs the distribution of electrons and thus distribution of negative charge. The hydrogen bond consists of a unique balance between classical electrostatics and quantum mechanical descriptions that yields a strong intermolecular interaction. Efforts have been taken into decomposing the full intermolecular interaction into well-known classical (electrostatic (ES), polarizability (PL), dispersion (DISP)) various quantum in nature (exchange repulsion (EX), charge transfer (CT)) interactions. Those efforts will be discussed here to prove the unique dichotomy of the classical and quantum nature of the hydrogen bond.

1.1.1 Energy Decomposition Analysis

The energy decomposition analysis (EDA) has been used to separate the electrostatic and quantum mechanical aspects of the hydrogen bond. EDA separates intermolecular interactions into various components by selecting or constraining molecular orbitals. [23–26] Early work by Kitaura and Morokuma defined ES as the "interaction between occupied molecular orbitals (MOs), which does not cause any mixing of MOs", PL as the "interaction which causes the mixing between the occupied and vacant MOs within each molecule", EX as the "interaction between occupied MOs, which causes electron exchange and delocalization between molecules", and CT as the "interaction which causes

intermolecular delocalization by mixing the occupied MOs of one molecule with the vacant MOs of the other". [27] Their definitions were based on the quantum mechanical description of the above energetic terms. Morokuma's subsequent work applied the EDA scheme to hydrogen bonded dimers. [25] Under this scheme, he reported that ES and EX play a major role, whereas CT and PL played a minor but significant role in interaction energy. Additionally, the ES and CT terms were basis set dependent. Although all terms weakened with intermolecular distance, EX was most sensitive to that coordinate. In their decomposition analysis, there is a bit of residual energy they call the "coupling term" and attribute it to "higher order interactions between [ES, PL, EX, and CT]". [25]

It should be noted that the way energy is decomposed is not unique. Decades after the introduction of Morokuma's decomposition scheme, Glendening and Streitwieser proposed a natural energy decomposition analysis (NEDA) based on the natural bond orbitals (NBO). [23] Under this scheme, CT is well defined by the difference in energy between the full molecular wavefunction and the anti-symmetric product of localized and orthogonal monomer wavefunctions that comes out naturally from the NBO analysis. The ES term is computed by taking the energy difference between the the anti-symmetric localized wavefunction and the monomer components separately. Under this scheme, the ES term also contains the attractive piece of PL. A final deformation (DEF) term describes the energetic cost of deforming the the monomer's wavefunction due to EX and PL.

The energy decomposition is also sensitive to methodology. For the water dimer (HF/4-31G) [23] the NEDA yields much stronger ES and CT contributions to the binding energy than Morokuma's EDA (-17.75 / -13.25 and -10.5 / -2.4 kcal/mol, respectively). In addition, DEF is much larger than its corresponding EX (24.79 vs. 6.2 kcal/mol). Although the NEDA calculates stronger terms that cancel, both analyses show qualitatively similar results. Another method, density based absolutely localized molecular orbital (ALMO-)EDA, [24, 28, 29] constrains the electron-density to produce ES, EX, PL, CT, and dispersion (DISP) terms. This form produced the same trend as the previous methods but predicts stronger contributions from PL than from CT and highlights the importance of DISP in stabilizing the interactions.

Despite EDA's non-unique definitions of the various components of the interaction energy, it still provides a useful perspective in separating they hydrogen bond from a pure electrostatic interaction. For example, various EDA methods were applied to chemical systems. [23, 24, 30] NEDA shows

small CT character in alkali metal halide (<20%) and large CT character in borane-Lewis base dimers (>1000%), suggesting the former are largely electrostatic, whereas the latter are largely covalent in nature. [23] For reference, NEDA predicts that CT is $\sim 170\%$ of the binding energy of the water dimer. [23] Similarly, ALMO-EDA suggests that CT contributions are more prominent in the ammonia-borane dimer than in the water dimer. [24] Finally, the structure of the water dimer has important consequences on the decomposition of its interaction energy. The linear water dimer, exhibiting the traditional hydrogen bond, is stabilized $\sim$30% by CT in the ALMO-EDA, whereas the bifurcated water dimer, which forms two highly bend hydrogen bonds, is much more electrostatic, with CT contributions being less than 10%. [24]

The EDA analysis of hydrogen bonded structures, discussed here mostly for water, show that hydrogen bonds are both classical and quantum in nature. Classical electrostatics make up a large portion of of the attractive character in hydrogen bonding, stemming largely from the attraction between one molecule's electrons and another molecule's nuclei. However, quantum mechanical contributions from the molecular orbital overlap become an important stabilizer of the hydrogen bond. The decomposition into the various physical interactions does not remain unique while maintaining the full description of the electronic wavefunction. Therefore, we will introduce a fully classical picture of the ES interactions in a hydrogen bond to discuss their unique PL. Then we will discuss the modern approaches used to recover the quantum mechanical EX and CT terms.

1.1.2 Electrostatics (ES) and Polarizability (PL)

The electrostatic picture of molecular interactions discussed here will abstract the sub-atomic particles into a molecular charge distribution that participates in electrostatic interactions. That distribution of charge creates an electric field that interacts with the charges on another molecule. This charge distribution can be represented by a multipole expansion. The full details of multipolar interactions have been studied exhaustively. [31] Here, we will focus on the dominant monopole (single charge) and dipole (separation of charge) interactions and how they strengthen hydrogen bonds.

In a charged molecule the leading multi-polar moment is the monopole, represented by a point

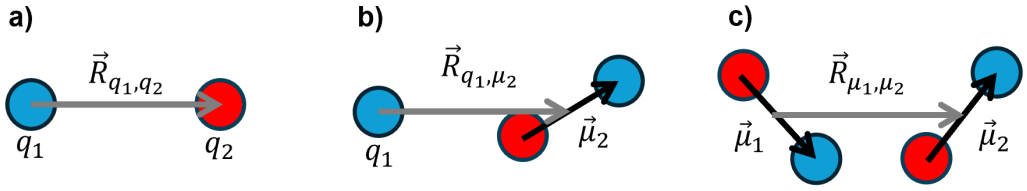

Figure 1.1: Schematics of the (a) monopole interaction, (b) monopole-dipole interaction, and (c) dipole-dipole interaction. Circles represent point charges (blue is positive, red is negative). The black arrow represents the dipole moment of two associated charges and the gray line represents the vector between the multipole origins.

charge at the center of charge for the molecule. The monopole-monopole interaction is

$$V_{q_1,q_2}(\vec{R}_{q_1,q_2}) = -\frac{1}{4\pi\varepsilon_0}\frac{q_1 q_2}{||\vec{R}_{q_1,q_2}||} \tag{1.1}$$

where q_1 and q_2 is the charge of the monopoles, and $\vec{R}_{q_1,q_2}$ is the separation of the monopoles. The locations of the monopole within each molecule is the average location of the charge. These interactions are uncommon between a hydrogen bond donor and acceptor, since proton transfer will occur to neutralize the charge. However, doubly ionic hydrogen bonds can occur (ex. NH_4^+–Cl^-) and are important for the hydrogen bonding interactions in some ionic liquids. [21, 22, 32] Even though the monopole-monopole interaction will dominate, nonuniform charge distribution throughout the molecule will still produce strong dipoles. Subsequent discussion on charge neutral interactions will still be applicable to these charged systems.

Most hydrogen bond donors and acceptors are charge neutral, meaning their leading multipole moment is the dipole. The dipole is created by the nonuniform separation of charge throughout the molecule, given by a sum over nuclei point charges and an integral over electron density,

$$\mu = \sum_i^N (r_i - r_0)q_i - \int_V \rho(\vec{r})(\vec{r} - r_0)d^3\vec{r} \tag{1.2}$$

where q_i is the charge of the i^{th} nuclei up to N nuclei, r_i is the position of nuclei r_0 is some reference point in the molecule commonly the center of charge, and $\rho(\vec{r})$ is the electron density. Note that electrons are not treated as point charges in quantum mechanical electronic structure calculations and therefore must be integrated over their density to produce a dipole moment. In singly charged

hydrogen bonds, such as NH_4^+–H_2O or H_2O–Cl^-, the dipole moment of one molecule interacts with the monopole on the other molecule through

$$V_{q_1,\mu_2}(\vec{R}_{q_1,\mu_2}) = -\frac{1}{4\pi\varepsilon_0}\frac{q_1\vec{\mu}_2 \cdot \hat{R}_{q_2,\mu_2}}{||\vec{R}_{q_1,\mu_2}||^2} \tag{1.3}$$

where q_1 is the monopole on one molecule, $\vec{\mu}_2$ is the dipole vector on the other molecule, $\vec{R}_{q_1,\mu_2}$ is the distance between the monopole and dipole origins, and $\hat{R}_{q_1,\mu_2}$ is the unit vector along $\vec{R}_{q_1,\mu_2}$, see figure 1.1. The directionality of $\vec{\mu}$ and $\vec{R}_{q_1,\mu_2}$ matters, and the sign of in eq. (1.3) reflects the directions shown in figure 1.1. Most commonly, the hydrogen bond donor and acceptor are charge neutral, meaning that the leading multipolar interaction is dipole-dipole. This interaction takes the form

$$V_{\mu_1,\mu_2}(\vec{R}_{\mu_1,\mu_2}) = \frac{1}{4\pi\varepsilon_0}\frac{\vec{\mu}_1 \cdot \vec{\mu}_2 - 3(\vec{\mu}_1 \cdot \hat{R}_{\mu_1,\mu_2})(\hat{R}_{\mu_1,\mu_2} \cdot \vec{\mu}_2)}{||\vec{R}_{\mu_1,\mu_2}||^3} \tag{1.4}$$

where $\vec{\mu}_1$ is the dipole on one molecule, $\vec{\mu}_2$ is the dipole vector on the other molecule, and $\vec{R}_{\mu_1,\mu_2}$ is the distance between the dipole origins, see figure 1.1.

Based on eqs. (1.1) – (1.4) it becomes apparent that structure and orientation are important in maximizing the ES interaction. Interactions get stronger as molecules get closer together, prevented from collapse by EX. If able to freely rotate, dipoles will align themselves to the most favorable orientation, placing positive charges in proximity to negative charges. This orientational and distance preference becomes important for molecular dipole-dipole interactions. For example, formaldehyde, which has a molecular dipole along the C=O axis, maximizes dimer interactions by aligning their dipoles in an anti-parallel fashion. [33, 34]. The hydrogen bond differs from pure dipole-dipole interactions by adopting different orientations that strengthen their interactions through other means.

The position of the electrons and nuclei in a hydrogen bond are distinct from the distribution in the respective isolated monomers. Each particle can and will respond to the electric field of the other molecule in order to strengthen the dipole moment and the electrostatic interaction. Most responsive are the electrons, who morph their density and change the molecular dipole through polarization. The polarizability tensor, which describes the proportional change of the dipole in response to an external field, is calculated as the change in the dipole moment divided by the change in external field. These are developed in the context of perturbation theory, [31] but are simply

expressed here. For an induced dipole, the polarizability tensor is a 3x3 matrix with elements that describe how the components of the dipole moment change with respect to the components of an electric field.

$$a_{\alpha,\beta} = \frac{\Delta\mu_\alpha}{\Delta E_\beta} \tag{1.5}$$

The energy due to induction is related to the field and polarizability. It is convenient to define the induced dipole from polarizability and add it to the static dipole,

$$\vec{\mu}_1 = \vec{\mu}_1^{(0)} + \frac{1}{4\pi\varepsilon_0}\frac{2\boldsymbol{a}_1 \cdot \vec{\mu}_2}{||\vec{R}_{\mu_1,\mu_2}||^3} \tag{1.6}$$

which can be easily easily substituted into eq. (1.4) to yield

$$V_{\mu_1,\mu_2,\alpha_1}(\vec{R}_{\mu_1,\mu_2}) = V_{\mu_1,\mu_2}(\vec{R}_{\mu_1,\mu_2}) + \frac{1}{8\pi^2\varepsilon_0^2}\frac{2\boldsymbol{a}_1 \cdot \vec{\mu}_2 \cdot \vec{\mu}_2 - 6(\boldsymbol{a}_1 \cdot \vec{\mu}_2 \cdot \hat{R}_{\mu_1,\mu_2})(\hat{R}_{\mu_1,\mu_2} \cdot \vec{\mu}_2)}{||\vec{R}_{\mu_1,\mu_2}||^6} \tag{1.7}$$

The energy contribution from polarizability decays with molecular distance as $1/R^6$. When either the polarizability or the electric fields from dipole moments are large, induced dipoles can further polarize the other dipoles. This must be solved self-consistently since the strengthening of one dipole via polarization will in turn further polarize the other.

In addition to the polarized electrons, the nuclei will respond to the electric field, changing bond lengths to increase dipoles. This is complicated, as the change in the nuclear coordinates of the molecules correspondingly changes the electron density and thus the dipole moment and polarizability. Structural changes are typically small, so it is safe to Taylor expand the dipole moments and polarizability tensors and substitute them into eqs. (1.3), (1.4), (1.6), and (1.7) making the electrostatic interaction potential not only a function of the intermolecular coordinates, but of the intramolecular degrees of freedom. This can get complicated and it is sometimes neglected in classical, non-polarizable force fields. Polarization is counter-balanced by the energetic cost of deforming the electronic wavefunction or the nuclei positions so that ES does not blow up.

This ES picture is sufficient in describing many of the properties of the hydrogen bond and is the basis for the most simple classical potentials. For simplicity, we will focus on some potentials for water. The simple point charge (SPC) [35, 36] and TIP3P [37] models for water places partial charges on each atom, which create the dipolar electric field and the electrostatic potential between water molecules. This type of model best matches other pairwise models for proteins and organic

molecules like Amber, Charmm and OPLS. [38–41] With the effective point charges, the dipole moment does not not need to be computed and instead the interaction can be treated as monopolar interactions, eq. (1.1), between atoms on different molecules. It is important to note that a general repulsive term is added to prevent the molecules from collapsing on themselves. This will be discussed later.

Additional pairwise potentials have been developed that modify the sites of the point charges to better describe the electrostatic interactions and hydrogen bonding of a water molecule. For example, in TIP4P the oxygen's charge center is moved along the bisector towards the center of mass (M site). [37, 42, 43] TIP4P and its variations can better describe the experimental properties of liquid water. TIP5P replaces the oxygen's atomic point charge with two point charges placed in tetrahedral positions relative to the oxygen, mimicking it's lone pairs. [44] The electrostatic charges in these models are not meant to reproduce the electric field of the water molecule. Instead they are fit to reproduce bulk-water behavior at some experimental conditions.

Atomic charges and nuclei position within water are typically held rigid. Therefore, there is no intramolecular response of charges in response to an external field. Advances have been made to allow the water molecule to deform throughout the simulation. [45–47] This permits the electrostatic attraction to strengthen through charge/nuclei separation, although the atomic displacement due to an electric field is generally much less than the vibrational displacement at the liquid temperatures. Electron polarization is not captured and is the major deficit in the transferability of these models.

Polarizable water models have been developed to allow the charge distribution of one molecule to respond to the field of another molecule. One of these is the Thole-type (TTM) family of polarizable water potentials. [48] In the TTM2.1-F and TTM3-F versions of this family of models, geometry-dependent charges and polarizability are incorporated into the electrostatic attraction of the model. It follows the 4-site model of TIP4P but with geometry-dependent charges on the hydrogen and the M site, and an inducible dipole on each atom. [49–51] The TTM2.1-F and TTM3-F models differ in using gas-phase or approximate solvent phase charges. Other potentials take the multipolar nature of the interactions further by performing a multipole expansion on each atomic site. [31, 52] AMOEBA adds an initial dipole capable of polarizability and quadrupole to each atomic site. [53] Each model tackles polarizability from a slightly different philosophy.

The ES of the hydrogen bond describes several of their properties. The strong dipoles create a

bulk of the attractive potential and polarizability strengthens the dipolar attraction and provides some selectivity for their unique geometry. ES, however do not capture the full picture. EX prevents molecular collapse by creating a large repulsive wall with respect to intermolecular distance. Furthermore, CT importantly stabilizes strong hydrogen bonds and seemingly further biases the non-dipole aligned geometry in water. Next, we will discuss how one can (or really, how one cannot) describe the quantum effects simply.

1.1.3 Electronic Structure Contributions, EX and CT

Outside of ES, EX and CT are two largely important interactions in the hydrogen bond. EX is the energetic cost of anti-symmetrizing each monomer's occupied molecular orbitals, and CT is the energetic benefit arising from mixing one monomer's occupied molecular orbital with a virtual orbital on the other monomer. Both originate from the overlap of molecular orbitals upon dimer formation. This is one of the biggest challenges in modern theoretical chemistry, with the outcome that a simple but complete description of the EX and CT terms will greatly improve classical modeling. I have not attempted in describing these terms fully, but will discuss how physical chemists have handled such terms in their potential models with relative success.

EX has been estimated in interaction potentials since it is the dominant repulsive term responsible for the spacial exclusivity of molecules.The earliest description for an EX function was in the development of the Lennard-Jones (LJ) potential. [54, 55] LJ-type particles are attractive with a $1/R^6$ dependence on interatomic distance, as an average of higher-order multipole and dispersion attraction. To model repulsion, a $1/R^{12}$ potential was chosen for computational simplicity ($(1/R^6)^2$). The LJ potential is

$$V_{LJ}(R) = \epsilon \left(\left(\frac{\sigma}{R} \right)^{12} - \left(\frac{\sigma}{R} \right)^{6} \right) \tag{1.8}$$

where ε describes the strength of the attraction, σ determines the intermolecular distance of R where $V_{LJ} = 0$. It should be recognized that $1/R^{12}$ was chosen for ease of computing, when in reality the repulsive wall of Lennard-Jones particles is steeper at $1/R^{14}$. [56] The former term qualitatively describes EX but is likely mathematically inappropriate. None-the-less, this potential is commonly implemented in pairwise force fields. [38–41] We finally note that the LJ potential is a special case of the more general Mie potential, [57] which casts the 2 terms of eq. (1.8) in terms

of general exponents n and m.

As EX stems from the overlap and resulting anti-symmetrizing of the occupied electronic wavefunctions, a function describing EX should be exponential in nature. Thus, the Buckingham potential [58] was introduced

$$V_B(R) = Ae^{-BR} - \frac{C}{R^6} \tag{1.9}$$

where A, B, and C are fit to the potential. This potential is more physically meaningful and describes EX at appropriate distances. However, the Buckingham potential turns over at very small distances since $\lim_{R\to\infty} V_B(R) = -\infty$. Therefore, a modified Buckingham potential [59] has been introduced

$$V_B(R) = \frac{C}{R^6}\left(Ae^{-BR} - 1\right) \tag{1.10}$$

that ensures stability over the range of R. [59, 60] The Buckingham potential has been implemented in many polarizable force fields [49–51] to describe the repulsive component of the hydrogen bond.

Although CT is adequately described by the overlap of occupied and unoccupied molecular orbitals, there is no physical description of the charge transfer interaction dependent only on intermolecular distances. Instead, modern potentials capture CT by mixing it into the short range part of other interactions. Pairwise potentials incorporate charge transfer into the functional form by fitting atomic charges to experimental observables. Other force fields choose to correct the near-range part of the potential by fitting to ab-initio energies. MB-Pol [61] uses permutationally invariant polynomials (PIPs) to fit the difference between the TTM4-F model and the ab-initio values of geometrically diverse dimers and trimers. This essentially is a fit to the charge transfer energy with respect to the intermolecular degrees of freedom. Other force fields fit PIPs to all interactions (ES, PL, EX, CT, & DISP) to fully describe the hydrogen bond interaction. [62]

1.2 Structural Definitions of Hydrogen Bonds

The energetics of the hydrogen bond makes a clear separation from weaker intermolecular interactions through a balance of ES and CT interactions. However, it is not feasible to identify every hydrogen bond via energy decomposition, especially in a multi-component, multi-molecular ensemble. In a molecular ensemble, geometries commonly exist away from equilibrium either due to competing interaction or temperature effects. The hydrogen bond donors and acceptors may be

present but no hydrogen bond may be formed. Instead, chemists have developed structural criteria for identifying a hydrogen bond. Under a structural definition, one can estimate the number of hydrogen bonds in a molecular cluster or bulk simulation. This provides important insight into the driving force of aggregation.

The hydrogen bond has unique structure. Instead of maximally aligning molecular dipole moments, the hydrogen bond places the hydrogen atom between the hydrogen bond donor and acceptor, see figure 1.2. This maximizes the PL and CT interactions. The hydrogen atom is close to collinear with the heavy atoms in order to maximize the molecular orbital overlap involved in charge transfer, such as the DH σ^* and the A lone pair (LP) orbitals. However, the hydrogen atom is not oftentimes exactly collinear with the D and A heavy atoms due to additional interactions. This restricted spacial capacity provides a means to define the hydrogen bond based on a set of intermolecular coordinates. The intermolecular distance can be described either by the donor-acceptor distance (R_{DA}) or the hydrogen-acceptor distance (R_{HA}). The hydrogen bond linearity can be described as the donor-hydrogen-acceptor or the hydrogen-donor-acceptor angle (θ_{DHA} or β_{HDA}, respectively). The former prefers to be around 180° whereas the latter prefers to be around 0°. Finally, the orientation of the acceptor molecule is important in placing the LP in a proper position to interact with the hydrogen atom, and must be accounted for in the definition of the hydrogen bonded structure. Other intermolecular degrees of freedom will not greatly effect the orbital overlap and therefore do not need to be described.

The structural definition of the hydrogen bond is going to depend heavily on the identity of the donor and acceptor, since the molecular orbitals will change significantly between each donor and acceptor pairing. There has been substantial work previously reported in defining the hydrogen bond geometry in water, which will be the focus of this section and parts of this thesis. However, we encourage investigations into defining the hydrogen bond geometries in other molecular systems, including mixed donor-acceptor pairs.

1.2.1 Skinner's $r - \psi$ Definition

A commonly used structual definition of the hydrogen bond, and the one used in this thesis, was developed by Kumar, Schmidt, and Skinner. [63] They combined the notion that the hydrogen

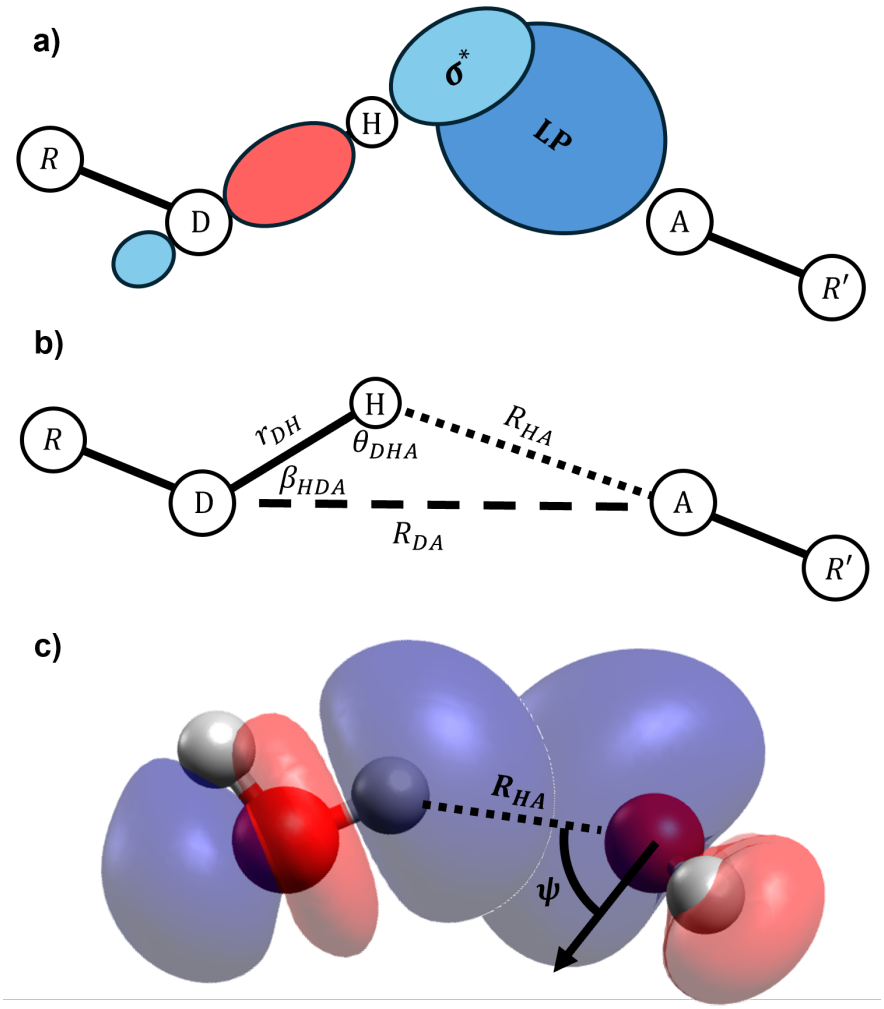

Figure 1.2: Orbital structure (a) and molecular structure (b) of the generic hydrogen bond, with coordinates and NBO orbitals for water in (c). NBO molecular orbitals (lone pairs (LP) nucleophiles and antibonding (σ^*) electrophiles) overlap based on the intermolecular coordinates.

bond can be energetically and structurally defined in order to create a consistent definition of the hydrogen bond. An earlier work [64] reported a bimodal pairwise energy distribution for liquid water molecules they attributed to the separation of hydrogen bonded (low energy, low entropy) and non-hydrogen bonded (high energy, high entropy) structures. Another two studies showed that a similar bimodal distribution exists for the electronic bond order between the hydrogen atom and the acceptor and also for the occupancy of the σ^* orbital. Noting all of these are dependent on the intermolecular geometries, Kumar et al. [63] proposed using a set of coordinates, consisting of one distance and one angle, to define the boundary between the hydrogen bonded and non-hydrogen bonded water molecules based on structural parameters.

They explored a large number of intermolecular coordinate combinations to define the hydrogen bond. Each set of coordinates produced a 2-D potential of mean force (PMF) with a transition state between hydrogen bonded and non-hydrogen bonded minima, which they used as the cutoff to structurally define the hydrogen bond. They determined the best pair of coordinates to be the hydrogen atom-acceptor distance (r in ref. [63], R_{HA} here) and the angle between $\vec{R}_{\mathrm{HA}}$ and the normal vector of the water molecule (ψ), see figure 1.2. ψ is rigorously less than 90°. The boundary of the contour follows a parabola and is easily evaluated (see eq. (3) in [63]). Furthermore, these coordinates predict the occupancy of the σ^* orbital (N) computed through NBO analysis at the B3LYP/aug-cc-pVDZ level of theory via. eq. 4 in [63], also included here

$$N(R_{\mathrm{HA}}, \psi) = e^{-R_{\mathrm{HA}}/0.343\ \text{Å}} \left(7.1 - 0.050\psi/\mathrm{deg} + 0.00021\psi^2/\mathrm{deg}^2\right) \tag{1.11}$$

The occupancy that separates the bimodal distribution is $N = 0.0085$, and the structural definition of hydrogen bond that predicts an occupancy above 0.0085 produces a similar boundary as that of the energetic contour. This occupancy definition is commonly used for it's connection back to CT.

1.2.2 Other Definitions

It should be noted that the structural definition of the hydrogen bond in water is arbitrary. Some definitions focus only on the intermolecular distances. Others incorporate the angle of the hydrogen bond, either β_{HDA} or θ_{DHA}. For example, Luzar [65] combined the structural distributions seen in liquid water ($R_{\mathrm{DA}} < 3.5$ Å) from the radial distribution function [66] and $\beta_{\mathrm{HDA}} < 30°$ from the libration mode [67]. Another work [68] used a parabolic boundary in the same coordinates

determined from sharp changes in the X-ray absorption spectrum across these two coordinates. Finally, a definition incorporating the hydrogen atom-acceptor distance (from radial distribution functions [69]) and donor-hydrogen-acceptor angles has been proposed. [70] Kumar [63] looked at these definitions and found none that followed the energetic contours of the PMF. Furthermore, the σ^* occupancy could be reasonably fitted to the variables discussed about 2x worse than in the $r - \psi$ definition. The fits remained quadratic with respect to the angles and thus only applicable to the parabolic $R - \beta$ definition of the hydrogen bond. [68]

Other definitions of the hydrogen bond do not restrict themselves to specific regions of the potential energy surface, but instead rely on some general criteria that is either experimentally or computationally accessible. For example work by Henchman [71] define a hydrogen bond as an $O_D H_D \cdots O_A$ pair with the smallest $H_D \cdots O_A$ distance. This, they claim, have benefits for allowing a diverse set of hydrogen bond structure and preventing two hydrogen bonds from forming between two molecules. Hammerich et.al. [72] further restricted this definition to include only the closest two $H_D \cdots O_A$ distances, limiting each water molecule in accepting at most two hydrogen bonds. Despite removing strict geometric constraints and focusing on generic hydrogen bond properties, these last two definitions are insufficient in describing non-traditional hydrogen bonds.

1.3 Applications of Hydrogen Bonding

With the definition of the hydrogen bond discussed in sections 1.1 and 1.2, we can now analyze the physical chemistry of the hydrogen bond. However, there is already a rich history of chemical investigations of hydrogen bonding. ES and CT produce an enthalpically rich but entropically poor interaction. This produces highly stable and highly ordered structures that are easily modulated with temperature.

In this section, we will go into greater detail about the chemical systems that contain hydrogen bonds. Throughout this thesis, we will focus on applications in biological and aqueous systems. We will discuss the importance of hydrogen bonding in peptide systems, including folding, molecular recognition, and enzymatic catalysis. Aqueous systems have robust and extensive hydrogen bonding environments that interact with and stabilize solutes.

1.3.1 Protein Folding and Catalysis

Proteins are biological polymers of functionalized amino acids. Protein structure and function is very diverse, with applications in catalysis, signaling, and structural support. [11] Some proteins are intrinsically disordered in order to respond to environmental conditions. [73] Other proteins are highly structured and fit specific roles in biological systems. The structure of proteins is governed by the weak intramolecular interactions, including hydrogen bonding. Every amino acid consists of two or more hydrogen bonding sites. The backbone of every amino acid contains an internal amide that can act as both a NH hydrogen bond donor and C=O hydrogen bond acceptor. Some amino acids have polar and aromatic substitutes that act as additional hydrogen bond donating and accepting sites.

Protein structure is separated into four levels. [11] Primary structure refers to the covalent attachment of the monomeric units, specifically on the order of amino acids. Secondary structure consists of small local patterns in peptides. These include α helices where hydrogen bonds form between backbone amides four residues away, and β sheets where strands of hydrogen bonds form between distant amide strands in parallel or anti-parallel fashions. Hydrogen bonds are important in providing stability and order in secondary structure, but side-chain identity is important for promoting or inhibiting the folding. [11, 74, 75] Tertiary structure describes the full three-dimensional structure of the poly-peptide chain influenced by the many competitive van-der-Waals, charge, and hydrogen bonding interactions. The tertiary structure creates the domains for specific protein function. The diversity of interactions and the range of interaction strength makes tertiary structure sensitive to environmental factors such as ion concentration, pH, and temperature. [11, 76–78] Quaternary structure describes how independent poly-peptide chains aggregate to form fully functional proteins. Although protein folding is a complicated process involving many types of interactions, hydrogen bonding plays a critical role in stabilizing structure.

Hydrogen bonds are also important in the interactions of proteins with substrate molecules. Receptor proteins have specific binding domains where substrates bind and cause downstream perturbations and signaling cascades. [11] The process of molecular recognition is complicated but dominated by the intermolecular interactions. Steric interactions describe the overall interactions stemming from molecular shape and dictate how molecules of a substrate can fit into the binding

domain of the receptor. This creates the "hand-in-glove" and "key-in-lock" models for ligand binding, where a ligand must have a shape corresponding to the binding domain for possible binding. [79] Stronger interactions, including hydrogen bonding, strengthen substrate binding and make receptors sensitive to specific molecules. The functional groups on the protein require corresponding hydrogen bond or charged sites on the substrate for favored binding. [80–82] Molecular docking attempts to describe both the steric hindrance and the strong electrostatic/hydrogen bond interactions to predict drug candidates for specific receptors. Sterics are weak and easily described through approximation methods, whereas subtle differences in hydrogen bonding are difficult to describe but are very important in receptor selectivity. [83]

The nicotinic acetylcholine receptor (nAChR) provides a useful case study to discuss the finer details of substrate binding. nAChRs are ligand gated ion channels that allow ions to enter the cell and cause intracellular signaling in response to a variety of ligands. [84–86] The natural agonist of these receptors is acetylcholine. Choline, which replaces the acetate functional group in acetylcholine with an alcohol, binds much less strongly to the nAChR do to changing sterics and the positioning of a hydrogen bonding site. [87] nAChRs consist of five subunits with slight modifications influencing the selectivity of the ion channels and the binding site. [86] For example, the α_7 receptor, which is related to memory, is sensitive to nornicotine (NOR), whereas the $\alpha_4\beta_2$ receptor, which is related to pleasure, is sensitive to nicotine (NIC). [88–92] NIC differs from NOR by N-methylation; NOR is able to form two strong N^+H hydrogen bonds whereas NIC forms only one N^+H hydrogen bond but can form weaker $C^{\delta+}H$ hydrogen bonds through the methyl group. The structures of the $\alpha_4\beta_2$ and α_7 receptors are similar with slight amino acid substitutions minimally impacting the shape of the receptor. [93, 94] The finer details in hydrogen bonding interactions are believed to separate the different selectivity of nAChRs, with a detailed analysis contrasting NIC and epibatidine. [82]

Hydrogen bonds also play an important role in the catalytic processes of many enzymes. Many metabolic processes are assisted by general acid/base catalysis, where the side chains of the amino acids act as proton donors and acceptors in hydrogen bonds and proton transfers. For example, serine proteases use three residues (serine, histidine, and aspartic acid) to cleave the stable amide bonds in amino acids. [95, 96] It first does so via. nucleophilic attack of serine to the amide coupled to the proton transfer from serine to the histidine. The C–N bond is cleaved through coupled proton

transfer from histidine, and water then serves to replace the proton and cleave the C–O ester bond between the substrate and serine. Hydrogen bonding also plays an important role in stabilizing the reaction intermediates, where the peptide backbones hydrogen bonds to the high energy oxyanion and the aspartate residue accepts a hydrogen bond from histidine, stabilizes its positive charge. Steric interactions play only a minor role in the cleavage, allowing the enzyme to selectively cleave specific amide bonds, lysine and arginine for trypsin, and tyrosine, tryptophan, phenylalanine, and leucine for chymotrypsin.

1.3.2 Bulk Water and Aqueous Solvation

The hydrogen bonds in water are of increased importance because of water's ubiquity as a biological and environmental solvent, a function that sustains life on Earth. Water's extensive hydrogen bonding network is associated with numerous anomalous macroscopic properties that result in unique chemical and biological functions. Hydrogen bonding interactions are crucial to the high solubility of polar and ionic solutes. Even if the solute is not a hydrogen bond donor or acceptor, positive charges serve to perturb and strengthen nearby hydrogen bonds, effectively increasing solubility. This has important implications in solution-based chemistry, ion solubility, and biological function. In fact, the ligand recognition process discussed earlier is commonly in competition with interactions with water. Describing these important biological process requires explicit consideration of the hydrogen bonds in liquid water and aqueous solutions.

For these reasons the hydrogen bond in water has been studied exhaustively. It's fundamental pairwise interactions have been studied exhaustively in the water dimer, with experiments yielding the water dimer's observable properties and theory determining what computational treatments are required to accurately describe its potential energy surface (PES). [97–109] In fact, a good part of our understanding of the hydrogen bond stems from our understanding of the water dimer PES. Studies have expanded to larger cluster systems to highlight the cooperative nature of hydrogen bonds. [110–112] Many-body (non pairwise additive) effects manifest in the strengthening of the hydrogen bonds and the increase of the enthalpic stability of the clusters. These effects are most pronounced in specific homodromic cycles that sacrifice entropy for stability. [112]

Our understanding of the hydrogen bonds in water clusters has significantly contributed to

our ability to understand bulk phase water. Water clusters constitute a manageable model for describing the hydrogen bond networks in water. Gas phase clusters are simple to simulate with high-level theory, where distinct configurations and individual hydrogen bonds can be disentangled. Experimentally measured properties of water clusters, such as spectral transitions, are distinct and used to explain the finer details (such as connectivity) of water networks. These systems have been used, among others, to study proton-transfer mechanisms in protonated water clusters, [113] to identify the origin of spectral signatures in IR spectra, [114–116] and to quantify strong cooperative effects in hydrogen bonding networks. [111, 117–119]

Ice provides a clearer picture of the hydrogen bonds in condensed phase water. Water forms nicely ordered, albeit many, structures in the solid phase. Each water molecule participates in four hydrogen bonds, two as a donor and two as an acceptor, obeying the Bernal-Fowler rules. [120] At one atmosphere, ice exists in it's Ih phase, where all water molecules are perfectly tetrahedral (OOO angles of 109.5°) and participate in six member cycles. [121, 122] This maximizes hydrogen bonding interactions but lowers the density (0.916 kg/L). [123] Pressure, which raises density, distorts water from tetrahedral order and weakens hydrogen bonds. [124, 125] At low temperatures, the hydrogen atoms remain ordered to form stronger hydrogen bonds by maximizing many-body interactions (i.e. in ice XI). [124] Temperature sacrifices many-body effects for entropy associated with proton ordering. At zero Celsius, the oxygen framework breaks down and ice melts.

In the liquid phase, water sacrifices strong hydrogen bonding interactions for entropy. Nonetheless liquid water is shown to retain a high degree of order, with distinct solvation shells and slow rearrangement. [5, 126, 127] Infrared (IR) studies suggest liquid water participates in only approximately 3.5 hydrogen bonds. [128] As a consequence, water molecules pack tighter and become more dense than ice (liquid density between 1.000 kg/L at 4 °C [129, 130] and 0.961 kg/L at 95 °C [130]). This is one of water's anomalous properties that is central to biological life. Other anomalous properties appear at ambient temperatures, including the non-linear character in both its heat capacity and isothermal compressibility. Each has been loosely attributed to the fleeting hydrogen bonding environment in water, with some evidence suggesting local inhomogeneity. [4] The latter properties occur in the middle of the temperature range, i.e. at 320 K, whereas the density anomaly occurs around 275 K. Understanding liquid water's anomalous properties, and how they relate to the structure of the hydrogen bond networks, is of utmost interest.

Water's importance is not limited to the homogeneous liquid. Aqueous solutions of polar and charged solutes are a common medium in chemistry and biochemistry. The interactions between charged ions and water networks are heavily studied. Heindel et al. reported that the interactions between ions and water sacrifice cooperative effects for strong charge-dipole pairwise interactions. [131] The typically cooperative 3-B interaction in water clusters becomes weak between water molecules and even repulsive between ions and water. Subsequent work by Herman et al. identified trends in the cooperativity and the structure-making/structure-breaking properties of aqueous ions in the Hofmeister series. [132] Ions that assist in protein folding and support structure, called kosmotropes (making order), exhibit stronger ion-water interactions that destabilize many-body effects, whereas chaotropes (making disorder), which denature proteins, maintain favorable water-water and water-water-water interactions. Changes in water's many-body terms originate from changes in the hydrogen bonding network structure due to the effect of ions. These effects are transferable to dilute solutions, which can be well modeled with classical potentials. Classical models, however, fail to describe concentrated solutions in which ions, and more appropriately the perturbed hydrogen bond networks, come into close contact. [133, 134]

The description of ion-water interactions is important not only in chemical physics, but has profound applications in battery research [135] and in biochemistry. [11] Per the scope of this thesis, we will focus our discussion on applications with biomolecules. The interaction of water with biomolecules competes with interactions between organic species, since cells are crowded aqueous environments. Protein folding, mentioned in an earlier section, is a competition between protein-protein and protein-water interactions. Additionally, ligand binding is a competition between ligand-water and protein-water interactions and protein-ligand interactions. The hydrogen bond networks in water stabilize organic ions, promoting or inhibiting protonation/deprotonation. The effects of water and pH on the protonation sites are important in producing bioacitve isomers of organic molecules for proper ligand recognition and binding. Water molecules both stabilize specific protonation sites and provide a means for hydrogen to transfer and respond to changing environments.

1.4 Final Introductory Remarks

This thesis will explore many of the properties of the hydrogen bond using specific examples. The intent is to report advancements in describing the hydrogen bond and its properties as they relate to nicotine and bulk water. Discussed throughout are the concepts of hydrogen bond energy, hydrogen bond structure, charge stabilization, proton transfer, and ligand binding. Beforehand, this thesis will discuss the quantum mechanical approaches used to investigate the hydrogen bond, centering around perturbation theory, which accurately describes the electronic and vibrational structure of hydrogen bonded complexes.

Chapter 2

QUANTUM MECHANICAL METHODOLOGY

The nature and finer details of the hydrogen bond can be modeled at the molecular level by quantum mechanics. Atoms and electrons are very light in mass and follow the principles of quantum mechanics, described by the Schrödinger equation

$$\hat{H}\,|\psi\rangle = i\hbar\frac{d}{dt}\,|\psi\rangle \tag{2.1}$$

where $|\Psi\rangle$ is the wavefunction, t is time, and $\hat{H}$ is the Hamiltonian operator, which is the sum of the kinetic energy operator, $\hat{T}$, and the potential energy operator, $\hat{V}$. This thesis focuses on the equilibrium behavior of the hydrogen bond, meaning that that desired observables are time-independent. Quantum states will be eigenstates of the Hamiltonian and observable phenomena will be related to these steady-state wavefunctions. This section of the thesis will discuss approaches of solving the time independent Schrödinger equation

$$\hat{H}\,|\psi\rangle = E\,|\psi\rangle \tag{2.2}$$

with a special emphasis on the perturbative methods.

In the one-dimensional picture, the kinetic and potential energy operators are simple, producing a simple 1-D Hamiltonian.

$$\hat{H} = \frac{1}{2m}\hat{p}^2 + V(\hat{x}) \tag{2.3}$$

where the momentum operator is defined as $\hat{p} = -i\hbar\frac{d}{dx}$ and the potential energy function V is purely a function of the position operator $\hat{x}$. In chemical systems, the Schrödinger equation differs largely based on the identity of the potential energy function.

Many forms of the one-dimensional Schrödinger equation are analytically solvable, such as the particle in a box, the harmonic oscillator, the Morse oscillator, and the hydrogen atom. Other systems have more complicated potential energy functions and remain solvable through various techniques such as perturbation theory and basis sets methods. In a basis set method, the 1-D

wavefunction, $|\psi\rangle$ is defined as a linear combination of basis functions $|\phi\rangle$ such that

$$|\psi\rangle = \sum_i c_i \, |\phi_i\rangle \tag{2.4}$$

and the Schrödinger equation equation can be rewritten as

$$\sum_i c_i \; \hat{H} \, |\phi_i\rangle = E \sum_i c_i \, |\phi_i\rangle \tag{2.5}$$

For a basis set approach, each function must be continuous, differentiable, and span the space of the problem at hand. When multiplied by $\langle\phi_j|$, one obtains the characteristic diagonalization equation

$$\sum_i c_i \, \langle\phi_j| \, \hat{H} \, |\phi_j\rangle = E \sum_i c_i \, \langle\phi_j| \, \phi_i\rangle = E c_j \tag{2.6}$$

Diagonalizing the matrix with elements of $\hat{H}$ in the basis of $|\phi\rangle$ yields a diagonal matrix, whose elements are the steady-state energies E of the Schrödinger equation, and a transformation matrix, whose vector components consist of the coefficients c_i in the linear combination of $|\psi\rangle$ from $|\phi\rangle$. The matrix representation requires a finite basis, so the basis set must nicely approximate the target wavefunctions to ensure accurate and fast convergence.

Although basis set methods are robust and can be applied to a number of differential equations, they can be expensive to solve. The Hamiltonian matrix consists of $n \times n$ elements where n is the number of basis function in the basis set. Each element must be evaluated through integration, which can be costly for some methods. Furthermore, the diagonalization of the Hamiltonian scales as $\mathcal{O}(n^3)$. This can become expensive very quickly and other methods must be used.

An alternative to diagonalization is perturbation theory, [136] in which one starts from a zero-order solution and perturbs the Hamiltonian to approximate the exact solution

$$\hat{H} = \hat{H}^{(0)} + \lambda \hat{H}^{(1)} \tag{2.7}$$

where $\hat{H}^{(0)}$ is the zero-order Hamiltonian, $\hat{H}^{(1)}$ is the first-order perturbation to the Hamiltonian operator, and λ is a bookkeeping factor describing the scale of the perturbation. In this form, the wavefunciton and energy takes a perturbative form

$$|\psi\rangle = \left|\psi^{(0)}\right\rangle + \lambda \left|\psi^{(1)}\right\rangle + \cdots \lambda^n \left|\psi^{(n)}\right\rangle \tag{2.8}$$

$$E = E^{(0)} + \lambda E^{(1)} + \cdots \lambda^n E^{(n)} \tag{2.9}$$

where

$$\hat{H}^{(0)} \left|\psi^{(0)}\right\rangle = E^{(0)} \left|\psi^{(0)}\right\rangle \tag{2.10}$$

making the zeroth order term to the wavefunctions orthonormal. Higher perturbative terms can be related back to the these zero-order terms. Since $\hat{H} |\psi\rangle = E |\psi\rangle$, it holds that all distributive terms with the same power of λ must also hold to the Schrödinger equation. I.e. for λ^1,

$$\lambda\hat{H}^{(1)} \left|\psi^{(0)}\right\rangle + \lambda\hat{H}^{(0)} \left|\psi^{(1)}\right\rangle = \lambda E^{(1)} \left|\psi^{(0)}\right\rangle + \lambda E^{(0)} \left|\psi^{(1)}\right\rangle \tag{2.11}$$

in which all λ's cancel. When multiplied by $\left\langle\psi_n^{(0)}\right|$, the energy of $E_n^{(1)}$ is solvable. This is done by evaluating $\left\langle\psi_n^{(0)}\right| \hat{H}^{(0)} \left|\psi_n^{(1)}\right\rangle = E_n^{(0)} \left\langle\psi_n^{(0)}\right| \left.\psi_n^{(1)}\right\rangle$ and subtracting this term from both sides, yielding

$$E_n^{(1)} = \left\langle\psi_n^{(0)}\right| \hat{H}^{(1)} \left|\psi_n^{(0)}\right\rangle \tag{2.12}$$

Corrections to the n^{th} wavefunction are found by multiplying eq. (2.11) by the m^{th} zero-order wavefunction, $\left\langle\psi_m^{(0)}\right|$ where $m \neq n$. This makes the first order correction to the wavefunction a linear combination of zeroth order wavefunctions.

$$\left|\psi_n^{(1)}\right\rangle = \sum_{m\neq n} \frac{\left\langle\psi_m^{(0)}\right| \hat{H}^{(1)} \left|\psi_n^{(0)}\right\rangle}{E_n^{(0)} - E_m^{(0)}} \left|\psi_m^{(0)}\right\rangle \tag{2.13}$$

The summation is carried out up to some maximum number of basis function. This method can be executed ad nauseam, where the truncation is determined by the order of perturbation. For example, $E^{(2)}$ which is important in many forms of perturbation, is expressed as

$$E_n^{(2)} = \left\langle\psi_n^{(0)}\right| \hat{H}^{(1)} - E_n^{(1)} \left|\psi_n^{(1)}\right\rangle \tag{2.14}$$

by multiplying the second order perturbation terms by $\left\langle\psi_n^{(0)}\right|$ and simplifying terms. The first order wavefunction can be substituted to leave this energy expression solely in terms of matrix elements of the perturbation operators in the basis of the zeroth order wavefunctions

$$E_n^{(2)} = \sum_{m\neq n} \frac{\left|\left\langle\psi_n^{(0)}\right| \hat{H}^{(1)} \left|\psi_m^{(0)}\right\rangle\right|^2}{E_n^{(0)} - E_m^{(0)}} \tag{2.15}$$

The scaling is much better and is limited by the evaluation of the terms in the Hamiltonian and the order of the perturbation expansion. The majority of the cost from perturbation theory comes from either computing the perturbation operator or the matrix elements. Perturbation theory can be applied in solving both the electronic and vibrational Schrödinger equations, both of which are used in this thesis.

2.1 The Multi-Electron Hamiltonian

The fundamental problem in computational and theoretical chemistry lies with the ability to solve the electronic Schrödinger equation. This is solvable for a one-electron system with the Hamiltonian

$$\hat{H} = -\frac{\hbar^2}{2m_e}\nabla^2 - \frac{Ze}{R} \tag{2.16}$$

where ∇^2 is the second derivative in three-dimensional space, Z is the nuclear charge, e is the charge of an electron, and R is the distance between the nucleus and the electron. Nuclei are much more massive than electrons. Through the Born-Oppenheimer approximation, [137] the system can be thought of a stationary nucleus and a mobile electron. This is because the electrons move many times faster than nuclei, and as such, can respond almost instantaneously to changes in the nuclear coordinates. The results is a fully separable Hamiltonian, consisting of the electronic Schrödinger equation, eq. (2.16) and (2.17), and the vibrational Schrödinger equation, eq. (2.53). For a single atom, the second derivative is best represented in spherical coordinates. Via separation of variables, this Schrödinger equation is solvable, with the wavefunction separated into angular and radial components. The former are the spherical harmonics, and the latter are Laguerre polynomials multiplied by an exponential.

The many-electron Hamiltonian becomes

$$\hat{H} = -\sum_i^n \frac{\hbar^2}{2m_e}\nabla_i^2 + V_{\mathrm{NN}} - \sum_i^n \sum_j^N \frac{Z_j e}{R_{ij}} + \frac{1}{2}\sum_i^n \sum_{j\neq i}^n \frac{e^2}{r_{ij}} \tag{2.17}$$

where R_{ij} is the nucleus-electron distance, r_{ij} is the electron-electron distance, N is the number of nuclei, n is the number of electrons, Z_j is the charge on nucleus j, in atomic units and V_{NN} is the nuclear repulsion energy, a function of nuclear coordinates. That final term, the electron-electron repulsion, makes the many-electron system solution non-analytic and the anti-symmetry of the electron wavefunction makes this term difficult to handle.

The first attempt to solve this many electron equation is through a basis set method. Since electrons are indistinguishable particles obeying the Fermi-Dirac statistics, their wavefunction must be antisymmetric with respect to exchange. For a two electron system, the wavefunction is written as

$$\Psi(\mathbf{x}_1, \mathbf{x}_2) = \frac{1}{\sqrt{2}}\left[\psi_1(\mathbf{x}_1)\psi_2(\mathbf{x}_2) - \psi_2(\mathbf{x}_1)\psi_1(\mathbf{x}_2)\right] \tag{2.18}$$

where ψ are the individual wavefunctions of electron 1 and 2, whose coordinates are denoted as $\mathbf{x}_1$ and $\mathbf{x}_2$. In higher dimensions, its convenient to represent the antisymmetric wavefunction of the electrons as a Slater determinate of molecular orbitals (ψ_i) for each electron.

$$\Psi(\mathbf{X}) = \hat{\mathcal{A}} \prod_i^n \psi_i(\mathbf{x}_i) = \frac{1}{\sqrt{n!}} \begin{vmatrix} \psi_1(\mathbf{x}_1) & \psi_2(\mathbf{x}_1) & \cdots & \psi_n(\mathbf{x}_1) \\ \psi_1(\mathbf{x}_2) & \psi_2(\mathbf{x}_2) & \cdots & \psi_n(\mathbf{x}_2) \\ \vdots & \vdots & & \vdots \\ \psi_1(\mathbf{x}_n) & \psi_2(\mathbf{x}_n) & \cdots & \psi_n(\mathbf{x}_n) \end{vmatrix} \tag{2.19}$$

where $\mathbf{X}$ are all the coordinates. A simpler way to represent this is with the antisymmetrizing operator, $\hat{\mathcal{A}}$. ψ_i are the molecular orbitals, defined as a linear combination of atomic orbitals (ϕ) in a basis set.

$$\psi_i = \sum_\mu c_{i\mu} \phi_\mu \tag{2.20}$$

All molecular systems in this work are closed shell. In the restricted closed shell framework, we describe one symmetric half of the electronic wavefunction, recognizing the second half is identical with opposite spin. As a result, the Hamiltonian becomes

$$\hat{H} = -\sum_i^{n/2} \frac{\hbar^2}{m} \nabla_i^2 + V_{\mathrm{NN}} - 2\sum_i^{n/2} \sum_j^N \frac{Z_j e}{R_{ij}} + \sum_i^{n/2} \sum_{j \neq i}^{n/2} \frac{e^2}{r_{ij}} \tag{2.21}$$

The construction of basis sets is a complicated topic that will not be discussed here. There exist various basis sets that were developed using different methodologies and are widely used nowadays. The one-electron pieces of eq. (2.17) are handled with the Slater product wavefunction, forming the $\hat{h}$ operator that includes the electron kinetic energy and the nuclei-electron attraction. The two-electron pieces must be approximated. Through the Hartree-Fock (HF) method, [138–140] the two electron interaction is separated into a mean-coulombic repulsion term $\hat{J}$ and an exchange term $\hat{K}$. These operators are best represented as operating on a molecular orbital

$$\hat{J}_j \psi_i(\mathbf{x}_k) = \int \frac{\psi_j^*(\mathbf{x}_l)\psi_j(\mathbf{x}_l)}{||\mathbf{x}_k - \mathbf{x}_l||} d\mathbf{x}_l \psi_i(\mathbf{x}_k) \tag{2.22}$$

and

$$\hat{K}_j \psi_i(\mathbf{x}_k) = \int \frac{\psi_j^*(\mathbf{x}_l)\psi_i(\mathbf{x}_l)}{||\mathbf{x}_k - \mathbf{x}_l||} d\mathbf{x}_l \psi_j(\mathbf{x}_k) \tag{2.23}$$

where $\mathbf{x}_k$ is one electron coordinate, and $\mathbf{x}_l$ is a different electron coordinate. Together, these produce the Fock operator

$$\hat{F}\psi_i = \hat{h}_k\psi_i + \hat{J}\psi_i - \hat{K}\psi_i = \varepsilon_i\psi_i \tag{2.24}$$

which is related to the energy of the molecular orbitals ε. With this operator, the coefficients of the atomic basis functions to constitute molecular orbitals are obtainable.

The matrix representation of these operators is in the basis of atomic orbitals from the basis set; the matrix elements of $\hat{F}$ is provided by

$$F_{\mu\nu} = h_{\mu\nu} + \sum_{\lambda}^{\mathbb{N}}\sum_{\sigma}^{\mathbb{N}} P_{\lambda\sigma}\left[(\mu\nu|\lambda\sigma) - \frac{1}{2}(\mu\lambda|\nu\sigma)\right] \tag{2.25}$$

where

$$h_{\mu\nu} = \int \phi_\mu^*(\mathbf{x}_i)\hat{h}\phi_\nu(\mathbf{x}_i)d\mathbf{x}_i \tag{2.26}$$

$$(\mu\nu|\lambda\sigma) = \int \frac{\phi_\mu^*(\mathbf{x}_i)\phi_\nu(\mathbf{x}_i)\phi_\lambda^*(\mathbf{x}_j)\phi_\sigma(\mathbf{x}_j)}{||\mathbf{x}_i - \mathbf{x}_j||}d\mathbf{x}_i d\mathbf{x}_j \tag{2.27}$$

and

$$P_{\lambda\sigma} = 2\sum_{i}^{n/2} c_{i\lambda}c_{i\sigma} \tag{2.28}$$

The summations in eq. (2.25) index the atomic basis functions in the basis set and go up the basis set size, $\mathbb{N}$. This matrix satisfies the Roothann-Hall matrix equation [141, 142]

$$\mathbf{FC} = \varepsilon\mathbf{SC} \tag{2.29}$$

where $\mathbf{C}$ are the coefficients of molecular orbitals and $\mathbf{S}$ is the overlap matrix,

$$S_{\mu\nu} = \int \phi_\mu^*(\mathbf{x}_i)\phi_\nu(\mathbf{x}_i)d\mathbf{x}_i \tag{2.30}$$

Eq. (2.29) must be solved iteratively since the elements in $\mathbf{F}$ depend on the coefficients of the molecular orbitals through λ and σ. Once converged, these orbitals can be used to describe the exact HF energy,

$$E_{\text{HF}} = V_{\text{NN}} + \frac{1}{2}\sum_{\mu,\nu} P_{\mu\nu}\left(h_{\mu\nu} + F_{\mu\nu}\right) \tag{2.31}$$

Computing each matrix element of $\hat{F}$ requires $\mathbb{N}^2$ two electron integrals, where $\mathbb{N}$ here is the number of basis functions. This makes the cost of HF formally $\mathcal{O}(\mathbb{N}^4)$.

This mean field method fails to account for electron correlation, which is the way one electron's wavefunction properly responds to another electron's wavefunction. A series of post-Hartree-Fock methods have been established in order to account for electron correlation. Multi-reference and Configuration Interaction (CI) methods, which define the wavefunction as a linear combination of Slater determinants are robust but too expensive for medium sized systems, such as those studied here. Alternatively, variants of perturbation theory are used to capture dynamic correlation, which will be discussed here.

2.1.1 Møller-Plesset Second-Order Perturbation Theory (MP2)

Møller-Plesset Second-Order Perturbation Theory (MP2) [143–147] uses perturbation theory as a simple way to account for dynamic electron correlation. The Hamiltonian is represented as the Fock matrix perturbed to produce the full Hamiltonian,

$$\hat{H}_{\mathrm{MP2}} = \hat{F} + \lambda \hat{H}^{(1)} \tag{2.32}$$

such that $\hat{H}^{(1)}$ is the difference between the exact electron-electron repulsion and the mean-field approximation in the Hartree-Fock method

$$\hat{H}^{(1)} = \sum_{k}^{n} \left[\sum_{j \neq k} \frac{e^2}{r_{k,j}} - \sum_{i}^{n} \left[\hat{J}_i(\mathbf{x}_k) - \frac{1}{2} \hat{K}_i(\mathbf{x}_k) \right] \right] \tag{2.33}$$

This Hamiltonian is processed through perturbation theory. Since $\hat{H}^{(0)} = \hat{F}$, $E^{(0)} = \sum_i \varepsilon_i$, or the sum of the orbital energies, and $\left|\Psi^{(0)}\right\rangle = |\Psi_{\mathrm{HF}}\rangle$. The first-order correction to energy, $E^{(1)}$ comes from the expectation value of the perturbation $\hat{H}^{(1)}$ and is rigorously the difference between the sum of the orbital energies and the Hartree-Fock energy. The second-order correction uses excited states of the Hartree-Fock wavefunction $\left|\Psi_d^{(0)}\right\rangle$, determined solely by moving electrons from a lower energy occupied molecular orbital to higher energy unoccupied molecular orbital. This comes directly from eq. (2.15) and is simplified for the ground state to

$$E_0^{(2)} = -\sum_{d \neq 0} \frac{\left| \left\langle \Psi_0^{(0)} \right| \hat{H}^{(1)} \left| \Psi_d^{(0)} \right\rangle \right|^2}{E_d^{(0)} - E_0^{(0)}} \tag{2.34}$$

where only double excitations are nonzero. When expressed back into the atomic basis functions the correction becomes a summation over two electron integrals

$$E_0^{(2)} = -\frac{1}{4}\sum_{i,j}\sum_{a,b}\frac{(ia|jb) - (ij|ab)}{\varepsilon_a + \varepsilon_b - \varepsilon_i - \varepsilon_j} \tag{2.35}$$

where i and j come from the occupied molecular orbitals and a and b come from virtual molecular orbitals. Note that these are the indices of molecular orbitals. The evaluation of the two electron integrals are over the basis functions, making this method scale approximately as $\mathcal{O}(\mathbb{N}^5)$.

This last term lowers the Hartree-Fock energy to account for electron correlation. This method can be sped up by limiting the number of excited states in the perturbation. Because of the high energy of core molecular orbitals, excited states that have core holes are generally too low in energy to participate in the perturbation and are commonly excluded, forming the frozen core approximation.

This method is commonly used to study the hydrogen bond. [148–150] HF greatly underestimates the binding energy in hydrogen bonded dimers, [151, 152] highlighting the importance of electron correlation in describing the hydrogen bond. The correlation energy is most important in the charge transfer and dispersion contributions to the interaction energy, [25, 153] which manifest most strongly as pairwise interactions. When performing and many-body decomposition of a hydrogen bonded cluster, the majority of electron correlation manifest in the two-body terms, and higher order molecular interactions remain well described by HF. [110] Although MP2 is generally sufficient in describing the hydrogen bond, this method does slightly over-predict the binding energy. [148] Coupled-cluster methods, such as CCSD(T), perform much better but are much more expensive, scaling as $\mathcal{O}(\mathbb{N}^7)$. Therefore, CCSD(T) is not used in this thesis. It is important to note, however, that CCSD(T) is considered the "gold standard" for chemical accuracy. For larger molecular systems, density functional theory (DFT) represents a viable alternative.

2.1.2 Density Functional Theory

DFT [154] tackles the many-electron Schrödinger equation by evaluating the electron density. The components of eq. (2.17) are expressed as functionals of the electron density, i.e. operations on a

function that produce an output. The kinetic energy of an ensemble of non-interacting electrons is

$$T_s[\rho] = \frac{3^{5/3}h^2}{40\pi^{2/3}m_e}\int \rho(\vec{r})^{5/3}d\vec{r} \tag{2.36}$$

and the electron-nuclei interactions are exact

$$V_{eN}[\rho] = \sum_i^N \int \frac{Z_i\rho(\vec{r})}{||\vec{r}_i - \vec{r}||}d\vec{r} \tag{2.37}$$

where the electron density is defined from the molecular orbitals

$$\rho(\vec{r}) = 2\sum_m \psi_m^*(\vec{r})\psi_m(\vec{r}) \tag{2.38}$$

where ψ are the spacial molecular wavefunctions. The electron density is evaluated in three dimensions, given by $\vec{r}$ and doesn't account for each individual electron as expressed by $\mathbf{x}_i$ in the basis set methods.

The coulombic interaction between electrons is exact

$$J[\rho] = \frac{1}{2}\int \frac{\rho(\vec{r}_1)\rho(\vec{r}_2)}{||\vec{r}_1 - \vec{r}_2||}d\vec{r}_1 d\vec{r}_2 \tag{2.39}$$

but the effects of electron exchange cannot be directly captured. This exchange functional is the fundamental problem for DFT, but is represented abstractly as

$$E_{\mathrm{XC}}[\rho] = T[\rho] - T_s[\rho] + V_{ee}[\rho] - J[\rho] \tag{2.40}$$

so that

$$E_{\mathrm{DFT}}[\rho] = T_s[\rho] + V_{eN}[\rho] + J[\rho] + E_{\mathrm{XC}}[\rho] \tag{2.41}$$

Excluding the approximation, the DFT energy is variational with respect to the density. Therefore, the energy can be can be minimized with respect to the electron density to obtain the DFT energy, electron density, and corresponding wavefunction. This is typically done by casting the functionals into the Kohn-Sham equation [155]

$$\left[\hat{h} + v_{\mathrm{XC}}(\mathbf{x}_i) + \int \frac{\rho(\mathbf{x}_j)}{||\mathbf{x}_i - \mathbf{x}_j||}d\mathbf{x}_j\right]\psi_k(\mathbf{x}_i) = \varepsilon_k\psi_k(\mathbf{x}_i) \tag{2.42}$$

in which the matrix elements can be obtained from equations similar to those in the HF method. In the above equation, $\vec{r}$ is replaced with $\mathbf{x}_j$ to cast this equation in the basis of the molecular orbitals.

Therefore, the Kohn-Sham equation can be solved iteratively in ways similar to the HF method and optimized molecular orbitals, which describe electron density, can be obtained. v_{XC} is provided by the functional derivative of the exchange function with respect to density. In the same way, $\hat{h}$ is the functional derivative of $T_{\mathrm{ref}}[\rho] + V_{eN}[\rho]$ and $\int \frac{\rho(\mathbf{x}_j)}{|\mathbf{x}_i - \mathbf{x}_j|} d\mathbf{x}_j$ is the functional derivative of $J[\rho]$, turning the double integral over two densities into an integral over one density. With each iteration of the Kohn-Sham equation, the linear combination of basis functions constructing atomic orbitals are updated and the density is re-evaluated. Once converged, the energy can be found through eq. (2.41). Since traditional DFT eliminates the expensive two-electron exchange integrals in HF, the method scales better, as $\mathcal{O}(\mathbb{N}^3)$.

Approximating $E_{\mathrm{XC}}[\rho]$ is the central question in the development of DFT. Approaches aimed at trying to express $E_{\mathrm{XC}}[\rho]$ have produced the "Jacob's Ladder" of functionals, [156] where improvements in the treatment of electron correlation are hierarchically constructed. The most primitive functional is the local density approximation (LDA), where the exchange-correlation is expressed as

$$E_{\mathrm{XC,LDA}}[\rho] = \frac{3^{4/3}}{4\pi^{1/3}} \int \rho(\vec{r})^{4/3} d\vec{r} \tag{2.43}$$

which attempts to model the two-point density rigorously required to define the exchange functional with a one-point density. [157] Functionals with the generalized gradient approximation (GGA) express the functionals as integrals over a function, F, of both the electron density and the gradient of the electron density.

$$E_{\mathrm{XC,GGA}}[\rho] = \int F(\rho(\vec{r}), \nabla\rho(\vec{r}), \nabla^2\rho(\vec{r})) d\vec{r} \tag{2.44}$$

These functions are diverse and non-descript, resulting in a large variety of functionals. Meta-GGA (mGGA) functionals improve upon GGA by incorporating the kinetic energy density, a function of the gradient of the molecular orbitals

$$\tau(\vec{r}) = \frac{\hbar^2}{2m_e} \sum_i \nabla\phi_i^*(\vec{r}) \cdot \nabla\phi_i(\vec{r}) \tag{2.45}$$

resulting in the functional

$$E_{\mathrm{XC,mGGA}}[\rho] = \int F(\rho(\vec{r}), \nabla\rho(\vec{r}), \nabla^2\rho(\vec{r}), \tau(\vec{r})) d\vec{r} \tag{2.46}$$

The functional form of F is developed and parameterized to best approximate the energy from exchange and correlation. Because of the complexity of these forms, it is common practice to

separate exchange and correlation energies

$$E_{\mathrm{XC}}[\rho] = E_{\mathrm{X}}[\rho] + E_{\mathrm{C}}[\rho] \tag{2.47}$$

and parameterize each functional separately. This produces a large library of density functionals that serve specific roles for the system they were parameterized on.

Many functionals used today are based on a hybrid approach where the exchange functional is a linear combination of other functionals and basis set methods like HF. The HF exchange operator is weighed and added explicitly to the functional and the Kohn-Sham equation. For example, B3LYP, [158] a common functional for studying organic molecules efficiently, is a linear combination of HF, LDA, and GGA functionals. The inclusion of HF exchange increases the the cost and accuracy of the method, since the exchange operator matrix $\hat{K}$ must be evaluated (cost $\mathcal{O}(\mathbb{N}^4)$). Hybrid functionals are not limited to HF exchange, with double hybrid functionals that include portions of MP2 correlation included explicitly at a steeper computational cost ($\mathcal{O}(\mathbb{N}^5)$). [159]

Two additional corrections can be applied to DFT for molecular systems. Exchange functionals generally exhibit improper long-range decay. To rectify this, the HF mixing in the exchange functionals becomes distance dependent where DFT's exchange correlation is dominant for short-range interactions while HF's exchange dominates for long-range interactions. This is called a range separated functional. [160] Furthermore, pure DFT cannot describe dispersion, the coherent and attractive fluctuations in electron distribution; basis set methods manage to describe dispersion adequately, albeit incompletely. To handle this, atom-atom distances are used to estimate the missing dispersion from DFT. [161, 162] C coefficients are parameterized for each functional to describe what is missing form the DFT and produce a proper answer. Dispersion corrections follow the general form

$$E_{\mathrm{disp}} = -s \sum_{i}^{N} \sum_{j>i}^{N} \frac{\sqrt{C_i C_j}}{R_{ij}^6} f(R_{ij}) \tag{2.48}$$

where the damping function (f) limits dispersion at close range. Implementations of dispersion corrections differ through the damping function.

In this thesis, the ωB97X-D functional was heavily used in the calculations. [163] This functional is an extension of the B97 functional [164] that incorporates long-range corrections and explicit dispersion corrections, which proves appropriate for describing the hydrogen bond. B97 is a hybrid-

GGA that separates exchange and correlation includes Hartree-Fock mixing.

$$E_{\mathrm{XC}}(\rho) = c_{\mathrm{HF}} E_{\mathrm{HF}} + (1 - c_{\mathrm{HF}}) E_{\mathrm{X}}[\rho] + E_{\mathrm{C}}[\rho] \tag{2.49}$$

Detailed descriptions of these functional forms can be found in the original work. [164] Chai and Head-Gordon [165] extended this functional through a gradual change in Hartree exchange. These ω versions of the functional, ωB97 and ωB97X, represent the full exchange functional as

$$E_{\mathrm{XC}} = E_{\mathrm{LR-HF}} + c_{\mathrm{HF}} E_{\mathrm{SR-HF}} + (1 - c_{\mathrm{HB}}) E_{\mathrm{X}} + E_{\mathrm{C}} \tag{2.50}$$

where the Hartree-Fock exchange operator, $\hat{K}$ is scaled up via an error function to produce $E_{\mathrm{LR-HF}}$ and $E_{\mathrm{SR-HF}}$. This function turns on long-range Hartree-Fock exchange at long range and dampens the density functional exchange in suite via the function

$$F(a) = 1 - \frac{8}{3} a \left[\sqrt{\pi}\ \mathrm{erf}(1/2\mathrm{a}) - 3\mathrm{a} + 4\mathrm{a}^3 + (2\mathrm{a} - 4\mathrm{a}^3)\mathrm{e}^{-1/4\mathrm{a}^2} \right] \tag{2.51}$$

where a is a parameter that depends on the density $\rho(\vec{r})$. In ωB97 the coefficient c_{HF} of the short range Hartree-Fock exchange is zero, making the short range asymptote for HF exchange zero. In ωB97X, this short range asymptote for HF mixing is described by the coefficient c_{HF}. This treatment of Hartree-Fock exchange proved sufficient in describing thermodynamic properties. [165] Chai and Head-Gordon [163] later re-parameterized this functional with the addition of dispersion corrections. This dispersion correction follows eq. (2.48) with a damping function

$$f(R_{ij}) = \frac{1}{1 + a(R_{ij}^{(0)}/R_{ij})^{12}} \tag{2.52}$$

where $R_{ij}^{(0)}$ is the sum of the van der Waal radii of atoms i and j. This produced the functional ωB97X-D that describes the important aspects of hydrogen bonding, including correlation through a GGA functional, exchange through a hybrid-GGA functional with Hartree-Fock mixing increasing from 22% to 100% with long-range corrections, and dispersion added empirically during the parameterization procedure.

2.2 The Polyatomic Vibrational Hamiltonian

Atoms similarly follow quantum mechanics and satisfy the Schrödinger equation. The full molecular Schrödinger equation includes a kinetic energy operator for all nuclei and wavefunctions that span

the nuclear coordinate space. However since nuclei are much more massive than electrons, the Born-Oppenheimer approximation [137] can be applied to separate nuclear and electronic degrees of freedom. In that vein, the electronic Schrödinger equation, eq. (2.17), can be solved independently for each nuclear arrangement. The leftover pieces are the nuclear Schrödinger equation

$$\hat{H} = -\sum_{i}^{N} \frac{\hbar^2}{2m_i} \nabla_i^2 + \hat{V}(\vec{x}) \tag{2.53}$$

where $\vec{x}$ are the Cartesian coordinates of each nuclei and $\hat{V}(\vec{x})$ is the solution to the electronic Schrödinger equation at coordinates $\vec{q}$

$$\hat{V}(\vec{x}) = \langle \Psi_{0,\text{elec}} | \, \hat{H}_{\text{elec}}(\vec{x}) \, | \Psi_{0,\text{elec}} \rangle \tag{2.54}$$

In the one dimensional case, eq. (2.53) is easily solvable through methods discussed earlier in this chapter. However, molecules become multidimensional very fast ($\sim$ 3 dimensions per atom) and basis set methods become difficult. In this section, we will discuss a few methods that solve the vibrational Schrödinger equation in polyatomic molecules.

In the vibration structure, the interaction of light and matter is fundamental in describing the spectra. In it's most general form, light interacts with mater through the time-dependent Schrödinger equation, eq. (2.1). For weak fields, this is handled through time-dependent perturbation theory, where the time dependent nature of the electric field is expressed as the first order perturbation.

$$\hat{H}^{(1)}(t) = 2\hat{H}^{(1)} \cos \omega t = -\vec{\mu} \cdot \vec{\text{E}}(e^{-i\omega t} + e^{-i\omega t}) \tag{2.55}$$

where the dipole moment $\vec{\mu}$ is dotted into the electric field of light $\vec{\text{E}}$. The zeroth order perturbation is the time independent vibrational Hamiltonian, yielding a basis $\left|\Psi^{(0)}\right\rangle$ such that the time-dependent wavefunction can be represented as the linear combination of basis functions

$$|\Psi(t)\rangle = \sum_{n} c_n(t) \Psi_n^{(0)} e^{-iE_n^{(0)} t/\hbar} \tag{2.56}$$

where the coefficients oscillate with respect to time due to the perturbation is the oscillation of each state due to the time dependent Schrödinger equation. Assuming that the molecular system starts fully in the ground state and changes in short time, as is true for systems considered in this thesis, the coefficients can be found as

$$c_f(t) = \frac{1}{i\hbar} \int_0^t H_{f,i}^{(1)}(t) e^{-i\omega_{f,i} t} dt \tag{2.57}$$

The probability of the state is the square of this coefficient and found to be, after substitution of the perturbation

$$P_f(t) = \frac{4\left|\left\langle \psi_f^{(0)}\right| \vec{\mu}\cdot\vec{\mathrm{E}}\left|\psi_i^{(0)}\right\rangle\right|^2}{\hbar^2}\frac{\sin^2(\frac{1}{2}(\omega_{f,i}-\omega)t)}{(\omega_{f,i}-\omega)^2} \tag{2.58}$$

The $\frac{\sin^2(\mathrm{x})}{\mathrm{x}^2}$ function to the right shows that only frequencies near that of the eigenstate transition will yield a probability of excitation. Due to the random distribution of molecular orientations in a liquid or gas phased system, the matrix elements simplify to $\frac{1}{3}|\mu_{f,i}|^2\mathrm{E}$, where $\mu_{f,i}$ is the transition dipole moment matrix element. The radiation is a function of all the frequencies, so the probability of transfer should integrate over all nearby frequencies that can cause the transition. This makes the transition rate

$$W_{i\to f} = \frac{1}{6\varepsilon_0\hbar^2}|\mu_{f,i}|^2\rho_R(\nu_{f,i}) \tag{2.59}$$

where $\nu = \omega/2\pi$. This prefactor on the energy density of states, $\rho(\nu_{f,i})$, is Einstein's coefficient for stimulated absorption.

$$B = \frac{|\mu_{f,i}|^2}{6\varepsilon_0\hbar^2} \tag{2.60}$$

Absorption spectra measure the amount of absorbed light, i.e. the decrease in light that passes through a sample. From the last paragraph, the change of energy density is

$$\frac{d\mathcal{U}}{dt} = -n(\nu)h\nu B\rho_R(\nu)d\nu \tag{2.61}$$

where $n(\nu)d\nu$ is the number density of molecules able to absorb light. This quantity is related to the change in intensity (dI)

$$dI = -\frac{n(\nu)h\nu}{c}BIdl = -\varepsilon(\nu)[\mathrm{A}]IdL \tag{2.62}$$

where $\rho_R = I/c$, ε is the molar absorption coefficient, and [A] is the concentration of absorbing molecules. Integrating with respect to $d\nu$ yields

$$\mathcal{A} = \frac{h\nu_{f,i}N_A}{c}B = \frac{\pi N_A}{3\varepsilon_0\hbar c}\nu_{j,i}|\mu_{j,i}|^2 \tag{2.63}$$

where $\mathcal{A}$ is the integrated absorption coefficient $\int\varepsilon(\nu)d\nu$. This shows that the absorption coefficients are related to the frequency of the transition and the transition dipole moment matrix element. For molecular vibrations, the transition dipole moment operator is the change in molecular dipole with respect to the nuclear coordinates.

Although this derivation comes directly from the absorption of light, many experiments, including those noted in this thesis, use action spectroscopy to detect light absorption for dilute gas phase molecules. The absorption of light is assumed to map evenly to the action detected (such as tag removal) across the range of frequencies observed. This assumption is carefully considered when designing the experiments and analyzing the results.

2.2.1 The N-Dimensional Harmonic Oscillator

The quintessential potential for vibrations is the harmonic oscillator. Covalent bonds are well approximated by a parabolic potential with a Hamiltonian

$$\hat{H} = -\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + \frac{\omega^2}{2m}\hat{x}^2 \tag{2.64}$$

where ω is the frequency of the motion in coordinate x and m is the mass of the particle moving in coordinate x such that the second derivative of the potential, k, is $\frac{\omega^2}{2m}$. The energy of each resulting state are $E_n = \hbar\omega(n + \frac{1}{2})$ and the wavefunctions are Hermite polynomials multiplied by a Gaussian. Matrix elements of polynomial operators of position and momentum are analytic, making computations easy.

Furthermore, the N-dimensional harmonic oscillator is separable and therefore solvable as N harmonic oscillators. If the Hamiltonian is a coupled harmonic oscillator,

$$\hat{H} = -\sum_{i}^{3N}\frac{\hbar^2}{2m_i}\frac{\partial^2}{\partial {x_i}^2} + \frac{1}{2}\sum_{i}^{3N}\sum_{j}^{3N}\frac{\partial^2 V}{\partial x_i \partial x_j}\hat{x}_i\hat{x}_j \tag{2.65}$$

it can always be expressed as separable uncoupled harmonic oscillators through a change in variables.

$$\hat{H} = -\sum_{i}^{3N-6}\frac{\hbar^2}{2}\frac{\partial^2}{\partial q_i^2} + \frac{1}{2}\sum_{i}^{3N-6}\frac{\partial^2 V}{\partial q_i^2}\hat{q}_i^2 \tag{2.66}$$

where q_i are the normal modes represented in as a linear combination of mass-weighted cartesian coordinates. The six coordinates associated with translations and rotations are projected out since they have zero potential and are not proper vibrations. In practice, this is done through the Hessian, comprised from all second derivatives of the potential in the mass-weighted Cartesian coordinates. The eigenvalues are the squares of the frequencies of each oscillator, and the eigenstates describe the decoupled coordinate, or normal mode, in it's Cartesian components. In this formulation, separation

of variables produces an easily solvable differential equation with a Hartree product wavefunction. The vibrational modes are distinguishable following Bose-Einstein statistics and thus represented by a symmetric wavefunction with respect to the exchange of any two particles

$$|\Psi\rangle = \prod_{i}^{3N-6} |\psi_i\rangle \tag{2.67}$$

This coordinate transformation can similarly be applied to necessary operators, such as the transition dipole moment operator generally constructed as a Taylor expansion of the molecular dipole. The matrix elements of the transition dipole are now simple. Since the dominant term in a transition dipole moment operator is usually linear, and quadratic terms cannot contribute to a fundamental harmonic transition, the dipole moment matrix elements is easily expressed as

$$\mu_{0,1_i} = \frac{\partial \vec{\mu}}{\partial q_i} \sqrt{\frac{\hbar}{2 m_i \omega_i}} \tag{2.68}$$

making the integrated absorption coefficient

$$\mathcal{A}_{0 \to 1_i} = \frac{N_A}{12 \varepsilon_0 c m_i} \left| \frac{\partial \vec{\mu}}{\partial q_i} \right|^2 \tag{2.69}$$

where 1_i indicates the fundamental transition with respect to mode i.

Although this solution is easy and sufficient for most applications, it fails to capture significant portions of vibrational structure stemming from the potential's anharmonicity. A true covalent bond is better modeled via a Morse potential, [60, 166, 167] which has non-negligible higher order derivatives. This makes the harmonic frequencies, on average, $\sim 5\%$ larger than their experimental values. This can be handled by scaling or shifting the frequencies. [168] Additionally, multi-quanta transitions, such as overtones and combination bands, can carry measurable intensity due to anharmonic effects that are not allowed under the harmonic approximation. For that reason, large efforts have been made to expand vibrational methods to describe anharmonicity in polyatomic systems.

2.2.2 Vibrational Perturbation Theory

Since the harmonic oscillator is an easily solvable problem and a good approximation of most vibrational potentials, it can be used as a good zeroth order Hamiltonian for perturbation. [169, 170] The full molecular Hamiltonian in Cartesian normal modes includes both anharmonic corrections

to the wavefunction and angular momentum couplings. [171] Due to the ease to compute matrix elements of polynomial operators with harmonic oscillator wavefunctions, the first- and second-order perturbative terms of the Hamiltonian encompass cubic and quartic functions of the position and momentum operators. These expressions have been worked out before [170, 172]

$$\hat{H}^{(1)} = \frac{1}{6}\sum_{i,j,k}^{3N-6} \frac{\partial^3 V}{\partial q_i \partial q_j \partial q_k} \hat{q}_i \hat{q}_j \hat{q}_k \tag{2.70}$$

$$\hat{H}^{(2)} = \frac{1}{24}\sum_{i,j,k,l}^{3N-6} \frac{\partial^4 V}{\partial q_i \partial q_j \partial q_k \partial q_l} \hat{q}_i \hat{q}_j \hat{q}_k \hat{q}_l + \sum_{\alpha} B_{\alpha}^{e} \sum_{i,j,k,l}^{3N-6} \zeta_{i,j}^{\alpha} \zeta_{k,l}^{\alpha} \hat{q}_i \hat{p}_j \hat{q}_k \hat{p}_l \tag{2.71}$$

where B_{α}^{e} is the rotational constant in rotational axis α, and $\zeta_{i,j}^{\alpha}$ is the Coriolis coupling constant in rotational axis α between normal modes i, j, k and l. A large portion of these perturbations, especially for bond stretches that are the focus of this thesis, originate from higher order derivatives of the potential around the minimum. Although second derivatives of most ab initio surfaces are available analytically in most electronic structure packages [173, 174], higher-order derivatives are not. A 3-point finite difference of the the Hessian is used to find all third-order derivatives and most fourth-order derivatives, meaning this method scales $\mathcal{O}(n^{2+l})$, where n is the number of atoms and l is the scaling in computing Hessian with the underlying electronic structure method with respect to the number of atoms. This perturbation is appropriate if the Taylor expansion converges over the range of the normal mode vibration and if the higher order derivatives are small compared to the second derivative. With this perturbed Hamiltonian, the vibrational energies and transition dipole moment matrix elements can be easily solved.

The first-order corrections to energy and wavefunctions were described abstractly prior and can be solved in this context here. Since the expectation values of the first order correction are zero, there is no first-order correction to the energy. Corrections to the wavefunction, however, are non-zero. For each wavefunction with mode $\mathcal{N}$, described by a set of 1-D quantum numbers, the correction is given by

$$\left|\Psi_{\mathcal{N}}^{(1)}\right\rangle = \frac{1}{6}\sum_{i,j,k}^{3N-6} \frac{\partial^3 V}{\partial q_i \partial q_j \partial q_k} \left[\sum_{\mathcal{M}\neq\mathcal{N}} \frac{\left\langle \Psi_{\mathcal{M}}^{(0)} \right| \hat{q}_i \hat{q}_j \hat{q}_j \left| \Psi_{\mathcal{N}}^{(0)} \right\rangle}{E_{\mathcal{N}}^{(0)} - E_{\mathcal{M}}^{(0)}} \left| \Psi_{\mathcal{M}}^{(0)} \right\rangle \right] \tag{2.72}$$

where $\mathcal{M}$ are all other sets of quantum numbers. Due to the selection rules of the integral of harmonic wavefunctions and position operators, only a few states can perturb the wavefunction,

specifically those that differ by one or three quanta from the original state. The expressions of these terms are analytic but cumbersome to fully write out.

The first correction to the energy appears at second-order. Due to selection rules, the summation over quartic terms can be simplified

$$\begin{aligned} E_{\mathcal{N}}^{(2)} =& \frac{1}{4}\sum_{i,j}^{3N-6}\frac{\partial^4 V}{\partial q_i^2 \partial q_j^2}\left\langle \Psi_{\mathcal{N}}^{(0)}\right| \hat{q}_i^2\hat{q}_j^2 \left|\Psi_{\mathcal{N}}^{(0)}\right\rangle + \sum_{\alpha} B_{\alpha}^{e}\sum_{i,j,k,l}^{3N-6}\zeta_{i,j}^{\alpha}\zeta_{k,l}^{\alpha}\left\langle \Psi_{\mathcal{N}}^{(0)}\right| \hat{q}_i\hat{p}_j\hat{q}_k\hat{p}_l \left|\Psi_{\mathcal{N}}^{(0)}\right\rangle \\ &+ \frac{1}{6}\sum_{i,j,k}^{3N-6}\frac{\partial^3 V}{\partial q_i \partial q_j \partial q_k}\left[\sum_{\mathcal{M}\neq\mathcal{N}}\frac{\left\langle \Psi_{\mathcal{M}}^{(0)}\right| \hat{q}_i\hat{q}_j\hat{q}_j \left|\Psi_{\mathcal{N}}^{(0)}\right\rangle^2}{E_{\mathcal{N}}^{(0)} - E_{\mathcal{M}}^{(0)}}\right] \end{aligned} \tag{2.73}$$

Only select fourth-order derivatives are required to compute the VPT2 energies. Second-order corrections to the wavefunctions are also important

$$\begin{aligned} \left|\Psi_{\mathcal{N}}^{(2)}\right\rangle =& \frac{1}{24}\sum_{i,j,k,l}^{3N-6}\frac{\partial^4 V}{\partial q_i \partial q_j \partial q_k \partial q_l}\left[\sum_{\mathcal{L}\neq\mathcal{N}}\frac{\left\langle \Psi_{\mathcal{L}}^{(0)}\right| \hat{q}_i\hat{q}_j\hat{q}_k\hat{q}_l \left|\Psi_{\mathcal{N}}^{(0)}\right\rangle}{E_{\mathcal{N}}^{(0)} - E_{\mathcal{L}}^{(0)}}\left|\Psi_{\mathcal{L}}^{(0)}\right\rangle\right] \\ &+ \sum_{\alpha} B_{\alpha}^{e}\sum_{i,j,k,l}^{3N-6}\zeta_{i,j}^{\alpha}\zeta_{k,l}^{\alpha}\left[\sum_{\mathcal{L}\neq\mathcal{N}}\frac{\left\langle \Psi_{\mathcal{L}}^{(0)}\right| \hat{q}_i\hat{p}_j\hat{q}_k\hat{p}_l \left|\Psi_{\mathcal{N}}^{(0)}\right\rangle}{E_{\mathcal{N}}^{(0)} - E_{\mathcal{L}}^{(0)}}\left|\Psi_{\mathcal{L}}^{(0)}\right\rangle\right] \\ &+ \frac{1}{36}\sum_{i,j,k}^{3N-6}\left(\frac{\partial^3 V}{\partial q_i \partial q_j \partial q_k}\right)^2\left[\sum_{\mathcal{L}\neq\mathcal{N}}\sum_{\mathcal{M}\neq\mathcal{L}\neq\mathcal{N}}\frac{\left\langle \Psi_{\mathcal{L}}^{(0)}\right| \hat{q}_i\hat{q}_j\hat{q}_k \left|\Psi_{\mathcal{M}}^{(0)}\right\rangle\left\langle \Psi_{\mathcal{M}}^{(0)}\right| \hat{q}_i\hat{q}_j\hat{q}_k \left|\Psi_{\mathcal{N}}^{(0)}\right\rangle}{\left(E_{\mathcal{N}}^{(0)} - E_{\mathcal{L}}^{(0)}\right)\left(E_{\mathcal{N}}^{(0)} - E_{\mathcal{M}}^{(0)}\right)}\left|\Psi_{\mathcal{L}}^{(0)}\right\rangle\right] \\ &- \frac{1}{2}\left\langle \Psi_{\mathcal{N}}^{(1)}\right| \left|\Psi_{\mathcal{N}}^{(1)}\right\rangle \left|\Psi_{\mathcal{N}}^{(0)}\right\rangle \end{aligned} \tag{2.74}$$

where that last term serves to normalize the correction to the wavefunction. [169] It is common to only include fourth derivatives over three coordinates since fully coupled fourth derivatives are not obtainable from a 3-point finite difference of the Hessian.

Expressions of the wavefunction are important in computing the intensity of anharmonic transitions. The transition dipole moment operator is commonly expanded to second-order in the perturbation expansion,

$$\hat{\mu} = \hat{\mu}^{(0)} + \lambda\hat{\mu}^{(1)} + \lambda^2\hat{\mu}^{(2)} \tag{2.75}$$

where the terms in the perturbation are come from a Taylor expansion so that

$$\hat{\mu}^{(0)} = \mu(0) + \sum_{i}^{3N-6}\left.\frac{\partial\vec{\mu}}{\partial q_i}\right|_{q_i=0}\hat{q}_i \tag{2.76}$$

$$\hat{\mu}^{(1)} = \frac{1}{2}\sum_{i,j}^{3N-6} \left.\frac{\partial^2 \vec{\mu}}{\partial q_i \partial q_j}\right|_{q_i=0,q_j=0} \hat{q}_i\hat{q}_j \tag{2.77}$$

and

$$\hat{\mu}^{(2)} = \frac{1}{6}\sum_{i,j,k}^{3N-6} \left.\frac{\partial^2 \vec{\mu}}{\partial q_i \partial q_j \partial q_k}\right|_{q_i=0,q_j=0,q_k=0} \hat{q}_i\hat{q}_j\hat{q}_k \tag{2.78}$$

The total transition dipole moment matrix element associated with a transition from the ground state to $\mathcal{M}$ is found through matrix elements of these operators with the corrections to the wavefunctions. The zeroth order transition dipole moment matrix element

$$\mu^{(0)}_{0\to\mathcal{F}} = \left\langle \Psi^{(0)}_{\mathcal{F}} \right| \hat{\mu}^{(0)} \left| \Psi^{(0)}_0 \right\rangle \tag{2.79}$$

corresponds to the harmonic approximation. First-order corrections to the transition dipole moment matrix element incorporate higher-order corrections in either the wavefunction or transition dipole operator.

$$\mu^{(1)}_{0\to\mathcal{F}} = \left\langle \Psi^{(1)}_{\mathcal{F}} \right| \hat{\mu}^{(0)} \left| \Psi^{(0)}_0 \right\rangle + \left\langle \Psi^{(0)}_{\mathcal{F}} \right| \hat{\mu}^{(0)} \left| \Psi^{(1)}_0 \right\rangle + \left\langle \Psi^{(0)}_{\mathcal{F}} \right| \hat{\mu}^{(1)} \left| \Psi^{(0)}_0 \right\rangle \tag{2.80}$$

Expressions for $\left|\Psi^{(1)}\right\rangle$ from eq. (2.72) are substituted into the above equation to yield the matrix elements analytically. Because of the selection rules of the position operator, only a few states of $\left|\Psi^{(0)}\right\rangle$ can produce intensity. For example, if one considers only a fundamental transition, which are the ones mostly studied in this work, all terms become zero. These corrections are most important for the overtone and combination transitions, which where are not studied in this thesis.

A similar expression for the second-order corrections to the transition dipole moment matrix element can be formulated as

$$\begin{aligned}\mu^{(2)}_{0\to\mathcal{F}} = &\left\langle \Psi^{(2)}_{\mathcal{F}} \right| \hat{\mu}^{(0)} \left| \Psi^{(0)}_0 \right\rangle + \left\langle \Psi^{(0)}_{\mathcal{F}} \right| \hat{\mu}^{(0)} \left| \Psi^{(2)}_0 \right\rangle + \left\langle \Psi^{(1)}_{\mathcal{F}} \right| \hat{\mu}^{(0)} \left| \Psi^{(1)}_0 \right\rangle \\ &+ \left\langle \Psi^{(1)}_{\mathcal{F}} \right| \hat{\mu}^{(1)} \left| \Psi^{(0)}_0 \right\rangle + \left\langle \Psi^{(0)}_{\mathcal{F}} \right| \hat{\mu}^{(1)} \left| \Psi^{(1)}_0 \right\rangle + \left\langle \Psi^{(0)}_{\mathcal{F}} \right| \hat{\mu}^{(2)} \left| \Psi^{(0)}_0 \right\rangle\end{aligned} \tag{2.81}$$

where several integrals can be non-zero for fundamental transitions and the intensity is modulated.

When the potential is well behaved, i.e. perturbations in the potential are small, this method works well. Perturbations in the energy and wavefunction become large when vibrational states are degenerate or near degenerate. This is because $E^{(0)}_{\mathcal{N}} - E^{(0)}_{\mathcal{M}} \approx 0$, making the corrections blow up. This can be done by checking the term in $E^{(2)}$ or $\left|\Psi^{(k)}\right\rangle$. [175, 176] The terms that cause large perturbations are projected out, solved for through a variational method, and reintegrated into the

analysis. This method, called degenerate perturbation theory, is the common implementation for VPT2 for polyatomic molecules. [173]

Degenerate perturbation theory is generally a good approximation in describing neutral hydrogen bonds. Anharmonicity is very important in accurately describing the experimental frequency of covalent bonds. Experimental frequencies are approximated from harmonic frequencies through scaling factors. However, anharmonicity generally increases with hydrogen bond strength so a proper description of the anharmonicity through perturbation theory is useful. Furthermore, hydrogen bond frequencies can be strongly perturbed by overtones in the bend frequencies, so resonance handling is important to include in order to describe experimental spectra. In charged systems, hydrogen bond stretches can become too anharmonic to be described through perturbation theory. In such systems, basis set methods are required to properly describe the vibrational states.

Chapter 3

EXTENDING BADGER'S RULE: I. THE RELATIONSHIP BETWEEN ENERGY AND STRUCTURE IN HYDROGEN BONDS [177]

3.1 Introduction

The hydrogen bond energy in simple gas phase hydrogen bonded dimers is defined by the binding energy or the energetic difference between the dimer and the two isolated monomers. The binding energies for some simple hydrogen-bonded dimers have been experimentally measured using sophisticated spectroscopic techniques and velocity map imaging [101, 178–181] and reported theoretically from first principles electronic structure calculations. [98, 99, 182] However, as systems become more complex, techniques to measure their binding energy and relate it to that of the underlying hydrogen bonds become less available. For example, doubly hydrogen-bonded complexes cannot have their total binding energies split equally among each hydrogen bond when no symmetry is present, while methylated species cannot separate the contribution of non-hydrogen bonded van-der-Waals dispersion interactions to the binding energy from those local to the hydrogen bond. Therefore, methods for measuring hydrogen bond strengths that focus on the individual hydrogen bonds are necessary.

Earlier work by Richard Badger in the early 20th century pioneered the correlation between local properties of covalent and hydrogen-bonded systems. [183–185] The so-called "Badger's rule", [183] relates the bond force constant to the covalent bond lengths, and has been extensively used by others in the field as a probe to describe chemical bonds via their spectroscopic signatures. [97, 114, 186–188] This relationship is robust, being extended to both polyatomic monomers and hydrogen-bonded clusters. [97, 114, 184] Subsequent reports by Badger suggested a linear relationship between the hydrogen bond enthalpies and the changes in the respective vibrational frequencies, called the "Badger-Bauer" rule. [185] This empirical relationship has been expanded, quantified, and utilized by chemists to estimate the enthalpies of hydrogen bonds using either the aforementioned linear or a square root relationship. [189, 190]

There is clearly a correlation between the strength of a hydrogen bond and the geometrical change (elongation) in the covalent DH bond that participates in it, which, in turn, induces a shift in the observed vibrational frequency. We will call this the *energetic–structural–spectral* correspondence of hydrogen bonds. Previous studies have linked Badger's rule to physical constants at the molecular level, creating a physically motivated *structural–spectral* correspondence. [114] By expressing the hydrogen bond interaction as a dipole interacting in an electric field, [114, 191] it was found that the frequency shift of the hydrogen bonds in water clusters depends on the force constant and anharmonicity of the participating covalent bond and the ratio of the second and first derivatives of the dipole moment. [114] Additional interactions, such as charge transfer and Pauli repulsion were not incorporated into these models but were not required to produce a predictive relationship that transferred well to a robust test set of aqueous clusters.

In the current study we develop the second piece of this correspondence, namely the *energetic–structural* correspondence, using several techniques employed in these earlier works. [114, 191] Note that neither the energy expression nor utilized interaction energy is a full interaction potential that spans the full dimensional coordinate space. Neither can be used in molecular dynamics simulations and, as such, it is not in competition with already developed potential energy functions that can compute the energy and gradient. [61, 62] The relationship rather focuses on a subset of the intra- and intermolecular coordinates, namely the elongation of the covalent bond of the donor fragment participating in the hydrogen bond (Δr_{HB}) and the inter-molecular coordinates R (separation between the heavy atoms) and θ (deviation of the hydrogen bond from linearity). It includes both electrostatics and electronic effects that are introduced and evaluated to incorporate and separate the classical and non-classical effects that contribute to the hydrogen bond energies. Six hydrogen-bonded dimers comprising of H_2O, HF, and NH_3 monomers and their combinations are used as models of systems that are bonded exclusively by hydrogen bonds. The physically motivated relationship using the newly derived energy expression is developed and tested against a robust set of dimer geometries to determine its performance and/or its limitations.

This chapter will outline the computational methodology used in section 3.2, viz. the choice of the hydrogen bond potential and the details of the scans of the Potential Energy Surfaces (PESs) of the hydrogen bonded dimers. In section 3.3 we present the results consisting of the fitting parameters of the potential using different expansions of the moments, derive the relationship

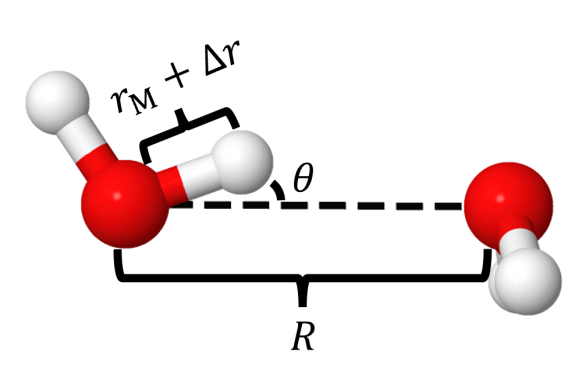

Figure 3.1: The three coordinates used in this work, illustrated for the water dimer. They include the heavy atom distance (R), the hydrogen bond angle (θ), and the hydrogen bond elongation (Δr). r_{M} is the covalent bond distance in the isolated monomer, used to define the bond elongation.

between energy and structure and further validate the model for non-linear hydrogen bonded dimer configurations. Final conclusions and an outlook of next efforts are contained in section 3.4.

3.2 Computational Methodology

3.2.1 The Hydrogen Bond Potential

The hydrogen bond expression, V_{HB}, is separated into the deformation (V_{Def}) and interaction (V_{Int}) components,

$$V_{\mathrm{HB}}(q_{\mathrm{intra}}, q_{\mathrm{inter}}) = V_{\mathrm{Def}}(q_{\mathrm{intra}}) + V_{\mathrm{Int}}(q_{\mathrm{intra}}, q_{\mathrm{inter}}) \tag{3.1}$$

where V_{Def} is a deformation energy term of the donor (D) molecule arising from the elongation of the covalent DH bond due to its participation in the hydrogen bond, V_{Int} is an interaction energy term, q_{intra} and q_{inter} are selected intra– and inter–molecular degrees of freedom.

The deformation energy (V_{Def}) is a function of just the intra–molecular coordinates of the donor molecule. Since the most consequential deformation is along the covalent bond of the donating hydrogen, it can be simplified as a single Morse potential [166] (V_{M}) along that coordinate

$$V_{\mathrm{Def}}(q_{\mathrm{intra}}) = V_{\mathrm{M}}(\Delta r) = D_{\mathrm{M}}(z-1)^2 = D_{\mathrm{M}}(e^{-\alpha\Delta r} - 1)^2 \tag{3.2}$$

where D_{M} is the dissociation energy of the covalent bond, z is the exponential of the Morse oscillator ($z = e^{-\alpha\Delta r}$), α is the range factor that relates to the dissociation energy D_{M} and the bond force

constant k_{M} via $k_{\mathrm{M}} = 2D_{\mathrm{M}}\alpha^2$, and Δr is the difference between the covalent bond distance (r) and its equilibrium distance (r_{M}) in the monomer, viz. $\Delta r{=}r - r_{\mathrm{M}}$. This coordinate ($\Delta r$) is shown in figure 3.1.

The inter–molecular component (V_{Int}) describes the interaction between the hydrogen bond donor and acceptor fragments. While this depends on all degrees of freedom of the dimer, it can be simplified into a form containing a subset of these coordinates that are most sensitive to the interaction. For the hydrogen-bonded systems considered in this work, we will in particular focus on one intra–molecular coordinate Δr (covalent bond participating in the hydrogen bonding interaction), and two inter–molecular coordinates R (separation between the heavy atoms) and θ (deviation of the hydrogen bond from linearity), as depicted in figure 3.1. This results in an interaction potential $V_{\mathrm{Int}}(\Delta r, R, \theta)$.

The energy expression is further extended in order to instill physics into the model. It incorporates the Pauli repulsive and the electrostatic attractive terms via

$$V_{\mathrm{Int.}}(\Delta r, R, \theta) = Ae^{-\beta(R-\gamma\Delta r\cos\theta)} - \frac{\varepsilon\mu(\Delta r)\cos\theta}{(R-d)^n}\left(1 + \frac{\varepsilon a(\Delta r)\cos\theta}{(R-d)^n}\right) \tag{3.3}$$

where A is the strength of the Pauli repulsion, β is the rate of the onset of the repulsion, γ describes how the repulsion varies via orbital deformation as the hydrogen atom moves across the hydrogen bond, ε describes the field strength created by the acceptor molecule, μ is the dipole moment on the donor molecule, a is the component of the polarizability tensor collinear to the dipole moment μ of the donor molecule, d describes the offset of the dipole origins from the heavy atoms on the donor and acceptor (see figure 3.2), and n is the order of decay of the electrostatic interaction, determined by the type of interaction. For example, $n{=}3$ for the dipole-dipole type interactions in neutral hydrogen-bonded systems, which are the focus of this study. Note that the expression in eq. (3.3) takes the form of a generalized Buckingham potential [58, 60] with respect to R with slight modifications. The intramolecular dependence in Δr is included to properly describe the coupling between the two degrees of freedom, Δr and R, with $\cos\theta$ modulating that coupling and scaling the electrostatic attraction. Physical parameters were used to more precisely attribute the attractive and repulsive terms in the generalized Buckingham potential to physical interactions. Finally, μ and a are functions of the intramolecular geometry and are Taylor expanded to various degrees to fit the interaction.

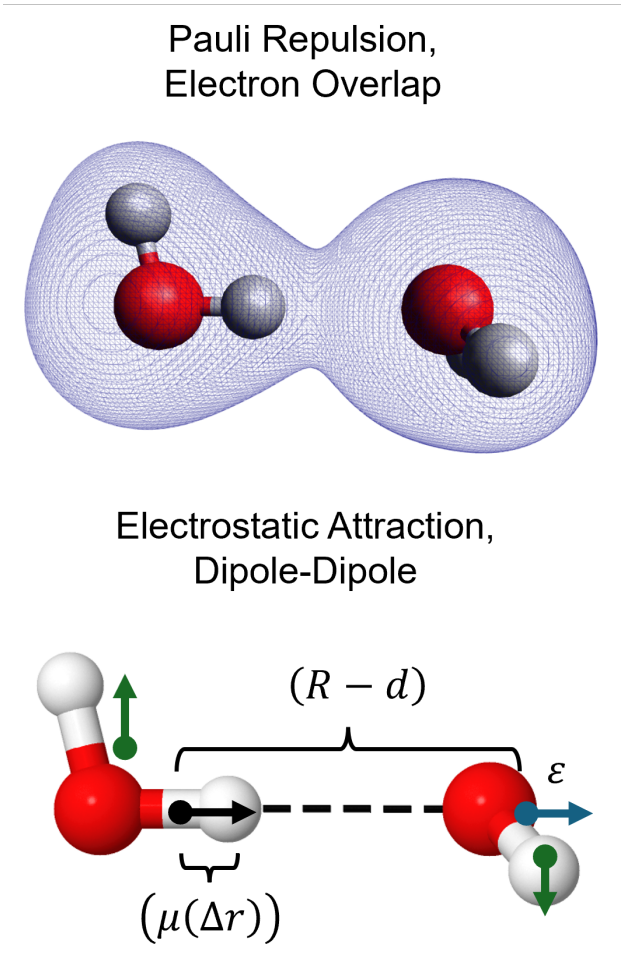

Figure 3.2: The two components of the energy expression for linear hydrogen bonds. Top: Pauli Repulsion resulting from the overlap of electron density (blue domains), which increases overlap as atoms get closer and changes shape with intramolecular geometry. Bottom: Electrostatics are broken down into two dipole moments (shown with black and blue arrows) that are co-linear with the heavy atom axis. The dipole moment on the hydrogen bond donor depends on the intramolecular geometry and the dipole moment on the acceptor. The "best" origins of the dipole moments may not be the heavy atoms themselves, so the distance is offset by d. Non-captured electrostatics may remain orthogonal to the heavy atom axis (green arrows) but are assumed to interact minimally.

Although both linear and non-linear hydrogen bonds are considered in this work through eq. (3.3), the development of the physical parameters in eq. (3.3) focused only on the linear hydrogen bond arrangements (θ=0) to further simplify the sampling of the configuration space, viz.

$$V_{\text{Int.}}(\Delta r, R, 0) = Ae^{-\beta(R-\gamma\Delta r)} - \frac{\varepsilon\mu(\Delta r)}{(R-d)^n}\left(1 + \frac{\varepsilon a(\Delta r)}{(R-d)^n}\right) \tag{3.4}$$

To decouple the dependence of the inter- and intra-molecular degrees of freedom, it becomes advantageous to Taylor expand V_{HB} in the intra–molecular degree of freedom (Δr). In this expansion, eq. (3.1) becomes

$$V_{\text{HB}}(\Delta r, R, \theta) = V_{\text{M}}(\Delta r) + V_{\text{Int}}(0, R, \theta) + \partial_{\Delta r} V_{\text{Int}}(0, R, \theta)\Delta r + \cdots \tag{3.5}$$

This approach is similar to that previously used by Hermasson [191] and Boyer et al., [114] who described V_{Int} by a dipole in an electric field and expanded that dipole with respect to Δr. For simplicity, in the following the partial derivatives of V_{Int} will be replaced with $c^{(i)}$, that is

$$c^{(i)}(R, \theta) = \partial^i_{\Delta r} V_{\text{Int}}(0, R, \theta) \tag{3.6}$$

which are further defined via. eqs. (3.3) and (3.4) as follows

$$c^{(i)}(R, \theta) = A\beta^i\gamma^i \cos^i\theta e^{-\beta R} - \frac{\varepsilon\mu^{(i)}\cos\theta}{(R-d)^3} \cdot \left(1 + \frac{\varepsilon\cos\theta\sum_{j=0}^{i} a^{(j)}\frac{\mu^{(i-j)}}{\mu(i)}}{(R-d)^3}\right) \tag{3.7}$$

and

$$c^{(i)}(R, 0) = A\beta^i\gamma^i e^{-\beta R} - \frac{\varepsilon\mu^{(i)}}{(R-d)^3} \cdot \left(1 + \frac{\varepsilon\sum_{j=0}^{i} a^{(j)}\frac{\mu^{(i-j)}}{\mu(i)}}{(R-d)^3}\right) \tag{3.8}$$

In this notation, eq. (3.5) becomes

$$V_{\text{HB}}(\Delta r, R, \theta) = V_{\text{M}}(\Delta r) + c^{(0)}(R, \theta) + c^{(1)}(R, \theta)\Delta r + \cdots \tag{3.9}$$

This equation will be used to derive the relationship between energy and underlying structure. Although this work uses our specific definition of V_{Int} to develop the energy-structure relationship, the definition in eq. (3.6) is general. Therefore any further analysis left in terms of $c^{(i)}$ is general and can be used for any different definition of the energy expression.

3.2.2 Scans of the Potential Energy Surfaces of Hydrogen Bonds

Six hydrogen-bonded dimers were examined in this work, namely $NH_3{-}NH_3$, $H_2O{-}H_2O$, HF−HF, $H_2O{-}NH_3$, HF−H_2O, and HF−NH_3. These were selected as the interaction is entirely due to hydrogen bonding. For each system two PESs were computed.

First, a relaxed hydrogen bond scan was performed to collect hydrogen bond energies and elongations for a diverse set of dimer geometries. The scans were performed from -0.5Å to 3Å in ΔR, where $\Delta R = R - R_e$ and R_e is the equilibrium heavy atom distance, and from $-25°$ to 25° in θ in an uneven grid. Spacing in the ΔR scan was 0.1Å for $\Delta R \leq 1$Å, and 0.2Å for $\Delta R \geq 1$Å. Spacing in θ was 1° for $|\theta| \leq 6°$, 2° for $6° \leq |\theta| \leq 10°$, 2.5° for $10° \leq |\theta| \leq 15°$, and 5° for $|\theta| \geq 15°$.

Second, scans for the linear hydrogen bond conformations ($\theta = 0$) were computed to obtain a PES in reduced configuration space. This was necessary in order to sample enough of the interaction potential in Δr and R to create reliable fits. Semi-relaxed 2-D scans, where some extra degrees of freedom were allowed to relax, were also performed and described here. An initial scan was performed over ΔR from -0.5Å to 18Å in an uneven grid. The spacing between the grid points was 0.1Å for $\Delta R \leq 1$Å, 0.2Å for 2Å$\leq \Delta R \leq 3$Å, 1Å for 3Å$\leq \Delta R \leq 8$Å, and 2Å for 8Å$\leq \Delta R \leq 18$Å. This grid spacing was chosen to bias fitting to the minimum of the potential ($\Delta R = 0$) while at the same time providing an extrapolation to the dimer dissociation. Hydrogen bonds were held linear and to C_s symmetry, but all other degrees of freedom were allowed to relax to permit the coupling to R. At each of these semi-optimized geometries, rigid scans of Δr, where all degrees of freedom were fixed, were performed from -0.1Å to 0.1Å on an even grid with a spacing of 0.01Å. The rigid scan of Δr was chosen to prevent coupling to the high-frequency mode of Δr, which is not included in the $c^{(i)}$ terms. Although the scans were performed to $\Delta R = -0.5$Å, the dimer surface was truncated to values of R where $V_{\mathrm{HB}}(0, R, 0) < 0$ to remove the high energy part of the repulsive wall which isn't well described by a Buckingham potential. [58–60]

The fitting was performed using SciPy's basin-hopping algorithm with a mean-squared error loss function. [192, 193] The interaction energy PES was obtained by subtracting the hydrogen bond donor's potential energy surface and the hydrogen bond acceptor's energy from the dimer's potential energy surface. The degree of Taylor expansions for μ and a was investigated starting with a simple static dipole model and subsequently adding derivatives to the dipole function and

polarizability tensor. The basin-hopping process was continued until no improvements to the loss function were made for 10,000 iterations.

All energy calculations were performed at the Møller Plesset second-order perturbation theory (MP2) [144–146] using Dunning's augmented correlation consistent triple zeta (aug-cc-pVTZ) basis set [194, 195] with the Gaussian 16 suite of codes. [173] All calculations as well as the PESs are deposited in `https://github.com/gds001/Energy-Structure`.

3.3 Results and Discussion

Using the hydrogen bond energy expression defined in section 3.2.1 and the ab initio surfaces computed in section 3.2.2, we proceed to derive a relationship between energy and structure. The procedure is as follows: first, the parameters from eq. (3.4) were determined for the linear hydrogen bond PESs. Using those parameters, the expression for the *energetic–structural* correspondence in hydrogen bonds was derived. Finally, this expression was validated for the non-linear hydrogen bonds in section 3.3.3.

3.3.1 Fitting the Parameters of Hydrogen Bonds using Expansions of the Dipole Moment and the Polarizability Tensor

The interaction energies of each linear hydrogen bond dimer system were fitted to the generalized Buckingham potential [60] expressed in eq. (3.3) for various degrees of expansion for both the dipole moment and polarizability tensor. The linear hydrogen bond surface was used to fit the parameters in order to properly capture the important effects of Δr and R on the interaction potential while minimizing the computational cost of the scans. Simpler models without containing the derivatives of either the polarizability tensor or the dipole moment fail to accurately fit the data. The errors for the simple models originate from a lack of attractive coupling between the R and Δr coordinates. High accuracy associated with a Mean Absolute Error (MAE) that is < 5.0 cm^{-1} is achieved by using a linear dipole moment and a quadratic polarizability tensor. There is a slight improvement (< 0.5 cm^{-1}) when incorporating third-order derivatives in the dipole moment; this improvement is minimal, and thus the linear dipole moment is used. Incorporating the second-order derivatives in the polarizability tensor proved to be important for accuracy, but improvements diminished by the cubic term. Therefore, this work will focus on using a linear expansion of the dipole moment and

a quadratic expansion of the polarization tensor, acting as a proper balance between the number of fitted parameters and the overall accuracy.

The optimized parameters for the linear dipole moment and quadratic polarizability produced inconsistent dipole moment derivatives across the same hydrogen bond donor. It is seen that the ratio of dipole moment derivatives (i.e. $\frac{\mu^{(0)}}{\mu^{(1)}}$) is not the same for the similar hydrogen bond donors, despite the fact that $\mu^{(i)}$ is a property specific to the hydrogen bond donor. Therefore, the systems were re-fit under the constraint that

$$\varepsilon\mu^{(1)}_{\mathrm{X-Y}} = \varepsilon\mu^{(0)}_{\mathrm{X-}} \frac{\mu^{(1)}_{\mathrm{X-X}}}{\mu^{(0)}_{\mathrm{X-X}}} \tag{3.10}$$

where $\mathrm{X-X}$ is the homo-dimer and $\mathrm{X-Y}$ is the hetero-dimer. This procedure was not performed on the polarizability terms because it is believed that these terms contain some of the undescribed charge transfer and polarizability of the hydrogen bond acceptor, both of which depend on the identity of the hydrogen bond acceptor. The fitting parameters and MAE are provided in table 3.1. With this new scheme, the dipole moment surface of each hydrogen bond donor is consistent.

The first section of table 3.1 (labeled as "Morse") contains the parameters that describe the covalent DH bond in the hydrogen bond donor; this is rigorously the same for each hydrogen bond donor as they were derived from a finite difference of the donor's monomer surface. The second section of table 3.1 (labeled as "Pauli") lists the Pauli repulsion term in eq. (3.3). For the stronger hydrogen bond acceptors (NH_3 >H_2O>HF), the onset of the repulsion becomes more gradual, with lower values of A and β. The value of β also sees slight gains with hydrogen bond donor strength, while γ is constant (0.425 ± 0.025) across all systems, and is generally smaller for stronger hydrogen bonds. Finally, the third section of table 3.1 (labeled as "Electrostatics") contains the electrostatic interaction between fragments, which is dependent on the dipole moment on the hydrogen bond donor (μ), the polarizability tensor (a), and the electric field produced by the hydrogen bond acceptor (ε). Each dipole term becomes larger for stronger hydrogen bonds, a trend arising from both the strengthening of the hydrogen bond donors (via. $\mu^{(i)}$) and acceptors (via. ε). The parameter d becomes more physically reasonable (near 0 Å) after constraining the dipole ratios. Errors in this final model are marginally worse than the unconstrained model but are well within spectroscopic accuracy. The MAE for the H_2O-NH_3 and $HF-H_2O$ systems increases by less than 0.5 cm^{-1}, but a larger loss in accuracy is found for $HF-NH_3$ (≈ 3 cm^{-1}). The values of these

Table 3.1: The fitting parameters for the linear dipole moment and quadratic polarizability tensor. Parameters are sectioned off by horizontal lines; the upper section contains parameters specific to the intra-molecular potential, the middle section the ones for the Pauli repulsion term of eq. (3.3), and finally the lower section the parameters for the electrostatic terms of eq. (3.3). The final section contains information about the dipole moment ratio and fitting errors.

	Parameters	NH_3−NH_3	H_2O−H_2O	HF−HF	H_2O−NH_3	HF−H_2O	HF−NH_3
Morse	r_A (Å)	1.0121	0.9613	0.9218	0.9613	0.9218	0.9218
	k_{M} (E_h/Å^2)	1.6547	1.9376	2.1991	1.9376	2.1991	2.1991
	D_{M} (E_h)	0.1892	0.1885	0.1918	0.1885	0.1918	0.1918
	α (1/Å)	2.0911	2.2670	2.3941	2.2670	2.3941	2.3941
Pauli	A (E_h)	188.53	222.77	384.03	187.96	294.12	190.65
	β (1/Å)	3.1230	3.4451	4.0843	3.2349	3.7691	3.3897
	γ	0.4413	0.4236	0.4292	0.4120	0.4040	0.4078
Electrostatic	$\varepsilon\mu^{(0)}$ (E_hÅ^3)	0.0687	0.1138	0.1181	0.1367	0.1878	0.2246
	$\varepsilon\mu^{(1)}$ (E_hÅ^2)	0.1476	0.1708	0.1746	0.2051	0.2776	0.3319
	$\varepsilon a^{(0)}$ (Å^3)	114.12	40.34	11.96	79.21	32.85	101.61
	$\varepsilon a^{(1)}$ (Å^2)	-100.71	0.77	7.73	24.34	25.38	75.80
	$\varepsilon a^{(2)}$ (Å)	898.60	211.26	94.28	418.05	214.55	519.85
	d (Å)	0.2033	0.2209	0.2093	0.0506	-0.0082	-0.2970
	$\frac{\mu^{(0)}}{\mu^{(1)}}$ (Å)	0.4654	0.6664	0.6766	0.6664	0.6766	0.6766
	MAE of fit (cm^{-1})	2.69	3.36	7.09	4.23	5.89	9.99

parameters and the corresponding errors in the performance of the mode are deemed acceptable for the development of the energy structural relationship and are used for the rest of this work.

3.3.2 The Relationship between Energy and Structure

The energy of the hydrogen bond can be determined via the hydrogen bond potential outlined in eq. (3.9). Since the hydrogen bond potential is zero at dissociation ($\lim_{R\to\infty}$), the hydrogen bond energy is the negative of the potential evaluated at the equilibrium geometry, viz. $\partial_r V_{\mathrm{HB}}(\Delta r_{\mathrm{HB}}, R, \theta) = 0$. This yields the equation

$$E_{\mathrm{HB}} = -V_{\mathrm{HB}}(\Delta r_{\mathrm{HB}}, R, \theta) = -D_{\mathrm{M}}(z_e - 1)^2 - c^{(0)}(R, \theta) - c^{(1)}(R, \theta)\Delta r_{\mathrm{HB}} - \cdots \tag{3.11}$$

where $z_e = e^{-\alpha \Delta r_{\mathrm{HB}}}$ and Δr_{HB} is the bond elongation at equilibrium. Without knowing how $c^{(i)}$ may relate to Δr_{HB}, the dependence of E_{HB} on Δr_{HB} cannot be determined from equation (3.11).

The expressions for $c^{(i)}$ depend only on R for linear hydrogen bonds. Therefore, any dependence of $c^{(i)}$ on Δr_{HB} comes from the dependence of Δr_{HB} on R, found via the chain rule.

$$\partial_{\Delta r_{\mathrm{HB}}} c^{(i)} = \partial_R c^{(i)} \partial_{\Delta r_{\mathrm{HB}}} R \tag{3.12}$$

The exclusion of θ from the chain rule in eq. (3.12) comes with the assumption of a linear hydrogen bond, which proves to be a good approximation (see section 3.3.3) that greatly simplifies the subsequent analysis. In this approximation $\partial_R c^{(i)}$ is solvable

$$\partial_R c^{(i)} = -A\beta^{i+1}\gamma^i e^{-\beta R} + \frac{3\varepsilon\mu^{(i)}}{(R-r_b)^4} \cdot \left(1 + \frac{2\varepsilon \sum_{j=0}^{i} a^{(j)} \frac{\mu^{(i-j)}}{\mu^{(i)}}}{(R-r_b)^3} \right) \tag{3.13}$$

The partial derivative $\partial_{\Delta r_{\mathrm{HB}}} R$ can be derived by setting the first derivative of the hydrogen bond potential equal to zero at the equilibrium geometry

$$\partial_{\Delta r} V_{\mathrm{HB}} = 0 = 2\alpha D_{\mathrm{M}}(z_e - z_e^2) + c^{(1)} + c^{(2)} \Delta r_{\mathrm{HB}} \tag{3.14}$$

The above equation is solved after Taylor expanding the Morse exponential to obtain an equation relating Δr_{HB} to R

$$\Delta r_{\mathrm{HB}} = \frac{k_{\mathrm{M}} + c^{(2)} - \sqrt{(k_{\mathrm{M}} + c^{(2)})^2 + 6\alpha k_{\mathrm{M}} c^{(1)}}}{3\alpha k_{\mathrm{M}}} \tag{3.15}$$

This can be inverted to obtain R as a function of Δr_{HB} only by Taylor expanding $c^{(i)}$ with respect to R. Such an expansion is not accurate over the long range of R observed. The derivative with respect to Δr_{HB} of eq. (3.14) yields the equation for $\partial_{\Delta r_{\mathrm{HB}}} R$, after the chain rule is applied to $\partial_{\Delta r_{\mathrm{HB}}} c^{(i)}$

$$\partial_{\Delta r_{\mathrm{HB}}} R = \frac{-k_{\mathrm{M}}(2z_e^2 - z_e) - c^{(2)}}{\partial_R c^{(1)} + \partial_R c^{(2)} \Delta r_{\mathrm{HB}}} \tag{3.16}$$

Without the linear hydrogen bond approximation used to create eq. (3.12), $\partial_{\Delta r_{\mathrm{HB}}} \theta$ would remain in eq. (3.16), reducing the usefulness of the analysis. A Taylor expansion was not necessary for the derivative.

Using equations (3.8) - (3.16), $\partial_{\Delta r_{\mathrm{HB}}} c^{(0)}$ can be obtained. By multiplying $\partial_R c^{(0)}$ and eq. (3.16), one finds

$$\partial_{\Delta r_{\mathrm{HB}}} c^{(0)} = \frac{-k_{\mathrm{M}}(2z_e^2 - z_e) - c^{(2)}}{\frac{\partial_R c^{(1)}}{\partial_R c^{(0)}} + \frac{\partial_R c^{(2)}}{\partial_R c^{(0)}} \Delta r_{\mathrm{HB}}} \tag{3.17}$$

This expression is general for any interaction potential, as $c^{(i)}$ and $\partial_R c^{(i)}$ have not been substituted. By substituting the expression for $c^{(i)}$ $\partial_R c^{(i)}$, eqs. (3.8) and (3.13), the relationship between $c^{(0)}$ and Δr_{HB} is expressed solely as parameters of the interaction potential. Although complete, such an expression is too complicated to be useful. Therefore, approximations to the derivative must be made.

The full derivative of $\partial_{\Delta r_{\mathrm{HB}}} c^{(0)}$ was separated into electrostatic (E), polar (P), and repulsive (R) pieces by negating various fitted parameters. For example, $c_{\mathrm{E}}^{(i)}$ lacks Pauli Repulsion and the polarizability tensor and is left as a simple dipole interaction

$$c_{\mathrm{E}}^{(i)}(R,0) = -\frac{\varepsilon \mu^{(i)}}{(R-d)^3} \tag{3.18}$$

These simpler electrostatic expansion coefficients can be used to stitch $\partial_{\Delta r_{\mathrm{HB}}} c_{\mathrm{E}}^{(0)}$. Since $\mu^{(2)} = 0$, $c_{\mathrm{E}}^{(2)} = 0$ and the derivative from eq. (3.17) using $c^{(i)}$ from eq. (3.18) simplifies to

$$\partial_{\Delta r_{\mathrm{HB}}} c_{\mathrm{E}}^{(0)} = -\, k_{\mathrm{M}} \frac{\mu^{(0)}}{\mu^{(1)}} \left(2 z_e^2 - z_e\right) \tag{3.19}$$

with no dependence on R remaining. The residual angular dependence also cancels in the division of $\partial_R c^{(0)}$ and $\partial_R c^{(1)}$, which is why the linear approximation works well. The antiderivative is trivial

$$c_{\mathrm{E}}^{(0)}(\Delta r_{\mathrm{HB}}) = 2\alpha D_{\mathrm{M}} \frac{\mu^{(0)}}{\mu^{(1)}} (z_e^2 - z_e) \tag{3.20}$$

The polar derivative, $\partial_{\Delta r_{\mathrm{HB}}} c_{\mathrm{P}}^{(0)}$, becomes more complicated, since pieces related to R do not cancel. Therefore, the derivative is left as

$$\partial_{\Delta r_{\mathrm{HB}}} c_{\mathrm{P}}^{(0)} = \frac{-k_{\mathrm{M}}(2 z_e^2 - z_e) - c_{\mathrm{P}}^{(2)}}{\frac{\partial_R c_{\mathrm{P}}^{(1)}}{\partial_R c_{\mathrm{P}}^{(0)}} + \frac{\partial_R c_{\mathrm{P}}^{(2)}}{\partial_R c_{\mathrm{P}}^{(0)}} \Delta r_{\mathrm{HB}}} + k_{\mathrm{M}} \frac{\mu^{(0)}}{\mu^{(1)}} \left(2 z_e^2 - z_e\right) \tag{3.21}$$

where the electrostatic term found earlier is subtracted from the new term that includes polarizability with the explicit R dependence remaining. This needs to be approximated as an integrable function of Δr_{HB}. Eq. (3.21) is displayed as the red line in figure 3.3. This derivative takes the shape of a quotient of polynomials, where the numerator determines the long-range behavior and the denominator induces massive slope changes near $\Delta r_{\mathrm{HB}} = 0$. The function chosen here is

$$\partial_{\Delta r_{\mathrm{HB}}} c_{\mathrm{P}}^{(0)} \approx \frac{b \frac{\Delta r_{\mathrm{HB}}^2}{2} + c \Delta r_{\mathrm{HB}}}{\Delta r_{\mathrm{HB}} + r_o} \tag{3.22}$$

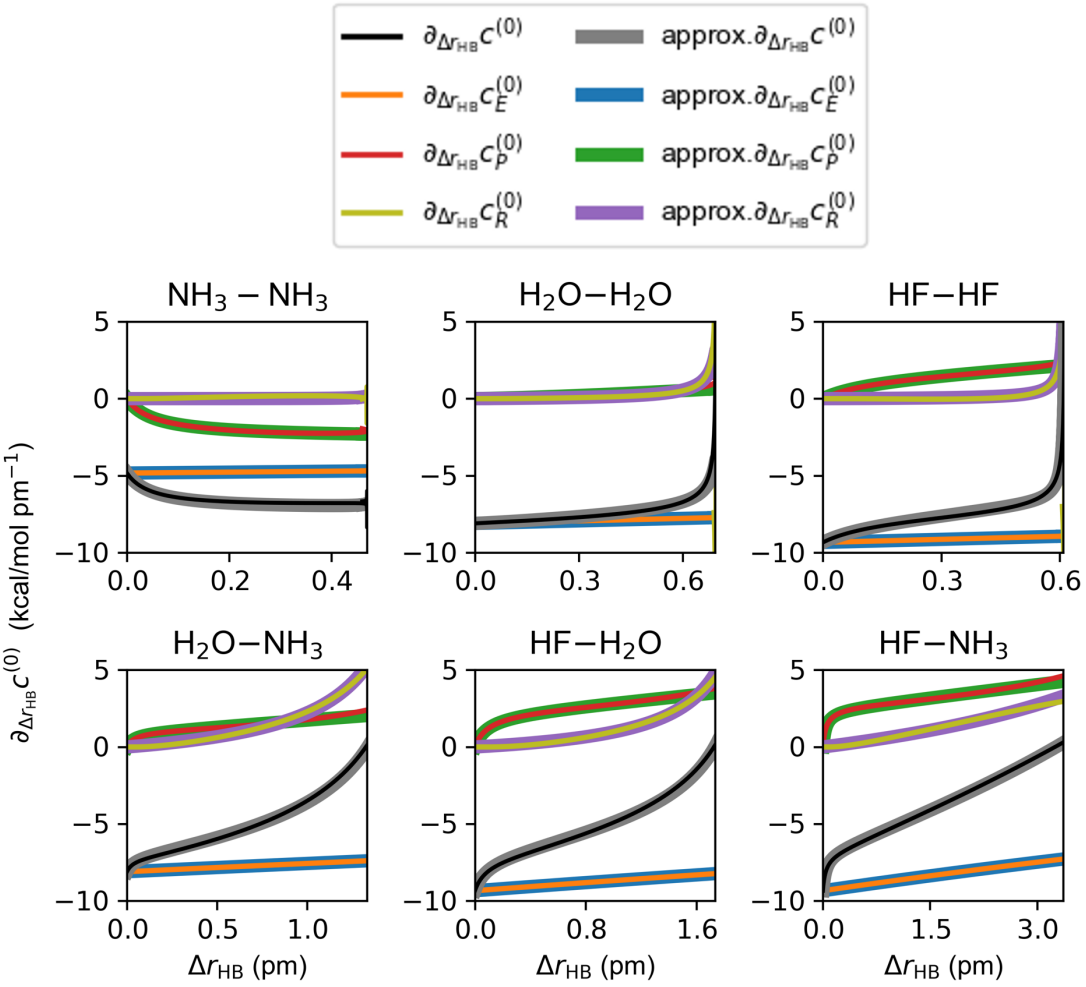

Figure 3.3: Derivative of $c^{(0)}$ with respect to Δr_{HB} for a linear dipole moment and quadratic polarizability tensor. Each panel is separated into electrostatic (E), polarizability (P) and repulsive (R) pieces.

where b describes the long-range slope of this term, c describes the intercept for a linear approximation, and r_o is a positive number describing the singularity causing the larger slope near $\Delta r_{\mathrm{HB}} = 0$. The best fit of eq. (3.22) to eq. (3.21) is denoted by the green curve in figure 3.3. The antiderivative of this function for positive Δr_{HB} is

$$c_{\mathrm{P}}^{(0)} \approx b\frac{\Delta r_{\mathrm{HB}}^2}{2} - (br_o - c)\Delta r_{\mathrm{HB}} + r_o(br_o - c)\ln\left[\frac{\Delta r_{\mathrm{HB}} + r_o}{r_o}\right] \tag{3.23}$$

A similar technique must be applied to the repulsive term in the derivative. The proper derivative is provided by eq. (3.24)

$$\partial_{\Delta r_{\mathrm{HB}}} c_{\mathrm{R}}^{(0)} = \frac{-k_{\mathrm{M}}(2z_e^2 - z_e) - c_{\mathrm{R}}^{(2)}}{\frac{\partial_R c^{(1)R}}{\partial_R c_{\mathrm{R}}^{(0)}} + \frac{\partial_R c_{\mathrm{R}}^{(2)}}{\partial_R c_{\mathrm{R}}^{(0)}}\Delta r_{\mathrm{HB}}} - \partial_{\Delta r_{\mathrm{HB}}} c_{\mathrm{P}}^{(0)} - k_{\mathrm{M}}\frac{\mu^{(0)}}{\mu^{(1)}}\left(2z_e^2 - z_e\right) \tag{3.24}$$

Eq. (3.24) is displayed in yellow in figure 3.3. This is best approximated as an inverse function.

$$\partial_{\Delta r_{\mathrm{HB}}} c_{\mathrm{R}}^{(0)} \approx \frac{B}{r_i - \Delta r_{\mathrm{HB}}} - \frac{B}{r_i} \tag{3.25}$$

where B is the magnitude of the repulsion and r_i is the maximum elongation distance. r_i comes directly from the calculated scans except for the $\mathrm{HF{-}NH_3}$ dimer due to the poor fitting in this region for that system. The antiderivative of this function for $\Delta r_{\mathrm{HB}} < r_i$ is

$$c_{\mathrm{R}}^{(0)} \approx -\frac{B}{r_i}\Delta r_{\mathrm{HB}} - B\ln\left[\frac{r_i - \Delta r_{\mathrm{HB}}}{r_i}\right] \tag{3.26}$$

The choice of approximate function for $\partial_{\Delta r_{\mathrm{HB}}} c_{\mathrm{R}}^{(0)}$ is not unique, as both inverse and inverse square root functions can model this asymptotic behavior well. The inverse function was chosen in this work for its marginally better performance, although the inverse square root function results in a final answer similar to the square-root relationship predicted by Iogansen. [189]

Summed together, these create a working equation for $c^{(0)}$

$$c^{(0)} \approx 2\alpha D_{\mathrm{M}}\frac{\mu^{(0)}}{\mu^{(1)}}\left(z_e^2 - z_e\right) + b\frac{\Delta r_{\mathrm{HB}}^2}{2} - \left(\frac{C}{r_o} + \frac{B}{r_i}\right)\Delta r_{\mathrm{HB}} + \ln\left[\frac{(\Delta r_{\mathrm{HB}} + r_o)^C}{(r_i - \Delta r_{\mathrm{HB}})^B}\frac{r_i^B}{r_o^C}\right] \tag{3.27}$$

where C replaces $r_o(br_o - c)$ for brevity. The logarithmic pieces cannot be simplified or approximated without loss in accuracy. The values of the coefficients are listed in table 3.2 and eq. (3.27) is compared against the computed values of $c^{(0)}$ in figure 3.4. The errors are small, with a maximum error of 0.31 kcal/mol. Errors exceed the MAE near $\Delta r_{\mathrm{HB}} = 0$ for most hydrogen-bonded systems

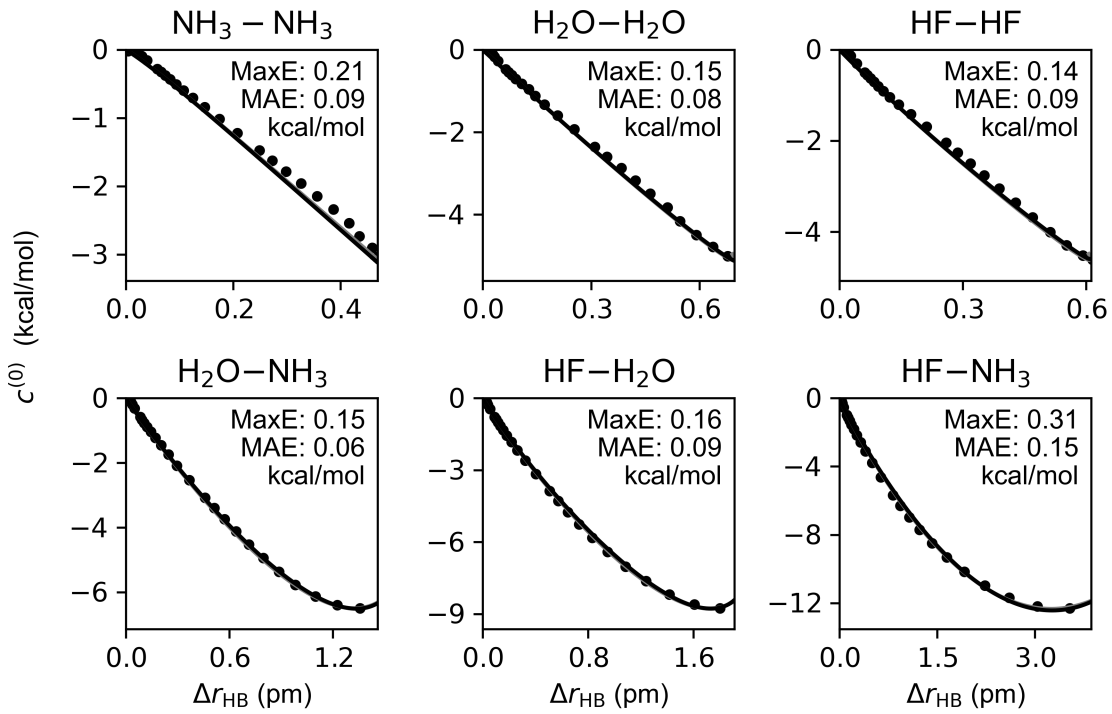

Figure 3.4: Plots of $c^{(0)}$ as a function of Δr_{HB}. Solid circles denote the values obtained from the ab-initio PESs around $r = r_{\mathrm{M}}$. The solid lines trace the proposed relationship from eq. (3.27). The mean absolute error (MAE) and the maximum absolute error (MaxE), are reported as a metric of the accuracy of eq. (3.27). The points and errors are limited to $\Delta R \geq 0$.

due to lack of precision in this region and around $\Delta r_{\mathrm{HB}} = \max(\Delta r_{\mathrm{HB}}/2)$ as the answer in eq. (3.27) approximates the curvature. With these small errors, eq. (3.27) will serve as the best approximation for $c^{(0)}$. It should be noted other fits of potential can yield answers with a different functional form.

The same analysis can be performed on higher orders of $c^{(i)}$; shown here is $c^{(1)}$. The partial derivatives of $c^{(1)}$ with respect to Δr_{HB} can similarly be separated into electrostatic ($\partial_{\Delta r_{\mathrm{HB}}} c_{\mathrm{E}}^{(1)}$), polarizabiltiy ($\partial_{\Delta r_{\mathrm{HB}}} c_{\mathrm{P}}^{(1)}$), and repulsive ($\partial_{\Delta r_{\mathrm{HB}}} c_{\mathrm{R}}^{(1)}$) pieces. The electrostatic piece retains its simplistic form from a similar derivation to eq. (3.20)

$$\partial_{\Delta r_{\mathrm{HB}}} c_{\mathrm{E}}^{(1)} = -k_{\mathrm{M}}(2z_e^2 - z_e) \tag{3.28}$$

$$c_{\mathrm{E}}^{(1)}(\Delta r_{\mathrm{HB}}) = 2\alpha D_{\mathrm{M}}(z_e^2 - z_e) \tag{3.29}$$

Table 3.2: The values of the coefficients used in eqs. (3.27), (3.30), and (3.31) using the fitted parameters from table 3.1.

Term	Unit	NH_3−NH_3	H_2O−H_2O	HF−HF	H_2O−NH_3	HF−H_2O	HF−NH_3
$2\alpha D_{\mathrm{M}}\frac{\mu^{(0)}}{\mu^{(1)}}$	(kcal/mol)	231.0680	357.4269	396.0229	357.4269	396.0229	396.0229
b	(kcal/mol pm^{-2})	-0.0838	1.7031	1.7858	0.8643	0.8737	0.5645
C	(kcal/mol)	0.1287	1.3304	-0.1825	-0.0350	-0.2024	-0.1023
r_o	(pm)	0.0512	1.0000	0.1190	0.0383	0.0877	0.0401
B	(kcal/mol)	0.0028	0.0814	0.0355	2.9264	2.2817	43.1086
r_i	(pm)	0.4847	0.7214	0.6164	1.7538	2.1313	8.5943
$2\alpha D_{\mathrm{M}}$	(kcal/mol pm^{-1})	4.9653	5.3633	5.763	5.3633	5.7639	5.7639
ζ	(kcal/mol pm^{-3})	0.0054	0.0064	0.0061	0.0073	0.0037	0.0060
ξ	(kcal/mol pm^{-4})	0.0000	0.0000	0.0000	0.0000	0.0090	0.0019
$k_{\mathrm{M}}\frac{\mu^{(0)}}{\mu^{(1)}}$	(kcal/mol pm^{-1})	4.8319	8.1027	9.4811	8.1027	9.4811	9.4811
x	(kcal/mol pm^{-2})	0.0577	1.0663	1.1645	0.6469	0.7084	0.5537
y	(kcal/mol pm^{-3})	0.0024	0.0011	0.0000	0.0015	-0.0012	-0.0001
z	(kcal/mol pm^{-4})	0.0000	0.0000	0.0001	0.0000	0.0016	0.0004

Only the stronger and more deformed hydrogen bonds have non-zero non-electrostatic contributions. Both are polynomial-like in nature, with $\partial_{\Delta r_{\mathrm{HB}}} c_{\mathrm{P}}^{(1)}$ being a true quadratic function and $\partial_{\Delta r_{\mathrm{HB}}} c_{\mathrm{R}}^{(1)}$ being quadratic until very close to r_i. The quadratic fits show good accuracy. Using simple antiderivaties, a complete expression for $c^{(1)}$ is

$$c^{(1)}(\Delta r_{\mathrm{HB}}) = 2\alpha D_{\mathrm{M}}(z_e^2 - z_e) + \zeta\frac{\Delta r_{\mathrm{HB}}^2}{2} + \xi\frac{\Delta r_{\mathrm{HB}}^3}{6} \tag{3.30}$$

where ζ and ξ arise from the sum of the linear and quadratic pieces of $\partial_{\Delta r_{\mathrm{HB}}} c_{\mathrm{P}}^{(1)}$ and $\partial_{\Delta r_{\mathrm{HB}}} c_{\mathrm{R}}^{(1)}$. The parameters for eq. (3.30) are reported in table 3.2.

The values of $c^{(1)}$ computed via finite difference of the ab-inito surface were compared against those predicted by eq. (3.30) in figure 3.5. Eq. (3.30) is quite accurate due to the simple behavior of $c^{(1)}$, with a MAE less than 0.0005 kcal/mol pm^{-1}. Errors are largest at the longest bond elongations but do not exceed 0.005 kcal/mol pm^{-1}. Therefore, eq. (3.30) suffices as an accurate approximation for $c^{(1)}$.

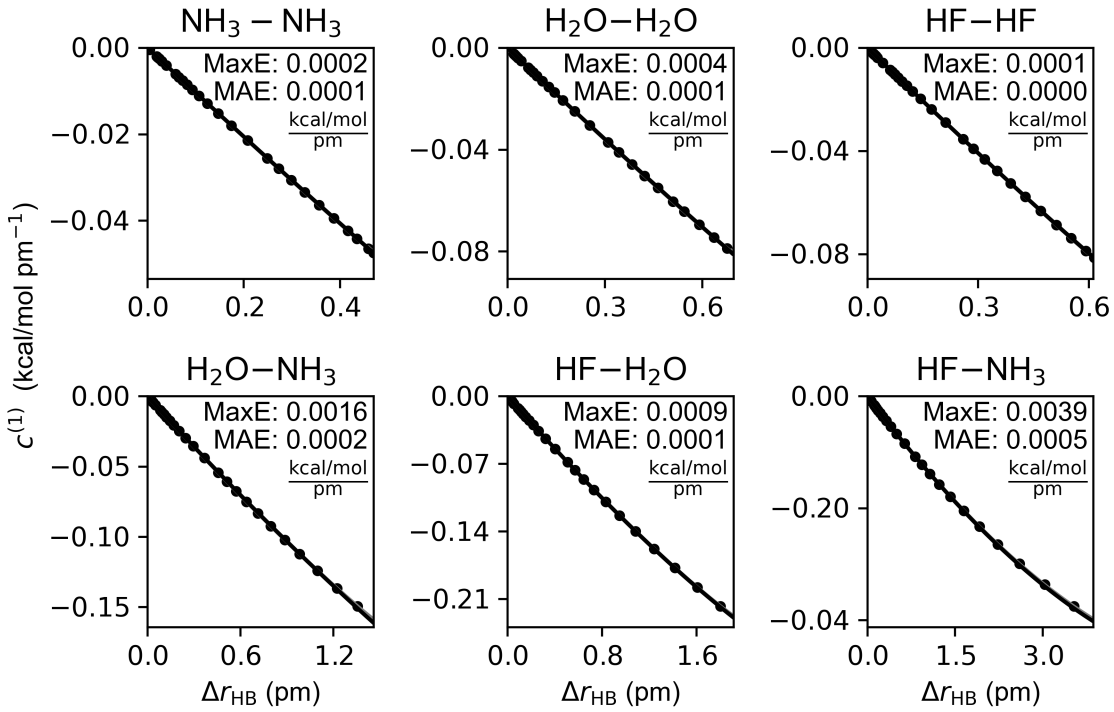

Figure 3.5: Plots of $c^{(1)}$ as a function of Δr_{HB}. Solid circles denote the values obtained from a finite difference of the *ab-initio* PESs around $r = r_{\mathrm{M}}$. The solid lines trace the proposed relationship from eq. (3.30). The mean absolute error (MAE) and the maximum absolute error (MaxE), are reported as a metric of the accuracy of eq. (3.30). The points and errors are limited to $\Delta R \geq 0$.

With expressions for $c^{(0)}$ and $c^{(1)}$ expressed as function of Δr_{HB}, it is now possible to express E_{HB} as a function of Δr_{HB} from eq. (3.11) as

$$\begin{aligned} E_{\mathrm{HB}}(\Delta r_{\mathrm{HB}}) = &- D_{\mathrm{M}}(z_e - 1)^2 - 2\alpha D_{\mathrm{M}} \frac{\mu^{(0)}}{\mu^{(1)}} \left(z_e^2 - z_e\right) - b\frac{\Delta r_{\mathrm{HB}}^2}{2} + \left(\frac{C}{r_o} + \frac{B}{r_i}\right) \Delta r_{\mathrm{HB}} \\ &- \ln\left[\frac{(\Delta r_{\mathrm{HB}} + r_o)^C}{(r_i - \Delta r_{\mathrm{HB}})^B} \frac{r_i^B}{r_o^C}\right] - 2\alpha D_{\mathrm{M}}(z_e^2 - z_e)\Delta r_{\mathrm{HB}} - \zeta\frac{\Delta r_{\mathrm{HB}}^3}{2} - \xi\frac{\Delta r_{\mathrm{HB}}^4}{6} \end{aligned} \tag{3.31}$$

in which all terms of this expression are defined previously in this chapter, and their values are provided throughout tables 3.1 and 3.2. Although the logarithm term caused by repulsion cannot be Taylor expanded without significant loss in accuracy, the exponential terms in a Morse oscillator

are nicely expandable. Doing so results in a terse form of the expression

$$\begin{aligned} E_{\mathrm{HB}}(\Delta r_{\mathrm{HB}}) =& k_{\mathrm{M}} \frac{\mu^{(0)}}{\mu^{(1)}} \Delta r_{\mathrm{HB}} - x \Delta r_{\mathrm{HB}}^2 - y \Delta r_{\mathrm{HB}}^3 - z \Delta r_{\mathrm{HB}}^4 \\ & - \ln \left[\frac{(\Delta r_{\mathrm{HB}} + r_o)^C}{(r_i - \Delta r_{\mathrm{HB}})^B} \frac{r_i^B}{r_o^C} \right] + \left(\frac{B}{r_i} + \frac{C}{r_o} \right) \Delta r_{\mathrm{HB}} \end{aligned} \tag{3.32}$$

where $x = \frac{3}{2}\alpha k_{\mathrm{M}} \frac{\mu^{(0)}}{\mu^{(1)}} + \frac{b}{2} - \frac{k_{\mathrm{M}}}{2}$, $y = 3\alpha k_{\mathrm{M}} - \frac{7}{6}\alpha^2 k_{\mathrm{M}} \frac{\mu^{(0)}}{\mu^{(1)}} + \frac{\zeta}{2}$, and $z = \frac{15}{24}\alpha^3 k_{\mathrm{M}} \frac{\mu^{(0)}}{\mu^{(1)}} - \frac{7}{8}\alpha^2 k_{\mathrm{M}} + \frac{\xi}{6}$. The linear B and C terms are not contracted since they serve to cancel the linear effect of the logarithm. The values of x, y, and z are reported in table 3.2. The values of y and z are very small and contribute minimally (<0.05 kcal/mol) to the model for the strongest hydrogen bonds. The value of y is smaller than its corresponding ζ, suggesting that the minimal impact of y comes from canceling terms between ζ and the cubic piece of $c_{\mathrm{E}}^{(0)}$. The x term remains sizable and contributes to the model's overall accuracy. The majority of the curvature in the relationship between energy and structure comes from the repulsion piece of the logarithmic term and thus should be investigated further using additional dimers to understand how this important term changes with different hydrogen bond acceptors.

Eq. (3.32) is compared against the minimum energies computed for each heavy atom distance in figure 3.6. Errors are comparable to those seen with $c^{(0)}$ in figure 3.4, showing the majority of errors in E_{HB} come from the approximation of $c^{(0)}$. A highly accurate analytic representation of hydrogen bond energies has been developed from a physical representation of the hydrogen bond potential. This representation needs to be tested against a realistic range of hydrogen bond geometries in order to validate it's general use and descriptive capabilities.

3.3.3 Validation of the Energy - Structure Relationship for Non-linear Hydrogen Bonds

Although the linear hydrogen bond configurations are used to develop an accurate model potential for hydrogen bonds and use it to derive a physically motivated relationship between hydrogen bond energy and changes in the underlying intramolecular geometry, most hydrogen bonds are not linear. The equilibrium geometries of all hydrogen-bonded dimers used in this work are a few degrees away from linearity (17.1, 5.5, 6.4, 6.8, and 1.5 degrees for $NH_3{-}NH_3$, $H_2O{-}H_2O$, HF−HF, $H_2O{-}NH_3$, and HF$-H_2O$, respectively), except for HF$-NH_3$ for which the hydrogen bond is linear (zero degrees), while the configuration space biases bent (non-linear) hydrogen bonds during molecular

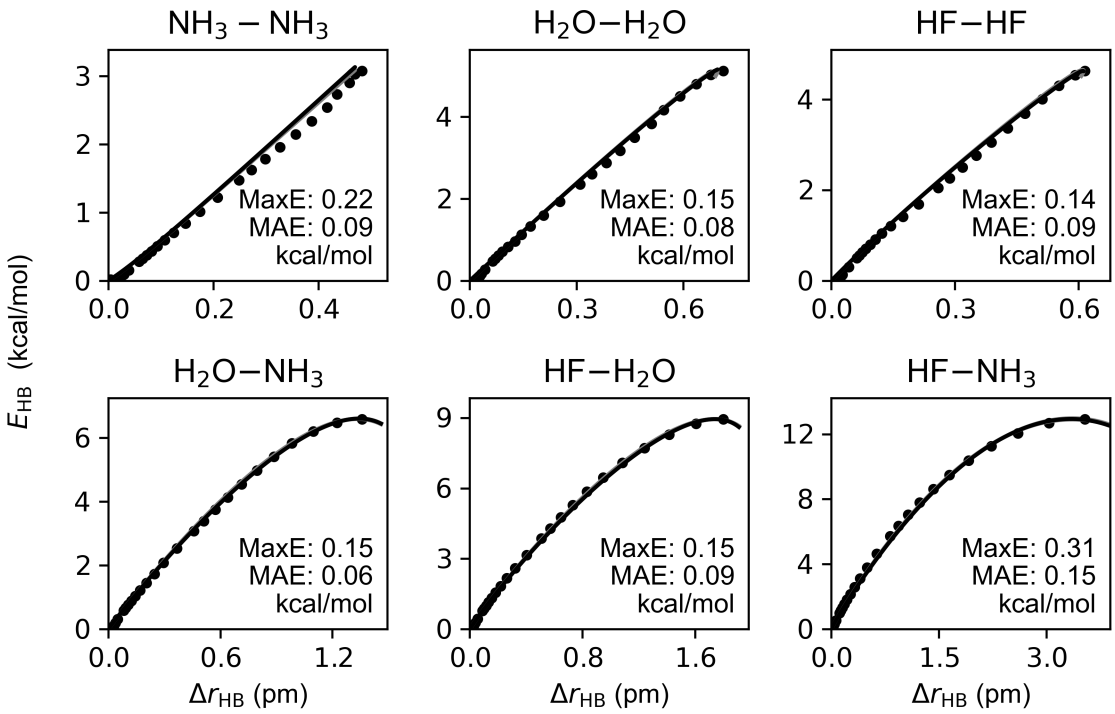

Figure 3.6: Plots of E_{HB} as a function of Δr_{HB}. The solid circles denote the values of the hydrogen bond energies from the linear scans. The solid lines trace the relation from eq. (3.31). The mean absolute error (MAE) and the maximum absolute error (MaxE), are reported as a metric of the accuracy of eq. (3.31).

dynamics simulations of water clusters and liquid water. Therefore, it is important to explore the validity of the relationships developed and reported in eq. (3.31) for non-linear hydrogen bonds.

The hydrogen bond energies for heavy atom distances between R_e and $R_e + 3$Å and the angles, as defined in figure 3.1, between -15° and 15° are plotted against hydrogen bond elongation in figure 3.7. This range in θ corresponds to 98% of the hydrogen bond angles observed in low energy water clusters. [196, 197] Overlayed is eq. (3.31). Non-linearity in hydrogen bonds instills minimum errors in the relationship, with the MAE doubling for stronger hydrogen bonds.

The transferability of the obtained relationship between energy and underlying structure is rationalized by considering the effect of angular dependence ($\cos\theta$) in eq. (3.3) on the derived relations in the previous sections. The division of $\partial_R c^{(i)}$ rigorously cancels the electrostatic angular

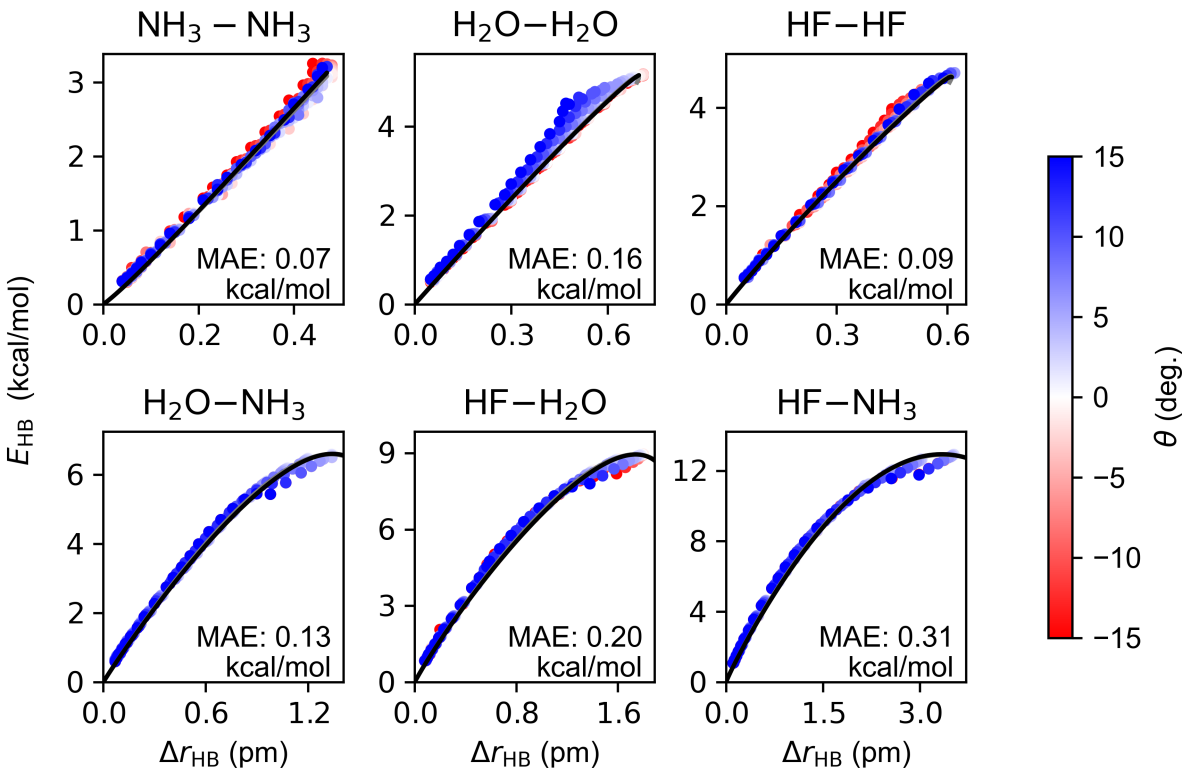

Figure 3.7: Plots of E_{HB} as a function of Δr_{HB}. The solid circles denote the values of the hydrogen bond energies from the linear scans. The solid lines trace the relation from eq. (3.31). The solio points are color-coded based on their deviation (in degrees) from linearity. The mean absolute error (MAE) is reported as a metric of the accuracy of eq. (3.31).

dependence in $\partial_{\Delta r_{\mathrm{HB}}} c_{\mathrm{E}}^{(i)}$. This is observed visually in figure 3.7, where all points are co-linear for short bond elongations. The non-canceling of the angular dependence occurs only with the onset of the non-linear terms in the expression of E_{HB}. In figure 3.7 the data points begin to spread above and below the predicted relationship introducing non-recoverable errors. However, the errors are still small, with the MAE being less than 0.30 kcal/mol.

Large errors appear for the positive angles in the water dimer. This is probably due to the water dimer achieving a chelating geometry, where the previously free OH on the donating water molecule forms a bent hydrogen bond with the other lone pair on the acceptor water molecule. Such interactions are not possible in the other dimers because of the lack of additional hydrogen-bond

accepting or donating sites. It should be noted that these angles are at the extreme range of hydrogen bond angles in water cluster minima. [197] Such structures also epitomize the complicated nature of hydrogen bonds, where multiple hydrogen bonding interactions cannot be disentangled via the binding energy. There is a slight overestimation of hydrogen bond energies in HF$-$H$_2$O and HF$-$NH$_3$ near $\Delta R = 0$ and for angles greater than 10°. This arises from the growing angular dependence as hydrogen bonds get stronger. These errors are small compared to the overall hydrogen bond energy. The relationship derived here combined with the energy expression proves to be a robust method for predicting the energy of a hydrogen bond from its structural properties.

3.4 Conclusions

The optimal structures of hydrogen bonded systems suggest that there is clearly a correspondence between the strength of a hydrogen bond DH$\cdots$A, where D denotes the donor and A the acceptor, and the change in the underlying covalent DH bond. The developed analytical model incorporates the appropriate physics that are needed to account for this relationship, viz. the PES of the covalent DH bond, the Pauli repulsion and electrostatic interactions. The parameters of the model were fitted using the PESs of six hydrogen bond dimers. Using this model we developed generalized forms of the relationship between hydrogen bond energy and hydrogen bond structure. The correlations were found to be largely electrostatic in nature, with non-linearity arising from the polarizability and the repulsive effects at close separations. The maximum absolute error in the interaction energy remained less than 10% of the total dimer binding energy except for H$_2$O$-$H$_2$O, for which deviations originated from superfluous interactions for highly non-linear hydrogen bonds. This further justifies the developed model as a reliable and predictive tool for calculating interaction energies for hydrogen bonded dimers. The presented analysis is quite general, allowing for the possibility to adding more terms to improve the model outlined in eq. (3.3) such as, for example, explicit charge transfer terms that may improve the fits and relationships of the stronger hydrogen bonds, especially those in HF$-$H$_2$O and HF$-$NH$_3$.

It is remarkable that a similar relationship for all dimers is found at the electrostatic limit. Most notably, the linear term ($k_{\mathrm{M}}\frac{\mu^{(0)}}{\mu^{(1)}}$), which dominates the relationship until the onset of Pauli repulsion is 8.1 and 9.5 kcal/mol pm^{-1} for all H$_2$O and HF donors, respectively, is due to similar force constants and dipole derivative ratios. The differences between NH$_3$, H$_2$O, and HF are rectified

by polarizability, yielding a slope of about 8.0 kcal/mol pm^{-1} (or 0.8 kcal/ mol per 0.001 Å) before the Pauli repulsion comes into play. Therefore, a single relationship based on the properties of the hydrogen bond donor serves as a good starting point for the overall relationship between energy and structure. This similarity between elements of the same row on the periodic table acting as hydrogen bond donors conceptually matches those in Badger's rule. [183] The hydrogen bond acceptor largely influences the maximum elongation Δr_{HB}, corresponding to r_i in eq. (3.32) and the onset of non-linearity from Pauli-repulsion for tightly bound dimers. A more general non-linear term to the relationship that describes the equilibrium geometries of each system could serve to unify the relationship for all first-row and possibly second-row donors, again in the spirit of the original relation proposed by Badger. [183]

The current effort has contributed to the understanding of the relationship between hydrogen bond strength and underlying structure by quantitatively connecting the energy of a hydrogen bond to a single structural coordinate that is highly perturbed by hydrogen bonding. To the best of our knowledge, no other previous work has reported such a quantitative relationship that incorporates physical origins. That structural coordinate (Δr_{HB}) can be easily calculated using standard computational techniques (i.e. electronic structure methods) and can serve as an accurate descriptor for predicting the strength of individual hydrogen bonds. In the next chapter of this thesis, we will turn our attention to the experimentally accessible spectroscopic relationship with energy in hydrogen bonds.

Chapter 4

EXTENDING BADGER'S RULE: II. THE RELATIONSHIP BETWEEN ENERGY AND VIBRATIONAL SPECTRA IN HYDROGEN BONDS [198]

4.1 Introduction

The strength of the hydrogen bond is important in describing the properties associated with their structure and stability. The previous chapter of this thesis established a physically motivated relationship between the energy of a hydrogen bond and structural changes in the participating donor covalent bond (DH). Although structural deformation is a logical descriptor of hydrogen bond strength, the DH bond length cannot be easily determined experimentally. X-ray diffraction cannot detect hydrogen and rotational spectroscopy is not sensitive enough to determined DH bond lengths in polyatomic clusters larger than a few atoms. Therefore, it is a goal in this thesis to extend the *energetic–structural* relationship to include vibrational spectroscopy, which is known to correlate with hydrogen bond strengths.

As discussed prior, work established by Richard Badger [183, 184] related the covalent bond force constants to the covalent bond distances, a correlation that has been coined as "Badger's rule." [183, 184] Although Badger's initial work focused on isolated molecules, recent work showed the relationship is valid when a molecule is perturbed by external interactions, such as a hydrogen bond. [114] This previous work found a linear relationship for aqueous ionic clusters

$$\Delta\omega_{\mathrm{HB}} = 19.1 \frac{\mathrm{cm}^{-1}}{0.001\ \text{Å}} \Delta r_{\mathrm{HB}} \tag{4.1}$$

where Δr_{HB} is the elongation of the OH covalent bond participating in the hydrogen bond and $\Delta\omega_{\mathrm{HB}}$ is the corresponding red-shift in its vibrational frequency from the isolated monomer values. The numerical value of the slope originates from both the mechanical anharmonicity of the covalent bond and the electrical anharmonicity in the dipole moment. [114] In the following, we will refer to this as the "beyond Badger" relationship.

Richard Badger also found evidence for a linear relationship between the hydrogen bond enthalpies and the change in vibrational frequencies, called the "Badger–Bauer relationship." [185]

Subsequent studies [190, 199–206] have quantified this correlation, providing values for the slope and intercept for specific hydrogen bond donors and solvents. For example, a 2014 study [190] concatenated the results for alcohols donating to various bases in carbon tetrachloride solvent suggesting the relationship

$$\Delta H_{\mathrm{HB}} = 0.013 \ \frac{\mathrm{kcal/mol}}{\mathrm{cm}^{-1}} \Delta\omega_{\mathrm{HB}} + 0.514 \ \mathrm{kcal/mol} \tag{4.2}$$

where ΔH_{HB} is the hydrogen bond enthalpy and $\Delta\omega_{\mathrm{HB}}$ is the red-shift in the OH frequency participating in the hydrogen bond. The slope of the relationship remained relatively similar across all donors while the intercept depended greatly on the functionalization of the alcohol. Iogansen also suggested a square-root relationship [189, 207]

$$\Delta H_{\mathrm{HB}} = 0.33 \ \frac{\mathrm{kcal/mol}}{\sqrt{\mathrm{cm}^{-1}}} \sqrt{\Delta\omega_{\mathrm{HB}} - 40 \ \mathrm{cm}^{-1}} \tag{4.3}$$

Both of the above relationships have been employed by physical-, organic- and bio-chemists to compare hydrogen bonds. [199–206, 208, 209] However, neither functional form was developed from rigorous derivation that included incorporating the underlying physics into the model. Instead, these previous relationships describe the general shape and curvature of the correlation.

The this chapter aims to provide the third and final step in the 3-way general correspondence, by further extending Badger's rule to derive the *energetic–spectral* relationship. The organization of this chapter is as follows. In Section 4.2 we outline the computational methodology, referencing chapter 3 often. In Section 4.3 we derive the *energetic–spectral* relationship for linear hydrogen bonds and validate it for parts of the PES associated with non-linear hydrogen bonds. In Section 4.4 we compare the results for E_{HB} predicted by the derived relationship to the ones obtained by either experiment or theory using the experimental or theoretical vibrational red-shifts for the six hydrogen bonded systems considered here. Conclusions and outlook are presented in Section 4.5.

4.2 Computational Methodology

The hydrogen bond potential, V_{HB}, is defined in the previous chapter, eqs. (3.1)–(3.9). Please refer to section 3.2.1 for details regarding fitting parameters and the expansion of the interaction terms. We used the data from the same six hydrogen-bonded dimers, namely $NH_3{-}NH_3$, $H_2O{-}H_2O$, $HF{-}HF$, $H_2O{-}NH_3$, $HF{-}H_2O$, and $HF{-}NH_3$, discuss in the previous chapter to fit each dimer's interaction potential to eq. (3.3). A similar set of relaxed hydrogen bond scans were performed at

Table 4.1: Mean Signed Error (MSE), Mean Absolute Error (MAE) and Maximum absolute Error (MaxE) for the relationships between Δr_{HB}, $\Delta\omega_{\mathrm{HB}}$, and E_{HB} derived in this work. For reference, the errors in the *energetic–structural* relationship [210] are -0.04, 0.09, and 0.31 kcal/mol for the MSE, MAE, and MaxE, respectively.

Relationship	Eqs.	$\Delta\omega_{\mathrm{HB}}(\Delta r_{\mathrm{HB}})$ (cm^{-1})			$\Delta r_{\mathrm{HB}}(\Delta\omega_{\mathrm{HB}})$ (pm)			$E_{\mathrm{HB}}(\Delta\omega_{\mathrm{HB}})$ (kcal/mol)		
		MSE	MAE	MaxE	MSE	MAE	MaxE	MSE	MAE	MaxE
Beyond Badger	(4.9-4.11)	-6.2	7.2	29.4	0.03	0.04	0.14	0.19	0.22	0.79
Extending Badger	(4.17-4.20)	-3.3	3.3	11.8	0.02	0.02	0.06	0.07	0.11	0.63

the Møller Plesset second-order perturbation theory (MP2) [144–146] using Dunning's augmented correlation consistent triple zeta (aug-cc-pVTZ) basis sets [194, 195] with the Gaussian 16 suite of codes. [173] The scans were performed on the same grid as in the prior work, [210] but an additional frequency analysis was performed at each partially optimized geometry to obtain the hydrogen bond harmonic frequency. All hydrogen atoms not participating in the hydrogen bond were deuterated to minimize the kinetic coupling between local modes.

4.3 Derivation of the Energetic–Spectral Relationship

The sought after *energetic–spectral* relationship for hydrogen bonds was developed via a composition of the previously derived *energetic–structural* [210] and *spectral–structural* [114] relationships. Specifically, the *spectral–structural* relationship was first extended to hydrogen bonds that do not involve water and compared against an expanded relationship that included Pauli repulsion as in the *energetic–structural* relationship. [210] Approximations were made to ensure the invertibility of the relationships and the two variations were compared against the ab initio PESs. With a selected representation of the *structural–spectral* relationship, the *energetic–spectral* relationship was obtained and evaluated against the ab initio PESs for both linear and non-linear hydrogen bonds.

4.3.1 A Revised General Relationship between Structure and Spectra

Although the relationship between structure and spectra has been previously derived for aqueous and ionic aqueous clusters, [114] it is re-investigated in this work under the fitted Buckingham po-

tentials to expand it for hydrogen bonds not involving water (NH and FH stretches) and also ensure a consistent treatment of all three relationships. A generalized relationship between frequency and the potential in eq. (3.3) is detailed here.

The harmonic frequency of the hydrogen bond stretch is related to the force constant, or the second derivative of the interaction potential with respect to the covalent bond

$$k_{\mathrm{HB}} = \partial^2_{\Delta r} V_{\mathrm{HB}}(\Delta r_{\mathrm{HB}}, R, \theta) = k_{\mathrm{M}}(2z_e^2 - z_e) + c^{(2)} + c^{(3)} \tag{4.4}$$

where k_{M} is the force constant of the covalent bond in the isolated hydrogen bond donor, k_{HB} is the force constant in the covalent bond in the hydrogen bounded donor, and z_e is the exponent of the Morse oscillator evaluated a the equilibrium dimer geometry ($e^{-\alpha \Delta r_{\mathrm{HB}}}$). To relate the force constant to frequency shifts, it is best to divide each side by k_{M} and subtract from one. Then eq. (4.4) becomes

$$1 - \frac{k_{\mathrm{HB}}}{k_{\mathrm{M}}} = 1 - (2z_e^2 - z_e) - \frac{c^{(2)}}{k_{\mathrm{M}}} - \frac{c^{(3)} \Delta r_{\mathrm{HB}}}{k_{\mathrm{M}}} \tag{4.5}$$

This factorization is convenient because the left side of eq. (4.5) is easily related to frequency shifts ($\Delta\omega_{\mathrm{HB}} = \omega_{\mathrm{M}} - \omega_{\mathrm{HB}}$), viz.

$$1 - \frac{k_{\mathrm{HB}}}{k_{\mathrm{M}}} = \frac{2\Delta\omega_{\mathrm{HB}}}{\omega_{\mathrm{M}}} - \frac{\Delta\omega^2_{\mathrm{HB}}}{\omega^2_{\mathrm{M}}} \approx \frac{2\Delta\omega_{\mathrm{HB}}}{\omega_{\mathrm{M}}} \tag{4.6}$$

assuming the reduced mass of ω_{HB} and ω_{M} are the same. It will also be assumed that the hydrogen bonded stretch is localized onto the Δr_{HB} coordinate, with minimal coupling to R. This is generally appropriate for weaker hydrogen bonds. This expression is actually quadratic, but is commonly approximated as linear since $\Delta\omega_{\mathrm{HB}} << \omega_{\mathrm{M}}$ for weaker hydrogen bonds. The right side of eq. (4.5) can be expressed as a function of hydrogen bond elongation (Δr_{HB}). In that way, $\Delta\omega_{\mathrm{HB}}$ can be expressed as function of Δr_{HB} as

$$\Delta\omega_{\mathrm{HB}}(\Delta r_{\mathrm{HB}}) = \frac{\omega_{\mathrm{M}}}{2}\left(1 - 2z_e^2 + z_e - \frac{c^{(2)}}{k_{\mathrm{M}}} - \frac{c^{(3)}\Delta r_{\mathrm{HB}}}{k_{\mathrm{M}}}\right) \tag{4.7}$$

for weaker hydrogen bonds and

$$\Delta\omega_{\mathrm{HB}}(\Delta r_{\mathrm{HB}}) = \omega_{\mathrm{M}}\left(1 - \sqrt{2z_e^2 - z_e + \frac{c^{(2)}}{k_{\mathrm{M}}} + \frac{c^{(3)}\Delta r_{\mathrm{HB}}}{k_{\mathrm{M}}}}\right) \tag{4.8}$$

for the full range of hydrogen bond strengths. To obtain a usable form for this relationship, $\frac{c^{(2)}}{k_{\mathrm{M}}}$ and $\frac{c^{(3)}}{k_{\mathrm{M}}}$ must be cast as functions of Δr_{HB}.

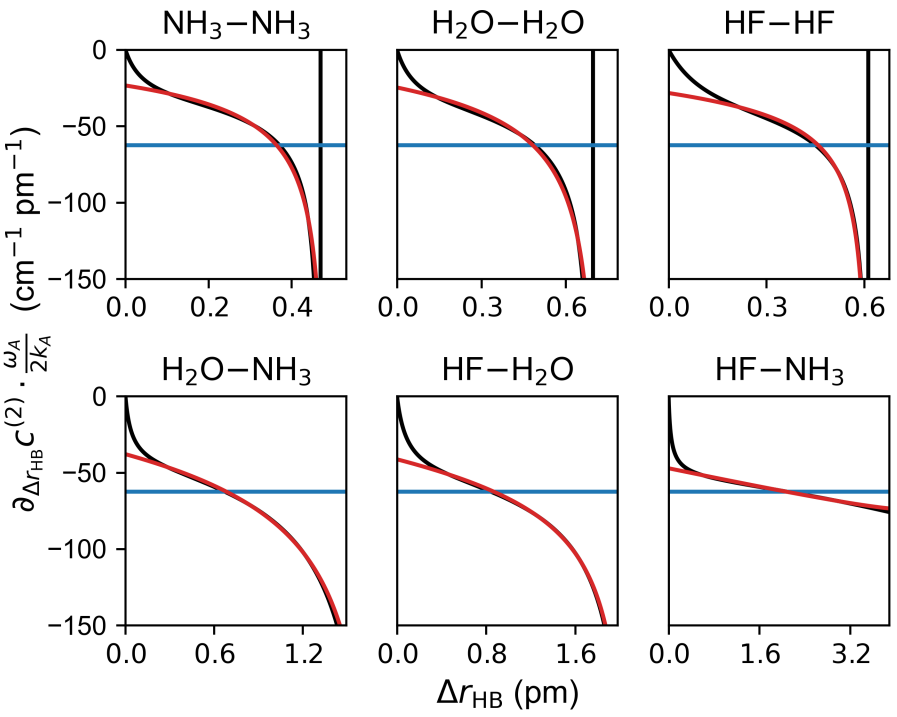

Figure 4.1: The exact expression of $\partial_{\Delta r_{\mathrm{HB}}} c^{(2)} \cdot \frac{\omega_{\mathrm{M}}}{2k_{\mathrm{M}}}$ compared against approximate expressions for each dimer system studied. The blue line expresses the constant derivative approximated in the "beyond Badger" relationship, [114] and the red line is an inverse square root approximation that accurately captures repulsion.

Table 4.2: The linear terms in the *spectral–structural*, *structural–spectral* (inverse of *spectral–structural*), and *energetic–spectral* relationships for each hydrogen bond donor. Units of "per 100 cm^{-1}" ensure appropriate magnitude of the linear terms as well as represent the general range of hydrogen bond red-shifts. "A" stands for the general hydrogen bond acceptor.

Parameter	Eq.	NH_3−A	H_2O−A	HF−A
$\frac{\omega_{\mathrm{M}}\left(3\alpha+\mu^{(2)}/\mu^{(1)}\right)}{2}$ (cm^{-1}/pm)	(4.9)	172.0	191.0	207.1
$\frac{2}{\omega_{\mathrm{M}}\left(3\alpha+\mu^{(2)}/\mu^{(1)}\right)}$ (pm/100 cm^{-1})	(4.10)	0.58	0.52	0.48
$\frac{2k_{\mathrm{M}}\mu^{(0)}/\mu^{(1)}}{\omega_{\mathrm{M}}\left(3\alpha+\mu^{(2)}/\mu^{(1)}\right)}$ (kcal/mol/100 cm^{-1})	(4.11)	2.81	4.24	4.51

4.3.2 Beyond Badger: The Simplest Solution

An invertible representation of the *spectral–structural* relationship has already been established by Boyer et al. [114] They reported the following relationship

$$\Delta\omega_{\mathrm{HB}} = \frac{\omega_{\mathrm{M}}}{2}\left(3\alpha + \frac{\mu^{(2)}}{\mu^{(1)}}\right)\Delta r_{\mathrm{HB}} \tag{4.9}$$

by modeling the hydrogen bond as a dipole in an electric field. In the above expression $\mu^{(1)}$ and $\mu^{(2)}$ are the first and second derivatives of the dipole moment parallel to the DH bond (D stands for the hydrogen bond donor) when the dipole moment is expanded about the equilibrium configuration in the presence of the hydrogen bond acceptor (i.e. polarized dipole derivatives). This expression matches the derivation earlier in this chapter (eq. (4.7), where $\frac{\mu^{(2)}}{\mu^{(1)}}\Delta r_{\mathrm{HB}} = -\frac{c^{(2)}}{k_{\mathrm{M}}}$, $c^{(3)}$=0, and a linear expansion of $2z_e^2 - z_e$ is used. This solution is trivially invertible, viz.

$$\Delta r_{\mathrm{HB}} = \frac{2}{\omega_{\mathrm{M}}\left(3\alpha + \frac{\mu^{(2)}}{\mu^{(1)}}\right)}\Delta\omega_{\mathrm{HB}} \tag{4.10}$$

With this invertible solution, Δr_{HB} can be substituted in the *energetic–structural* developed and established in the prior chapter. This solution is displayed in eq. (4.11) below

$$\begin{aligned} E_{\mathrm{HB}} = {} & \frac{2k_{\mathrm{M}}\frac{\mu^{(0)}}{\mu^{(1)}}}{\omega_{\mathrm{M}}\left(3\alpha + \frac{\mu^{(2)}}{\mu^{(1)}}\right)}\Delta\omega_{\mathrm{HB}} - \frac{6\alpha k_{\mathrm{M}}\frac{\mu^{(0)}}{\mu^{(1)}} + 2b - 2k_{\mathrm{M}}}{\omega_{\mathrm{M}}^2\left(3\alpha + \frac{\mu^{(2)}}{\mu^{(1)}}\right)^2}\Delta\omega_{\mathrm{HB}}^2 \\ & - \ln\left[\frac{\left(\frac{2\Delta\omega_{\mathrm{HB}}}{\omega_{\mathrm{M}}\left(3\alpha+\mu^{(2)}/\mu^{(1)}\right)} + r_o\right)^C}{\left(r_i - \frac{2\Delta\omega_{\mathrm{HB}}}{\omega_{\mathrm{M}}\left(3\alpha+\mu^{(2)}/\mu^{(1)}\right)}\right)^B}\frac{r_i^B}{r_o^C}\right] + \frac{2\left(\frac{B}{r_i} + \frac{C}{r_o}\right)}{\omega_{\mathrm{M}}\left(3\alpha + \frac{\mu^{(2)}}{\mu^{(1)}}\right)}\Delta\omega_{\mathrm{HB}} \end{aligned} \tag{4.11}$$

where variables such as r_o, r_i,b, B, and C were defined the prior chapter, eq. (3.31 and 3.32). x from 3.32) is expressed in terms of α, $\mu^{(i)}$, b, and k_M since higher order terms of Δr_{HB} (y and z from eq. (3.32)) are neglected.

By fitting a data set of hydrogen bond donor covalent bond elongations and associated redshifts in water and aqueous ionic clusters, Boyer et al. found the slope of $\frac{\omega_{\mathrm{M}}}{2}\left(3\alpha + \frac{\mu^{(2)}}{\mu^{(1)}}\right)$ to be approximately 191 cm^{-1}/pm (19.1 cm^{-1}/1000 Å). [114] This solution, put in the context of the model, means that $\frac{\mu^{(2)}}{\mu^{(1)}} \cdot \frac{\omega_{\mathrm{M}}}{2} \approx -62.5$ cm^{-1}/pm. This value is compared against $\frac{\partial_{\Delta r_{\mathrm{HB}}} c^{(2)}}{k_{\mathrm{M}}} \cdot \frac{\omega_{\mathrm{M}}}{2}$ as the blue line in figure 4.1. This approximation appears to average the exact partial derivative, shown in

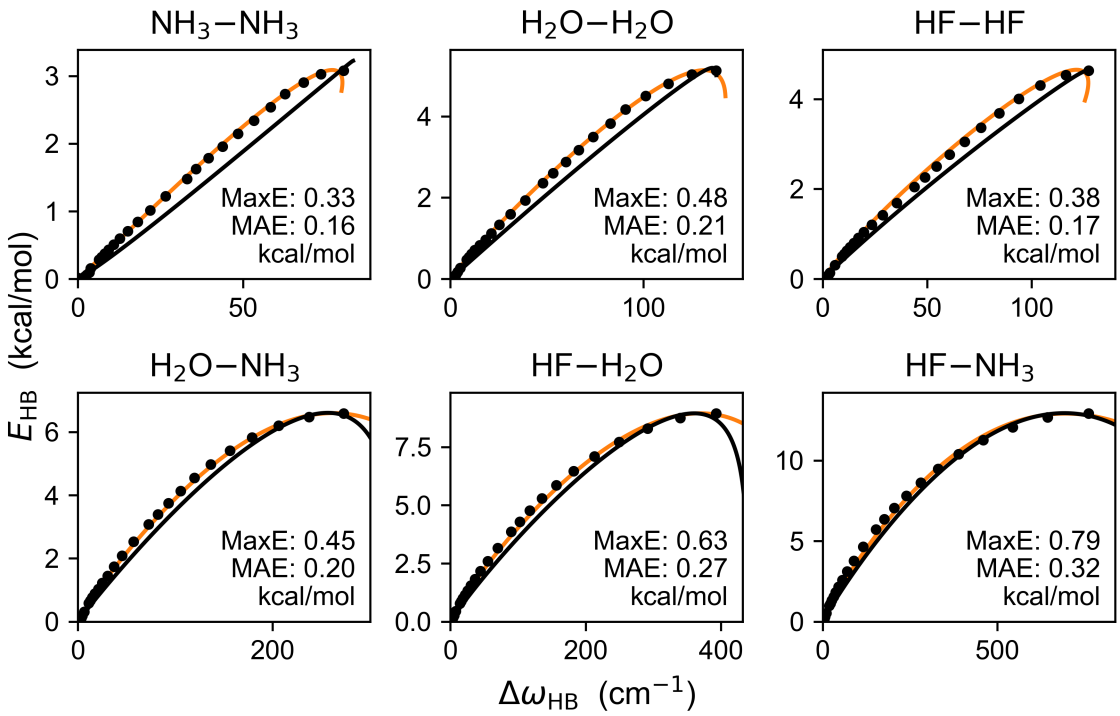

Figure 4.2: Plots of E_{HB} as a function of $\Delta\omega_{\mathrm{HB}}$ assuming the "beyond Badger" *structural–spectral* relationship. [114] Solid circles trace the values of the hydrogen bond energy from the linear scans. The black lines show the relationship from eq. (4.11), whereas the orange lines trace a parametric relationship. The values of the MaxE and MAE errors are shown to compare the performance of the relationship for different hydrogen-bonded dimers.

black, by underestimating it at low Δr_{HB} and overestimating it at high Δr_{HB}. The assumption that $\frac{\mu^{(2)}}{\mu^{(1)}}$ remains constant for all six systems appears reasonable. Errors are sizable (see figure 4.2 and table 4.1) but the solution remains usable. The linear terms of eq. (4.11) are 2.81, 4.24, 4.51, kcal/mol/100 cm^{-1} for the NH_3, H_2O, and HF donors respectively, reported in table 4.2.

4.3.3 Extending "Beyond Badger": Incorporating Non-linear Effects

In the Beyond Badger approximation, [114] the relationship was found to be linear, meaning that $c^{(2)}$ was linear. However, the $c^{(2)}$ computed from the interaction potential, eq. (3.3), is far from linear, as shown by the black line in figure 4.1. The singularity observed earlier in $\partial_{\Delta r_{\mathrm{HB}}} c^{(0)}$ at r_i is

observed here for $\partial_{\Delta r_{\mathrm{HB}}} c^{(2)}$. Therefore, improvements to eq. (4.11) can be achieved by incorporating non-linear effects.

In the following, $c^{(2)}$ will be expressed as a function of Δr_{HB} using the chain rule, similar to the approach used in the prior chapter. This yields an expression for $\partial_{\Delta r_{\mathrm{HB}}} c^{(2)}$

$$\partial_{\Delta r_{\mathrm{HB}}} c^{(2)} = \frac{-k_{\mathrm{M}} - c^{(2)}}{\frac{\partial_R c^{(1)}}{\partial_R c^{(2)}} + \Delta r_{\mathrm{HB}}} \tag{4.12}$$

Although the derivative of $c^{(2)}$ does depend on $c^{(2)}$ itself, this formulation was found to produce more reliable results with respect to an invertible form. This derivative can be separated into electrostatic, polarizability, and repulsive pieces. The electrostatic contribution to the derivative is rigorously zero since the dipole is assumed to be linear. Therefore, expressing $c^{(2)}$ as a function of Δr_{HB} will rely on the representation of the polarizable and repulsive contributions to the derivative, $\partial_{\Delta r_{\mathrm{HB}}} c_{\mathrm{P}}^{(2)}$ and $\partial_{\Delta r_{\mathrm{HB}}} c_{\mathrm{R}}^{(2)}$, respectively.

As already highlighted for $\partial_{\Delta r_{\mathrm{HB}}} c_{\mathrm{R}}^{(0)}$ in the previous chapter, the approximations for each of these terms are not unique; the only requirement is that each partial derivative be integrable and produce a relatively accurate fit. The preliminary representation of $\partial_{\Delta r_{\mathrm{HB}}} c_{\mathrm{P}}^{(2)}$ and $\partial_{\Delta r_{\mathrm{HB}}} c_{\mathrm{R}}^{(2)}$ similar to that of $\partial_{\Delta r_{\mathrm{HB}}} c_{\mathrm{P}}^{(0)}$ and $\partial_{\Delta r_{\mathrm{HB}}} c_{\mathrm{R}}^{(0)}$ produces a highly accurate *spectral–structural*. Errors in predicting $\Delta\omega$ are summarized in table 4.1. Although these functions represent $c^{(2)}$ well, they produce an non-invertible solution for the *spectral–structural* relationship, which is necessary to find the *energetic–spectral* relationship.

Modeling the polarizability and repulsive terms with high accuracy prevents an invertible solution. Instead, the singularity caused by $\partial_{\Delta r_{\mathrm{HB}}} c_R^{(2)}$ was modeled accurately with an inverse square root function while polarizability was modeled as a constant derivative. Under the constraints that eq. (4.5) be inverted and solved for Δr_{HB}, a limited number of forms is possible. The best solution was found to be

$$7\alpha^2 \Delta r_{\mathrm{HB}} + \frac{\partial_{\Delta r_{\mathrm{HB}}} c^{(2)}}{k_{\mathrm{M}}} = \frac{\mathcal{R}}{\sqrt{r_i - \Delta r_{\mathrm{HB}}}} - \frac{\mathcal{R}}{\sqrt{r_i}} - \mathcal{P} \tag{4.13}$$

where the inverse root function is chosen for an invertible antiderivative when combined with a linear function. $\mathcal{R}$ and $\mathcal{P}$ are fitted together to the full derivative in order to incorporate the minor non-linear effects of polarization into the non-linear term. Under this perspective, $\mathcal{P}$ describes an average effect of the polarized dipole, and $\mathcal{R}$ describes the rate of onset of the derivative due to

repulsion. The fitted values of each parameter is supplied in the first section of table 4.4. The value for r_i will come directly from the *energetic–structural* relationship, in which it assumed its largest Δr_{HB} (longest bond elongation) for the linear hydrogen bond arrangements. Consistency in r_i between the *energetic–structural* and the *spectral–structural* relationships is important in describing a consistent range in Δr_{HB}. The approximate fit for eq. (4.13) is shown in figure 4.1. An inverse square function also achieves a solvable solution but does not fit the partial derivative as well. The quadratic expansion term of $2z_e^2 - z_e$ ($7\alpha^2 \frac{\Delta r_{\mathrm{HB}}^2}{2}$)is lumped into this substitution since this term is important for maintaining accuracy but would make the resulting expression unsolvable for Δr_{HB}.

The antiderivative of eq. (4.13) is

$$7\alpha^2 \frac{\Delta r_{\mathrm{HB}}^2}{2} + \frac{c^{(2)}}{k_{\mathrm{M}}} \approx -2\mathcal{R}\sqrt{r_i - \Delta r_{\mathrm{HB}}} + 2\mathcal{R}\sqrt{r_i} - \left(\mathcal{P} + \frac{\mathcal{R}}{\sqrt{r_i}}\right)\Delta r_{\mathrm{HB}} \tag{4.14}$$

where the $2\mathcal{R}\sqrt{r_i}$ term ensures the function passes through the origin; $c^{(3)}$ will be approximated as a constant so that $\frac{c^{(3)}}{k_{\mathrm{M}}}\Delta r_{\mathrm{HB}}$ is linear and the solution is invertible

$$\frac{c^{(3)}}{k_{\mathrm{M}}}\Delta r_{\mathrm{HB}} \approx -\mathcal{T}\Delta r_{\mathrm{HB}} \tag{4.15}$$

The parameters in table 4.3 provide important insights into how the fits of $c^{(2)}$ and $c^{(3)}\Delta r_{\mathrm{HB}}$ change between the various dimers. The first section of the table gives the physical parameters specific to the hydrogen bond donor. The second section provides the parameters for eqs. (4.14) and (4.15). In general, $\mathcal{R}$ is negative and small except for HF$-$NH$_3$, which exhibits long hydrogen bond elongations. On the other hand, $\mathcal{P}$, which describes the average effect of polarizability, increases from 0.0134 to 0.0236 1/pm with hydrogen bond strength. Finally, $\mathcal{T}$ is much smaller but follows a similar trend to $\mathcal{P}$, increasing from 0.0009 to 0.0021 1/pm. These terms provide the basis for representing the *spectral–structural* and its inverse *structural–spectral* relationships.

Eqs. (4.14) and (4.15) can be incorporated into eqs. (4.5) and (4.6) and solved for $\Delta\omega_{\mathrm{HB}}$ and Δr_{HB} yielding

$$\frac{2\Delta\omega_{\mathrm{HB}}}{\omega_{\mathrm{M}}} - \frac{\Delta\omega_{\mathrm{HB}}^2}{\omega_{\mathrm{M}}^2} = \left(3\alpha + \mathcal{P} + \mathcal{T} + \frac{\mathcal{R}}{\sqrt{r_i}}\right)\Delta r_{\mathrm{HB}} + 2\mathcal{R}\sqrt{r_i - \Delta r_{\mathrm{HB}}} - 2\mathcal{R}\sqrt{r_i} \tag{4.16}$$

Table 4.3: Fitted parameters for extending Badger's rule to the *structural–spectral* relationship.

Parameter	NH_3−NH_3	H_2O−H_2O	HF−HF	H_2O−NH_3	HF−H_2O	HF−NH_3
α (1/Å)	2.0911	2.2670	2.3941	2.2670	2.3941	2.3941
ω_{M} (cm^{-1})	3492.32	3779.13	4026.06	3779.13	4026.06	4026.06
$\mathcal{R}$ ($1/\mathrm{pm}^{1/2}$)	-0.0148	-0.0210	-0.0122	-0.0503	-0.0354	0.0177
$\mathcal{P}$ (1/pm)	0.0134	0.0131	0.0140	0.0200	0.0203	0.0236
$\mathcal{T}$ (1/pm)	0.0010	0.0006	0.0003	0.0011	0.0011	0.0022
r_i (pm)	0.4847	0.7214	0.6156	1.7540	2.1316	8.4825
$\mathcal{L}$ (pm/100 cm^{-1})	0.7423	0.6483	0.5767	0.5940	0.5327	0.5091
$\mathcal{Q}$ (pm/(100 cm^{-1})^2)	0.0887	0.0527	0.0316	0.0293	0.0153	0.0058
$\mathcal{S}$ (pm)	-0.5065	-0.9002	-0.3313	-4.5512	-2.0197	0.9373
$\mathcal{U}$ (1/100 cm^{-1})	1.1107	0.6263	0.7676	0.1942	0.1850	0.0638
$\mathcal{V}$ (1/(100 cm^{-1})^2)	0.0159	0.0083	0.0095	0.0026	0.0023	0.0008

$$\Delta\omega_{\mathrm{HB}} = \omega_{\mathrm{M}}\left(1 - \sqrt{1 - \left(3\alpha + \mathcal{P} + \mathcal{T} + \frac{\mathcal{R}}{\sqrt{r_i}}\right)\Delta r_{\mathrm{HB}} - 2\mathcal{R}\sqrt{r_i - \Delta r_{\mathrm{HB}}} + 2\mathcal{R}\sqrt{r_i}}\right) \tag{4.17}$$

$$\begin{aligned}\Delta r_{\mathrm{HB}} =& \frac{\frac{2\Delta\omega_{\mathrm{HB}}}{\omega_{\mathrm{M}}} - \frac{\Delta\omega_{\mathrm{HB}}^2}{\omega_{\mathrm{M}}^2}}{3\alpha + \mathcal{P} + \mathcal{T} + \mathcal{R}/\sqrt{r_i}} + \frac{2\mathcal{R}\sqrt{r_i}(3\alpha + \mathcal{P} + \mathcal{T})}{\left(3\alpha + \mathcal{P} + \mathcal{T} + \mathcal{R}/\sqrt{r_i}\right)^2} \\ &\cdot \left(1 - \sqrt{1 - \frac{3\alpha + \mathcal{P} + \mathcal{T} + \mathcal{R}/\sqrt{r_i}}{(3\alpha + \mathcal{P} + \mathcal{T})^2 r_i}\left(\frac{2\Delta\omega_{\mathrm{HB}}}{\omega_{\mathrm{M}}} - \frac{\Delta\omega_{\mathrm{HB}}^2}{\omega_{\mathrm{M}}^2}\right)}\right)\end{aligned} \tag{4.18}$$

$\Delta\omega_{\mathrm{HB}}$ is found by solving the quadratic for $\Delta\omega_{\mathrm{HB}}$ similar to eq. (4.8). Δr_{HB} is also solved via the quadratic of $\sqrt{r_i - \Delta r_{\mathrm{HB}}}$, from which Δr_{HB} can be found. The entire process starts with completing the square and ends with simplifying like terms.

This complicated expression for Δr_{HB} is displayed in figure 4.3, where the proper representation of the repulsive wall creates a large curvature near r_i. The radical term has a non-zero linear and quadratic contribution at $\Delta\omega_{\mathrm{HB}} = 0$. To isolate the weak hydrogen bond relationship (polynomial-like) from the relationship near r_i, the linear and quadratic expansions of the square root term will

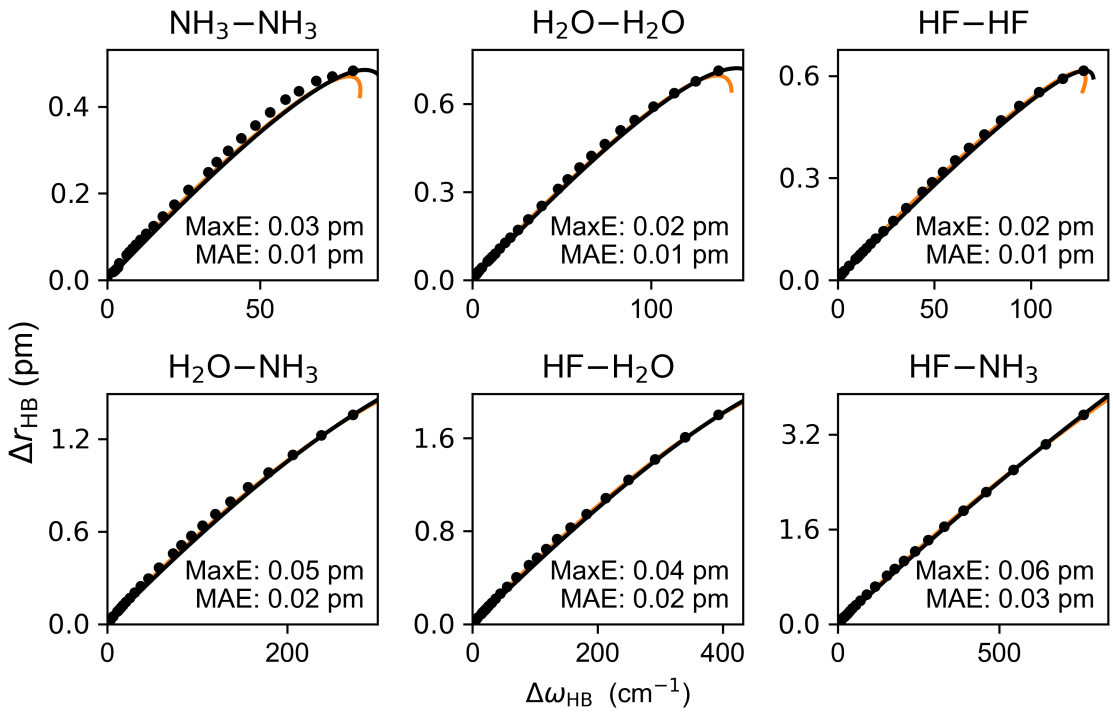

Figure 4.3: Plots of Δr_{HB} as a function of $\Delta\omega_{\mathrm{HB}}$ using eq. (4.19). Solid circles denote the values of Δr_{HB} and $\Delta\omega_{\mathrm{HB}}$ from the linear ab initio scans. The black lines show eq. (4.19) whereas the orange lines trace a parametric relationship, which shows a sharper turn-around than eq. (4.19) past $\Delta r_{\mathrm{HB}} = r_i$. The values of the MaxE and MAE errors, are shown to compare the performance of the relationship for different hydrogen-bonded dimers.

be combined with the linear and quadratic polynomial terms. The tedious variable substitution is outlined in Appendix A of ref. [211]. The final result is

$$\Delta r_{\mathrm{HB}} = \mathcal{L}\Delta\omega_{\mathrm{HB}} - \mathcal{Q}\Delta\omega_{\mathrm{HB}}^2 + \mathcal{S}\left(1 - \sqrt{1 - \mathcal{U}\Delta\omega_{\mathrm{HB}} + \mathcal{V}\Delta\omega_{\mathrm{HB}}^2} - \frac{\mathcal{U}}{2}\Delta\omega_{\mathrm{HB}} - \left(\frac{\mathcal{U}^2}{4} - \mathcal{V}\right)\frac{\Delta\omega_{\mathrm{HB}}^2}{2}\right) \tag{4.19}$$

In the above relationship $\mathcal{L}$ and Q are the linear and quadratic expansions of the *structural–spectral* relationship around $\Delta\omega_{\mathrm{HB}}$, respectively. Values for the new shortened parameters in the third part of table 4.3.

The values of $\mathcal{L}$ are relatively similar for each dimer, ranging between 0.65 and 0.51 pm/100 cm^{-1} for H_2O and HF donors. The value of $\mathcal{L}$ for $NH_3{-}NH_3$ is about 15% higher than that for

Table 4.4: Fitted parameters for extending Badger's rule to the *energetic–spectral* relationship.

Parameter	NH_3−NH_3	H_2O−H_2O	HF−HF	H_2O−NH_3	HF−H_2O	HF−NH_3
k_M (E_h/Å^2)	1.6547	1.9376	2.1991	1.9376	2.1991	2.1991
$\frac{\mu^{(0)}}{\mu^{(1)}}$ (Å)	0.4654	0.6664	0.6766	0.6664	0.6766	0.6766
ω_M (cm^{-1})	3492.32	3779.13	4026.06	3779.13	4026.06	4026.06
B (kcal/mol)	0.00	0.08	0.04	2.93	2.34	42.59
C (kcal/mol)	0.13	1.31	-0.16	-0.03	-0.19	-0.10
r_o (pm)	0.05	1.00	0.12	0.04	0.09	0.04
r_i (pm)	0.48	0.72	0.62	1.75	2.13	8.48
$\mathcal{A}$ (kcal/mol/100 cm^{-1})	3.5922	5.2607	5.3852	4.8119	4.9735	4.7536
$\mathcal{B}$ (kcal/mol/(100 cm^{-1})^2)	0.4610	0.8750	0.6730	0.4659	0.3392	0.1963
$\mathcal{C}$ (kcal/mol/(100 cm^{-1})^3)	-0.0073	-0.0727	-0.0414	-0.0225	-0.0113	-0.0033
$\mathcal{D}$ (kcal/mol/(100 cm^{-1})^4)	0.0004	0.0030	0.0011	0.0006	0.0002	0.0000
$\mathcal{X}$ (kcal/mol)	-2.4823	-9.0399	-3.3436	-63.6838	-24.5186	7.7907
$\mathcal{Y}$ (kcal/mol/100 cm^{-1})	0.0574	1.7814	0.5304	6.1047	2.0157	-0.4911
$\mathcal{Z}$ (kcal/mol/(100 cm^{-1})^2)	-0.0008	-0.0236	-0.0066	-0.0808	-0.0250	0.0061
$\mathcal{G}$ (kcal/mol/100 cm^{-1})	-1.3798	-2.8332	-1.2834	-6.1826	-2.2681	0.2484
$\mathcal{H}$ (kcal/mol/(100 cm^{-1})^2)	-0.3318	0.1518	-0.0268	0.3743	0.1097	-0.0148
$\mathcal{J}$ (kcal/mol/(100 cm^{-1})^3)	0.0056	0.0484	0.0274	0.0042	0.0024	0.0001
$\mathcal{K}$ (kcal/mol/(100 cm^{-1})^4)	-0.0001	-0.0011	-0.0005	-0.0003	-0.0001	0.0000

H_2O−H_2O. The quadratic term, $\mathcal{Q}$, correlates with the linear term decreasing from 0.089 to 0.006 with hydrogen bond strength. The negative curvature caused by $\mathcal{Q}$ and the square root term diminishes changes in the linear term between systems, creating a uniform relationship especially for each of the equilibrium dimer Δr_{HB} values.

The solution for Δr_{HB} in eq. (4.19) can be substituted into the *energetic–structural* relationship to yield an *energetic–spectral* relationship that is more robust than the one derived from the "beyond Badger" relationship, eq. (4.11). This solution is quite cumbersome and can be cast in terms of simplified parameters as

$$
\begin{aligned}
E_{\mathrm{HB}} &= \mathcal{A}\Delta\omega_{\mathrm{HB}} - \mathcal{B}\Delta\omega_{\mathrm{HB}}^2 - \mathcal{C}\Delta\omega_{\mathrm{HB}}^3 - \mathcal{D}\Delta\omega_{\mathrm{HB}}^4 \\
&+ \left(\mathcal{X} + \mathcal{Y}\Delta\omega_{\mathrm{HB}} + \mathcal{Z}\Delta\omega_{\mathrm{HB}}^2\right)\left(1 - \sqrt{1 - \mathcal{U}\Delta\omega_{\mathrm{HB}} + \mathcal{V}\Delta\omega_{\mathrm{HB}}^2}\right) - \mathcal{G}\Delta\omega_{\mathrm{HB}} - \mathcal{H}\Delta\omega_{\mathrm{HB}}^2 - \mathcal{J}\Delta\omega_{\mathrm{HB}}^3 - \mathcal{K}\Delta\omega_{\mathrm{HB}}^4 \\
&- \ln\left[\frac{r_i^B}{r_o^C}\frac{\left(r_o + \left(\mathcal{L} - \frac{\mathcal{SU}}{2}\right)\Delta\omega_{\mathrm{HB}} - \left(\mathcal{Q} + \frac{\mathcal{S}\left(\mathcal{U}^2 - \mathcal{V}\right)}{8}\right)\Delta\omega_{\mathrm{HB}}^2 + \mathcal{S}\left(1 - \sqrt{1 - \mathcal{U}\Delta\omega_{\mathrm{HB}} + \mathcal{V}\Delta\omega_{\mathrm{HB}}^2}\right)\right)^C}{\left(r_i - \left(\mathcal{L} - \frac{\mathcal{SU}}{2}\right)\Delta\omega_{\mathrm{HB}} + \left(\mathcal{Q} - \frac{\mathcal{S}\left(\mathcal{U}^2 - \mathcal{V}\right)}{8}\right)\Delta\omega_{\mathrm{HB}}^2 + \mathcal{S}\left(1 - \sqrt{1 - \mathcal{U}\Delta\omega_{\mathrm{HB}} + \mathcal{V}\Delta\omega_{\mathrm{HB}}^2}\right)\right)^B}\right] \\
&+ \left(\frac{B}{r_i} + \frac{C}{r_o}\right)\left(\left(\mathcal{L} - \frac{\mathcal{SU}}{2}\right)\Delta\omega_{\mathrm{HB}} - \left(\mathcal{Q} + \frac{\mathcal{S}\left(\mathcal{U}^2 - \mathcal{V}\right)}{8}\right)\Delta\omega_{\mathrm{HB}}^2 + \mathcal{S}\left(1 - \sqrt{1 - \mathcal{U}\Delta\omega_{\mathrm{HB}} + \mathcal{V}\Delta\omega_{\mathrm{HB}}^2}\right)\right)
\end{aligned}
\tag{4.20}
$$

The substitution of eq. (4.19) into the *energetic–structural* relationship and the simplification of like terms is outlined in Appendix B of ref. [211] with parameters supplied in table 4.4. In summary, this is done so that $\mathcal{A}$ is the purely linear relationship around $\Delta\omega_{\mathrm{HB}} = 0$ since linear terms in the square root and logarithmic terms are canceled by $\mathcal{G}$ and $\left(\frac{B}{r_i} + \frac{C}{r_o}\right)\mathcal{L}$. The term $\mathcal{B}$ and higher terms in the polynomial do not represent the full polynomial relationship around $\Delta\omega_{\mathrm{HB}} = 0$ since the logarithmic term in the *energetic–structural* relationship carries quadratic and higher derivatives around $\Delta\omega_{\mathrm{HB}} = 0$.

This expression for the *energetic–spectral* relationship is clearly more accurate than the solution that uses the "beyond Badger" relationship, as seen by comparing the results in figure 4.2 with the ones in figure 4.4. Errors are reported in table 4.1, showing the increase in accuracy obtained by using the more robust *structural–spectral* relationship. Incorporating accurate Pauli repulsion resulted in more than a two-fold increase in accuracy (decreases in MSE and MAE) in the *spectral–structural* relationship, causing the two-fold increase in accuracy seen in the *energetic–spectral* relationship. The Maximum absolute Error (MaxE) surprisingly remains relatively unchanged between the "beyond Badger" and extending Badger to the *energetic–spectral* relationship, decreasing from 0.8 kcal/mol to 0.6 kcal/mol despite the large decrease in the maximum error for the *structural–spectral* relationship. The good performance of the *energetic–spectral* relationship comes mostly from the cancellation of consistently negative errors in the *structural–spectral* relationship (due to the constant approximation of polarizability through $\mathcal{P}$) and the usually positive errors in the *energetic–structural* relationship. The largest errors in the *energetic–structural* are negative, which compound with the largest negative errors in the *structural–spectral* relationship to produce similar maximum errors despite the error cancellation elsewhere.

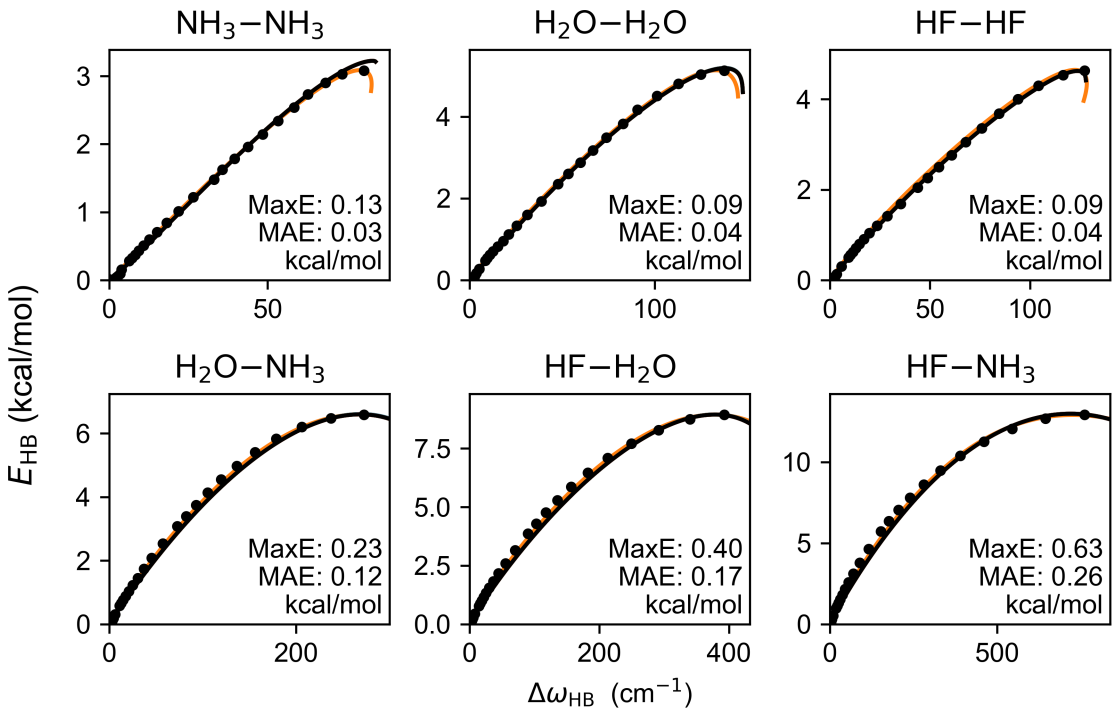

Figure 4.4: Plots of E_{HB} as a function of $\Delta\omega_{\mathrm{HB}}$ assuming the extended Badger *structural–spectral* relationship. [114] Solid circles trace the values of the hydrogen bond energy for the linear scans. The black lines show the relationship from eq. (4.20), whereas the orange lines trace a parametric relationship. The values of the MaxE and MAE errors are shown to compare the performance of the relationship for different hydrogen-bonded dimers.

The linear term $\mathcal{A}$ is the dominant term for the polynomial expression, which governs behavior for small red-shifts. It is consistent for all H_2O and HF donors, with values between 4.75 – 5.38 kcal/mol/100 cm^{-1}. The quadratic term $\mathcal{B}$ neglects quadratic contributions from the polarizability effect on energy but maintains a sizable contribution to the hydrogen bond energy at moderate red-shifts (100 cm^{-1}). The higher order terms $\mathcal{C}$ an $\mathcal{D}$ grow increasingly negligible for smaller red-shifts and only marginally affect hydrogen bond energies for more red-shifted hydrogen bonds (>500 cm^{-1}) as the curvature in this region is more strongly affected by the square root and logarithm terms. The magnitude of each non-linear polynomial term decreases with hydrogen bond strength (except NH_3−NH_3), a fact that arises from the increased linearity of the *structural–*

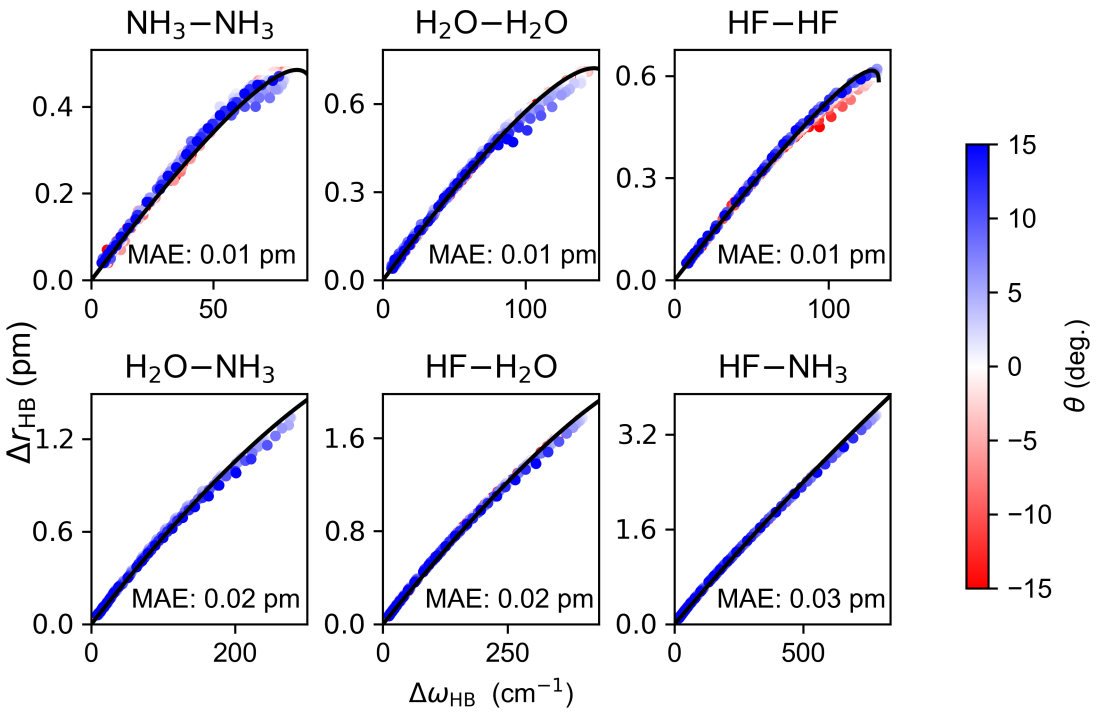

Figure 4.5: Plots of Δr_{HB} as a function of $\Delta\omega_{\mathrm{HB}}$. The solid circles denote the values of the hydrogen bond energies from the non-linear scans. The solid lines trace the relation from eq. (4.19). The solid circles are color-coded based on their deviation (in degrees) from linearity. The MAE is reported as a metric of the accuracy of eq. (4.19).

spectral relationship. In contrast, the *energetic–spectral* relationship is highly non-linear, with the weak hydrogen bonds having large quadratic contributions and sharp turnovers near the equilibrium geometry, and the strong hydrogen bonds having large gradual curvature from Pauli repulsion.

4.3.4 Validation of the Energetic–Spectral Relationship for Non-linear Hydrogen Bonds

The previous analysis based on the PESs restricted to linear hydrogen bonds can be extended to non-linear ones. The *structural–spectral* relationship, eq. (4.18), also holds for non-linear hydrogen bonds. As shown in figures 4.3 and 4.5, the MAEs are comparable between linear and non-linear hydrogen bonds. There is a small reduction in the accuracy for the most deformed hydrogen bonds, but the MAE remains the same. This is not surprising since the previous study [114] established the

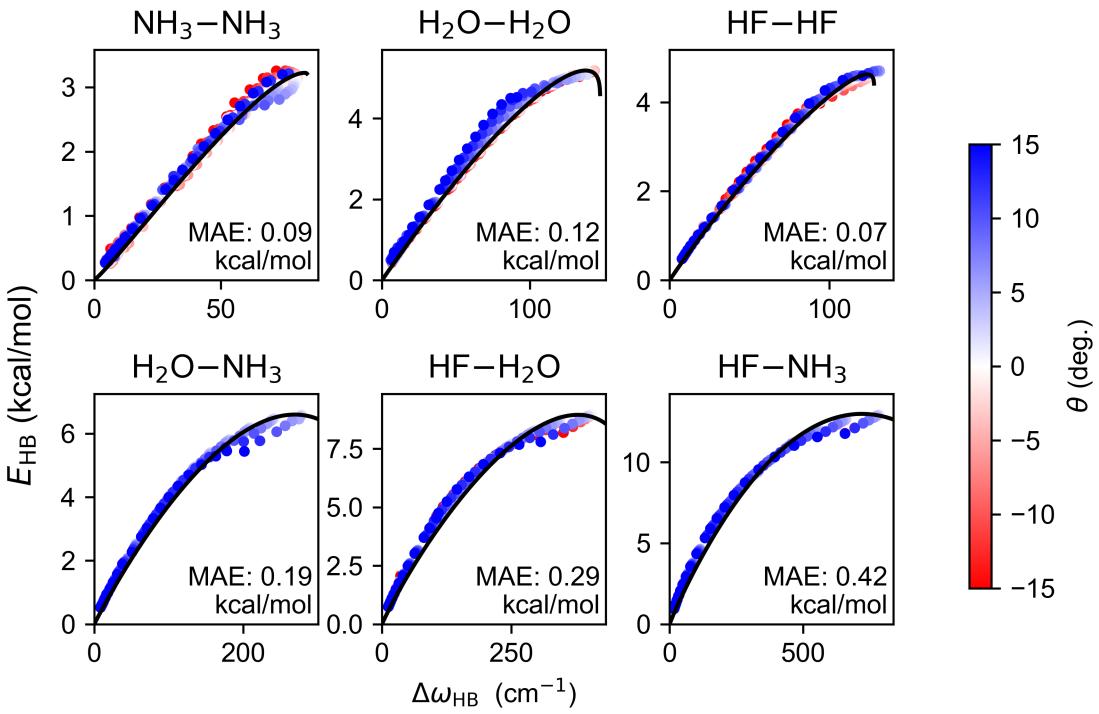

Figure 4.6: Plots of E_{HB} as a function of $\Delta\omega_{\mathrm{HB}}$. The solid circles denote the values of the hydrogen bond energies from the non-linear scans. The solid lines trace the relation from eq. (4.20). The solid circles are color-coded based on their deviation (in degrees) from linearity. The MAE is reported as a metric of the accuracy of eq. (4.20).

"beyond Badger" relationship by considering aqueous ionic clusters that also contained non-linear hydrogen bonds.

Without a significant loss of accuracy in the *structural–spectral* relationship with hydrogen bond angle, the *energetic–spectral* relationship remains accurate for non-linear hydrogen bonds, as can be seen from figure 4.6. The relationship is slightly less accurate when considering non-linear hydrogen bonds than the one when restricting it to linear hydrogen bonds, with an approximate doubling in the MAE. This was also seen previously for the *energetic–structural* relationship in the prior chapter, for which the error increases where the relationship was non-linear. The MAE in predicting the hydrogen bond energy via the *energetic–spectral* relationship is slightly larger than when predicting it via the *energetic–structural* relationship for both linear and non-linear hydrogen bonds, a fact

Table 4.5: Experimental (exp.) and theoretical (calc.) harmonic $\Delta\omega_{\mathrm{HB}}$, anharmonic $\Delta\nu_{\mathrm{HB}}$ vibrational shifts and bond energies E_{HB} (ΔD_{e}) for the various dimers. The MP2/aug-cc-pVTZ values obtained in this work are indicated. Bond energies including zero-point corrections (ΔD_0) are also listed.

	NH_3-NH_3	H_2O-H_2O	HF−HF	H_2O-NH_3	HF−H_2O	HF−NH_3
$\Delta\nu_{\mathrm{HB,exp.}}$ (cm^{-1})	60[a] (46)[212]	105 [213]	95 [180]	221 [178]	355 [181]	748 [214]
	41 (in He) [215]	120 (in Ne) [213]	98 (in He) [216]	242 (in He) [217]	n/a	n/a
$\Delta\nu_{\mathrm{HB,calc.}}$ (cm^{-1})	36 [179]	99 [218, 219]	94 [220]	(288 [221])[b]	330 [222]	740 [223]
$\Delta\omega_{\mathrm{HB,calc.}}$ (cm^{-1})	58 [215]	134 [102]	135 [220]	245 [224]	359 [225]	797 [226]
$E_{\mathrm{HB,exp.}}$ (ΔD_{e}, kcal/mol)	3.32 [179]	5.06 [101, 213]	4.59 [180, 227]	6.63 [178, 224]	10.25±0.19 [228]	n/a
$E_{\mathrm{HB,calc.}}$ (ΔD_{e}, kcal/mol)	3.25 [229]	4.99±0.04 [103, 230]	4.58 [227]	6.44 [224]	8.51 [222]	12.3 [226]
$\Delta\omega_{\mathrm{HB,this\ work}}$ (cm^{-1})	63	168	134	297	406	785
$E_{\mathrm{HB,this\ work}}$ (kcal/mol)	3.29	5.18	4.71	6.71	8.96	12.93
$\Delta D_{0,\mathrm{exp.}}$ (kcal/mol)	1.88± 0.06 [179]	3.16± 0.03 [101]	3.03 [180]	4.4 ± 0.03 [178]	8.20 [228]	n/a
$\Delta D_{0,\mathrm{calc.}}$ (kcal/mol)	1.85 [229]	3.21 [218, 230]	3.02 [220]	4.20 [224]	6.3 [222]	9.15 [214]

a) De-perturbed by estimating the effects of Fermi resonances.

b) Using a reduced dimensionality model for the dimer.

that originates from compounding errors in the *structural–spectral* relationship. The errors seen for the non-linear hydrogen bonds are generally acceptable and provide a predictive, physically motivated relationship for estimating hydrogen bond energies from vibrational frequencies.

4.4 Estimation of Hydrogen Bond energies from Experimental Spectra

Using the developed *energetic–spectral* relationship, eq. (4.20), hydrogen bond energies can be estimated from experimental spectra and in addition theoretically calculated bond energies can be correlated to the calculated red-shifts in the DH vibrational frequency. The anharmonic vibrational frequencies associated with the hydrogen bond for all six dimers considered here have been experimentally measured, however the bond energies of only five out of the six are available experimentally, [101, 178–181] as shown in table 4.5. Naturally, the full set of harmonic and anharmonic vibrational frequencies and bond energies are available theoretically for all six dimers.

The comparison between the experimental/theoretical hydrogen bond energies obtained from table 4.5 and those from eq. (4.20) using the experimental or calculated $\Delta\omega_{\mathrm{HB}}$ shifts is shown in figure 4.7. To estimate bond energies from vibrational frequencies when the harmonic monomer and

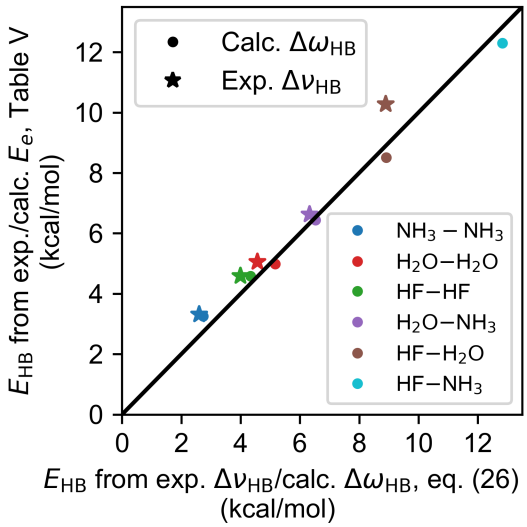

Figure 4.7: Correlation between predicted and observed hydrogen bond energies. Each $\Delta\omega_{\mathrm{HB,calc.}}$ (spheres) or $\Delta\nu_{\mathrm{HB,exp.}}$ (stars) from table 4.5 is converted to an E_{HB} and compared against the corresponding dimer binding energies (ΔD_e from table 4.5).

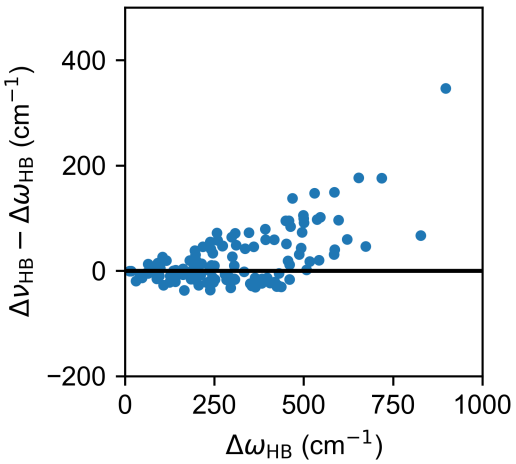

Figure 4.8: The differences in anharmonic redshifts compared against harmonic redshifts. On average, anharmonic redshifts are slightly larger, but differences remain small below 500 cm^{-1}. Above 500 cm^{-1}, errors grow.

dimer vibrational frequencies are not available experimentally, we will rely on the approximation that the shifts with respect to the monomer for the measured anharmonic frequencies are comparable to the shifts of the corresponding harmonic frequencies. Past results for water clusters show only small differences between harmonic and anharmonic frequency shifts for those shifts less than 500 cm^{-1}. [114] The difference in harmonic and anharmonic frequency shifts are plotted against harmonic frequency shifts for all reported stretches from Boyer et al. in figure 4.8. [114] For all shifts less than 500 cm^{-1}, the mean absolute difference between the harmonic and anharmonic shifts is 26.5 cm^{-1} and the root mean square difference is 37.1 cm^{-1}. As can be seen from the calculated values listed in table 4.5 this approximation ($\Delta\omega_{\mathrm{HB}} \approx \Delta\nu_{\mathrm{HB}}$) is reasonable for the systems considered. Therefore, we can assume small errors when using measure anharmonic shifts in the place of harmonic ones to estimate the bond energies.

The bond energies of five of the six dimers considered here have been experimentally measured. [101, 178–180, 213, 224, 227, 228] Vibrational shifts are collected from their spectra and used to predict hydrogen bond energies via eq. (4.20) and the parameters listed in tables 4.3 and 4.4. The reference (monomer) vibration for H_2O donating dimers is the average of the antisymmetric and symmetric OH stretches. The reference (monomer) vibration for NH_3 donating dimers is the average of the symmetric and two antisymmetric NH stretches. For the ammonia dimer we used the average of the two hydrogen bonded NH frequencies, since two normal modes show large bound NH stretching motions.

Experimentally estimated dimer energies (D_0) incorporate zero-point effects, which are not captured in the current model and therefore they are not directly comparable to the ones derived using our scheme. Although Δr and R capture some important changes in the zero-point energy, vibrational frequencies in other degrees of freedom contribute significantly to changes in the zero-point energy. Previous studies estimated the zero point effects and produced electronic dissociation energies (D_e), which are reported in table 4.5. For the $H_2O{-}H_2O$, HF−HF, and $H_2O{-}NH_3$ dimers, changes in zero point corrections were estimated from the spectral frequencies and applied to the experimentally deduced bond energies (D_0). To determine the D_e of the $NH_3{-}NH_3$ dimer, a 42% zero-point contribution to the binding energy is assumed from diffusion Monte Carlo simulations. [231] The experimentally measured D_e and estimated D_0 values for $NH_3{-}NH_3$, $H_2O{-}H_2O$, HF−HF, and $H_2O{-}NH_3$ agree well with the results of accurate electronic structure calculations.

The hydrogen bond energy from eq. (4.20) derived using the experimental and theoretical values of $\Delta\omega_{HB}$ for HF$-$H$_2$O agrees well with the CCSD(T)/CBS result by Tschumper and co-workers. [222] This value is smaller by ∼20% from the previously reported experimental estimates of Thomas [181] and Legon *et al.*, [228] who suggested a much stronger binding energy ($D_0 = 8.20$ kcal/mol, $D_e = 10.25$ kcal/mol) listed in table 4.5. However, these experimental estimates are old and not based on vibrational predissociation, possibly casting doubt in the accuracy of the reported energetics.

The ammonia dimer hydrogen bond energy predicted from experimental frequencies is further away from the measured or the calculated values compared to the other dimers. The experimental NH frequency for NH$_3-$NH$_3$ is known to be in Fermi resonance with the NH bends, causing a 38% reduction in the vibrational shift. [215, 232] The derivation in Section III is based on harmonic frequencies and thus large changes in the frequencies due to anharmonic effects can cause significant errors. To address this issue, we estimated a de-perturbed NH frequency, i.e. one without Fermi resonances, through a weighted average of the two strongest transition in resonance (B: 3315 cm^{-1} and C: 3255 cm^{-1}). [212] Since the lower frequency transition is ∼25% less intense than the higher frequency one, it's assumed that the higher transition is ∼ 57% of the NH stretch and ∼ 43% of the NH bend. [233] This produces a 10 cm^{-1} splitting for the de-perturbed frequencies, compared to 60 cm^{-1} for the perturbed ones. The smaller splitting yields a de-perturbed NH frequency of 3290 cm^{-1} and a $\Delta\nu_{HB} = 60$ cm^{-1}, which accurately predicts a hydrogen bond energy of 2.6 kcal/mol in much better agreement with experiment. In contrast, the use of $\Delta\nu_{HB} = 46$ cm^{-1} based on the perturbed frequencies yields a less accurate hydrogen bond energy (2.0 kcal/mol).

4.5 Conclusions

This chapter describes the development of the third and final part of the proposed *energetic–structural–spectral* relationship for hydrogen bonds. The first part consisted of going "beyond Badger's" rule to develop the *spectral–structural* relationship. [114] The second part "extended Badger's" rule to develop the *energetic–structural* relationship, [210] where in the present study the extension of Badger's rule was applied to develop the final *energetic–spectral* relationship.

Following the procedure introduced in the prior chapter, we relied on the ab initio PESs of six hydrogen bonded systems to develop the *energetic–spectral* relationship for hydrogen bonds between

charge neutral monomers initially under the constrain of linear hydrogen bonds and subsequently verified it for the non-linear hydrogen bonds that are present in the respective minimum energy structures. The relationship is fairly uniform for the six dimers studied here, consisting of first-row hydrides. The linear region of the relationship (for low values of $\Delta\omega_{\mathrm{HB}}$) that describes the purely electrostatic hydrogen bond has the same slope regardless of the identity of the hydrogen bond donor and the hydrogen bond acceptor due to the similar value of $k_{\mathrm{M}}\frac{\mu^{(0)}}{\mu^{(1)}}$ for the six dimers considered here as seen in the *energetic–structural* relationship, and a similar value of $\frac{2}{\omega_{\mathrm{M}}(3\alpha+\mathcal{L})}$ in the *structural–spectral* relationship. However, the developed relationship does deviate for the different dimers near the equilibrium geometry due to the logarithmic terms. In contrast, the *structural–spectral* relationship does not deviate much for the six dimers since $\mathcal{L}$ remains similar and non-linear effects are minimal for all dimers considered here.

The final form of the developed relationship is quite complex, since it describes energy changes as the hydrogen bond forms and the geometry approaches the minimum. The overall shape of the relationship is formed by the linear term ($\mathcal{A}$), the quadratic term ($\mathcal{B}$), and the logarithmic-repulsive term (containing r_i and B) in eq. (4.20). Other terms are small and serve only to improve the accuracy of the relationship.

We applied the developed relationship to estimate hydrogen bond energies using either theoretical or experimental vibrational shifts and compared them with the reported experimental/theoretical predictions. Overall, the agreement is quite satisfactory. However we reiterate the concerns arising from large anharmonic effects, which can be difficult to model especially when Fermi resonances result in the splitting of the spectral peaks. This splitting can be estimated from the intensity sharing in the experimental spectra and can be used to obtain a de-perturbed anharmonic frequency, which was found to produce accurate results via the developed model.

It should be noted that the developed relationship works well for areas around the equilibrium heavy atom separation (R). Extensions will need to be considered when the dimer bond energy is perturbed by an external field due to a nearby molecule or ion that ultimately results in a large perturbation of the equilibrium heavy atom separation due to, for instance, the existence of large non-additive effects. Except for highly constrained intramolecular hydrogen bonds, in which other interactions force large heavy atom distances, most gas phase hydrogen bonds exist close to the equilibrium heavy atom distance. Intramolecular, which is studied in the next chapter, constraints

may force the acceptor and donor motifs to be distant, and it is expected that this model will also describe those weaker hydrogen bonds quite well. However, electrostatic perturbations, such as external fields, induction effects, additional hydrogen bonds or strong hydrogen bonds in ionic aqueous clusters [234–236] can effect the hydrogen bond potential and consequently the hydrogen bond geometry, and thus alter the parameters in this relationship. Additional hydrogen bonds and ionic hydrogen bond acceptors in the water clusters used to derive the "beyond Badger" relationship [114] in many instances produced red-shifts beyond what is allowed by eq. (4.17), viz. 147 cm^{-1} for $H_2O{-}H_2O$ and 447.8 cm^{-1} for $H_2O{-}NH_3$). Changes to the repulsive term of the *energetic–structural* relationship allow for larger red-shifts in the presence of a stronger hydrogen bond acceptor, which is currently under investigation. Preliminary results suggest additional parameterization is required, specifically for hydrogen bonds greater than 8 kcal/mol but less than 12 kcal/mol. Therefore, describing how these changes to the PESs systematically change the equilibrium structure, equilibrium vibrations, and equilibrium energy may contribute in unifying the changes seen for the non-linear hydrogen bonds.

Chapter 5

ON THE STRENGTH OF *OUROBOROS* HYDROGEN BONDS

5.1 Introduction

Hydrogen bonds (HBs) are one of the most fundamental and ubiquitous non-covalent interactions in nature. They greatly influence the structure of hydrogen bonded networks by dominating the physics behind cluster formation, [237] influencing the structures in phase-separated micelles and bi-lipid membranes, [238] and creating elusive, *ouroboros* (*intra*-molecular) HBs. This class of HBs is known to compete with steric interactions causing structural and functional changes in medicinal organic molecules, [239] stabilizing the structure of bio-polymers such as proteins and DNA, [11, 15] as well as determining the secondary and tertiary structures in peptides including fibrils. [11, 240] Intramolecular HBs (IMHBs) also play a role in self-healing thermoplastic polymers and elastomers by affecting their strength and reversibility. [16, 241, 242] Understanding the nature and most importantly the strength of these non-traditional HBs will enhance the molecular level understanding of important phenomena and processes observed in biochemistry and materials engineering.

The strength of traditional or classical, *inter*-molecular HBs, is well defined, as seen in the prior two chapters. For other systems exhibiting more complex interactions such as strong dispersion, π–π, or additional electrostatic interactions in addition to single or multiple HBs, the HB strength is more difficult to be singled out as a fraction of the total dissociation energy. This approach cannot be directly applied to non-traditional, *ouroboros* HBs as these cannot be uniquely separated into distinct fragments without the breaking of covalent bonds. In other words, the zero reference energy level is not uniquely defined and consequently the strength of the IMHB cannot be uniquely evaluated. [208] The strengths of these elusive HBs are of paramount importance in determining the properties of organic chemical systems as well as proteins and polymers.

Various approaches have been previously suggested to estimate the strength of IMHBs with varying success as described in a comprehensive recent review. [208] The review succinctly sum-

marizes the methods used to estimate the strength of IMHBs and their classification based on the approach used, viz. conformational methods and their variants, rotation barrier methods (RBM), [243], dimer models, isodesmic reaction methods, Quantum Theory of Atoms in Molecules (QTAIM) based methods [244–246] and finally empirically based methods.

Among the above methods, the RBM, [243] one of the most popular approaches, approximates the strength of an intramolecular hydrogen bond from the rotational barriers of the donor and/or the acceptor groups involved. The stationary points, if any, along the rotational coordinate can be indirectly attributed to the HB energy and torsional effects. Torsional effects are subtracted through calculations of a reference molecule, and the HB energy is estimated. Errors are introduced through differences between the molecule and its reference. Although some IMHB configurations for the RBM can be accessible experimentally, the application of the RBM is greatly limited due to the inability to experimentally measure high energy "broken" structures. [247]

The approach based on Quasi-Atomic Orbitals (QUAOs), which uses the kinetic energy interaction between the orbitals involved in the bonding to estimate the energy of that interaction, previously employed to estimate covalent and hydrogen bond energies, [248–251] has also been recently used to compute the strength of intramolecular HBs. [252] The above approaches as well as others [253, 254] require numerous expensive energy evaluations and are based on ad-hoc approximations with have no direct experimental verification or direct correspondence to an experimental observable. As explicitly noted in the previous review [208], "unlike in its intermolecular counterpart, breaking the intramolecular interaction is impossible without disturbing the structure of the molecule. Because of this fact, not only it is impossible to find a strict definition of the intramolecular interaction energy, but what is more, this energy is not even strictly definable."

This thesis has focused on the last class of empirical methods described in the previous review. [208] Specifically, we seek to identify relations between the strength of a HB and the shift in the corresponding vibrational frequency, in the spirit of Iogansen's relationship [189], viz.

$$E_{\mathrm{HB}} = 0.33 \text{ kcal/mol}\sqrt{\Delta\nu_{\mathrm{HB}}/\mathrm{cm}^{-1} - 40} \tag{5.1}$$

where $\Delta\nu_{\mathrm{HB}}$ is the shift in of the hydrogen bonded frequency from its unbound value. This work contrasts other approaches [243, 248–250, 253] by using a single experimental observable, namely the vibrational frequency of the HB ($\Delta\nu_{\mathrm{HB}}$), to construct a scale that is used to estimate the

strength of elusive hydrogen bonds from the corresponding frequency shifts.

Before Iogansen, Badger has related the length and identity of covalent bonds to their force constant, a property which is directly related to spectroscopy, creating a *spectral–structural* relationship. [183, 184] Recent extensions of this idea for hydrogen bonded systems have found a linear relationship between the HB covalent distances and corresponding vibrational frequencies. [97, 114, 186, 255, 256] Badger and Bauer extended past work to develop a linear relationship between the shift in the frequency and the enthalpy of the corresponding HB, a *energetic–spectral* relationship. [185] This relationship has been heavily explored by others where different trends are found for different systems in different environments. [190, 200, 202–204, 209, 255–262] For the general alcohol, the following linear relationship was found to be valid.

$$\Delta H_{\mathrm{HB}} = 1.27 \ \frac{\mathrm{kcal/mol}}{100 \ \mathrm{cm}^{-1}} \Delta\nu_{\mathrm{HB}} + 0.51 \ \mathrm{kcal/mol} \tag{5.2}$$

With both the Badger-Bauer and the Iogansen relationships proposed, there is debate on the appropriate functional form of the relationship [185, 189]. Iogansen's version of this relationship [189] has been applied to IMHBs in nucleotides, [263–265] but the relationship was never validated for either IMHBs or the hydrogen bonding interactions in nucleotides. [189, 208] However, the change in the DH frequency has also been shown to correlate linearly with the Hammett parameters [266, 267] and the orbital populations [268] for systems with IMHBs, both being qualitative metrics of the strength of the HB and supporting the plausibility of the *energetic–spectral* relationship being applicable to intramolecular, *ouroboros* HBs as well.

Chapters 3 and 4 have explored the physical origins of the *energetic–structural–spectral* relationship by creating a simple model of the hydrogen bond and deriving a relationship from the parameters of the model. These works studied six simple hydrogen bonded dimers, namely $NH_3{-}NH_3$, $H_2O{-}H_2O$, HF−HF, H_2O-NH_3, HF−H_2O, and HF−NH_3. The electrostatic contributions to the *energetic–structural* and *energetic–spectral* relationships were mostly linear and uniform between the hydrogen bonded systems, being 8.0 kcal/mol per pm and 4.5 kcal/mol per 100 cm^{-1}. Non-linear effects from exchange repulsion were important for strong hydrogen bonds, but non-linear terms differed for each dimer. These relationships require individual parametrization to describe the repulsive part of the potential, making it untenable to describe the complicated and hard to parameterize *ouroboros* hydrogen bonds. The empirical relationships by Badger and Iogansen

suggest that a universal relationship should exist. The non-linear effects are believed to be unified through the Badger-Bauer relationship, where the non-linear effects averaged into a less steep linear term; the result is a non-physical, non-zero intercept. The same should hold true for the Iogansen relationship.

This chapter explores the empirical correlation between the energy of a HB and its underlying structural and spectral features to estimate their strength. We have relied on a set of inter-molecular OH, NH and FH donors with various N, O, and F acceptors, in order to arrive at an empirical the relationship between strength and underlying frequency shift, viz. an *energetic–spectral* relationship, that is uniform across different hydrogen bond donors noting that the different analytic relationships from chapters 3 and 4 are specific to each individual hydrogen bond. We use those analytic relationships to justify the choice of our empirical function, and we fitted the function to a set of 100 hydrogen bonded dimers. We discussed ways to extend the relationship to a wide variety of hydrogen bond donors by looking at redshifts in isolated monomers. To validate the relationship for the more elusive IMHBs, we have compared the relationship's predicted HB energy with the results of RBM [243] for two sets of IMBs for which there is no controversy regarding the definition of the reference energy with respect to which the strength of the intra-molecular HB is evaluated. This validated relationship forms the basis for developing a scale to estimate the strength of *ouroboros* HBs, such as the ones between biomolecules and DNA base pairs, from either the measured experimental spectra.

5.2 Computational Details

The training set consisted of the optimized geometries and vibrational frequencies of 100 inter-molecular hydrogen bonded dimers computed at the MP2 level of theory [144, 145, 147] with the aug-cc-pVDZ and aug-cc-pVTZ basis sets [194, 195] using Gaussian-16. [173] These dimers include OH, NH, and FH donors with various N, O, and F acceptors. Each dimer was carefully chosen to be interacting only through a single HB with minimal or no additional interactions. Harmonic (not scaled) and anharmonic frequencies at the second order of vibrational perturbation theory (VPT2) [269, 270] were calculated. The average of the symmetric and asymmetric monomer frequencies was used as a reference for symmetric monomers (eg. H_2O). Dimers where the hydrogen bonding stretch was involved in multiple stretching normal modes (such as NH_3 and CH_3NH_2 donor dimers)

were also averaged.

The validation of the *energetic–structural–spectral* relationship obtained from the set of inter-molecular HBs was further validated for the intra-molecular HBs present in a set of 10 different substituted enolones and 10 different substituted aminoalcohols by comparing the prediction with the corresponding result obtained from the RBM. [243] The calculations for these systems were carried out at the MP2/aug-cc-pVDZ [144, 145, 147, 194, 195] level of theory using Gaussian-16. [173] The HB strengths from RBM [243] were determined by considering a 180° rotation of the alcohol group in the enolone/enol and aminoethanol/ethanol. Structures were fully optimized and both electronic and zero-point corrected hydrogen bond energies were computed viz

$$E_{\mathrm{RBM}} = E_{\mathrm{open}} - E_{\mathrm{closed}} - (E_{R_{\mathrm{open}}} - E_{R_{\mathrm{closed}}}) \tag{5.3}$$

where R_{open} and R_{closed} are the reference conformations.

5.3 Results and Discussion

5.3.1 Development of an Empirical Relationship from Dimers

With the aim of creating a universal *energetic–spectral* relationship for intra-molecular hydrogen bonds, the linear Badger-Bauer and square-root Iogansen relationships for dimers were revisited. As shown throughout this work, the Iogansen relationships provides a more accurate description of the *energetic–structural–spectral* relationship, likely because it better models the non-linear shape of the relationships found in the prior chapters. In this work, we will fix the y-intercept of the Iogansen relationship to be zero, viz.

$$\Delta E = a\sqrt{\Delta r(\text{or } \Delta\omega) + b} - a\sqrt{b} \tag{5.4}$$

where a and b are fitted to the ab initio data collected for gas-phase dimers. We will also examine the linear relationships for comparison, since a linear relationship is more interpretable and provides insights into how to unify it. The elongation of the covalent bond (Δr_{HB}), the redshift in harmonic frequencies ($\Delta\omega_{\mathrm{HB}}$), and energy of the hydrogen bond (E_{HB} and $E_{0,\mathrm{HB}}$) were investigated for gas phase dimers in this section.

The same *energetic–structural–spectral* relationship was found to hold for all dimers consisting of H_2O, NH_3, and HF independently of hydrogen bond acceptors. For the linear relationship,

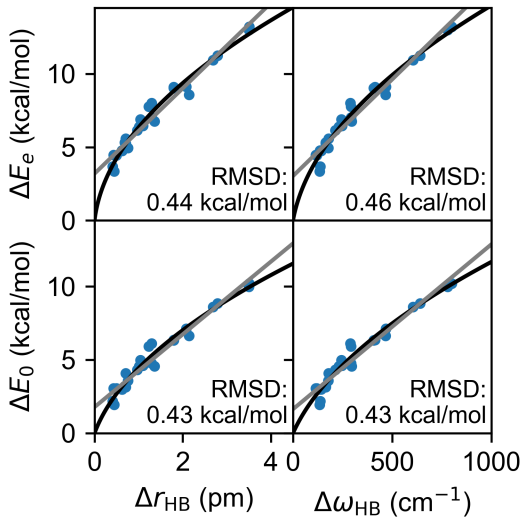

Figure 5.1: The linear (gray) and square-root (black) *energetic–structural–spectral* relationship for intermolecular hydrogen bonds with H_2O, NH_3 and HF donors. RMSDs describe the square-root relationships.

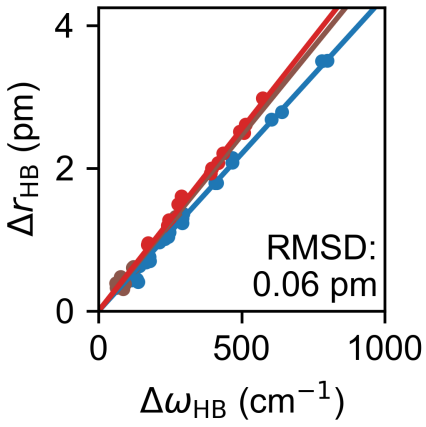

Figure 5.2: The linear *structural–spectral* relationship for intermolecular hydrogen bonds. The blue points and line describe H_2O, NH_3 and HF donors. The brown points and line describe MeOH, and $NH_X(Me)_Y$ donors. The red points and line describe electron poor OH donors.

the hydrogen bond energy increases by 2.91 kcal/mol every pm or 1.32 kcal/mol every 100 cm^{-1}. The zero-point corrected energies increase slightly lower, 2.48 kcal/mol per pm or 1.12 kcal/mol per 100 cm^{-1}, because it is known that the zero-point corrections increase as hydrogen bonds get stronger. The Iogansen relationship shows a similar trend, where zero-point corrections lowered a and increase b, both flattening the square root relationship. The *structural–spectral* relationship was also found to be uniform for this dataset (blue in figure 5.2), with a slope of 0.44 pm per 100 cm^{-1} or 22.7 cm^{-1} per 1000 Å. This value is in agreement with the previous study of the beyond Badger *structural–spectral* relationship in water clusters. [114] The uniformity of a linear relationship between H_2O, NH_3, and HF suggests that the non-linear terms in the extending Badger relationships can be combined when focusing on the equilibrium geometries. The *energetic–spectral* relationship for H_2O, NH_3, and HF donors found here

$$\Delta E_0 = 0.44 \text{ kcal/mol}\sqrt{\Delta\omega_{\text{HB}}/\text{cm}^{-1} + 39} - 2.75 \text{ kcal/mol} \tag{5.5}$$

is similar to that found by Iogansen,eq. (5.1); differences likely stem from forcing this relationship through the origin.

When extending the linear relationship to more hydrogen bond donors, the same slope was observed with larger intercepts. The intercept corresponding to various OH hydrogen bonds correlates strongly to the monomer covalent bond distance and vibrational frequency of the hydrogen bond donor. The perturbation in the electronic structure of the OH bond by an electron withdrawing or donating group that shifts the vibrational frequency and changes the covalent bond distances similarly increases the intercept of the linear Badger-Bauer relationship. It is unclear if this results in increasing the electrostatic contributions to the *energetic–structural–spectral* relationships ($k_{\text{M}}\frac{\mu^{(0)}}{\mu^{(1)}}$) or in weakening the non-linear repulsion terms. This chapter will utilize the changing properties of the hydrogen bond donor to unify the relationship. The relationship between the hydrogen bond energies and the changes in the hydrogen bond donor frequencies and associated structure can be added to the linear *energetic–structural–spectral* relationship to create a universal linear relationship. Doing so creates an "adjusted" hydrogen bond elongation and an "adjusted" hydrogen bond redshift that can be applied to the square root Iogansen-like relationship.

Hydrogen bond dimers with donors including H_2O, NH_3, HF, MeOH, $NMeH_2$, NMe_2, $(CH_2)_2NH$, FOH, H_2O_2, $HO_2^{\cdot}$, HNO_3, $HClO_4$, ethenol, 3-hydroxy-2-propenal, formic acid, acetic acid, for-

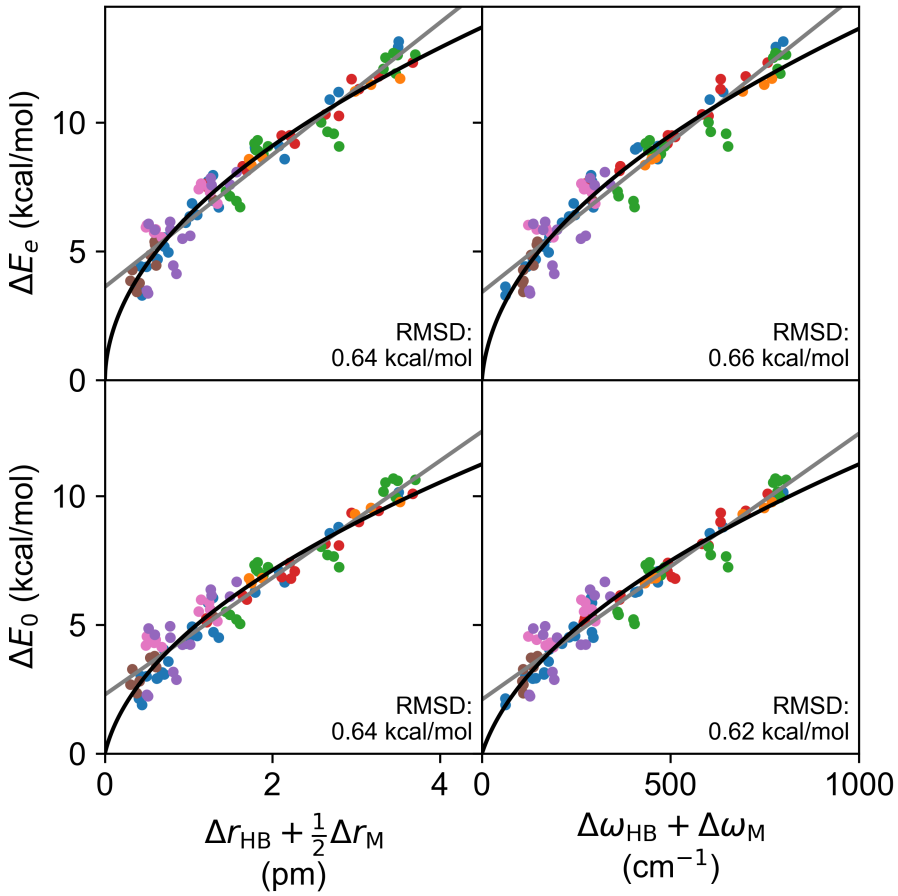

Figure 5.3: The linear (gray) and square-root (black) *energetic–structural–spectral* relationship for intermolecular hydrogen bonds. Included in the training of this relationship are H_2O, NH_3, and HF donors (blue), MeOH and $NH_X(Me)_Y$ donors (brown), misc. electron poor OH donors (red), enol and enolone donors (green), carboxylic acid donors (orange), amide donors (pink), and imine, vinyl amines, and pyrrole donors (purple).

mamide, N-methyl-formamide, acetamide, N-methyl-acetamide, methanimine (azomethine), ethenamine (vinyl amine), diethenamine, and pyrrole were used to parameterize the final relationships, shown in figure 5.3. This set is robust, with a wide variety of donors with diverse functionalization. Many functional groups from important IMHBs are included, showing that this relationship should be transferable to the organic systems that form intermolecular hydrogen bonds. The square-root *energetic–structural* relationships are

$$\Delta E_e = 6.54 \ \frac{\text{kcal/mol}}{\sqrt{\text{pm}}} \sqrt{\Delta r_{\text{HB}} + 0.5 \Delta r_{\text{M}}} \tag{5.6}$$

and

$$\Delta E_0 = 5.91 \ \frac{\text{kcal/mol}}{\sqrt{\text{pm}}} \sqrt{\Delta r_{\text{HB}} + 0.5 \Delta r_{\text{M}} + 0.05 \text{ pm}} - 1.35 \text{ kcal/mol} \tag{5.7}$$

where Δr_{M} is the elongation of the OH and NH bond of the hydrogen bond donor compared to water and ammonia respectively. The scale of Δr_{M} was held at 0.5 for simplicity; the optimal value was 0.37, and minimal errors are induced by assuming a value of 0.5. Δr_{M} of enol, enolone, carboxylic acid, amide, and enamines donors are shifted by an additional 1.1 pm to account for the conjugation with the donating group in the hydrogen bond donor. This value was determined by the average elongation in the CH bond lengths between ethylene and ethane, ethanol and enol, and ethylamine and vinylamine.

The square-root *energetic–spectral* relationships are

$$\Delta E_e = 0.45 \ \frac{\text{kcal/mol}}{\sqrt{\text{cm}^{-1}}} \sqrt{\Delta \omega_{\text{HB}} + \Delta \omega_{\text{M}} + 1.5 \text{ cm}^{-1}} - 0.55 \text{ kcal/mol} \tag{5.8}$$

and

$$\Delta E_0 = 0.41 \ \frac{\text{kcal/mol}}{\sqrt{\text{cm}^{-1}}} \sqrt{\Delta \omega_{\text{HB}} + \Delta \omega_{\text{M}} + 22 \text{ cm}^{-1}} - 1.95 \text{ kcal/mol} \tag{5.9}$$

where $\Delta \omega_{\text{M}}$ is the redshift of the OH and NH bond of the hydrogen bond donor compared to water and ammonia respectively. The scale of $\Delta \omega_{\text{M}}$ was held at 1 for simplicity; the optimal value was 0.8. $\Delta \omega_{\text{M}}$ of enol, enolone, and carboxylic acid, amide, imine, and vinyl amine donors are shifted by an additional 150 cm^{-1} to account for changes in the orbital character of the hydrogen bond donor, also derived from the CH stretches between ethylene and ethane, ethanol and enol, and ethylamine and vinylamine. The relationships are strong, with an RMSD around 0.64 kcal/mol.

The past two chapters showed that the *energetic–structural–spectral* relationship deviates for strong hydrogen bonds, figures 3.7 and 4.6. The dimers of *cis*-acetic acid and *cis*-formic acid form

highly bent hydrogen bonds due to interactions with the carbonyl group. In strong hydrogen bonds, angular effects lower the hydrogen bond energy more than the DH bonds shortens and consequently blue-shifts. Additionally, the O–O or O–N distance is ∼0.1 Å smaller in the *cis*-carboxylic acids than the *trans* carboxylic acids, again due to the interactions with the carbonyl group. In strong hydrogen bonds, the DH bonds continue to lengthen and redshift as D–A distances decreases beyond equilibrium and the HBs get weaker. Either of these can be the source for the overestimation of the energies in *cis*-carboxylic acids and *cis*-amides.

Although our training set consisted of harmonic vibrations and harmonic zero-point corrections, past work has shown that the *energetic–structural–spectral* relationship is transferable to experimental systems with anharmonic considerations. [114] For most dimer systems, harmonic an anharmonic values offered little deviation, virtually no changes in zero-point corrections. However, more systems contain Fermi resonances that strongly perturb the spectral positions. A good relationship was retained if deperturbed VPT2 frequencies were used, which is a good description of the localized vibration and a worse description of experimental spectra. Therefore, we propose using an intensity weighted average of coupled frequencies when applying this relationship to experimental spectra, [233] similar to what was done for the ammonia dimer in chapter 4.

For the first time, the Badger-Bauer relationship was generalized for a variety of hydrogen bond donors that are common in organic *ouroboros* molecules. This was achieved by re-referencing the "monomer" frequency to be the same across OH and NH donors separately. This appears to perfectly counteract the expected stronger interaction of acidic OH donors. Notably, the electronic structure and vibrational approximations do not affect the trend. The previous chapter has applied a more complicated but physically motivated version of the *energetic–spectral* to experimental measurements with good agreement. Therefore, we are confident in the ability of this more simplistic relationship in describing experimental measurements.

5.3.2 Validation for IMHBs via the RBM

This empirical method was deemed appropriate for predicting the strength of intermolecular hydrogen bonds with many different functional groups and was subsequently tested on intramolecular hydrogen bonds. Enolones have been a prime example for the RBM, since the conjugated character

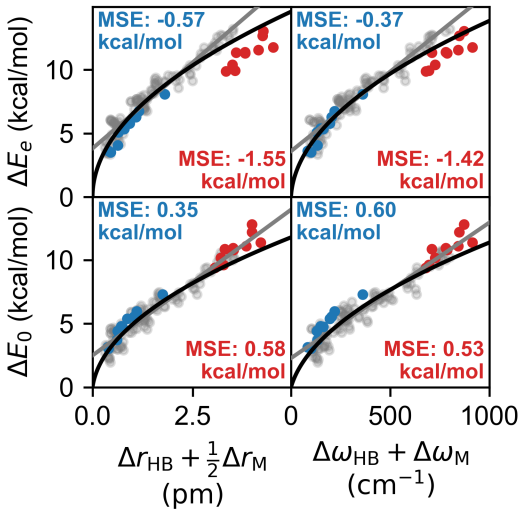

Figure 5.4: The linear (gray) and square-root (black) *energetic–structural–spectral* relationship (black line) compared against the hydrogen bond energies computed from the rotational bond method for substituted aminoethanols (blue) and substituted enolones (red). The MSEs for the square-root relationship of each set of IMHBs indicated the challenges of this method.

results in simple rotational surfaces that are well preserved in the reference enol molecules. The enolones previously studied by Buemi et al. [243] were re-calculated at a higher level of theory (MP2/aug-cc-pVDZ) and zero-point corrections were applied. Additionally, the hydrogen bond strengths of aminoethanols were calculated to span a wider range of hydrogen bond energies.

In figure 5.4, the hydrogen bond energy predicted through the RBM was compared against the hydrogen bond elongation and spectral shifts. The *energetic–structural–spectral* relationships nicely predicted the electronic energy of the hydrogen bonds in substituted aminoalcohols and slightly under-predicted the zero-point corrected hydrogen bond energy. This is because the zero-point correction in the aminoalcohols is systematically smaller than in gas-phase dimers, due to a blue-shift of the torsional vibrational modes instead of a conversion of inter-molecular modes to intra-molecular modes. For example, the OH–N out-of-plane bending mode in amino-ethanol is 547 cm^{-1}. This is only a $\sim$220cm^{-1} increase (0.6 kcal/mol) from the open amino-ethanol OH torsion (227 cm^{-1}), much less than the 638cm^{-1}(1.8 kcal/mol) out-of-plane bend in the water dimer (similar strength). Errors remained small, with an RMSD less than 1 kcal/mol, and we conclude

that the relationships established in this work works for weaker hydrogen bonds.

The linear Badger-Bauer relationships greatly over-predicted the hydrogen bond energy in enolones (>2 kcal/mol) and the square-root relationships did slightly better (∼1.5 kcal/mol). This was believed to be because the hydrogen bonds in these systems were deformed due to the intramolecular constraints. In the enolone-water dimer, the HOO angle was ∼3.4^{o} and the O-O distance was ∼ 2.77 Å. In the intramolecular enolone the hydrogen bond was 19.7^{o} and 2.58 Å. The same concerns shown in the prior chapters as well as for *cis*-carboxylic acids and *cis*-amides were present in these enolones. Namely, frequencies redshift while hydrogen bonds weaken due to intramolecular structural issues associated with the close proximity between the donor and acceptor molecules. These same deformation considerations did not affect weak hydrogen bonds, such as the aminoethanols which were also highly bent (∼46^{o}). The past chapters showed that the *energetic–structural–spectral* relationship is less sensitive to these considerations for weak hydrogen bonds. Without explicitly accounting for hydrogen bond geometries stemming from intramolecular constraints, we cannot reliably describe intramolecular hydrogen bonds over 10 kcal/mol. Instead, we get an upper estimate of the intramolecular hydrogen bond energies.

Zero-point corrections appeared to rectify the overestimation of hydrogen bond energies in enolones. The zero-point correction in enolones was even smaller than seen in amino-ethanols because of the larger red-shift in the OH frequency without significant differences in the blue-shift in low frequency vibrations. Interestingly, this smaller corrections made the Iogansen relationship accurate for the zero-point corrected hydrogen bond energies in enolones. It is unclear if this correction is unique for enolones or is transferable for strong hydrogen bonds. None-the-less, it appears safe to assume that for strong intramolecular hydrogen bonds, the electronic relationships are an upper estimate for hydrogen bond energies and the zero-point corrected relationship is a lower estimate Further exploration on other strained intramolecular hydrogen bonded systems is recommended, but this is not needed for the systems studied for the rest of this chapter.

5.3.3 Application to some esoteric IMHBs

The novelty in this method is its ability predict the individual hydrogen bond strengths of a wide variety of systems using experimental spectra. A number of important molecules studied by

prominent researchers contain intramolecular hydrogen bonds and have well-resolved spectra in the OH and NH stretching regions. [255, 256, 271–273] The relationship derived in this chapter has the power to quantify the qualitative conclusions that certain hydrogen bonds are stronger by giving that difference magnitude.

For many intramolecular hydrogen bond donors, it may be intractable to measure the free stretch vibration of the hydrogen bond donor. For the *energetic–spectral* relationship, $\Delta\omega_{\mathrm{HB}}$ and $\Delta\omega_{\mathrm{M}}$ carry approximately the same weight in the redshift. Therefore, ω_{M} (and corresponding anharmonic ν_{M}) will be assumed to be the average DH frequency of water or ammonia, experimentally 3706 and 3407 cm^{-1} respectively. Now, $\Delta\nu_{\mathrm{M}} = 0$ and $\Delta\nu_{\mathrm{HB}} = \Delta\nu_{\mathrm{OH}} = \nu_{\mathrm{H_2O}} - \nu_{\mathrm{HB}}$ or $\Delta\nu_{\mathrm{HB}} = \Delta\nu_{\mathrm{NH}} = \nu_{\mathrm{NH_3}} - \nu_{\mathrm{HB}}$.

Table 5.1: The hydrogen bond strengths of aminoethanols through the *energetic–spectral* Badger-Bauer relationship. ν_{HB} is the bound OH stretch, $\Delta\nu_{\mathrm{OH}}$ is the difference in frequency, all in cm^{-1}. ΔE_e and ΔE_0 is the predicted hydrogen bond energy in kcal/mol.

Molecule	ν_{HB} [272]	$\Delta\nu_{\mathrm{OH}}$	ΔE_e	ΔE_0
AE	3570	136	4.7	3.3
BMAE	3530	176	5.4	3.8
BFMAE	3333	373	8.1	6.3

Kjaergaard has studied the infrared spectra of important atmospheric molecules, many of which contain intramolecular hydrogen bonds. Here, we looked Kjaergaard's work on amino-ethanols, which are charge neutral and applicable to this methodology. [272] With these vibrational frequencies of the hydrogen bonded OH, the IMHB strength was estimated using experimental information for 2-aminoethanol (AE), 1-amino-2-methyl,2-propanol (BMAE), and 1-amino-2,2-bistrifluromethyl-2-ethanol (BFMAE). The energy of these hydrogen bonds were calculated and placed in table 5.1. Notably, BMAE was only slightly stronger than the AE, suggesting there is minimal influence of the methyl groups on hydrogen bond energy. In contrast, the electron withdrawing character of fluoromethyl groups strengthened the hydrogen bond by about 2-3 kcal/mol. Although it is known that the hydrogen bond is stronger in BFMAE than in BMAE or AE, [272] this was the first time the difference in energy has been quantified.

The power of the unified *energetic–spectral* relationship is its ability to measure intramolecular

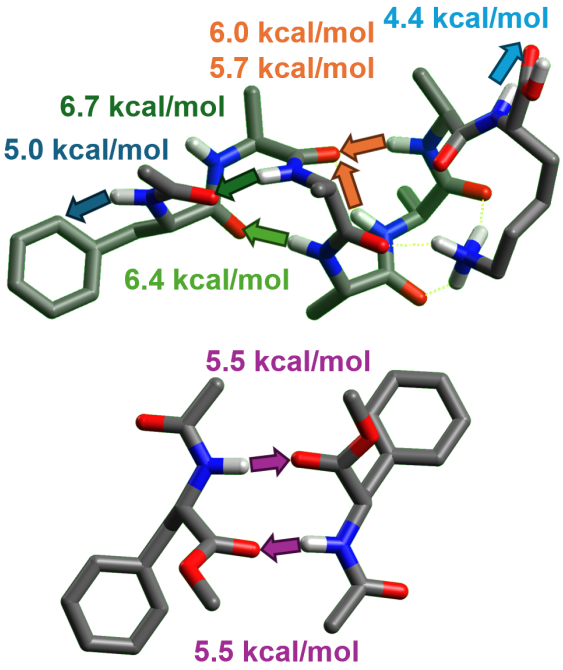

Figure 5.5: The hydrogen bond strengths of a peptide helix (top) [274] and a beta sheet model (bottom) [275].

Table 5.2: The hydrogen bond strengths of peptides through the *energetic – spectral* Badger-Bauer relationship. ν_{HB} is the bound NH stretch, $\Delta\nu_{\mathrm{NH}}$ is the difference in frequency, all in cm^{-1}. ΔE_e and ΔE_0 is the predicted hydrogen bond energy in kcal/mol.

	ν_{HB} [274, 275]	$\Delta\nu_{\mathrm{NH}}$	ΔE_e	ΔE_0
	Phe-$(\mathrm{Ala})_5$-Lys			
N-term	3298	259	6.7	5.0
	3320	237	6.4	4.7
Int.	3351	206	6.0	4.3
	3362	195	5.7	4.1
NH-π	3404	153	5.0	3.5
C-term	3434	123	4.4	3.0
	$(\mathrm{Ac–Phe–OMe})_2$			
NH	3376	181	5.5	3.9

hydrogen bond strength in complicated molecules of biological relevance. Well-resolved vibrational spectra of proteins are rare because of the high number of NH stretches and isomers overlapping in the hydrogen bond region. However, Stearns et al. isolated gas-phase helices for penta-alanine and deca-alanine and collected well resolved NH stretches. [274, 276] It's important to note that the structures studied by Stearns et al. [274, 276] are a combination of 3_{10} and α helical structures, but they serves as good models for the general helical secondary structure. Each peptide chain was capped with phenylalanine at the N-terminus and lysine at the C-terminus, which appear responsible for the helix's stability in the gas phase. Stretches in the amide backbone of penta-alanine ranged from 3300 to 3365 cm^{-1}, which correspond to electronic hydrogen bond energies between 4.8 and 6.4 kcal/mol. Individual assignments are in figure 5.5 and table 5.2. Notably, backbone hydrogen bonds are stronger at the N-terminus than the C-terminus. It is unclear if this is because of the termini or the residues at each termini. Deca-alanine, which adopted the same general structure as penta-alanine, was not computed so peaks were not individually assigned. However, the amide stretches ranged between 3295 and 3400 cm^{-1}, suggesting the hydrogen bond energies were still between 6.4 and 5.2 kcal/mol. Therefore, we can conclude that each hydrogen bond in an alpha helix has an energy of about 5.8 kcal/mol. These are in agreement with the range of peptide hydrogen bond energies [277] and could explain the preference of the hydrogen bonding arrangement in peptide secondary structure.

This technique is also applicable to intermolecular systems that have additional interactions. Although intrinsically intramolecular, gas-phase β-sheets are commonly studied through dimer motifs. Gerhards and Unterberg collected the vibrational spectrum of the Ac-Phe-OMe dimer, which forms hydrogen bonds in a β-sheet type fashion. [275] The experimental transition, 3363 cm^{-1}, was assigned as the IR active asymmetric stretch, although the dark symmetric transition was computed at a similar frequency. This frequency is in strong fermi resonance with the overtone of the C=O stretch, 3389 cm^{-1}, so the average of the two frequencies, 3376 cm^{-1} is used. Using this frequency, we estimated the hydrogen bonds to each have a strength of 5.5 kcal/mol. The structure of this β-sheet motif is added to figure 5.5 and table 5.2 for comparison. The predicted hydrogen bond energy was similar to those in the helical structures. The *energetic–spectral* relationship will continue to serve useful as additional peptide structures are characterized with high resolution vibrational spectroscopy.

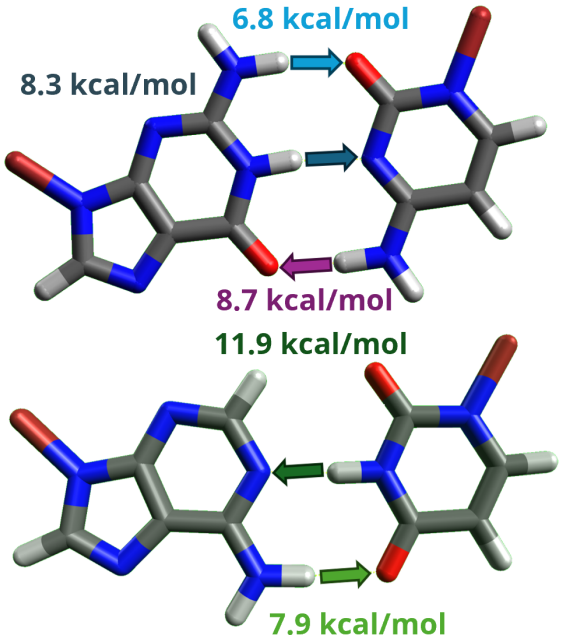

Figure 5.6: The hydrogen bond energies for the cytosine-guanine dimer (top) and the the adenine-thymine dimer (bottom), assigned from experimental spectra. [278, 279]

Table 5.3: The hydrogen bond strengths of DNA base pairs through the *energetic – spectral* Badger-Bauer relationship. ν_{HB} is the bound NH stretch, $\Delta\nu_{\mathrm{NH}}$ is the difference in frequency, all in cm^{-1}. ΔE_e and ΔE_0 is the predicted hydrogen bond energy in kcal/mol.

	ν_{HB} [274, 275]	$\Delta\nu_{\mathrm{NH}}$	ΔE_e	ΔE_0
		C-G		
$\mathrm{N_G} \rightarrow \mathrm{O_C}$	3290	267	6.8	5.1
$\mathrm{N_G} \rightarrow \mathrm{N_C}$	3165	392	8.3	6.4
$\mathrm{N_C} \rightarrow \mathrm{O_G}$	3130	427	8.7	6.8
		A-T		
$\mathrm{N_A} \rightarrow \mathrm{O_T}$	3205	352	7.9	6.0
$\mathrm{N_T} \rightarrow \mathrm{N_A}$	2784	773	11.9	9.7

The spectral-energy relationship was further applied to other biological molecules, such as the nuclease in DNA and RNA. DNA and RNA are complicated macromolecules contain both inter- and intramolecular hydrogen bonds, so measuring the strengths of such hydrogen bonds from spectroscopy is of interest. Because of the complexities in the ribose-phosphate backbone, nucleobase dimer motifs are commonly studied when focusing on base pairs. Hydrogen bonds are the energetic driver for the DNA base-pair matching. The cytosine-guanine (C-G) dimer has three hydrogen bonds; therefore, the measurable dissociation energy should be approximately the sum of the three hydrogen bonds. The affinity of the C-G base pair has been quantified, [280, 281] but the contributions from the hydrogen bonds have not been individually analyzed. Using the experimental spectrum of the cytosine-guanine (C-G) dimer (3130 cm^{-1}, 3165 cm^{-1}, and 3290 cm^{-1}), [278] the hydrogen bond strengths in the C-G dimer were found to be 6.5 (5.0), 8.0 (6.3), and 8.4 (6.7) kcal/mol (assignments in figure 5.6 and table 5.3). Their sum, 22.9 (17.9) kcal/mol, is about the experimental binding energy of C-G (21.0 kcal/mol), [281] suggesting the dimer is mostly hydrogen bonded with minor non-hydrogen bonding interactions stabilizing the base-pairing. From the spectral assignments, the hydrogen bond originating from the cytosine is the strongest and its complementary hydrogen bond ($NH_2 \cdots O$) on the guanine is more than $\sim$2 kcal/mol weaker.

The experimental spectrum for solvated or gas-phase adenine-thymine (A-T) is more complex with many competing minima. [282] However careful experimentation has captured the Watson-Crick geometry seen in helical DNA. [282] This brings up important considerations in the energetic landscape of the A-T dimer, addressed by Plutzer et.al. [282] In the gas-phase spectrum of the A-T dimer, [279] the hydrogen-bond stretches corresponding to the Watson-Crick geometry were believed to be 2784 and 3205 cm^{-1}, corresponding to hydrogen bond energies of 12.4 (10.4) and 7.5 (5.9) kcal/mol (assignments in figure 5.6 and table 5.3). The $NH \cdots N$ hydrogen bond appears much stronger in the A-T pairing than the C-G pairing, allowing for the total hydrogen bond strength to be similar between the base pairings (19.9 kcal/mol vs. 22.9 kcal/mol). The breakdown of the hydrogen bond strength in DNA nucleotide pairings allow for a detailed analysis of the interactions. Continued exploration of the hydrogen bonds in DNA/RNA can explain the energetic origins for both RNA hairpin structures, DNA miss-matches, and RNA acid/base catalysis. [283] Vibrational frequencies now provide a quantitative method for describing the strength of these hydrogen bonds.

5.4 Conclusions

For the first time, the intramolecular hydrogen bond strengths of experimental molecules have been estimated. The complications in the *energetic–structural–spectral* relationship of individual neutral hydrogen bonds were investigated so that a single protocol can be used on a diverse set of hydrogen bonds, including intermolecular hydrogen bonds. It was found that by including redshifts and structural changes in the isolated hydrogen bond donor monomer unified all OH, NH, and HF donors. This means that referencing $\Delta\omega_{\mathrm{HB}}$ to the stretches in H_2O and NH_3 provides a good estimate of hydrogen bond strength. This makes this method easier to use computationally and experimentally, where finding the free stretch in the hydrogen bond donor may prove difficult. The effects of Fermi-resonances were also discussed. With this trend, the hydrogen bonds strengths in biological molecules have been experimentally estimated and the consequences of their strength discussed.

With this new tool that can obtain experimental hydrogen bond energies for complicated systems, we hope more spectra will be measured and the energetic consequences of intramolecular hydrogen bonds can be quantified. Example uses of this relationship were outlined in this chapter, including establishing a range of peptide hydrogen bond strengths and dissecting the binding energy of DNA base-pairs into individual interactions. This is not the exhaustive use of this relationship and we encourage others spectroscopist to apply this relationship to their systems. Many gas-phase vibrational spectra are collected as ions. [113, 284–287] Additional parameterization must be done in order to apply this relationship to charged hydrogen bonds because these bonds are much stronger and likely have a differing balance of linear-electrostatic and non-linear repulsion effects. Re-parameterization of the extending Badger relationships must be done before predicting the energy of charge hydrogen bonds.

Chapter 6

DESCRIPTORS OF WATER AGGREGATION: I. WATER CLUSTERS [288]

6.1 Introduction

Water clusters are useful models for understanding hydrogen bonding and chemical transformations in aqueous environments via microsolvation approaches. [289–293] These systems have been used, among others, to study proton-transfer mechanisms in protonated water clusters, [113] to identify the origin of spectral signatures in IR spectra, [114–116] and to quantify strong cooperative effects in hydrogen bonding networks. [111, 117–119] While water clusters have claimed their space and significance in the chemical physics and simulation communities, the question of how the information gained from small water clusters translates to the liquid and/or ice phases still remains an active field of research. For example, water clusters typically have far more 2- and 3-coordinated water molecules whereas the condensed phase has mostly 4-coordinated molecules. For the smaller clusters, the need to maximize hydrogen bonding results in all atoms lying on the surface of the cluster, [294] whereas the smallest water cluster with a "fully solvated" water molecule appears at $n\sim$17. [295, 296] In addition to the different connectivity, small water clusters have a smaller extended hydrogen bonding network than liquid water and ice (in which the long-range structure is ordered). While it is known that some features of water clusters evolve with increasing cluster size, [297] it is not fully understood at what point the clusters begin to exhibit sub-structures and/or structural parameters akin to those found in liquid water and ice. Improving our knowledge of this transition via appropriate descriptors could aid in our fundamental understanding of water aggregation while at the same time assisting in the development of model systems to effectively probe the physical phenomena in condensed-phase aqueous environments.

Previous studies have reported and used descriptors of the local and extended hydrogen bond environment in bulk systems to distinguish ice-like from liquid-like structures [298–300] and also to identify unique types of solvation environments during liquid water simulations. [301–303] These descriptors of bulk environments typically take into account the relative orientations of numerous

surrounding molecules, such as the orientational tetrahedrality order parameter (q), [304, 305] the translational tetrahedrality order parameter (s_k), [304] the Steinhardt order parameters, [306–309] and the local structure index (LSI) [310]. Other studies have noted the importance of extended structural descriptors, [311] such as incorporating a measure of translational order between the first and second solvation shells (ζ) [312] and the node total communicability (NTC). [302] Manserrat et al. [313] used the smooth overlap of atomic positions (SOAP) descriptor to demonstrate the presence of ice-like sub-structures in liquid simulations, based on similarities in their local environment. However, a subsequent study by Donkor et al. [314] demonstrated the dissimilarity of the SOAP parameter and other commonly used local structure parameters in detecting similar molecular environments. However, there is still much to understand about the local and extended structure of bulk systems, especially the emergence of bulk-like qualities upon aggregation. Furthermore, while structural and connectivity descriptors have been applied widely in bulk (ice and liquid) systems, there has been little work to develop descriptors of aggregation in smaller cluster systems which have much smaller hydrogen bond networks with properties that are further away from the ones for the condensed environments.

The goal of this work, detailed in this and the next chapter, is to develop and apply structural and connectivity descriptors for water; this section will build the descriptor frame work cluster minima ranging from n=6–25 contained in a recent database of 4,948,959 structures that was created as a result of extensive Monte Carlo Temperature Basin-Paving (MC-TBP) simulations [196] with the TTM2.1-F polarizable potential. [50] We specifically focus on a subset of the full database containing 162,892 water cluster minima that are within 5 kcal/mol of the putative minima for water cluster sizes n=6–25. The restriction of the existing dataset to low-energy structures is likely sufficient in describing minimum-like structures. However, caution should be taken when using the descriptors to describe high energy or non-stationary point structures that lie outside of the ranges used in the training and analysis. First we define the descriptors we use and analyze the evolution of these descriptors for increasing sizes of water clusters in an attempt to provide insight into the transition from small to larger water clusters. While the binding energy (stability) of a cluster naturally evolves as a function of the size of the cluster as an extensive quantity, other structural descriptors are found to correlate strongly with the binding energy, lending insight into emerging stable features upon aggregation. Additionally, the correlations between descriptors are examined to

better understand how these evolve in tandem upon aggregation. Lastly, an unsupervised learning approach was implemented to group water clusters of different sizes in the database based on similarities in their structural and connectivity descriptors. This presents a unique way of classifying water clusters of different sizes with the goal of identifying similar patterns across the size regime n=6–25. The current study aims at developing descriptors of water aggregation up to ∼30 molecules so they can be used to understand the formation of a critical nucleus for water aggregation. We certainly do not claim that these descriptors are transferable to liquid water and/or ice, since the networks found in these smaller clusters (formed by the need to maximize hydrogen bonding, resulting to most atoms lying on the surface of the water cluster up to n∼17) are quite different than those present in the condensed phase environments. Since many computational techniques suffer from poor scaling with respect to the system size, it will never be possible to apply them for very large clusters or the bulk. The current work offers the possibility to use the developed descriptors to compare the similarity between clusters regardless of size. Altogether, this work identifies descriptors of water aggregation using data science to better understand how hydrogen bonding structures change during the transition from small gas-phase systems to larger aggregates of water molecules.

6.2 Definition of Descriptors

The descriptors used in this work are defined below. It is intuitive to rely on the many-body concept to categorize the various descriptors as one-body (intramolecular), two-body (intermolecular and pairwise), and many-body (intermolecular and cooperative). Utilizing the Cartesian geometry of each cluster, one can collect a large set of descriptors for localized geometries that will become the one-, two-, three-, and many-body structural descriptors. Graphs were constructed as previously described, [196] where each node is a water molecule and each edge is a hydrogen bond satisfying the $r-\psi$ definition of the hydrogen bond outlined by Kumar et al. [63] These graphs are undirected and contain adjacency information as labels of each node. More delocalized properties were derived from the graph representation of each cluster. These descriptors are termed connectivity descriptors because they contain information about the hydrogen bond network. Certain descriptors can have multiple values across a whole cluster. In cases where a single value per cluster is required, the average of the descriptor over the cluster will be used and will be denoted with angled brackets,

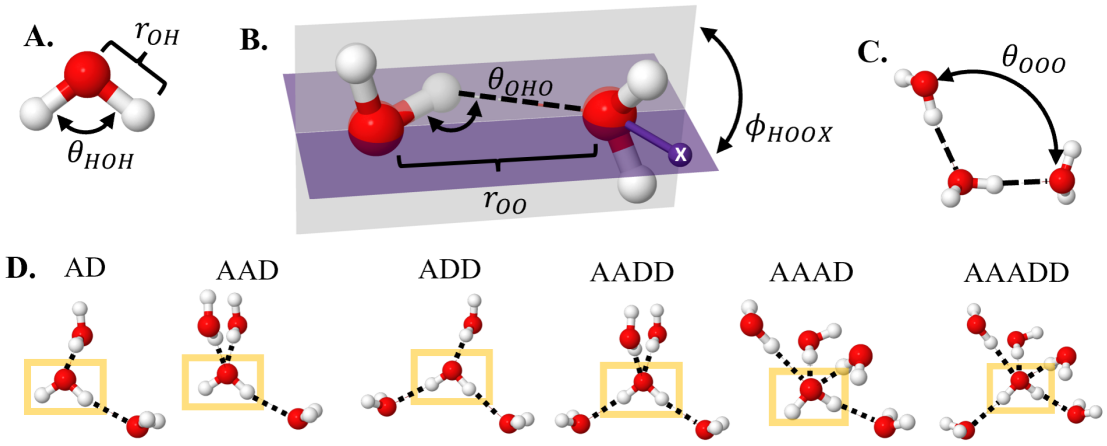

Figure 6.1: Structural descriptors. Water molecules are depicted as balls and sticks, red is used for the oxygen and white for the hydrogen atoms. X denotes the bisector of the HOH angle. A) the two intramolecular (1-body) descriptors. B) three intermolecular (2-body) descriptors of a hydrogen bond between adjacent molecules: hydrogen bond distance (r_{OO}), hydrogen bond angle (θ_{OHO}), and ϕ_{HOOX}. The gray vertical plane contains the free hydrogen of the donor and the two oxygen atoms, while the purple one contains the bisector of the hydrogen atoms on the hydrogen bond acceptor and the two oxygen atoms. The dihedral angle, ϕ_{HOOX}, is the angle between these two planes. D) the O-O-O angle (3-body) descriptor (θ_{OOO}). E) shows examples of the most common adjacencies, ranging from AD to AAADD, for the highlighted central water molecule (A=acceptor, D=donor).

viz. $\langle \cdots \rangle$.

6.2.1 Structural Descriptors

Two one-body descriptors were utilized to describe the intramolecular geometry. These are the OH bond lengths (r_{OH}) and the HOH bond angle (θ_{HOH}) shown in figure 6.1A. Three two-body descriptors were used to describe each hydrogen bond. From the 6 intermolecular degrees of freedom in a water dimer, the three most important coordinates for a hydrogen bond — the OO distance (r_{OO}), the OHO angle (θ_{OHO}), and the HOOX dihedral angle (ϕ_{HOOX}), where X lies on the bisector of the HOH angle — were chosen. These are depicted in figure 6.1B. The r_{OO} has previously been shown to have a strong relationship with the vibrational frequencies as well as the hydrogen bond energy. [315, 316] The θ_{OHO} angle describes the linearity of each hydrogen bond; close-to-linear hydrogen bonds are known to be stronger. [315, 316] The ϕ_{HOOX} angle is used to quantify the

cis-trans dihedral angle that is important for the "strong –weak" nearest-neighbor hydrogen bond definition introduced by Kirov et al. [317] These authors argued that *trans* (ϕ_{HOOX} near 180°) hydrogen bonds are more favorable than *cis* (ϕ_{HOOX} near 60°) hydrogen bonds in polyhedral water clusters, [317] since in the gas-phase water dimer the *trans* conformation is about 1 kcal/mol more stable than the *cis* conformation. [318] This descriptor is defined as the dihedral angle between the hydrogen on the donating water that is not participating in the hydrogen bond, the two oxygen atoms, and the bisector of the two hydrogen atoms on the accepting water molecule (Figure 6.1B). Three- and many-body descriptors become increasingly complex, with 21 total degrees of freedom for a trimer. It, therefore, becomes intangible to analyze all three-body degrees of freedom. Instead, we will only focus on the OOO angle (θ_{OOO}) (Figure 6.1C). This descriptor is the angle between a central oxygen atom and two oxygen atoms participating in a hydrogen bond with the central molecule.

The tetrahedral order parameter (q) is a combinatorial metric of the distribution of bonds around an atom or molecule. [304] Our analysis uses the definition adopted by Errington and Debenedetti, [305]

$$q_4 = 1 - \frac{3}{8}\sum_{i}^{4}\sum_{j>i}^{4}\Big(\cos(\theta_{i,c,j}) - \cos(109.5^\circ)\Big)^2 \tag{6.1}$$

where $\theta_{i,c,j}$ is the angle formed by two water molecules, i and j, around a central, c, water molecule. eq. (6.1), sets a range between q=0 (stochastic distribution observed in an ideal gas) and q=1 (completely tetrahedral order). The original and subsequent work using the tetrahedral order parameter was based on the four closest water molecules to define q. [304, 305, 319] Water clusters, however, may lack four neighbors in the first solvation shell, especially at the surface of the cluster . Therefore, neighbors will be determined through hydrogen bonds. The definition for tetrahedrality has been expanded in this work to accommodate a water molecule with 2, 3, and 5 (in addition to 4 for the standard definition) hydrogen bonds. Following the procedure outlined by Errington et al. [305] the coefficient of the summation is modified to hold that the expectation value is still zero for an ideal gas. That is

$$\langle q_n \rangle = 1 - A_n \sum_{i}^{n}\sum_{j>i}^{n}\int_0^{\pi}\Big(\cos(\theta_{i,c,j}) - \cos(109.5^\circ)\Big)^2 \sin(\theta_{i,c,j})d\theta_{i,c,j} = 0 \tag{6.2}$$

Solving for the coefficient yields $A_n = \frac{9}{4\binom{n}{2}} = \frac{9}{2n^2-2n}$, where n is the number of water molecules

bound to the central water of interest. A_4 reverts back to $\frac{3}{8}$. Therefore, the general tetrahedral order parameter (q_n) that will be used in this work is

$$q_n = 1 - \frac{9}{2n(n-1)} \sum_{i}^{n} \sum_{j>i}^{n} \Big(\cos(\theta_{i,c,j}) - \cos(109.5^\circ) \Big)^2 \tag{6.3}$$

The newly defined q_n value still quantifies how close the orientations of the hydrogen bonds are to 109.5° (tetrahedral), but can also account for a different number of hydrogen bonds to the central atom. For example, q_3 can be thought of as how far a three-coordinated molecule is from trigonal pyramidal. In the following q will be used to denote the modified tetrahedrality order parameter regardless of coordination (i.e., it incorporates q_m, m=2–5).

An additional many-body structural descriptor is the radius of gyration (R_g), which describes the radial distribution of particles away from the center of mass. It is defined as

$$R_g = \sqrt{\frac{\sum_i^n \left|\left|\vec{r}_i - \vec{r}_{\text{com}}\right|\right|^2}{n}} \tag{6.4}$$

where n is the number of water molecules, $\vec{r}_i$ is the position of each water molecule, and $\vec{r}_{\text{com}}$ is the position of the center of mass.

6.2.2 Connectivity Descriptors

Many-body connectivity descriptors are derived from the graph representation of the clusters and generally describe the connectivity of the hydrogen bonding network. Specifically, each cluster is represented by an undirected graph where each node is a water molecule and each edge is a hydrogen bond. Several of these descriptors have been the focus of previous studies, [297] but these will be revisited and investigated more in-depth. One such descriptor is the hydrogen bond saturation (% HB), or the percent of hydrogen atoms participating in a hydrogen bond. It is simply $\frac{\text{edges}}{2 \cdot \text{nodes}}$, where the number of edges and nodes come from the graph representation of each cluster, or $\frac{n_{\text{HB}}}{2 \cdot n_{\text{H}_2\text{O}}}$, where n_{HB} is the number of hydrogen bonds and $n_{\text{H}_2\text{O}}$ is the number of water molecules.

Another connectivity descriptor is the adjacency – the number of hydrogen bonds accepted (A) and donated (D) by each water molecule. The concept of adjacency has been extensively used by infrared (IR) spectroscopists to correlate with shifts in the observed vibrational spectral bands. [320] These adjacencies range from D and A, or single donors and single acceptors respectively, to

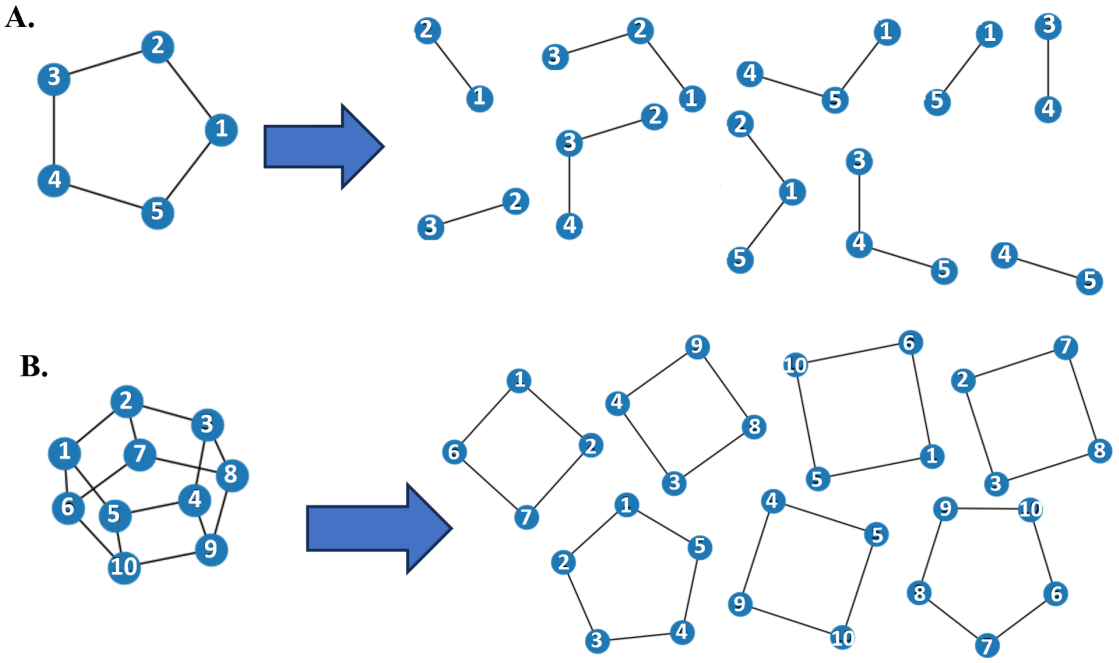

Figure 6.2: Descriptors that are visual in nature are depicted for clarity. A) the shortest paths for a pentagonal graph of $(H_2O)_5$, which are enumerated to compute the values of the Wiener index (15) and the ASPL (1.5). B) enumerates the non-short-circuited cycles in the pentagonal prism $(H_2O)_{10}$.

AAADD, which refers to a penta-coordinated water molecule that accepts three and donates two hydrogen bonds (figure 6.1D). Although adjacency can be easily computed by finding the directionality of each hydrogen bond, adjacencies were provided as node labels in the graph representation of each cluster. To minimize size extensive behavior, the percentage of water molecules in each adjacency class will be used.

Other connectivity descriptors are slightly more complex. The Wiener topological index is the sum of the shortest paths from all nodes to all other nodes. [321] It describes how connected a graph is, with lower values describing graphs with stronger connections and shorter paths. Wiener introduced this parameter to correlate the boiling points of alkanes [321], but applications have been expanded its use to predict other thermochemical properties and toxicity of bioactive molecules.

[322] Closely related is the Average Shortest Path Length (ASPL), which is the Wiener index divided by the number of paths $\left[\binom{n}{2}\right]$, where n is the number of water molecules], making the ASPL less size extensive than the Wiener index. An example graph and corresponding shortest paths are shown in figure 6.2A to illustrate the computation.

Non-short-circuited cycles are the final descriptor in the connectivity category considered in this study. Non-short-circuited cycles, as described by Rahman and Stillinger, are hydrogen-bonded rings in water that are not cut short by additional hydrogen bonds. [323] Mathematically, that means the shortest paths (or tied for shortest paths) between all sets of two nodes in the cycle is along that cycle. The non-short-circuited cycles for the pentagonal prism are depicted in figure 6.2B. This produces a non-trivial problem where the brute force method for computation, i.e., finding all cycles and checking for short-circuited behavior, is too expensive. Instead, we used an optimized algorithm to find good candidates for non-short-circuited cycles via King's shortest paths, [324] and then checked for short-circuited behavior. The code to perform such searches can be found at `https://github.com/gds001/Descriptors-for-Water-Aggregation.git`. To minimize size extensive behavior, the percentage of the cycles of each size were used.

6.3 Development of Machine Learning Models

To explore and utilize the correlation between the various descriptors, various supervised and unsupervised machine learning (ML) techniques were employed. Regression, the simplest supervised machine learning model, was widely used to explore the interplay of various descriptors and properties of water clusters. Regression can not only create predictive models but can also provide insight into the correlations and relationships between descriptors. Regression models were created by minimizing the RMSE between a fitting function and the descriptors using the Python library SciPy. [192] Linear regressions utilize a linear function,

$$F(x_1, x_2, \cdots, x_N) = \sum_{i=1}^{N} a_i \frac{x_i - \mu_i}{\sigma_i} \tag{6.5}$$

where x_i are the features of the model (descriptors or products of descriptors), μ_i is the average value of x_i, σ_i is the standard deviation of x_i, and a_i is the fitting weight. The square of the correlation coefficient (R^2) served as a good metric for the strength of the relationship. When

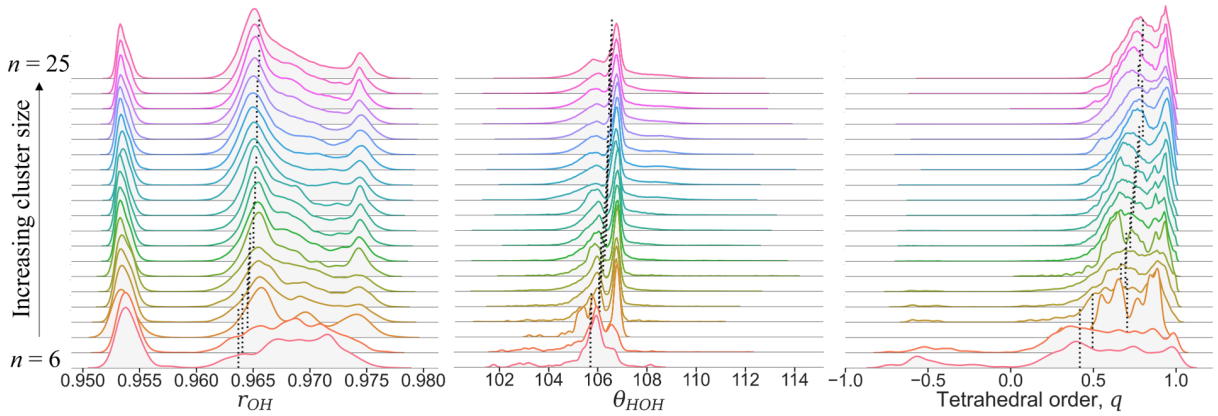

Figure 6.3: Distributions of the r_{OH} (Å, left), θ_{HOH} (middle), and q descriptors (right) for water cluster sizes n=6 (bottom of each panel) to n=25 (top of each panel). Colors represent different clusters and are used for clarity. Dashed vertical lines denote the arithmetic mean of each distribution.

used for prediction, multi-dimensional and polynomial regressions were created in a similar fashion, using initial weights determined through low dimensional principal component (PC) models.

To simplify the configuration space and create useful clustering models, principle component analysis (PCA) was employed using scikit-learn [325]. PCA has been widely used for a variety of chemical applications, ranging from property prediction to group clustering. [326–330] In summary, PCA linearly transforms descriptors to pull as much variance into early principal components (PCs) and reduce the dimensions required for a successful model. Mean-centering and norm-scaling were employed on the descriptors prior to PCA analysis. PCA can be utilized with all descriptors, which can be useful when attempting to analyze the interplay between the descriptors as a whole. However, certain descriptors are not good candidates for a PCA model, such as those that are uniform across the data set and those that do not contain information about desired properties. Including such descriptors will pull unnecessary information into higher PCs and require more dimensions for a successful model. Thus, the PCA model used in this chapter was constructed by selecting a subset of potential descriptors and inspecting for visual separation. Descriptors were added and removed one by one until no improvements were found.

6.4 Results and Discussion

6.4.1 Evolution of Descriptors with Cluster Size

By examining how the descriptors correlate to cluster size (n), we gain insight into how these descriptors (associated with stable clusters) evolve upon aggregation.

We will first focus on the changes in the water intramolecular geometry upon aggregation via the 1-body descriptors. Figure 6.3 depicts the distributions of r_{OH} (left) and θ_{HOH} (middle) for water cluster sizes n=6 (bottom of each panel) to n=25 (top of each panel). For these finite clusters we observe a relatively constant peak around 0.950-0.955 Å for the free r_{OH} across all cluster sizes in the r_{OH} distributions. However, upon increasing the cluster size, we see a splitting of the peak between 0.965 and 0.975 Å upon aggregation. This splitting suggests that those molecules exist in differing hydrogen bonding environments within the water clusters. In general, the average r_{OH} value (dashed line) lengthens for larger cluster sizes (R^2=0.59) in a manner that is akin to the increase in % HB. Furthermore, in the θ_{HOH} distributions, we see that the small peak between 102-104° disappears quickly around n=12. The peak around 106° decreases as the cluster size increases, which is also associated with the introduction of a peak around 107°. At about n=12 a small shoulder starts to grow in around 109°. This expansion of the θ_{HOH} relative to the gas-phase monomer value (104.5°) upon aggregation is well-documented in the literature. [331–337] However, the overall correlation between θ_{HOH} and cluster size is weak (R^2=0.30) on average, as larger, bulk-like angles start to appear and the smaller gas-like angles disappear quickly upon increasing the cluster size.

Figure 6.4 traces the evolution of (A) the hydrogen bond saturation, (B) the radius of gyration and ASPL, (C) the θ_{OHO}, (D) the θ_{OOO}, (E) the percentage of molecule type, and (F) the percentage of cycle counts as a function of water cluster size (n=6-25). A few of the many-body descriptors are inherently size-extensive. These include the Wiener index, [297] the number of hydrogen bonds, and the total binding energy ($R^2 > 0.97$). In addition, there are many descriptors that intuitively change with increasing cluster size. For example, the % HB increases for larger cluster sizes and trends toward 100% (R^2=0.62, figure 6.4A). This is expected because the molecules in the interior of a cluster are more coordinated and, consequently, there is an on average smaller fraction of molecules with a free r_{OH}. While we expect % HB $\simeq$ 100 for the bulk phase, it does not exceed

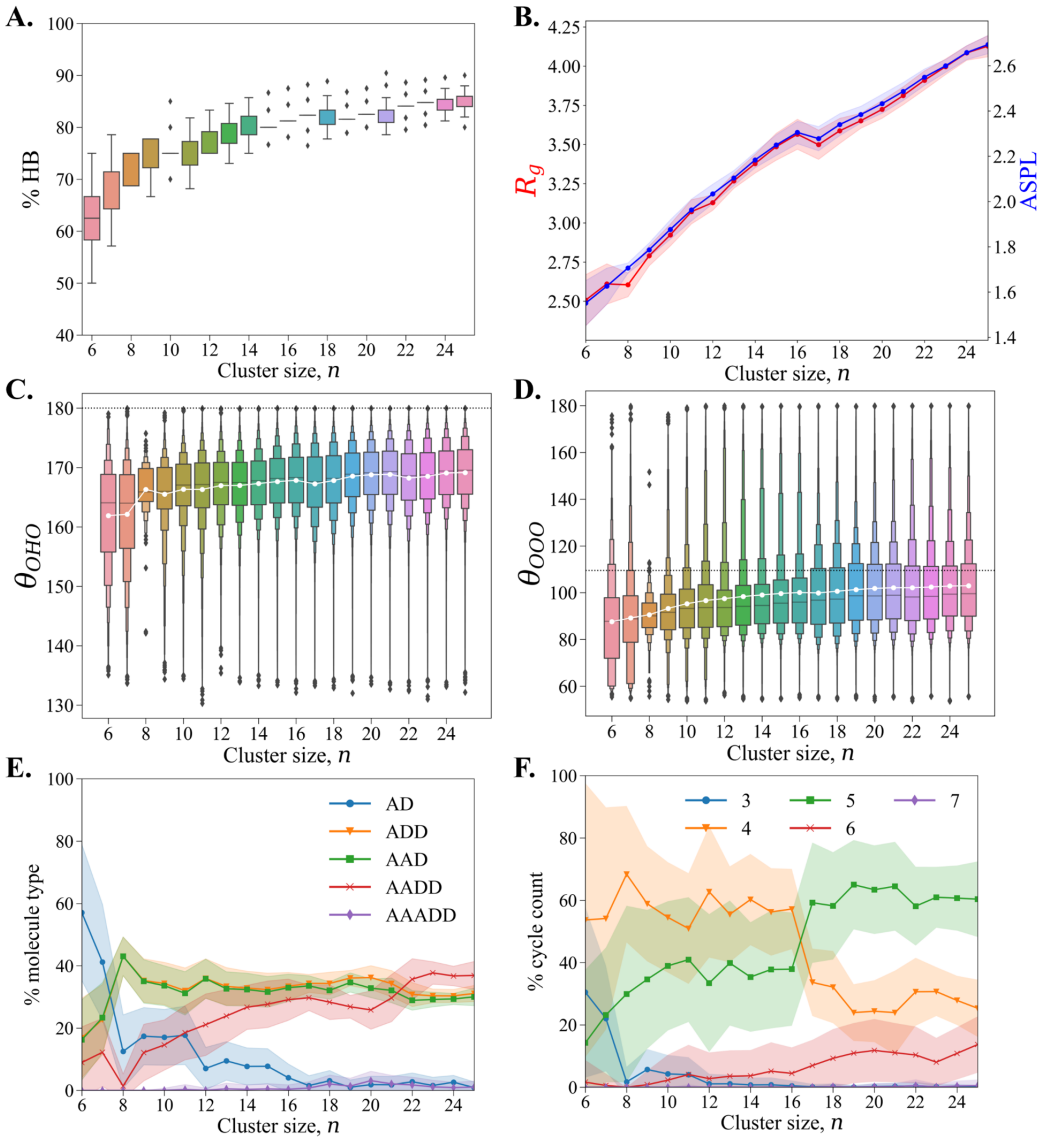

Figure 6.4: The evolution of the hydrogen bond saturation (A), radius of gyration and ASPL (B), θ_{OHO} (C), θ_{OOO} (D), percentage of molecule type (E), and percentage of cycle counts (F) as a function of water cluster size (n=6-25). The shaded regions in panels B, E, and F represent the standard deviation across the clusters of a given size. The connected markers in panels B, C, D, E, and F represent the average across a given cluster size. The boxes in panels C and D represent each decile of the distribution. Scattered points in panels A, C, and D are statistical outliers in the distributions. The dotted line in panel C represents the perfectly linear hydrogen bond. The dotted line in panel D represents the perfectly tetrahedral θ_{OOO}.

90% for clusters up to n=25. The radius of gyration and ASPL are also strongly correlated with the cluster size ($R^2 > 0.94$, figure 6.4B). This is not surprising given that both metrics quantify the size and shape of a cluster, despite being derived from different representations of the cluster (Cartesian coordinates vs. a graph).

In general, 2-body descriptors (i.e., r_{OO}, θ_{OHO}, and ϕ_{HOOX}) have weak correlations with cluster size. The dihedral angles and resulting ratios of *cis*- vs. *trans*- hydrogen bonds stays relatively consistent at $30-40\%$ *cis*- (0-90°) and $60-70\%$ *trans*-hydrogen bonds (90-180°). The r_{OO} distance, on average, is smaller for large cluster sizes (R^2=0.38) while the θ_{OHO} (R^2=0.31, figure 6.4C) gradually becomes more linear with increasing cluster size. In addition, there exists a correlation between θ_{OOO} and the size of the cluster (R^2=0.79, figure 6.4D). On average, the θ_{OOO} increases from an average of 161.9° for n=6 to an average of 169.2° for n=25. However, the correlation with tetrahedral order (q) is weaker (R^2=0.31). Since distribution plots for that descriptor have been reported for liquid simulations, [304, 305, 319, 338–340] we provide the distributions of the tetrahedral order (q) for cluster sizes n=6-25, for comparison (Figure 6.3, right panel). As observed in liquid simulations, [305, 319, 338–340] the distribution is bimodal. However, the peak existing at lower q-values is more prominent than the corresponding one in liquid water, indicating the large ratio of molecules exhibiting a weaker local tetrahedral ordering in the clusters compared to liquid water. However, there is a substantial amount of molecules approaching q=1, indicating the emergence of fully solvated molecules having 4 nearest neighbors with a near tetrahedral local structure.

The evolution of the hydrogen bond network (connectivity descriptors) with cluster size is of particular interest. Figure 6.4E traces the changes in the fraction of molecule types (local bonding information) as a function of cluster size. For small water clusters (n=6–7), the AD-type water molecules dominate with small contributions from ADD and AAD. The fraction of AD water molecules sharply decreases up to size n=8 and continues to decrease for larger cluster sizes. It is at this size regime that the energetically stable n=8 cube appears, which is entirely ADD/AAD. This structural motif persists through larger clusters and is likely the cause for the low % of AD molecules in the low-energy subset of the full cluster database used in this study. The fractions of ADD- and AAD-type water molecules grow at nearly the same rate, reaching a maximum at n=8 and staying somewhat constant beyond that point. Eventually, at around n=22 the AADD-type

water molecules become the most prevalent (approaching 40 %, on average) while the ADD- and AAD-type ones decrease slightly. This demonstrates the point at which the AADD-type molecules become the most prevalent type in the cluster. Regardless, even at n=25 we see that roughly 60% of the water molecules are still 3-coordinated, on average.

The average number of hydrogen bonds in liquid water is about 3.5, [63] suggesting limiting values of AADD-type molecules in liquid water. Furthermore, a small percentage of water molecules are penta-coordinated (AAADD-type) for n>10. The AAADD molecules become more prevalent for larger water clusters albeit never exceeding 17% for a single cluster. The emergence of the AAADD molecules is accompanied by a simultaneous loss of AAD molecules, demonstrating the role of the AAADD molecules in balancing the number of hydrogen bond donors in the cluster.

Figure 6.4E traces the changes in the fraction of molecule types (local bonding information) and the fraction of cycle counts as a function of cluster size. We observe significant changes in the cycles that are produced as part of the extended hydrogen bond environment (Figure 6.3F). For small water clusters, the small cycles (3 and 4) naturally dominate. While these ring sizes are often strained, they help to maximize the number of hydrogen bonds for small clusters. However, the 5-cycles abruptly become the dominant cycle size for n>16 with the 4-cycles steadily decreasing. The 6-membered cycles very slowly increase in prevalence but never surpass roughly 20 % of the cycles in a cluster for n=25. Given that hexagonal ice (ice Ih) has only 6-membered rings, [122] we anticipate the prevalence of 6-cycles in stable water cluster minima to increase for larger cluster sizes (n>25).

6.4.2 Correlation between Descriptors

It is expected that several descriptors will correlate with one another. Some of these correlations are an artifact of the way each descriptor is defined, whereas others provide insight into the structural and connectivity patterns present in the water cluster minima. The various correlations are summarized by the corresponding strengths (correlation coefficients) traced by colors in figure 6.5. Descriptors that have no correlations, such as $\langle\phi_{\mathrm{HOOX}}\rangle$, % AAADD, and % of 6-cycles, are excluded for clarity.

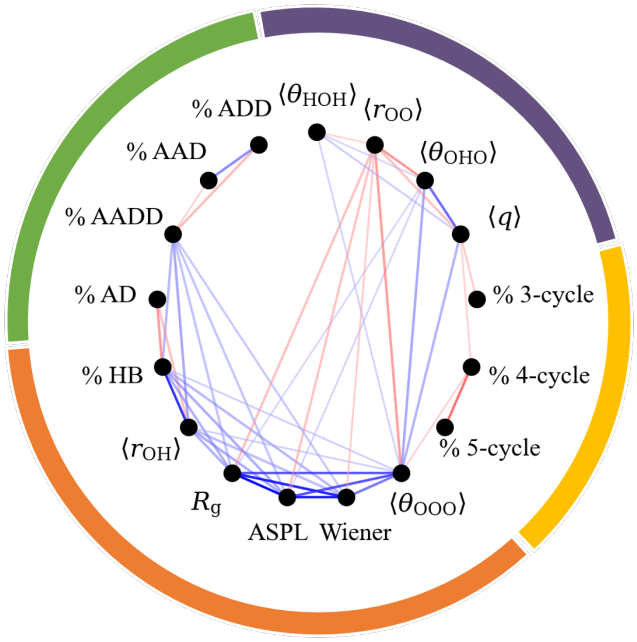

Figure 6.5: The strengths of the correlations between the various descriptors are depicted in the chord plot. Blue lines denote positive correlations, whereas red lines denote negative correlations. The darkness of the colors encodes the strength of the correlations, with dark blue and dark red having R^2 values near 1, and lighter (pale) lines representing R^2 values near 0.5. Correlations with $R^2 < 0.33$ are excluded for clarity. The outer circle shows what the descriptors best describe — energy (purple), cycles (yellow), size (orange), and adjacencies (green).

The descriptors appear to fall into two groupings that are related to size and hydrogen bond strength. There is a strong correlation between ASPL and R_g ($R^2 = 0.98$). There is no mathematical reasoning for the correlation in the formulas for these descriptors, suggesting there are two distinct ways in describing a similar property using either the Cartesian coordinates (R_g) or graphs (ASPL). The correlation is, however, logical, since both ASPL and R_g increase with cluster size and decrease with the spherical nature of the cluster. Closely correlated to ASPL and R_g is the Wiener index (R^2=0.93), which is similar in computation to the ASPL. Another correlation exists with the aforementioned descriptors and $\langle\theta_{OOO}\rangle$ ($R^2 \approx 0.8$). The origin of this latter correlation is not clear, but it serves as yet another link between structural and connectivity descriptors. A natural correlation exists between % HB and $\langle r_{OH}\rangle$ (R^2=0.9). This arises from the lengthening of r_{OH} due to the formation of hydrogen bonds. [114] In the presence of additional hydrogen bonds

(higher % HB), more r_{OH} will be elongated, therefore $\langle r_{\mathrm{OH}} \rangle$ will be greater. Altogether these six descriptors form the core of one group of correlations. As discussed earlier, all these descriptors have strong correlations with size.

Another set of correlations exists between the rest of the structural descriptors. The strongest correlation in this group is between $\langle q \rangle$ and $\langle \theta_{\mathrm{OHO}} \rangle$ (R^2=0.76). It translates to the realization that water molecules in a water cluster are on average more tetrahedral when the hydrogen bonds are more linear. This is reasonable since hydrogen bonds that are far from 109.5° (tetrahedral) generally bend their hydrogen bond to increase hydrogen-lone-pair interactions. [341] It is interesting that the correlation between $\langle q \rangle$ and $\langle \theta_{\mathrm{OHO}} \rangle$ is stronger (R^2=0.76) than the correlation between $\langle q \rangle$ and $\langle \theta_{\mathrm{OOO}} \rangle$ (R^2=0.54), which are mathematically related to one another. Since the function between q and θ_{OOO} is not one-to-one, the correlation of their average values is weak. Another strong correlation is the one between $\langle \theta_{\mathrm{OHO}} \rangle$ and $\langle r_{\mathrm{OO}} \rangle$ (R^2=0.63). This is quite reasonable, as both descriptors are well-known metrics of hydrogen bond strength. [315, 316] This implies that the correlations between $\langle r_{\mathrm{OO}} \rangle$, $\langle \theta_{\mathrm{OHO}} \rangle$, and $\langle q \rangle$ (all descriptors focusing on the local hydrogen bonding environment) stem from the relationship between such descriptors and the underlying hydrogen bond strengths. Oddly, $\langle \theta_{\mathrm{HOH}} \rangle$ shows correlations only with the descriptors mentioned in this paragraph ($R^2 \approx 0.35$). Therefore, it will be grouped accordingly despite the weakness of the correlations. It is important to note that r_{OH} is also known to correlate with the hydrogen bond strength, with a longer r_{OH} associated with stronger hydrogen bonds. [184, 185] This relation is not seen in this analysis, as it is skewed by the inclusion of the free r_{OH} in the averaging of the value of the descriptor.

Weak cross-correlations exist between the two groups. Notably, the strongest correlations between subsets of the two groups are between $\langle \theta_{\mathrm{OOO}} \rangle$ and descriptors of hydrogen bond strength ($R^2 \approx 0.57$) and weaker correlations are between $\langle r_{\mathrm{OO}} \rangle$ and descriptors of size ($\mathrm{R}^2 \approx 0.47$). This relationship is not surprising because of the existence of a weak correlation between θ_{OOO}, q, and hydrogen bond strength. The latter, however, is surprising since the correlations between the two-body descriptors and the cluster size are quite weak. This suggests that additional information besides size and hydrogen bond energy is encoded in these descriptors.

Correlations between adjacency descriptors provide important insight between the stability of hydrogen bond networks and the balance of hydrogen bond acceptors (A) and donors (D). Strong

correlations exist between the % of AD, AAD, ADD, and AADD, which are most prevalent in the size regime of the water clusters examined in this study. The correlation between those descriptors is also shown in figure 6.5. The anti-correlation between AD and AAD/ADD is weak ($R^2 \approx 0.17$) due to the rapid disappearance of AD molecules. The correlation is stronger between AAD/ADD and AADD ($R^2 \approx 0.42$), a fact that demonstrates the interplay between AAD/ADD and AADD in large, highly connected clusters. The strong correlation between AAD and ADD (R^2=0.67) arises from the natural constraint that for each acceptor there is a donor. As seen previously, there is a disappearance of AD and the appearance of AADD water molecules upon increasing cluster size. This loose correlation with cluster size manifests as weak ($0.3<R^2<0.6$), while correlations between AD and AADD descriptors are found to strongly correlate with size.

Correlations in the % cycles exist because more of one cycle means less of another cycle. As most cycles consist of 4 and 5 molecules, there is a strong correlation bounded by the anti-correlation line. Correlations between 3- and 6-cycles and other cycles are extremely weak because their populations are small throughout the dataset. Although correlations with other descriptors are weak, interesting trends emerge with descriptors that predict hydrogen bond strength. There exist weak correlations ($R^2 \approx 0.3$) between the 3-, 4- and 5-cycles and $\langle q \rangle$, $\langle r_{\mathrm{OO}} \rangle$, $\langle \theta_{\mathrm{OHO}} \rangle$, and $\langle \theta_{\mathrm{OOO}} \rangle$. Notably, values with increased hydrogen bond strength correspond to more 5-cycles and less 3- and 4-cycles. This is reasonable as 3- and 4-cycles are more strained ($\langle \theta_{\mathrm{OOO}} \rangle \leq 90^\circ$) and therefore are associated with weaker hydrogen bonds.

6.4.3 Partition of Descriptors via Principal Component Analysis

The PCA model constructed in this study utilizes descriptors that can effectively translate to the bulk phase. For this reason, size, ASPL, Wiener index, and R_g are excluded from the model. The resulting best model utilized the descriptors $\langle r_{\mathrm{OO}} \rangle$, $\langle \theta_{\mathrm{OHO}} \rangle$,$\langle \theta_{\mathrm{OOO}} \rangle$, $\langle q \rangle$, % HB, % AD, % ADD, % AAD, % AADD, % AAADD, % 4-cycles, and % 6-cycles. The results of the PCA analysis are shown in figure 6.6, suggesting that interesting structural patterns emerge warranting further analysis. It was found that the one-body descriptors ($\langle r_{\mathrm{OH}} \rangle$ and $\langle \theta_{\mathrm{HOH}} \rangle$) muddled the model by smearing groups together and were thus not used. Additionally, the average dihedral angles of hydrogen bonds ($\langle \phi_{\mathrm{HOOX}} \rangle$) and the % 5-cycles combined seemingly meaningful separations; thus

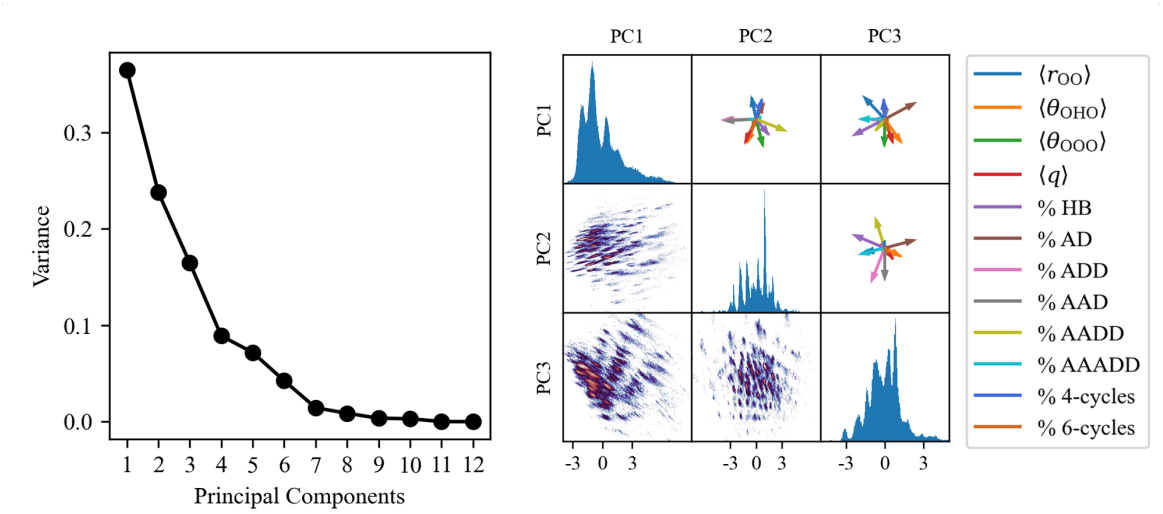

Figure 6.6: The variance (left), the distribution as a histogram (right, lower triangle and diagonal), and the components (right, upper triangle) of the PCA model used. The colors in the histograms represent the number of clusters in each bin with blue in the tens and red in the hundreds.

⟨ϕ_{HOOX}⟩ and % 5-cycles were not used. The descriptors % HB, % 4-cycles, and % 6-cycles had minor effects on the separation of principal components but served to increase the visible separation. Notably, % HB improved the clustering of isomorphs with fewer PCs. The latter two are included in the model for potential benefit when using multiple PCs for analysis, but will not be discussed in detail here.

An initial analysis examined the component matrices produced by the PCA to describe how the initial descriptors map onto this variance-separated space. PCs 4, 5, and 6 are dominated by % AAADD, % 6-cycles, and % 4-cycles respectively. It is noticeable, however, that a significant proportion of % AAADD contributes to PC3. PCs 1-3, shown in figure 6.6, have a good distribution of the contribution from each of the remaining descriptors. The strongest descriptors contributing to PC1 are ⟨r_{OO}⟩, ⟨θ_{OHO}⟩, ⟨θ_{OOO}⟩, and ⟨q⟩. PC3 is similar to PC1, but includes % HB and % AD information. PC2 is unique in that it contains information about adjacencies (% ADD, % AAD and % AADD). These differences, combined with the structural patterns that appear and the variances in figure 6.6, make these PCs good candidates for analyzing water clusters.

It is found that PCA succeeds in separating clusters based on descriptors not included in the

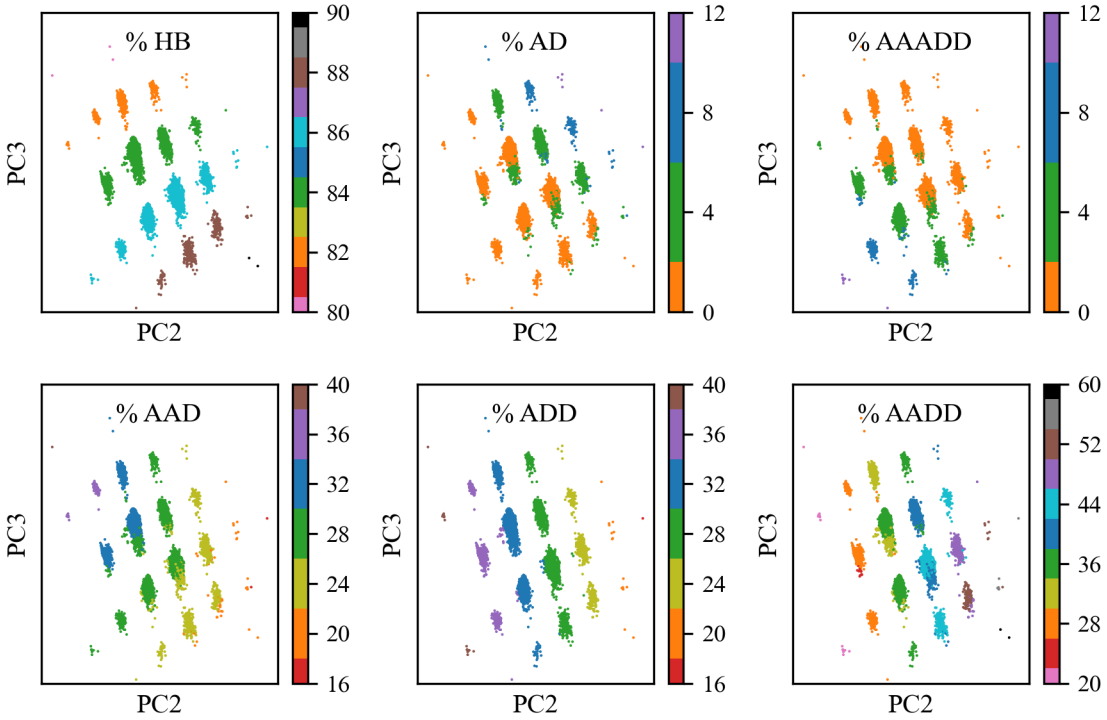

Figure 6.7: The separation of descriptors important for PC2 and PC3 are colored for clusters with n=25 water molecules across PC2 and PC3. The vertical separation is determined by hydrogen bond saturation, whereas the horizontal separation is determined by the combination of adjacencies.

model. Size correlates strongly with PC1, (R^2=0.76), with similar correlations between PC1 and ASPL, Wiener index, and R_g. This is likely because PC1 contains information about $\langle\theta_{\mathrm{OOO}}\rangle$, which is closely related to size. The combination of $\langle r_{\mathrm{OO}}\rangle$, $\langle\theta_{\mathrm{OHO}}\rangle$, $\langle\theta_{\mathrm{OOO}}\rangle$, and $\langle q\rangle$ in PC1 seems to pull the size extensive behavior of $\langle\theta_{\mathrm{OOO}}\rangle$ into PC1, while leaving the size-independent character of $\langle\theta_{\mathrm{OOO}}\rangle$ in lower PCs. The implicit separation of size is especially important as it allows aspects of the information related to size to be drawn up to PC1 and aspects of information unrelated to size to remain pure in the lower PCs. When targeting attributes that must remain size-independent, it will be beneficial to remove PC1 from the model.

PCA provides a useful tool to separate and group similar data points, and this technique was employed to classify and compare certain water clusters. Since PC2 and PC3 provided the best visual separation, they will be used heavily in this section. For quantitative metrics, PC 2-7 will be

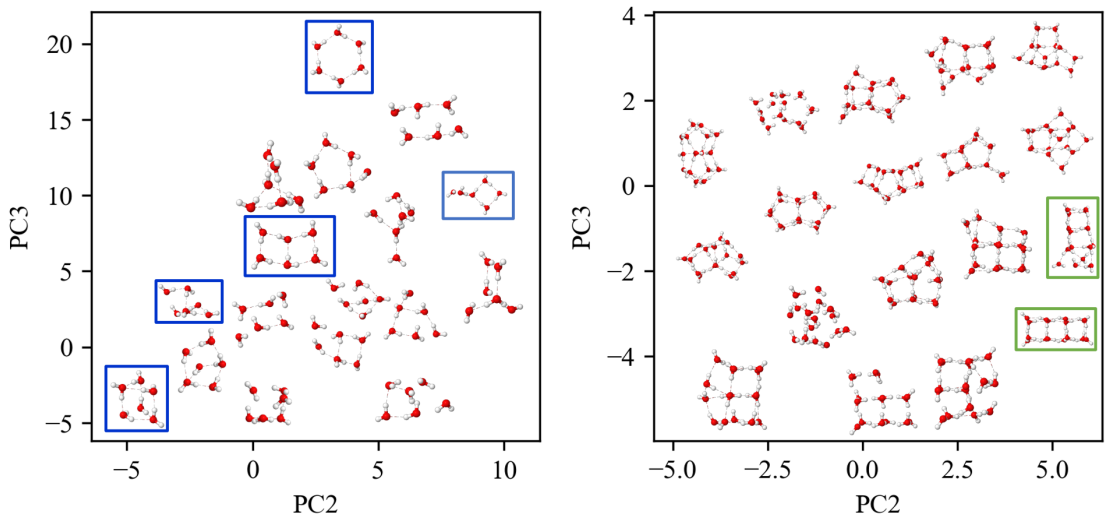

Figure 6.8: Representative structures for the water hexamer (left) and the water 16-mer (right) are shown as examples of how PC 2 and PC 3 can separate clusters based on structural differences. Common hexamer isomorphs are identified by blue borders in the left panel. An example of a "defect" for n=16 (see text) is indicated by a green border in the right panel. PC3 consistently separates clusters based on their degree of connectivity, but PC2 changes separations based on structures in the size group.

used. PC1 will not be used because it largely encodes cluster size, and this study seeks to classify patterns in water clusters independent of size. As discussed earlier in this section, PC2 separates mainly on adjacency and PC3 separates mainly on % HB. These descriptors have discrete values, arising from the division of small integers whose magnitudes are <60. Therefore, their distinct separation is natural but further analysis shows it is more subtle. Figure 6.7 depicts the separation of the aforementioned descriptors across PC2 and PC3 for water clusters of size n=25. PC3 nicely separates hydrogen bond saturation into distinct strata, with negative PC3 representing highly connected structures and positive PC3 representing less connected structures. Clusters along PC2 contain mixtures of the various % adjacencies encoding the detailed information on the hydrogen bond networks. Much of the information across PC2 is separated by % ADD. Nonetheless, the separation provided by PC2 and PC3 in this model provides a great way to describe similarly connected water clusters using descriptors that have measurable values in bulk water.

Another application of the separation is to group similar clusters, such as families of isomorphs. The water hexamer have been widely studied for the emergence of a complex, non-quasiplanar landscape of low-energy isomorphs. [342–344] The PC model developed here appropriately separates the isomorph family (Figure 6.8 left). Notably, the common families of the water hexamer, delineated by blue borders, appear in distinct groupings. Prism and cage-like clusters appear in quadrant III, whereas ring structures appear in quadrant I. The other isomorphs, shown with no outlines, group around similarly structured clusters. For example, the traditional "cage" is grouped with a similar but distinct cage-like structure. More clearly, the "book" structure is grouped among other fused ring structures. Similar patterns emerge for larger water cluster sizes. Isomorphs remain separated across PCs 2-3. Figure 6.8 (right) shows the separation for the water 16-mer, where the lowest energy isomer in each grouping is shown. Although each grouping is expected to be diverse for the medium-sized clusters, the structures shown in figure 6.8 are good representatives of the structures that separate via PCA.

The power of this technique rests in its ability to identify water clusters of different sizes with similar structural motifs. As discussed earlier, water-6's "prism" (trigonal prism) structure appears in the far negative quadrant across PC2 and PC3. Similar structures, such as water-8's cube and water-10's pentagonal prism, appear in a similar area, although slightly higher in PC3 due to differences in $\langle q \rangle$ and $\langle \theta_{\mathrm{OHO}} \rangle$. Fused prisms appear similarly in the negative region of PC3, but are spread out across PC2, with positive PC2 having more shared faces. Defects to fused prisms appear in proximal groupings to their "parent" cluster. An example of a "defect" (indicated by a green border in figure 6.8), involves breaking the arrangement of one end and the shifting of that cluster "upwards" in PC3 (because of fewer hydrogen bonds) and "left" in PC2 (because of fewer ADD/AAD molecules).

To exemplify the PCA classification, the clusters closest to the lowest energy 25-mer in the data set were identified and shown in figure 6.9. Clusters were found by proximity (norm 2) across PC 2-7. The three closest structures interestingly belong to cluster sizes 22, 19, and 16. Each structure shows similar structural motifs, a nucleating pentagonal prism-like structure (shown on the right side of each image), and the beginnings of an ordered cage surrounding a centered water molecule (shown on the left). As these are the most structurally similar clusters, they would make great candidates as starting structures when needing to study the 25-mer. This analysis can be done for

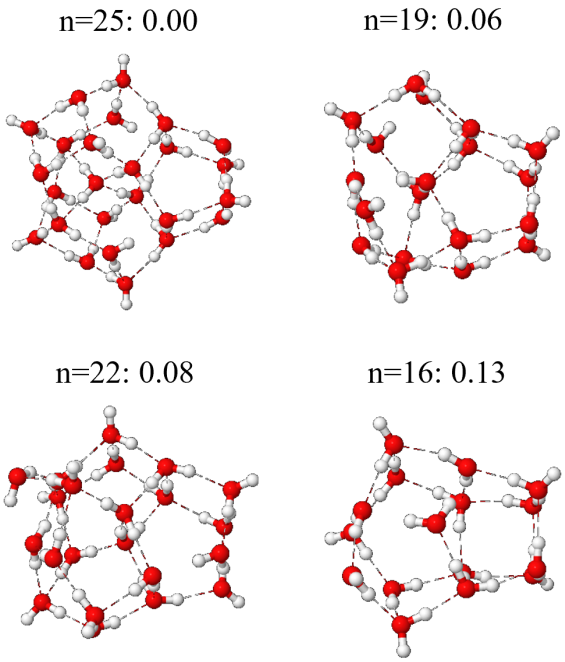

Figure 6.9: The putative minima of the 25-mer water cluster is depicted in the top left. The three closest water clusters from the PCA analysis are also shown, with their distances along the PCA indicated in the figure. Each serves as a potential good model for studying the water 25-mer.

any water cluster, including structures such as clathrates, which were not included in the subset of the full database that was used for the analysis (see the discussion in the Introduction of this chapter).

6.4.4 Prediction of Energetic Stability

It is finally of interest to examine which descriptors are predictive of the energetic stability (binding energy per water molecule, (BE_n) of the cluster. As discussed earlier, BE_n naturally evolves with the cluster size. More specifically, there exists a strong correlation between the binding energy per water molecule and the cluster size (R^2=0.99), as seen in the left panel of figure 6.10. Note that the limiting value of BE_n for $n \rightarrow \infty$ for the fit shown in the right panel of figure 6.10 is –12.7 kcal/mol when the exponent of n is considered a variable (with a value of 0.82). Incidentally, this limit is near the lattice energy of ice (14.07 kcal/mol per molecule). [345] The cluster sizes (up to n=25)

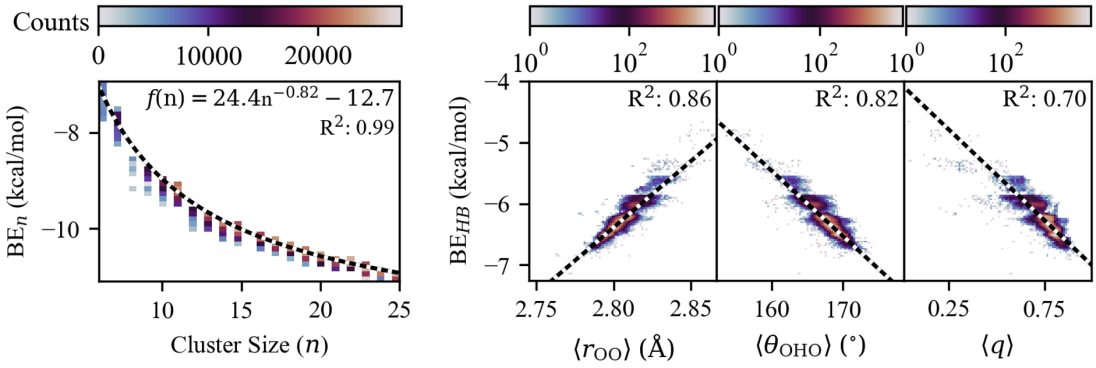

Figure 6.10: (Left) The strongest single predictor for the binding energy per molecule (BE_n) is cluster size. It follows an inverse n function, where n is the number of water molecules, and an asymptotic value of –12.7 kcal/mol for $n \to \infty$. (Right) $\langle r_{OO} \rangle$, $\langle \theta_{OHO} \rangle$, and $\langle q \rangle$ are predictive of the binding energy per hydrogen bond (BE_{HB}) in the cluster. Both panels are 2-D histograms, where the color indicates the density of points.

are too small to be used in conjunction with a fit of the volume element $n^{-1/3}$, which produces a less accurate regression for this cluster regime than when the exponent is a variable.

Descriptors that correlate with size (r_{OH}, θ_{OOO}, ASPL, Wiener index, % HB, and R_g) also show correlations with the binding energy per water molecule ($R^2 \approx 0.95$ for θ_{OOO}, ASPL, Wiener index, and R_g, $R^2 \approx 0.7$ for r_{OH} and % HB). However, none of these descriptors are predictive of relative stability between isomers within a particular water cluster size. The descriptor that most improved the accuracy of the prediction (albeit by only 5%) was % HB. Therefore, it is inferred that the correlation with the aforementioned descriptors comes mostly from their relationship with size and size's relationship with the cluster binding energy.

Another way to analyze energetic stability is to scale the cluster binding energy by the number of hydrogen bonds. This metric of energy does not scale with cluster size but instead scales strongly with structural many-body descriptors, such as r_{OO}, θ_{OHO}, and q, which are known to greatly influence the energy of the hydrogen bond. [315, 316, 346] It is therefore expected that they are predictive of the average energy of a hydrogen bond in a cluster. Each has relatively strong correlations (R^2 between 0.70 and 0.86) with the binding energy per hydrogen bond as can be seen from the right panel of figure 6.10. Strikingly, these three descriptors combine linearly for an

Quantity	Regression Model	All		Bulk		PC 1-3		PC 1-7	
		RMSE	k	RMSE	k	RMSE	k	RMSE	k
BE_n	Linear	0.036	22	0.050	18	0.075	3	0.062	7
	Quadratic (uncoupled)	0.032	44	0.047	36	0.073	6	0.061	14
	Quadratic (coupled)	0.028	275	0.041	189	0.072	9	0.056	35
BE_{HB}	Linear	0.022	22	0.030	18	0.043	3	0.034	7
	Quadratic (uncoupled)	0.020	44	0.028	36	0.043	6	0.034	14
	Quadratic (coupled)	0.018	275	0.025	189	0.041	9	0.032	35

Table 6.1: The errors (RMSE, kcal/mol) of various regression models to predict the binding energy per molecule (BE_n) and per hydrogen bond (BE_{HB}) for various sets of descriptors and regression styles. "All" indicates using all 22 descriptors, whereas "Bulk" indicates a fit using the 18 descriptors relevant to the bulk phase (see text). Additionally reported are the number of fitting parameters (k) for each of the regression models.

improved fit (R^2=0.95), suggesting each descriptor encodes distinct properties of hydrogen bond strengths.

Alternatively, more descriptors can also be combined to predict the cluster binding energy. There is strong predictive power for the energy per molecule and the energy per hydrogen bond from the combination of all descriptors with an RMSE of 0.036 and 0.022 kcal/mol per molecule, respectively (Table 6.1 and figure 6.11). The largest regression weights are associated with size, ASPL, and Wiener Index, suggesting that these descriptors together are most predictive of the binding energy per water molecule. Since these descriptors are not transferable to bulk water, the linear regression was refitted using only the descriptors that can be computed under periodic boundary conditions (Table 6.1), namely $\langle r_{\mathrm{OH}} \rangle$, $\langle r_{\mathrm{OO}} \rangle$, $\langle \theta_{\mathrm{HOH}} \rangle$, $\langle \theta_{\mathrm{OHO}} \rangle$, $\langle \theta_{\mathrm{OOO}} \rangle$, $\langle q \rangle$, % AD, % AAD, % ADD, % AADD, % AAAD, % AAADD, % 3-cycles, % 4-cycles, % 5-cycles, % 6-cycles. The fit is less predictive and begins to incorporate cycle counting as predictive features.

The developed regression models were applied to structures lying outside the dataset used for training to examine predictive behavior. A database of water clusters larger than n=25 was constructed using the same methods as outlined by Rakshit et.al. [196] Although the configuration space of these larger water clusters was not sampled at the same level of detail as the existing

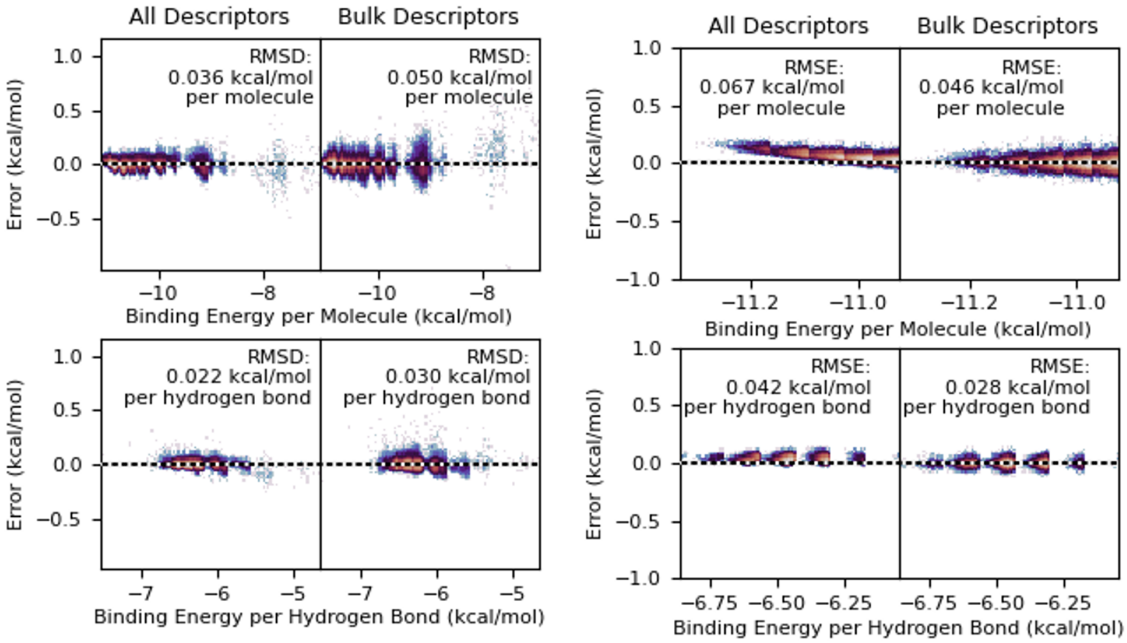

Figure 6.11: The errors of the linear regression for each model as 2-D histograms. Colors indicate the density of points: blue in the tens and red in the hundreds. Left panels trace the training error of the models for all structures $n \leq 25$ and those 5 kcal/mol from the punitive minima. Right panels display an extrapolation error for structures $26 \leq n \leq 30$, still lying within 5 kcal/mol from the punitive minima.

database for n=3–25, it nonetheless can provide useful information about any divergent behavior in the models. It was found that models based solely on bulk-like descriptors were successful in predicting the energy to a comparable accuracy for clusters inside and outside the training set. The RMSE values decrease, viz. from 0.050 kcal/mol per molecule to 0.046 kcal/mol per molecule, and from 0.030 kcal/mol per hydrogen bond to 0.028 kcal/mol per hydrogen bond. On the other hand, models that incorporate the size, ASPL, Wiener index and R_g, were found to over-estimate the energies for larger clusters, likely due to the non-linear correlation between energy and size (Figure 6.10 left). Complete extrapolation should never be assumed, as structures with descriptors not seen by the model may diverge.

The PCs of the PCA can also serve as indicators to predict stability. When using PCA, fewer parameters are required as the data has been reduced in dimensionality. The RMSE errors for

predicting the binding energy per molecule and per hydrogen bond are listed in table 6.1. Incorporating the first seven PCs produces relatively good fits. Although the fits are worse than the regression with bulk descriptors by 13-24 %, the number of parameters required is 61 % less. This current model is an excellent starting point in using reduced dimensions to predict energetic stability, but incorporating additional descriptors and higher energy local minima into the training set could improve the ability of the model to predict the binding energy.

The strong predictive power of the simplistic linear regressions is promising in that these cluster-averaged descriptors can be used to predict the energy of water cluster minima. Water cluster properties that strongly affect stability, such as proton ordering in the polyhedral cage water clusters $(H_2O)_n$, n=20, 24, 28, [317, 347, 348] are not explicitly incorporated in the model. Proton ordering might be implicitly described by other descriptors such as r_{OO}, resulting in predictive fits. With the incorporation of descriptors that explicitly account for proton ordering, the model is likely to improve. This assumption would be consistent with better-predicting models that use geometries explicitly as descriptors. [61, 62] Furthermore, more predictive models can be developed by using non-linear machine learning techniques. Polynomial regressions improve the fits by between 16 and 18 % (Table 6.1). Naturally, additional improvement is anticipated with more sophisticated models, such as neural networks, that are not the focus of this work. [297, 349, 350]

6.5 Conclusions

The database of water cluster minima generated by Rakshit et al. [196] has provided a useful resource to analyze structural and connectivity descriptors in water cluster minima (n=6-25). It was found that some descriptors, such as r_{OH} and % HB, correlated strongly with cluster size, while others, such as r_{OO}, θ_{OHO}, and q, correlate with one another due to their relationships with hydrogen bond strength. Cluster size was found to be the strongest predictor of stability (binding energy per molecule), while when descriptors are grouped together results in providing sufficient information to predict cluster stability. Energy prediction appears better when scaled for each hydrogen bond, as most descriptors directly affect the strength of individual hydrogen bonds.

The principal component analysis demonstrated the use of descriptor decomposition in classifying and predicting properties of water clusters. The PCA predicts the energy of each water cluster with ∼60 % fewer free parameters than linear regressions with a 13 % and 24 % loss in accuracy for

the binding energy per molecule and the binding energy per hydrogen bond, respectively. The PCA also shows promise in identifying similarly structured clusters of different sizes, allowing for the use of smaller clusters to model computationally expensive large clusters. The PCA model developed in this study is publicly available and can be used to further study the structural and connectivity patterns in either larger clusters or different aqueous systems. Improvements to the model, such as additional descriptors, additional isomers, and explicit/robust tests on bulk systems, will be the focus of future studies. Although the models described in this work were not trained on large clusters or liquid water/ice configurations, they should be transferable to structures with similar descriptors (containing 4, 5, 6 cycles, and mostly ADD, AAD, AADD adjacencies). Clusters and bulk structures that have descriptors well outside the ranges considered in this work will require additional training information to create a reliable model. Although some of the conclusions are expected, especially those regarding changes in cluster size and correlations with hydrogen bond strength (albeit not previously reported since the water cluster database has been made available only recently), many others are novel. These include the drastic transitions in the % of 5- and 4-cycles at n=17 and the correlation between ASPL and R_g. An additional novelty in this work is the synthesis of these structural and connectivity descriptors in describing these clusters, the structure classification of the PCA, and the energy regression from cluster-averaged descriptors. These findings set the stage for future investigations, as well as provides evidence that these traits persist throughout a wide range of water clusters.

Since the described approach is general, it can be easily extended to other training sets and other hydrogen-bonded systems to further assist in their structure prediction and structural motif analysis and classification, a field that has seen great improvements with the use of simplistic and descriptive machine-learning techniques. [351–354] In the next chapter, the descriptors and protocol developed here will be used to analyze the structural motifs for liquid water at different temperatures.

Chapter 7

DESCRIPTORS OF WATER AGGREGATION: II. LIQUID WATER

7.1 Introduction

The prior chapter provided a clear analysis of the hydrogen bond framework in water clusters, highlighting the use of simple machine learning methodologies that identified structural and energetic trends in these clusters. Their results showed promise that these set of descriptors, many of which are highly studied by physical chemists, can describe the extensive hydrogen bond properties of water. It remains to be seen how the introduction of temperature will influence the correlations and predictive power of these descriptors.

Liquid water is of extreme interest because of its anomalous properties, some of which were discussed in the introduction chapter. Water has extraordinarily high thermodynamic properties, including the enthalpy of fusion, enthalpy of vaporization, and heat capacity. The former two are directly related to the strength of the intermolecular interactions that are altered during a phase transition; they are large due to both the high strength of hydrogen bonds in water as well as the numerous hydrogen bonds per molecule (~ 2). The high heat capacity has been attributed to the intermolecular vibrations of the hydrogen bond through comparisons with hydrogen sulfide. [355] It is unclear how much of a role the breaking of hydrogen bonds play in the changing internal energy of liquid water.

Other anomalous properties regard non-linear behavior across the temperature regime. One of the most important anomalous properties is the density maximum. Unlike most other liquids, water is not most dense at the freezing point, instead it's density maximum is at 4°C. [130] Other deviations from typical liquid behavior is found in the isothermal compressibility around 320 K. [4, 356] Isothermal compressibility is closely related to the temperature derivative of density [356] suggesting a coherent, single cause for water's odd temperature dependent behavior. More recent studies have presented evidence for a liquid-liquid transition around 330 K. [357] Using the descriptor of tetrahedrality, they found that the well defined low-density (LDL) (tetrahedral) phase

of supercooled water is present in ambient water, and they observed a decrease in the LDL phase at 330 K. It is evident that the structure of liquid water changes throughout it's temperature range, and connecting these changes to water's anomalous observable properties is an active and challenging area of research.

In this chapter, our preliminary analysis of applying the descriptor framework to liquid water trajectories is presented, with an explicit emphasis on how these descriptors change with temperature from the melting to the boiling points. In section 7.3, changes with respect to single descriptors will be discussed in great detail. Later, in section 7.4, the descriptor framework will be applied to more traditional techniques for analyzing molecular dynamics (MD), namely the radial distribution function (RDF). Connections between the descriptors and oddities in water's observable and non-observable behavior will be made in order to attribute some of water's anomalous properties back to its physical origins in hydrogen bonding.

7.2 Methodology

The database of liquid trajectories was constructed by running canonical MD simulations using the TTM2.1-F flexible, polarizable, many-body interaction potential for water. [50] This potential was chosen as a balance between accuracy (properly describing the structure of liquid water) and cost. Simulations were performed using periodic boundary conditions with 256 water molecules at $T =$ 280, 290, 300, 315, 330, 345, and 360 K, at the experimental densities. [130] Simulations were run for 1.2 nanoseconds with a time step of 0.5 femtoseconds. Hydrogen bonds were identified through Skinner's $r - \psi$ definition. [63] The resulting simulations total nearly 1 terabyte in Cartesian geometries. Despite the large size, the following analysis was performed on the entire dataset.

Descriptors were computed for each frame using the Cartesian geometries and the graph representations of the hydrogen bonding network. We focused on the descriptors from the prior chapter that proved to be important in describing structural and energetic character in water cluster, including r_{OH}, θ_{HOH}, r_{OO}, θ_{OHO}, θ_{OOO}, q, percent adjacency, percent cycles, and percent hydrogen bonding. Certain descriptors, such as the Wiener index, are not defined for the bulk phases and thus cannot be used in the context of this work. The incorporation of temperature increased the stochasticity to the structural descriptors, making bulk-averaged values less informational. Thus, bulk averaged descriptors were not used in this work.

The relationships between descriptors were analyzed using a variety of methods. Linear regressions were performed using the SKLearn package. [325] Uneven sampling in temperature lead to a bias in high hydrogen bond saturation. To minimize over sampling of the low temperature/highly connected frames, 1/3 of the 290 K frames were selected for regression analysis of descriptors against the number of hydrogen bonds.

Radial distribution functions, $g(r)$, were computed as

$$g(r) = \frac{\rho(r)}{\rho} \tag{7.1}$$

where ρ is the density of water molecules, r is the distance between two particles, and $\rho(r)$ is the particle density at some distance r away from a source particle. $\rho(r)$ is computed by counting the number of particles in binned distances between 1 and 9 Å by 0.0001 Å, and scaling by the volume of the radial bin. The distribution was computed for each particle in each frame and averaged to produce a single distribution function. A sliding boxcar averaging was applied for $\Delta r = 0.01$ Å. Therefore, the precision in the RDF is approximately 1,000$^{\text{th}}$ of an angstrom. A descriptor selected distribution function, $g_{\alpha,\beta}(r)$, was developed to describe the distribution functions of water molecules with descriptor β around water molecules with descriptor α. The radial density, $\rho_{\alpha,\beta}(r)$, is scaled by the bulk density of particles with descriptor β, ρ_β, in order to ensure $g_{\alpha,\beta}(r)$ goes to 1 as r approaches infinity.

7.3 Changes in Descriptors with Temperature

Hydrogen bonds break as temperature increases. This has been known before with correlations of hydrogen bonds and temperature seen in the high-pressure, high-temperature region of the phase diagram and small changes seen in the ambient temperature region of the phase diagram (280-380 K). [128, 358, 360] The simulations in this chapter qualitatively agree with the past trends for the % hydrogen bonds, with systematic deviations appearing because of changes in the definition of hydrogen bonding. For example, the work by Suresha and Naik [358] reproduce the dielectric constant for water as a function of temperature, and use statistical mechanics to back calculate the thermodynamic properties of hydrogen bonds as well as the number of hydrogen bonds per water molecule. This assumes that the majority of interactions comes from hydrogen bonds alone. Our method provides support to their claims that hydrogen bond networks change with temperature by

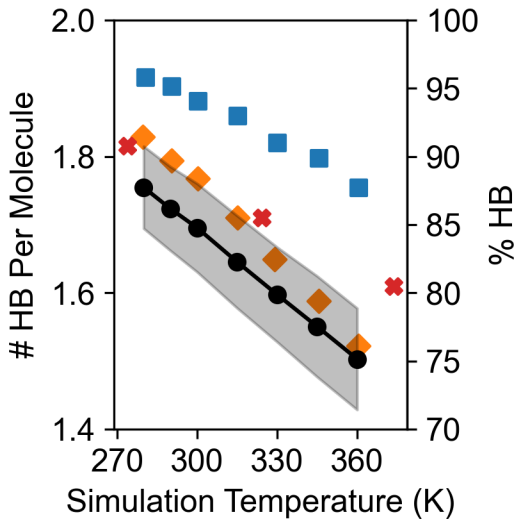

Figure 7.1: Anti-correlation between temperature and the number of hydrogen bonds per molecule. Black circles show the average number of hydrogen bonds for each simulation temperature, with the gray error bars showing two standard deviations. Blue squares come from ref. [358], orange diamonds from ref. [359], and red X's from measured IR spectra in ref. [128].

evaluating the hydrogen bond networks precisely and classifying hydrogen bonds from structure, instead of changes to the dielectric constant or infrared vibrational bands.

Under this structural definition, hydrogen bond saturation is 87.5% near freezing and 75% near boiling, see figure 7.1. This is in the range seen in water clusters, although the free OH environment is distinctly different. In clusters, free OHs appear on the surface of the clusters, whereas in liquid water OHs are surrounded by distant water molecules. On average, each water molecule participates in between 3.5 and 3 hydrogen bonds (1.75 to 1.5 hydrogen bonds per molecule). Approximately one third of water molecules will break a hydrogen bond break for every 100 K. The relationship is strong, with $R^2 = 0.87$. The hydrogen bond saturation found here agrees will with reports by Luzar et al. [359] and Luck et al. [128], and slightly under-counts the hydrogen bonds compared to Suresha and Naik. [358] It should be noted that the structural definition of the hydrogen bond is not unique, and variations in the definition will systematically adjust the hydrogen bond counts.

The average enthalpy of the simulations increase with simulation temperature, with an isobaric heat capacity 27.5 cal/mol K, or 1.53 cal/g K ($R^2 = 0.96$), see figure 7.2. This value agrees well

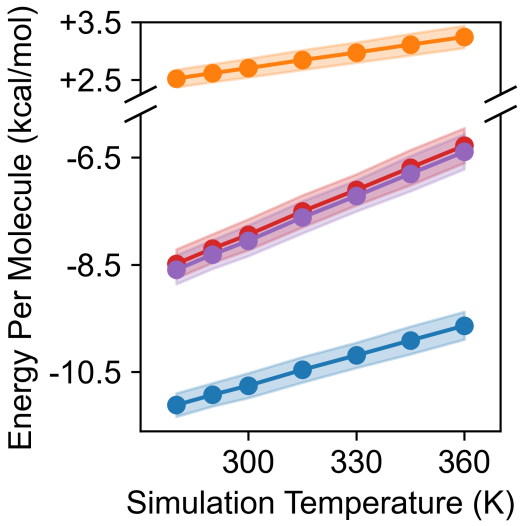

Figure 7.2: The enthalpy (red), internal energy (purple), potential energy (blue), and kinetic energy (orange) per molecule increases with simulation temperature.

with the isobaric heat capacity of water simulated with MB-pol (27.6 cal/mol K average from 278 K to 368 K). [361] Since the simulation volumes remain about the same for all temperatures (3% increase over 100 K), the isochoric heat capacity is found to be the same as the isobaric heat capacity. The internal energy is plotted against simulation temperature in figure 7.2 ($R^2 = 0.96$). Both metrics over-estimate the experimental heat capacity of water (≈1 cal/g K) due to the lack of nuclear quantum effects. [362] When a correction of 6 cal/mol K is applied, corresponding to differences seen in classical and path-integral simulations, [363] the heat capacity is lowered to 1.19 cal/g K.

The internal energy can be further decomposed into the kinetic and potential energy terms, see figure 7.2. The kinetic energy increases by 9.0 cal/mol K, whereas the potential energy increases by 18.5 cal/mol K. The potential energy contributes about 67% to the classical heat capacity of liquid water, whereas the kinetic energy contributes only 33%. Increases in the kinetic energy can be attributed to thermally accessible low energy degrees of freedom, namely the intramolecular librations. Increases in the potential energy can be attributed to changes in the strong interactions between water molecules, namely the hydrogen bonds.

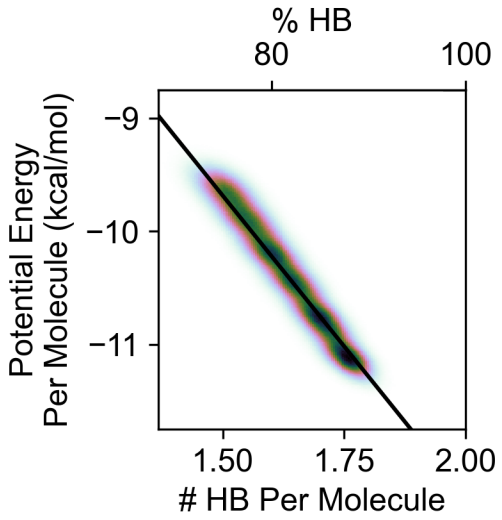

Figure 7.3: The potential energy of the simulation decreases with the number of hydrogen bonds. The slope of this relationship describes the average energy of each hydrogen bond, and the intercept describes residual non-hydrogen bonding interactions.

The potential energy of each frame correlates strongly with the number of hydrogen bonds in the simulation ($R^2 = 0.91$), see figure 7.3. For each hydrogen bond broken, the potential energy raises by 5.34 kcal/mol, meaning that each hydrogen bond is about 5.34 kcal/mol. The intercept of this relationship is -1.66 kcal/mol and describes non-hydrogen bonding contributions to potential energy, such as long range electrostatics.

When we consider that 0.32 hydrogen bonds break per molecule every 100 K, and that each hydrogen bond is about 5.34 kcal/mol in potential energy, we can find that the potential energy increases by 16.9 cal/mol K from the breaking of hydrogen bonds. This is about 90% of the 18.5 cal/mol K for the full simulation, meaning the majority of water's high heat capacity comes from the breaking of hydrogen bonds.

The adjacency environment was shown to correlate with hydrogen bond saturation in water clusters; therefore a relationship with temperature is expected as well, figure 7.4. The number of AADD water molecules decreases steadily with temperature, being between 60% and 30% for all temperatures and hydrogen bond saturations. Under coordinated water molecules (AAD, ADD,

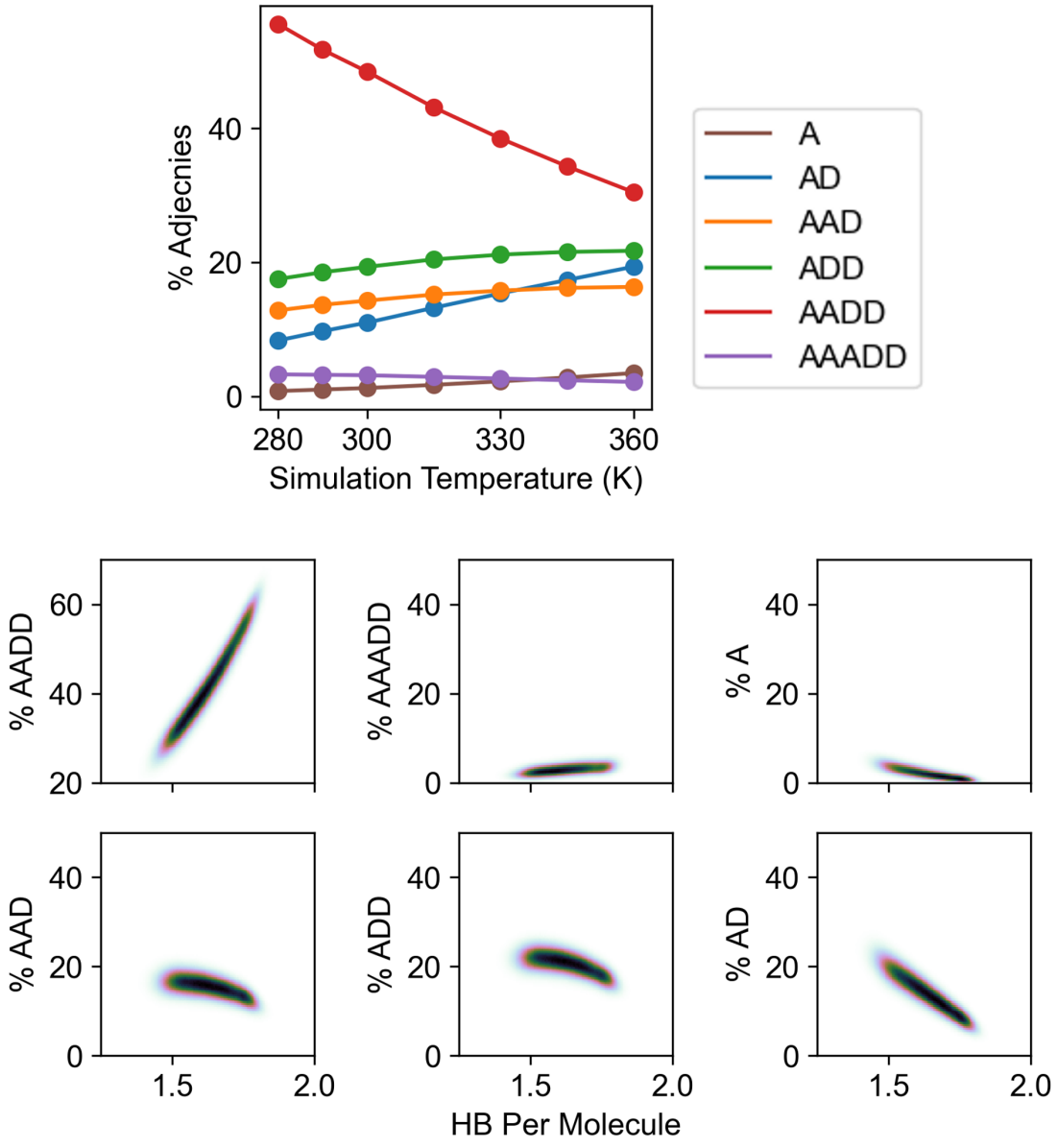

Figure 7.4: The percent of each water molecule in each adjacency class as a function of temperature (top) and the number of hydrogen bonds per molecule (bottom).

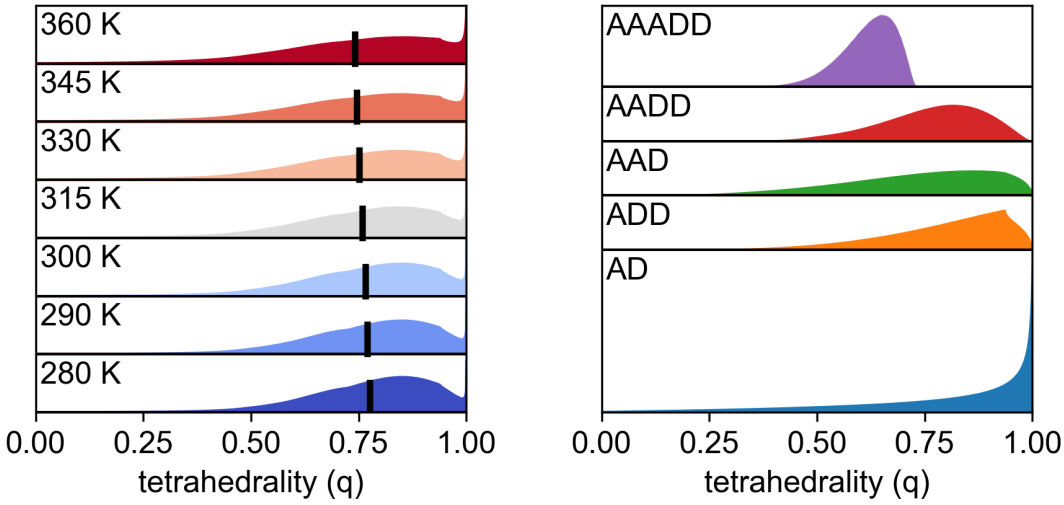

Figure 7.5: Distributions of tetrahedrality (q) with respect to temperature (left) and adjacency (right).

and AD) increase with temperature and decrease with hydrogen bonds, with AD water molecules ranging most drastically, from 5% to 25 %. AAADD and A water molecules make up a sizable portion of the dataset, being up to 5% of the water molecules in each simulation frame. These ranges (except for A) are all in the general ballpark of the water clusters seen in the last chapter. In contrast to ice phases, liquid water has a diverse adjacency environment that is sensitive to temperature. The properties of liquid water may change throughout the temperature range due to changes in the local properties of each adjacency.

Tetrahedrality is one such descriptor that changes in response to adjacency. The average tetrahedrality of each frame decrease weakly with temperature ($R^2 = 0.59$). The tetrahedrality of each individual water molecule strongly depends on the adjacency environment of that water molecule (figure 7.5). AADD, ADD, and AAD molecules all have a similar range of adjacency, with a maximum less than 1 due to the low probability that multiple O-O-O angles will be at 109.5°. AAD molecules have a longer tail to lower adjacency due to the delocalization of the accepting hydrogen bonds. Conversely, ADD molecules have a sharp peak at higher adjacency due to the rigid positioning of donating hydrogen bonds.

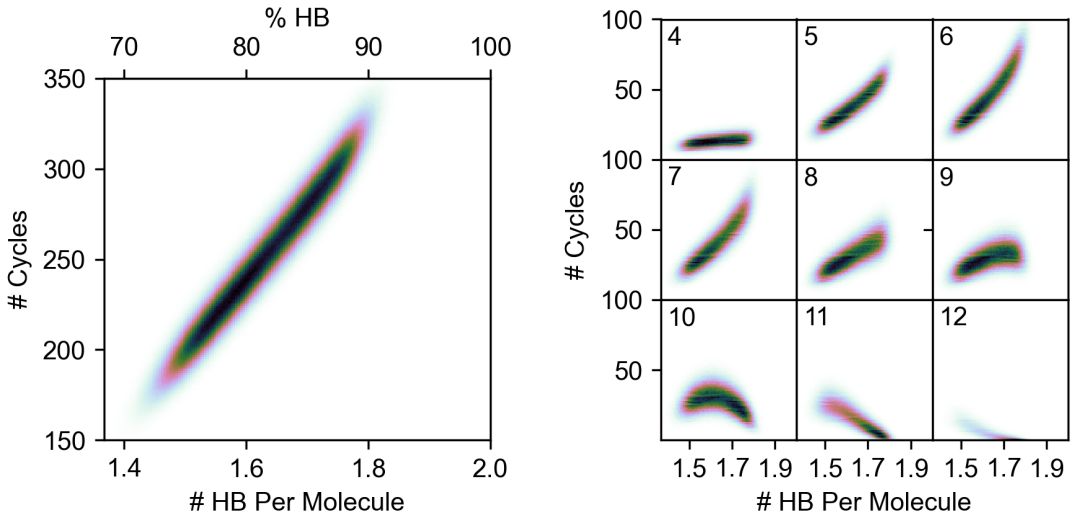

Figure 7.6: Number of cycles per hydrogen bond in the simulation. Total number of cycles are on the left, whereas the number of cycles of a particular size is displayed on the right.

AD water molecules have a peculiar distribution which is most "tetrahedral". This is mathematically because there is only one O-O-O angle in AD water molecules, and it vibrates around 109.5°. Therefore, snapshots of AD molecules will see the "tetrahedral" (more appropriately bent) geometry of the water network. AAADD water molecules are not tetrahedral, since there are too many molecules to have O-O-O angles be near 109.5°. Instead these defective water molecules have a maximum tetrahedrality near 0.7 and a similarly shaped Poisson distribution to low tetrahedral geometries. The increase in AD water molecules with temperature is seen in the distribution of each water molecule's tetrahedrality with temperature, with a peak growing near 1 in figure 7.5. Average tetrahedrality decreases with temperature because of the decrease of AADD and increase in AAD.

Non-short-circuited cycles are known to be important in the stability of water clusters and are characteristically different between the phases of ice. [112, 125] In liquid water, however, the distribution of cycles is very large, with most cycles containing 5, 6, or 7 water molecules, but some cycles increasing to size 11, 12 or 13 water molecules at high temperatures. We found that the number of cycles strongly decrease with temperature ($R^2 = 0.81$) but more strongly increase

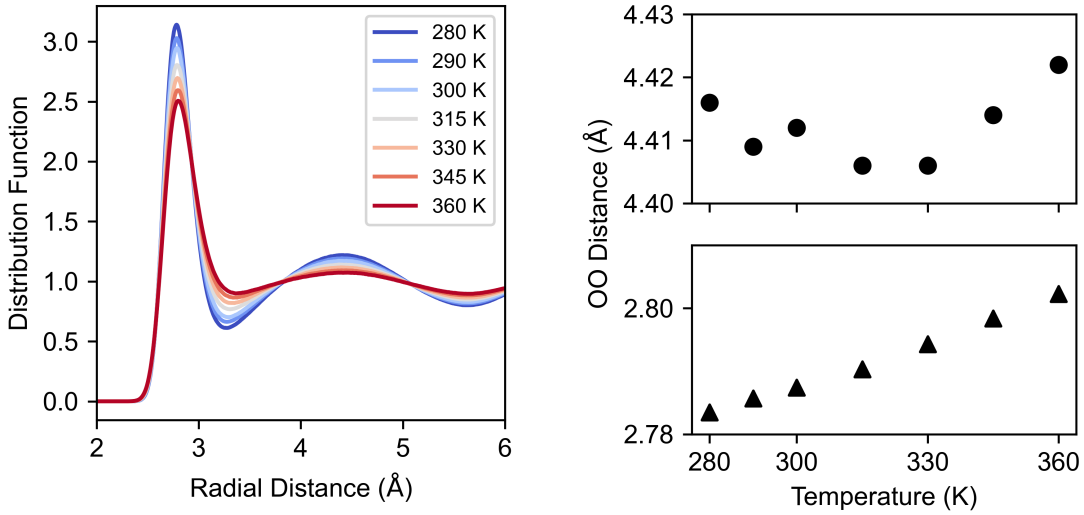

Figure 7.7: The radial distribution function of the simulations with the TTM2.1-F potential at different temperatures (left). The positions of the peak maxima are plotted on the right.

with the number of hydrogen bonds ($R^2 = 0.94$), suggesting that the changes in cycles come from the breaking of hydrogen bonds, see figure 7.6. Although it is true that small cycles disappear as hydrogen bonds are broken, larger cycles appear. This suggests a mechanism in which fused rings have their shared edge broken, forming a larger non-short-circuited cycle. Water, as it approaches it's boiling point, adopts large entropy-rich rings instead of becoming a disconnected network.

Other descriptors established in the prior chapter do not show large changes with temperature due to the influence of MD. Descriptors such as r_{OO} and θ_{OHO} contained information in water clusters because geometries were restricted to local minima, and very subtle differences could be observed. In an MD simulation, these descriptors oscillate around their expected values with high variance. Therefore changes with respect to temperature and correlations with other descriptors were weak. Furthermore, temperature effects induce a shift in the average values of descriptors, such as r_{OO}, away from their 0 K equilibrium values, pushing these liquid frames structurally far away from the water clusters investigated in the prior chapter and making a comparative analysis with the PCA impossible. Instead, we can rely on techniques traditional to MD to analyze metrics that correspond to these descriptors

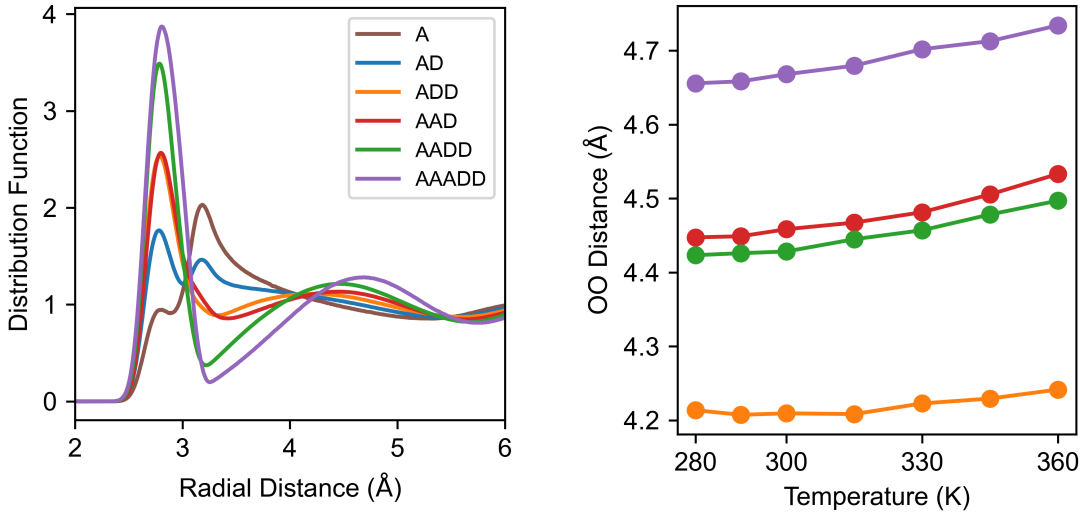

Figure 7.8: The radial distribution function of the simulations with the TTM2.1-F potential originating from different adjacencies (left). The positions of the peak maxima are plotted on the right

7.4 Adjacency Selected Radial Distribution Functions

The radial distribution function is a traditional metric to describe the radial structure of an ensemble of molecules and has been widely studied for liquid water. In order to describe r_{OO} which has high variance frame-to-frame, we have computed the radial distribution function of the simulations at each temperature, figure 7.7. As seen earlier, [361] the first solvation shell appears near 2.8 Å, and increases with temperature. The second solvation shell initially gets closer as temperature increases and then grows to further distances. This trend matches what is seen in experimental and theoretical RDFs but to a lesser extent. [126, 127, 361] Furthermore, the peak positions are systematically at closer distances. Nonetheless, the trends exist and we can use the descriptor framework to understand changes in radial distribution functions.

The radial distribution functions were decomposed based on the adjacency of water molecules, figure 7.8. The peaks of the first solvation shell were fit to the a gaussian to quantify peak positions and height, reported in table 7.1. The peak position changes minimally with adjacency, being between 2.78 Å and 2.81 Å. Within each adjacency, the peak position increases with temperature. The increases in first solvation shell position with respect to temperature is attributed to the

Table 7.1: Peak positions (in Å) and heights of the adjacency selected radial distribution functions.

Adjacency	1^{st} Peak Distance	2^{nd} Peak Distance	1^{st} Peak Height	2^{nd} Peak Height
A	2.803	3.186	0.945	2.028
AD	2.783	3.181	1.766	1.463
ADD	2.789	4.217	2.534	1.110
AAD	2.801	4.447	2.567	1.133
AADD	2.787	4.467	3.488	1.214
AAADD	2.811	4.680	3.869	1.280

decreasing density instead of the changing adjacency. The peak height increases linearly with the number of hydrogen bonds from A to AADD; each additional hydrogen bond increases uniformly in population in the first solvation shell. This suggests that adjacency changes and the hydrogen bonds more commonly break through increasing O-O distances rather than changing (intermolecular) hydrogen bond angles. The peak height only slightly increases for AAADD due to a much broader solvation shell stemming from the more-unequal O-O distances between the donating hydrogen bonds and the excessive number of accepting hydrogen bonds.

The position of the second solvation shell (right panel of figure 7.8) depends greatly on the number of accepting hydrogen bonds, being upwards of 4.68Å for AAADD, down to 4.22Å for ADD water molecules. The presence of AAADD water molecules at low temperatures and the conversion to ADD water molecules in the middle of the temperature range is a reason for the distant second solvation shell below room temperature. Increases in the distance of the second solvation shell return at high temperature due to the typical increase in solvation shell position with respect to temperature. This highlights the role that adjacency plays on the anomalous behavior of liquid water, specifically in the structure of the second solvation shell.

The second solvation shell of A and AD water molecules is anomalously close, being only ∼0.4 Å from the first solvation shell. An earlier study [364] identified this odd feature and hypothesized it to be associated with a metastable intermediate in the process of hydrogen bond breaking. It again appears in these trajectories and with this structural definition of the hydrogen bond. Therefore,

this warrants further investigation. The height of this peak is not correlated with temperature, initially suggesting it may not be fully a consequence of hydrogen bond breaking, which increases in frequency with respect to temperature. To test this, we tracked the position of water molecules as they passed through OO distances of 3.18 Å.

For pairs of adjacencies, we identified water molecules between 3.18 and 3.19 Å apart and computed the descriptors of the two water molecules at 3000 fs before and after the pair entered this interstitial region. Selected trajectories are shown in figure 7.9, focusing on three pairs. AD-AD pairs showed the most stark peak in the radial distribution function at 3.18Å, so became our prominent sampling of this "metastable structure". AAD-ADD pairs also showed a peak in the radial distribution function at 3.18Å, and they were used to identify trends that are not characteristic of adjacency. Finally, AADD-AADD pairs were identified as negative controls, detailing how water molecules move between solvation shells without forming the "metastable structure".

Diverse paths of trajectories were observed, and it became clear that this peak is not purely associated with the hydrogen bond breaking process. Many instances exist where O-O vibrations from the first solvation shell extend into this "1.5" solvation shell region (center column of figure 7.9). Although this is commonly joined with the hydrogen bond being broken, the hydrogen bond is re-formed throughout the course of the oscillation. In other instances, there appears a stable, long-lived oscillation of the O-O distance around 3.18 Å (left column of figure 7.9). This pattern was not connected to a formal hydrogen bond breaking process. Rather, when a pair of water molecules passes through this region to move from one solvation shell to the other, fully breaking the hydrogen bond (right column of figure 7.9), they do so in a single step. In these events, there does not appear to be a significant and concerted slowdown in distance around 3.18 Å, and therefore there is no evidence for an intermediate structure associated with hydrogen bond breaking. In fact, many instances of hydrogen bond breaking show steady first solvation shell oscillation before rapid dissociation, suggesting that the majority of the population seen in this region comes from solvation shell oscillations instead of a metastable intermediate of hydrogen bond breaking.

The cause of these anomalous peaks comes from the definition of the hydrogen bond. As r_{OO} increases, it becomes increasingly likely that the hydrogen bond breaks temporarily. This is seen in figure 7.9 where the hydrogen bond readily breaks around r_{OO} =3Å. This hydrogen bond breaking converts AADD-AADD pairs into AAD-ADD pairs and AAD-ADD pairs into AD-AD pairs, biasing

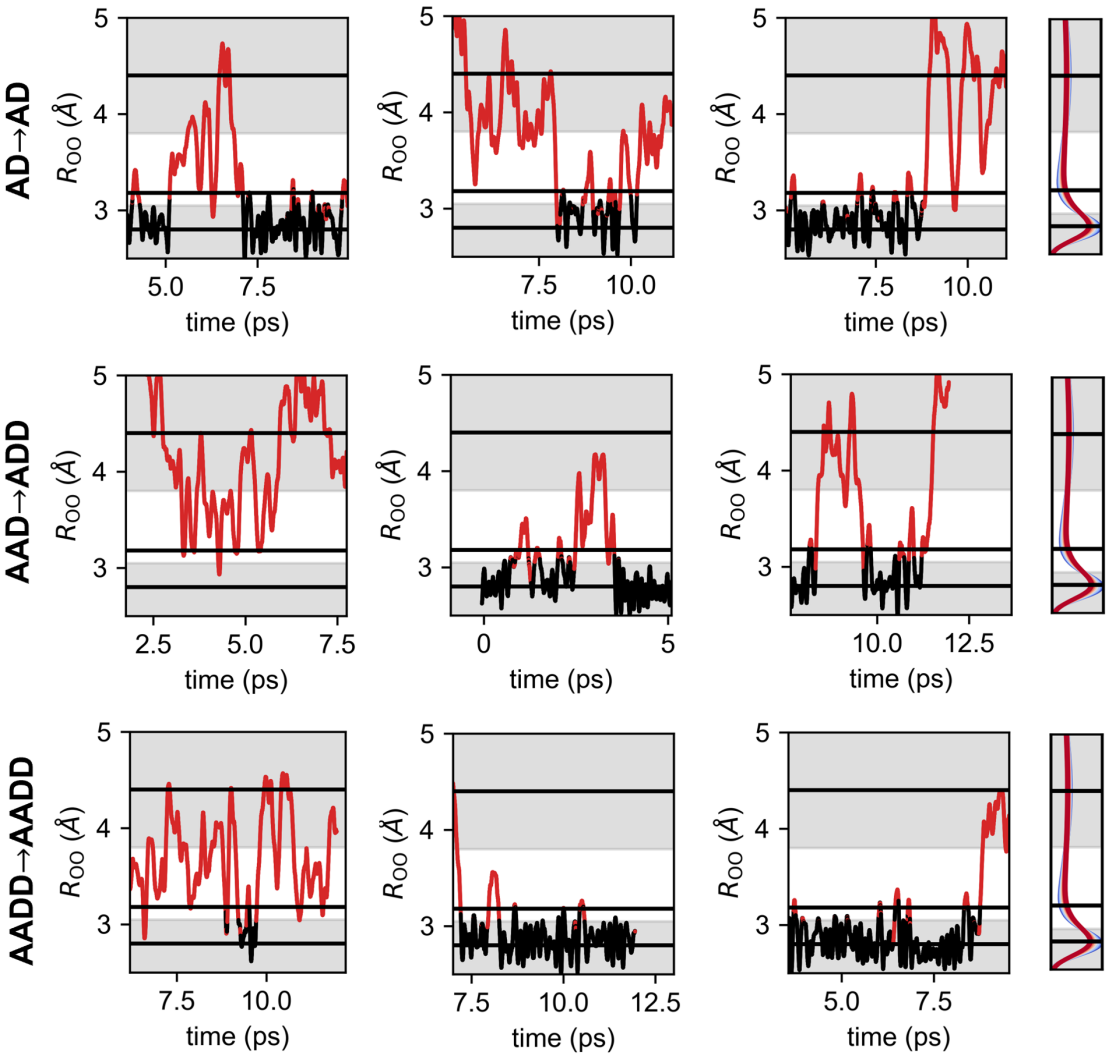

Figure 7.9: r_{OO} distances for selected water molecule pairs. The pairs are hydrogen bonded when the line is black, and not hydrogen bonded when the line is red. Solvation shells are highlighted with gray backgrounds. Approximate peaks in the radial distribution functions are marked through horizontal black lines. On the far left are the RDFs.

this region to lower adjacencies. Therefore, it is our belief that this feature in the RDFs comes directly from defining the hydrogen bond and not from the physical properties of water. This is further supported by comparisons to the radial distribution functions from ref. [364]. Using a different definition of the hydrogen bond, they observe more sharp peaks in the AD, AAD, and ADD radial distribution function at 3.18 Å. If these peaks come from the boundary between bound and unbound water molecules, then they should manifest differently when the hydrogen bond is defined differently.

7.5 Conclusions

In this chapter, the descriptor framework was applied to liquid water trajectories. Although most information was "averaged out" through the addition of temperature, certain descriptors provided insight into important changes in hydrogen bond structure and corresponding properties in liquid water through the temperature regime. The number of hydrogen bonds decreased linearly with temperature and had important effects on the energy of liquid water. Notably, 90% of water's heat capacity contribution from potential energy could be attributed to the breaking of hydrogen bonds, supporting past claims that water's high heat capacity stems from the energy of the hydrogen bond. [355] The number of hydrogen bonds had a direct effect on the distribution of adjacencies of each water molecule with downstream effects observed.

Tetrahedrality decreased with temperature due to a complicated shift in adjacencies that had distinct tetrahedrality distributions. More importantly, adjacencies greatly effected the position of the second solvation shell, with AAADD molecules having significantly distance solvation shells and ADD molecules has a slightly closer second solvation shell. Decreases in the number of AAADD molecules explain the initial and anomalous contraction of the second solvation shell between 280 and 315 K, and provide a framework to analyze the temperature dependent anomalous properties of water. It should be noted that AAADD molecules are characteristic of the high density phase of water, whereas past works suggested the low density phase is the cause of the liquid-liquid phase transition. [357] These two should be investigated in tandem in order to identify potential cooperative descriptors for the density maximum in bulk water.

Finally, the results presented herein refute the claims of a metastable stable intermediate in the hydrogen bond breaking process. The peaks in the adjacency selected radial distribution functions

at 3.18 Å for A and AD water molecules is a consequence of the changing adjacency of water molecules and not a minimum in the potential of mean force. Hydrogen bonds break and reform interchangeably in this region, producing increased probability of low adjacency water molecules. This becomes an artifact of the hydrogen bond definition and as such it does not provide any insight into the properties of bulk water.

Chapter 8

AFFINITY OF NICOTINOIDS TO A MODEL NICOTINIC ACETYLCHOLINE RECEPTOR (NACHR) BINDING POCKET IN THE HUMAN BRAIN [365]

8.1 Introduction

Nicotine addiction is a global health crisis and has been perpetuated by the evolution of tobacco products in response to social and regulatory shifts. [91, 366, 367] Recently, nicotine consumption through flavored vape products has revitalized tobacco use in the youngest generation. [368, 369] In response, the US Food and Drug Administration (FDA) is taking regulatory action, announcing the plan to limit the amount of nicotine in all products. [370] In the last couple of years, the FDA announced that it would lower the amount of nicotine allowed in cigarettes "to minimally addictive or non-addictive levels." [371] The tobacco industry has a mandate to lower the amount of nicotine in their products and the incentive to introduce nicotine-like analogues (nicotinoids) to their products to abide by the upcoming regulations. Understanding the mechanism of nicotine addiction from a molecular level and predicting how similar molecules will act in this capacity is therefore of great importance in the tobacco industry and tobacco regulation.

The biochemical mechanism for nicotine addiction is well understood. Nicotine and similar alkaloids (nicotinoids) are known to interact with a family of neuroreceptors known as nicotinic acetylcholine receptors (nAChR). [84, 91, 92, 372] These receptors were discussed briefly in the introduction and will be discussed in further detail here. These receptors are found throughout the muscular and nervous systems with roles in diseases ranging from Alzheimer's, schizophrenia, and addiction. [88–92] Each receptor is a pentameric ligand-gated ion channel consisting of various subunits, α-ϵ. [86] Receptors in the central nervous system, including the one responsible for addiction and pleasure, consist of only α and β subunits. Agonists bind to one side of each α subunit, which opens the ion-gate that causes intracellular signaling and eventual dopamine release. [84, 85] Antagonists are also known to bind without opening the ion channel.

Nicotine is known to be most selective to the specific nAChR $\alpha_4\beta_2$ variant, found in the central

nervous system and tied to the pleasure response. [91, 93, 373–377] It is through this interaction that nicotine becomes addictive in humans. The crystal structures of nAChR and similar acetylcholine binding proteins (AChBP) show that nicotine binds at the interface between the α and β strands through strong interactions between the pyrrolidine nitrogen and the tryptophan residue. [93, 373, 374, 377] It is assumed that nicotine remains protonated at the pyrrolidine nitrogen due to the nature of the interaction with the binding pocket. The energetics of the protonation of nicotine have been the focus of previous studies. [378–380] With two basic nitrogen atoms, two singly-protonated isomers (protomers) are possible, protonated at either the pyridine nitrogen (Pyri-H^+) or at the pyrrolidine nitrogen (Pyrro-H^+), forming a pyridinium and pyrrolidinium ion, respectively. It is known that at biological pH, nicotine and similar compounds exist as Pyrro-H^+. [381–383] Recent studies, however, show that a nearly equal mixture of the two nicotine protomers is present in the gas phase, whereas Pyri-H^+ dominates in the gas phase for nornicotine, the demethylated form of nicotine. [378, 384] Therefore, the stability of both protonation sites in non-aqueous environments, such as the gas phase and peptide pockets, is an important consideration, despite the strong evidence that a single protomer is present in biological systems.

This chapter focuses on quantifying and analyzing the nature of the interactions between nicotinoids and the nAChR $\alpha_4\beta_2$ receptor in an effort to correlate the binding affinity to the experimentally determined level of addiction. The addictive thresholds of nicotine and other similar nicotinoids have been previously reported through intracranial self-stimulation (ICSS) threshold studies. [367] These studies examined 6 nicotinoids found naturally in tobacco leaves, viz. nicotine (NIC), nornicotine (NOR), anabasine (ANB), anatabine (ANT), myosmine (MYO), and cotinine (COT), with the results displayed in figure 8.1 together with the structures of their bioactive protonated forms. The neutral (not protonated) forms of these compounds are composed of two heterocyclic rings that contain four or five carbon atoms and one nitrogen atom each. Specifically, they all have a common basic 6-membered pyridine ring (located at the top right of each structure in figure 8.1), a structural motif that is well preserved in the nicotinoids examined in this study. Each nicotinoid also has a second heterocyclic 5- or 6-membered ring, connected to the pyridine ring through a single bond (located at the bottom left of each structure in figure 8.1) as follows: NIC has *N*-methyl-pyrrolidine, NOR has pyrrolidine, ANB has piperidine, MYO has 1-pyrroline and ANT has 4-piperidene. Finally, COT has *N*-methyl-2-pyrrolidone, is a five-membered, *N*-methylated

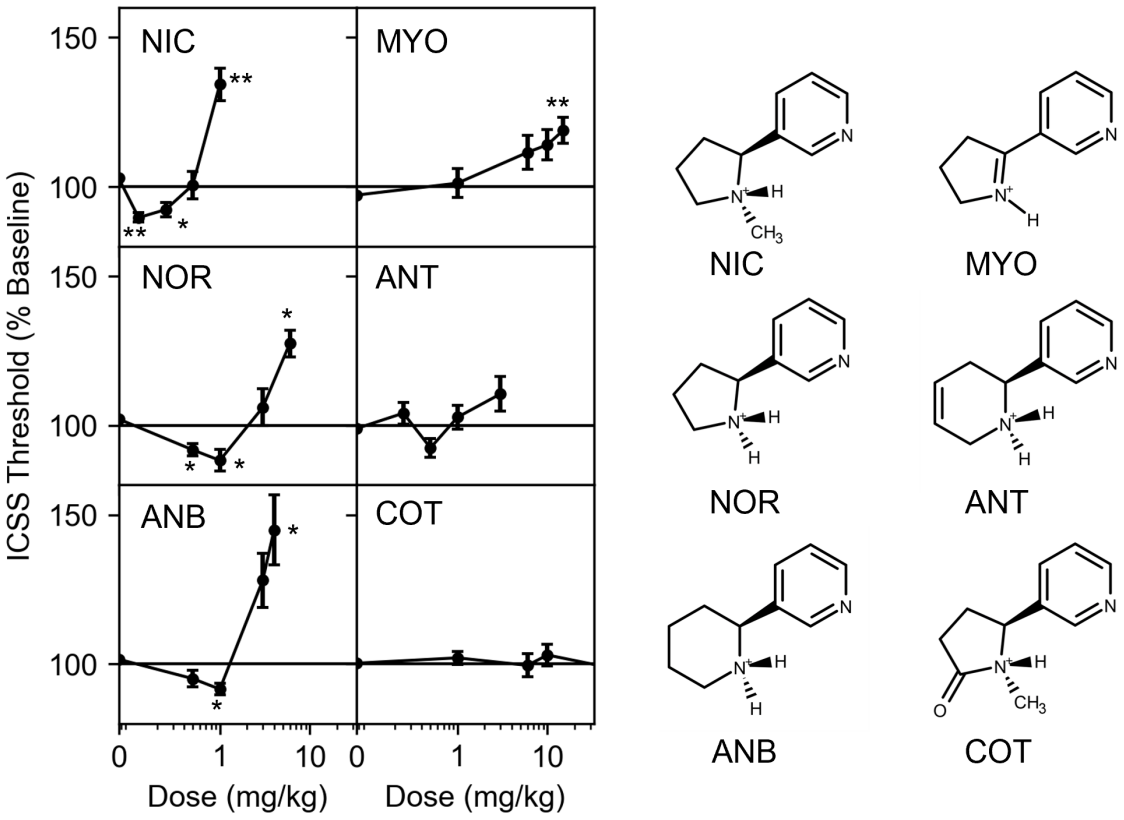

Figure 8.1: The six nicotinoids that have been studied through ICSS threshold experiments by Harris et al., reproduced with their permission [367]. Copyright 2015 Elsevier. Stars (*) represent p-values < 0.05 and double stars (**) p-values < 0.01, describing the probability that the population at that dose level is different than the baseline.

lactam. All except for MYO have a chiral center in the S configuration. The orientation of the rings is in the anti-syn configuration as suggested by Yoshida et al. and the binding structure reported by Morales et al. [93, 379] The protonation of these nicotinoids can occur either on the pyridine nitrogen (Pyri-H^+ protomer) – a common protomer for all 6 nicotinoids since they all have a pyridine ring – or on the nitrogen atom of the other-than-pyridine ring. Since this second ring is different for the various nicotinoids, we will hereafter refer to that second protomer as the *non*-Pyri-H^+ one. This protomer corresponds to the bioactive form and it is the one shown for all nicotinoids in the right panel of figure 8.1. It should be noted that *non*-Pyri-H^+ MYO hydrolyzes in aqueous solution with an equilibrium constant of 1.88. [385] Therefore, both forms are present in biological systems and will be considered. Additionally, the *non*-Pyri-H^+ protomer of COT is highly unstable, and only the Pyri-H^+ exists in aqueous solution. Although this and other protomers maybe more favorable, they may not be able to bind to the activated receptor; therefore, only the *non*-Pyri-H^+ protomer used in this study. The left panel of figure 8.1 is composed using the raw data in figure 1 of reference [367] that was provided by one of the authors of that study. In the rodent model, it is found that NIC is the most addictive naturally occurring nicotinoid, NOR and ANB are less addictive, MYO and ANT could be marginally addictive, and COT is virtually non-addictive. [367] These ICSS studies provide an indispensable benchmark that quantifies the potency of these six nicotinoids to elicit addictive behaviors.

We computed the interaction between models of the nAChR $\alpha_4\beta_2$ binding pocket (p) and the 6 nicotinoids (X) from ab-initio electronic structure calculations. These models were built with progressively increasing sizes with the objective of progressively incorporating the relevant interactions between the nicotinoids and the environment of the binding pocket. Our intent was to investigate whether the binding affinity of nicotinoids correlates with their reported ICSS threshold increasing/decreasing doses, [367] in other words whether the binding energy to the pocket is a potential descriptor for addiction that can be used for the fast and efficient screening of future nicotine analogs. Additionally, the various energetic components of the total binding were analyzed in an effort to provide insight into predicting the binding affinity from molecular properties of the various pieces comprising the binding pocket and to develop a computationally efficient protocol for screening future nicotine analogs for addiction.

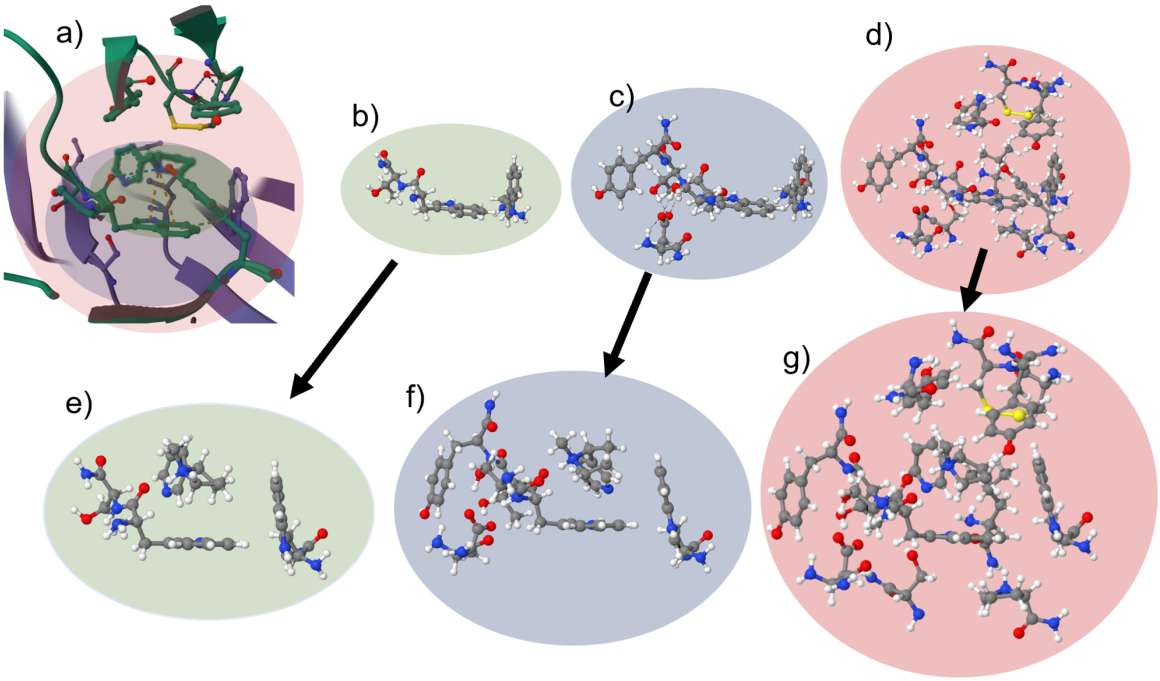

Figure 8.2: Model pockets were constructed from the crystal structure reported in PDB 5KXI [393]. (a) the binding domain. The three model pockets containing (b) 3, (c) 6 and (d) 14 amino acids used in this study. (e)-(g) nicotine bound to the three model pockets. The Tyr100 residue is removed from (g) to more easily see nicotine.

8.2 Methodology

The electronic structure calculations were performed at the ωB97X-D [163]/6-31++G(d,p) [386, 387] Density Functional level of theory [154] using Gaussian 16. [173] To speed up the optimization process, geometries were pre-optimized with the smaller 3-21+G basis set [388, 389] and then refined with the larger 6-31++G(d,p) basis set. The PCM solvent model was used to describe the fully solvated state of each nicotinoid in water for comparisons to more biologically relevant energy differences. [390] The calculations of the binding energy between nicotinoids and the model pockets remained in the gas phase. The ωB97X-D functional was chosen for its ability to describe long-range interactions ranging from hydrogen bonding to π-type interactions, [163, 391, 392] whereas the 6-31++G(d,p) basis set was chosen as a compromise between efficiency and accuracy, viz. minimizing the computational cost while including diffuse and polarization functions.

We constructed three models for the binding pocket based on the reported positions of the heavy atoms of the $\alpha_4\beta_2$ nAChR. [93] Amino acids were selected from the crystal structure in the human $\alpha_4\beta_2$ nAChR (PDB 5KXI). [393] This structure was chosen over the more recently reported structure by Walsh et.al. [377] due to the potential effects that antibodies could induce on the local geometry in the latter study. The model pockets were built by selecting amino acids deemed to be important for the interaction with the nicotinoids. The C-terminus was capped with a terminal amide whereas the N-terminus was capped with a neutral amine to model the interactions of each functional group while minimizing the model size and computational cost. The hydrogen atoms were subsequently added using chemical intuition, meaning that they were added to simply satisfy the valency of each heavy atom. There is only one case where a protonation site can occur in a residue, namely when considering the sole acidic/basic amino acid (Asp96). For this case, based on the pKa of the residue and the many hydrogen bond donors and few hydrogen bond acceptors, we can safely assume that Asp96 exists in the carboxylate form, and this was the one used in our calculations. As regards the orientation of the hydrogen atoms in the alcohol groups, they were originally directed towards hydrogen bond acceptors, and the ensuing optimization steps confirmed this choice for their placement. The nitrogen and carbon atoms in the backbone of each model pocket were held rigid to maintain the structure of the ligand-bound receptor and ion channel reported in the desensitized state. [93] Side chains and hydrogen atoms were allowed to relax. These model pockets were constructed to capture specific interactions between the nicotinoid and the pocket (Figure 8.2). The smallest model (3AA) consisted of the three closest amino acids and captured the direct molecular orbital and hydrogen bonding interactions (Trp156, Thr157, and Trpβ57). The next model (6AA) added the Asp96 residue (as a carboxylate ion) and two hydrogen bonding amino acids (Ser155 and Tyr158) for a total of 6 amino acids to model the electrostatics of the binding pocket and to increase nucleophilicity. Finally, to capture steric effects, the largest model (14AA) added 8 more free amino acid residues (Tyr100, Tyr197, Cys199, Cys200, Tyr204, Serβ108, Leuβ121, and Proβ123) whose side chains constrain the nicotinoids in the pocket (total of 14 amino acids). The total number of basis functions used in the geometry optimizations of the model pockets with the 6-31++G(d,p) basis set were 901 (3AA), 1540 (6AA), and 3466 (14AA), respectively. The total number of basis functions increases by up to 325 upon inclusion of a nicotinoid to the respective model pockets. The largest calculations involved geometry optimizations for the COT-14AA combined nicotinoid-pocket

system with 3791 basis functions with 735 unconstrained nuclear degrees of freedom.

The addictive thresholds previously reported by ICSS studies in a rat model [367] are statistical measurements that were translated to single values for the analysis of this chapter. Two important values are chosen from the experiments, namely (i) the ICSS threshold-decreasing dosages, where the ICSS threshold falls statistically below the baseline, and (ii) an ICSS threshold-increasing dosage where the ICSS threshold jumps statistically above the baseline. The ICSS threshold-decreasing dosage is selected to be 0.125, 0.5, and 1.0 mg/kg for NIC, NOR, and ANB. The ICSS threshold-increasing dosage is selected to be 1.0, 6.0, 3.0, and 15.0 mg/kg for NIC, NOR, ANB, and MYO respectively. These values were chosen as the lowest dosage, which results in p values less than 0.05 and the ICSS threshold is below or above the baseline. Both metrics are important to describe the potential addictive character of nicotinoids, since the ICSS decrease is associated with the abuse potential, whereas the ICSS increase is associated with withdrawal. [394] We focused on ICSS increasing doses in this chapter for simplicity. ANT and COT were not assigned addictive values because no measured dosages elicited ICSS thresholds statistically different from the baseline. It should be noted that COT is considered completely non-addictive (threshold dosage >100 mg/kg), whereas ANT may still be addictive outside the range of the dosages tested. [367]

The binding energies reported in this chapter were considered in the gas phase because the constructed smaller pockets do not include amino acid residues that could shield the bound ligand from solvent interactions. The binding energies of the nicotinoids to the respective pocket were computed and analyzed following the scheme shown in figure 8.3. Specifically, the total binding energy of nicotinoid X in pocket p_i, $\Delta E(X,\, p_i)$, is

$$\Delta E(X, p_i) = E_X^{p_i} - [E_0^{p_i} + E_X] \tag{8.1}$$

where subscripts "0" and "X" indicate an empty and filled (with nicotinoid X) pocket, respectively, $E_X^{p_i}$ is the energy of the optimized pocket with the nicotinoid X=NIC, NOR, ANB, ANT, MYO, COT, $E_0^{p_i}$ is the energy of the constrained and optimized (empty) pocket p_i=3AA, 6AA, 14AA, and E_X is the energy of the most stable gas phase protomer of nicotinoid. All relative energies are reported with respect to NIC.

Following the scheme shown in figure 8.3, the process of inserting nicotinoid X in the pocket can be thought of proceeding via the following steps: (i) nicotinoid X is promoted from the most stable

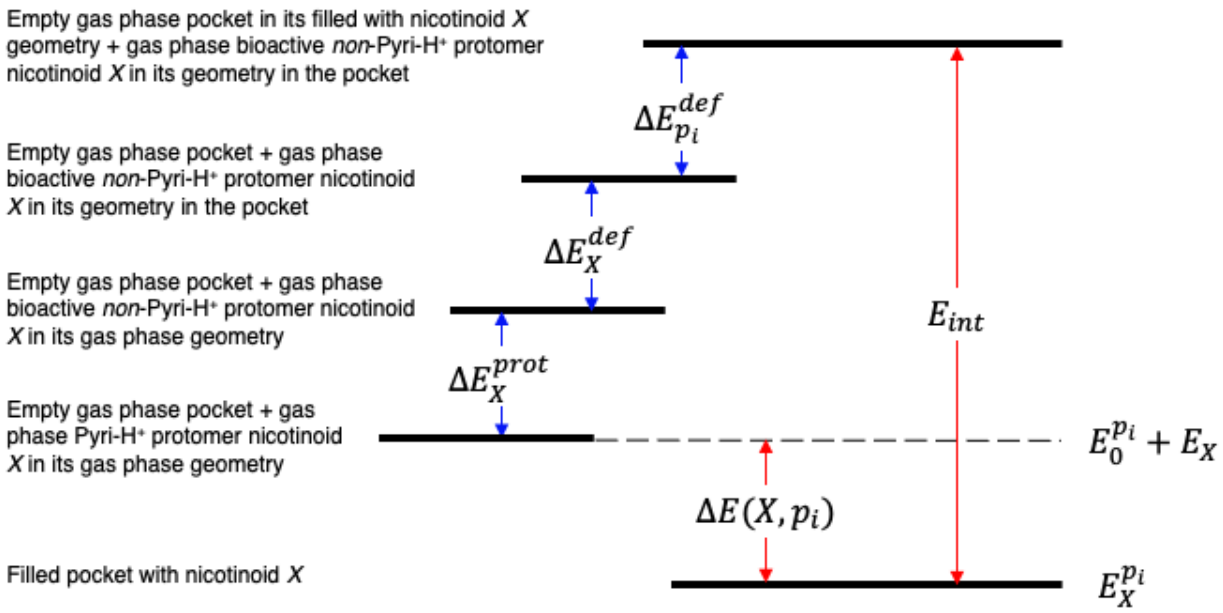

Figure 8.3: Definition of the various components of the binding energy, eqs. (8.1) and (8.2). Blue/red vertical lines indicate positive/negative energies. The dashed line indicates the reference of zero energy.

gas phase structure (Pyri-H$^+$ except for MYO for which the *non*-Pyri-H$^+$ protomer is lowest in energy) to its bioactive *non*-Pyri-H$^+$ protomer with an energy penalty (positive) corresponding to the the difference in the protonation energies between the two protomers, $\Delta E_X^{\rm prot}$, (ii) the bioactive *non*-Pyri-H$^+$ protomer of nicotinoid X is deformed from the gas phase geometry to assume its geometry in the pocket with an energy penalty (positive) of $\Delta E_X^{\rm def}$, (iii) the pocket is deformed from its gas phase geometry to assume its geometry when accommodating nicotinoid X with an energy penalty (positive) of $\Delta E_{p_i}^{\rm def}$, (iv) the deformed bioactive *non*-Pyri-H$^+$ protomer (Pyri-H$^+$ for MYO) of nicotinoid X interacts with the deformed pocket with an interaction energy (negative) $E^{\rm int}$. Therefore the total binding energy (negative) $\Delta E(X,\, p_i)$ of eq. (8.1) can be further decomposed into the following terms

$$\Delta E(X, p_i) = E^{\rm int} + \Delta E_X^{\rm prot} + \Delta E_X^{\rm def} + \Delta E_{p_{\rm i}}^{\rm def} \tag{8.2}$$

Additionally, we relied on the many-body expansion (MBE) [97, 395, 396] to identify important nicotinoid-peptide interactions and illustrate how such interactions change between nicotinoids. In

the MBE, an observable (in this case the binding energy) is cast as the sum of many-body terms truncated to the third-order, viz.

$$E_B = \sum_i E_{1\mathrm{B}_i} + \sum_{i,j} E_{2\mathrm{B}_{i,j}} + \sum_{i,j,k} E_{3\mathrm{B}_{i,j,k}} \tag{8.3}$$

where $E_{1\mathrm{B}_i}$ is the deformation energy of fragment i, $E_{2\mathrm{B}_{i,j}}$ is the pairwise interaction between fragments i and j, and $E_{3\mathrm{B}_{i,j,k}}$ is the non-pairwise additive interaction that emerges from the cooperative interaction between the three fragments i, j, and k. In equation (8.2), $\Delta E_X^{\mathrm{prot}}$, $\Delta E_X^{\mathrm{def}}$, $\Delta E_{p_\mathrm{i}}^{\mathrm{def}}$, are parts of $\sum_i E_{1\mathrm{B}_i}$, and E^{int} is decomposed in $E_{2\mathrm{B}_{i,j}}$ and $E_{3\mathrm{B}_{i,j,k}}$. The terms in the MBE were compared across the different nicotinoids to identify the source of differences in the total binding.

The individual 2B components of the interaction energy were computed from ab initio calculations and compared with the ones obtained classically via

$$E_{\mathrm{class}} = \sum_{i=1}^{N}\sum_{j=1}^{M} k\frac{q_i q_j}{r_{i,j}} + 4\epsilon_{i,j}\left(\left(\frac{\sigma_{i,j}}{r_{i,j}}\right)^{12} - \left(\frac{\sigma_{i,j}}{r_{i,j}}\right)^{6}\right) \tag{8.4}$$

where i and j index the atoms on each fragment, k is Coulomb's constant, q_i is the charge on atom i, $r_{i,j}$ is the distance between atoms i and j, and $\epsilon_{i,j}$ and $\sigma_{i,j}$ are parameters to describe the dispersion and Pauli-repulsion through a Lennard-Jones type interaction. The charges q_i were determined from Natural Bond Order (NBO) populations. [397] A single value of $\epsilon_{i,j}$ and $\sigma_{i,j}$ was used for each element for simplicity, derived from the most common values in the OPLS-AA force-field. [40] Hydrogens bound to electronegative atoms were assigned an ϵ value of zero, while hydrogens in the α position relative to the positive non-pyrrolidine nitrogen were assigned values of $\sigma = 0.500$, as in the OPLS3a force field. [41]

8.3 Results and discussion

8.3.1 Binding of Nicotinoids to the Pocket

NIC binds to the nAChR model binding pockets similar to that reported in the crystal structures, [93, 377] as indicated in figure 8.2 (e)-(f). The binding of the alkaloid to the pocket is mainly facilitated by a strong cationic hydrogen bonding interaction between the protic nitrogen and the carbonyl on residue Trp156, which is in agreement with NIC-nAChR interactions established in the

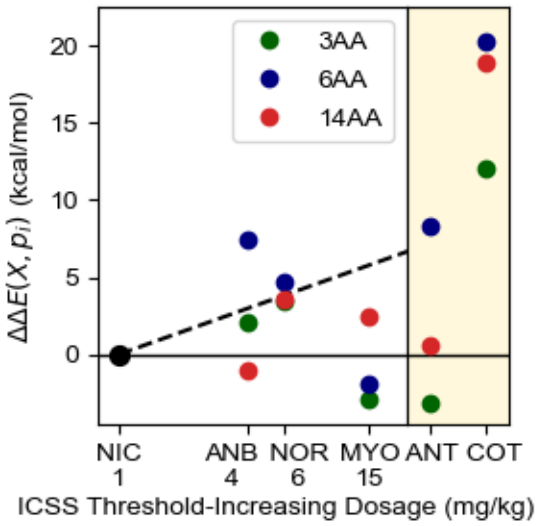

Figure 8.4: The binding energy (kcal/mol, relative to NIC) shown against the ICSS threshold-increasing dosages for the nicotinoids. In general, binding becomes less favorable for less addictive molecules. Nicotinoids in the yellow part of the figure are separated due to the lack of ICSS threshold-decreasing doses.

crystal structure. Similar binding motifs are found for the other nicotinoids. For NOR, ANB, and ANT, the secondary protonated amine was forced to bind through the *cis* hydrogen in a similar manner to NIC. Although binding through the *trans* hydrogen was lower in energy for the 3AA and 6AA model pockets, this structure was not found in the larger 14AA model pocket. Therefore, only the *cis* binding is considered. COT binds similarly to NIC, while MYO deforms both itself and the pocket in order to bind, an effect that is caused by the surrounding steric residues in the larger model pockets. The ring closed form of MYO was found to bind more strongly to the pocket than the ring open form. Therefore, the closed form will be used throughout the rest of this work. Additionally, the more favorable protonation isomers of COT were found to have large steric interactions with the neighboring amino acid residues in the pocket, particularly with Leuβ121 and Cys200. The binding of these other protomers of COT will come at the expense of the receptor's structure and therefore, its ability to function. For this reason, only the *non*-Pyri-H$^+$ isomer of COT will be considered.

The relative–to–NIC binding energies of the 5 other naturally occurring nicotinoids to the

Table 8.1: The binding energies and the various components of equation 8.2 (kcal/mol) for the different nicotinoids interacting with the three model pockets. The BSSE corrected binding energies (in parentheses) show an almost uniform correction that does not alter the relative interactions of the nicotinoids with the pocket. The percentage of the counterpoise correction to ΔE (% CP) is also listed.

Pocket (p_i)	Nicotinoid (X)	ΔE (ΔE_{CP})	% CP	$\Delta\Delta E$ ($\Delta\Delta E_{CP}$)	ΔE_X^{prot}	ΔE_X^{def}	$\Delta E_{p_i}^{def}$	ΔE_{int}
3AA	NIC	-36.67 (-34.27)	7.01	0.00 (0.00)	0.56	1.20	18.67	-57.10
	NOR	-33.20 (-30.83)	7.68	3.47 (3.44)	5.28	0.96	18.68	-58.11
	ANB	-34.56 (-32.27)	7.10	2.11 (2.00)	2.21	0.78	18.72	-56.26
	ANT	-39.80 (-37.31)	6.68	-3.12 (-3.04)	3.20	1.07	19.14	-59.91
	MYO	-39.60 (-37.02)	7.26	-3.03 (-2.75)	n/a	1.29	15.34	-59.53
	COT	-24.68 (-21.93)	12.57	11.99 (12.34)	21.10	2.21	15.61	-63.60
6AA	NIC	-96.84 (-94.08)	2.93	0.00 (0.00)	0.56	3.31	21.17	-121.88
	NOR	-92.16 (-89.56)	2.90	4.68 (4.51)	5.28	2.94	21.18	-121.55
	ANB	-89.42 (-86.79)	3.03	7.42 (7.29)	2.21	3.71	21.51	-116.85
	ANT	-88.49 (-85.80)	3.14	8.35 (8.28)	3.20	2.22	19.92	-121.02
	MYO	-98.78 (-95.90)	3.11	-2.05 (-1.82)	n/a	5.45	21.29	-118.33
	COT	-76.56 (-73.25)	4.52	20.27 (20.82)	21.10	8.81	20.68	-127.16
14AA	NIC	-113.09 (-108.05)	4.66	0.00 (0.00)	0.56	2.91	31.65	-148.22
	NOR	-109.47 (-104.71)	4.55	3.63 (3.35)	5.28	1.86	35.82	-152.42
	ANB	-114.10 (-109.50)	4.20	-1.01 (-1.44)	2.21	3.50	31.83	-151.64
	ANT	-112.48 (-107.77)	4.36	0.62 (0.28)	3.20	4.03	32.09	-146.82
	MYO	-110.71 (-105.88)	4.56	2.39 (2.18)	n/a	2.84	32.12	-150.54
	COT	-94.24 (-88.93)	5.98	18.85 (19.13)	21.10	3.73	30.00	-149.07

binding pockets were compared against their addictive thresholds in figure 8.4. These binding energies, listed in table 8.1, suggest that in general the less addictive nicotinoids bind less strongly with a higher (less negative) binding energy. The trend is strong and predictive for NIC, ANB, and NOR. In the 3AA model pocket, the correspondence between the binding energy and the ICSS threshold-increasing dosage is strong and predictive. BSSE corrections for the 3AA, 6AA, 14AA pockets, respectively, listed in parentheses in table 8.1, were found to lower the absolute binding energies almost uniformly by 2.45, 2.83, and 4.87 kcal/mol respectively, with a standard deviation of 0.18, 0.27, and 0.28 kcal/mol respectively. However, they do not alter the relative order of the interaction between the nicotinoids and the pocket. The similar BSSE corrections seen for these binding energies mean that the correction has little effect on differences in binding, which is the predicted descriptor for addiction. This is probably a special case for the molecules and pockets in this study, as the nicotinoids are of similar size and bind similarly. It may not be the case for all nicotinoids, especially those that are very different in size or different in binding structure. In the 6AA model pocket the model incorrectly orders ANB and NOR. The 14AA model pocket predicts a more favorable binding for ANB than NIC, which is inconsistent with their ICSS threshold-increasing dosages. Further investigations (*vide infra*) explain and rectify the ordering seen in the larger model pockets.

Not all nicotinoids clearly follow the trend produced by NIC, ANB NOR, and COT. From figure 8.4 it is seen that MYO binds more strongly than NIC in the smaller model pockets, despite being less addictive. Binding is a complex, multi-step process, involving dehydration, deformation, and interaction, while ion channel gating is an even more complicated process. To better understand the sources of the differences in the binding energies, and identify potentially useful descriptors of addictive thresholds, in the following, we examine the individual contributions of the various components of equation (8.2) to the total binding.

8.3.2 Contribution of Monomer Deformation to the Binding Energy

The *non*-pyridine protomer (*non*-Pyri-H$^+$) is responsible for binding to the active nAChR. In the gas phase, this protomer is formed through proton transfer (ΔE_X^{prot}), except for MYO for which the *non*-Pyri-H$^+$ protomer is the gas phase global minimum. These proton exchange energies are

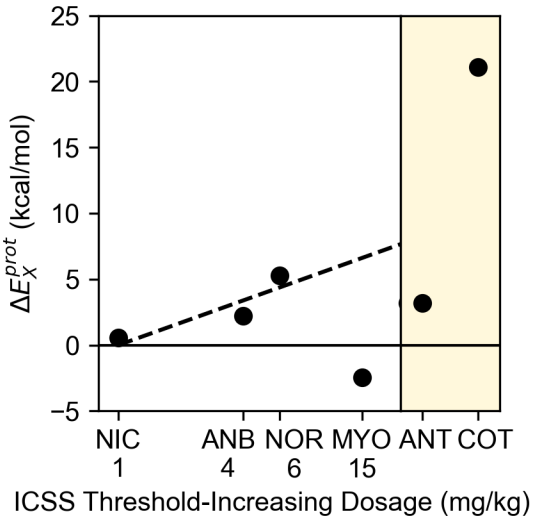

Figure 8.5: The proton exchange energy (energy difference between the Pyri-H^+ and the *non*-Pyri-H^+ protomers) correlates with the ICSS threshold-increasing dosage of nicotinoids for the strongly addictive molecules.

plotted in figure 8.5. The same addictive trend previously observed for the binding energy (Figure 8.4) is also observed for the monomer deformation piece for NIC, NOR, and ANB, suggesting that the majority of the predictive trend in the proposed model may actually come from the proton exchange energies.

It should be noted that in a real biological environment nicotinoids bind to the pocket in the human brain by moving from the aqueous state to the pocket (i.e., not from the gas phase to the pocket). The gas phase reference was chosen for this work for its simplicity, but a quick comparison to implicit solvent models was made to justify this choice. After all, the nicotinoids need to be desolvated in order to enter the pocket in the human brain as there is no water in the bound nicotinoid in the pocket. The desolvation energy of each nicotinoid was computed with PCM. Although desolvation is not linearly related to the proton exchange energy, the relationship is logarithmic and one-to-one. This shouldn't be surprising, as proton exchange energy and hydrogen bond strength are related to proton affinity. [398] Therefore, the same qualitative trend between binding energy and addiction can be produced with the reference being either the gas-phase Pyri-H^+

or the aqueous state *non*-Pyri-H^+ protomers.

After assuming the local minima corresponding to *non*-Pyrro-H^+ protomers, the nicotinoids bind to the pocket. This process is naturally broken into two parts, deformation (for both the nicotinoid and the pocket) and interaction (between the deformed nicotinoid and the deformed pocket, see figure 8.3), both of which influence the final geometries. The deformation energy (positive) is the energetic cost of moving from a local minimum to some deformed geometry of higher energy that the nicotinoid assumes in the binding pocket due to the interaction with its immediate environment. Although this energy contribution is important and sizable, there is no clear correlation between the nicotinoid's addictiveness and their deformation energies in the pockets. Note that the deformation of the pockets upon binding of a nicotinoid in order to maximize the interactions is minimal. These deformations are important, however, as small deformations in the protein can have a significant impact on the ability of the ion channel to open. The largest structural differences are observed when MYO is bound to the pocket.

It was found that MYO had some of the largest deformations as it entered the binding pocket, that were attributed mostly to the change in the dihedral angle between the two rings, χ, defined in prior works. [379] For all the nicotinoids besides MYO, the minimum energy structures have χ angles near -7°. The conjugation in MYO naturally prefers a flattened structure with χ being -79°. The preferred angle χ for most nicotinoids inside the pocket exists between 0 and -30 degrees. However, MYO must either deform strongly or relax in the pocket in some other way. It is found that MYO prefers the latter in the smaller pockets with deviations in the intermolecular degrees of freedom. This is of serious concern since the mechanism of binding can have a significant impact on the activation of the receptor pore, which constitutes an important aspect of receptor activation and the overall addiction mechanism.

8.3.3 Interactions with Amino Acids

The last but arguably more complicated aspect of binding is the interaction of the ligand and the receptor (E^{int}). This interaction can be viewed through fragment many-body energies, a direct consequence of the respective intermolecular geometry. All of these will be utilized in this section to understand more fundamentally how nicotinoids interact with the binding pocket of nAChRs

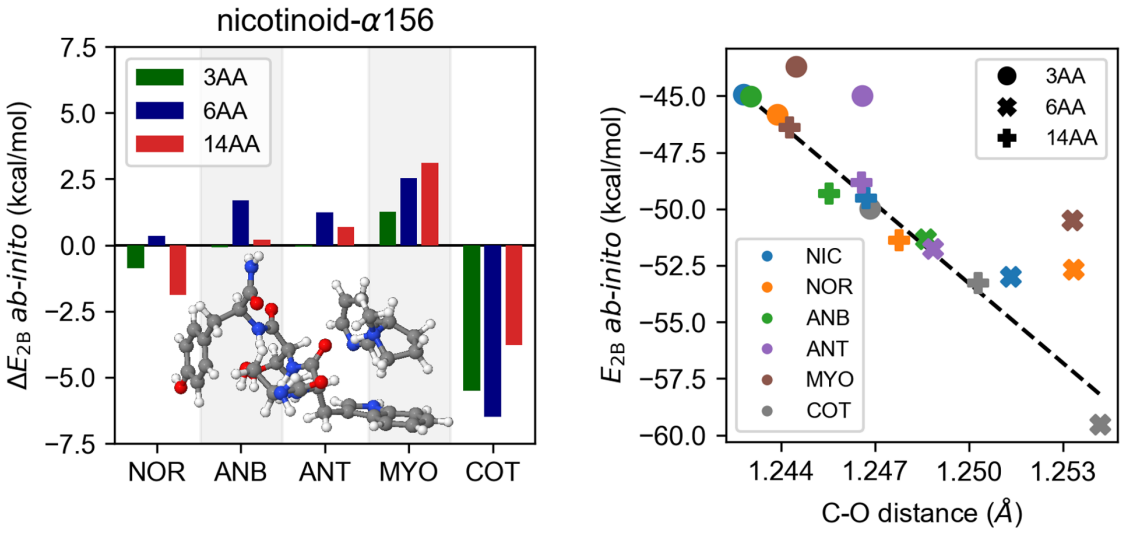

Figure 8.6: Left panel: the two body energy for the nicotinoid-α156 chain relative to that of NIC is plotted for the various nicotinoids and model pockets. Right panel: the attractive (negative) two body energies are plotted against a metric of the length of the carbonyl in the Trp156 residue, which constitutes a metric of hydrogen bond strength.

as well as illuminate how different aspects of this binding support the addictive behavior of these molecules.

A beneficial consequence of the approach of fragmented model pockets used in this study is that the MBE naturally allows for the ability to separate distinct nicotinoid-peptide interactions. Arguably, one of the most important interactions included in all model pockets is the interaction between the nicotinoid and Trp156; since the Trp156 amino acid is modeled as a polypeptide chain in our model pockets, this fragment will be called α156. This is a strong interaction (between -43.69 and -59.51 kcal/mol) coming from a charged hydrogen bond, cation-pi interactions, and additional van-der-Waals interactions. Subtle differences in this pairwise interaction energy penalize nicotinoids ANB, ANT, and MYO, shown by the positive bars in figure 8.6. This interaction is dominated by the cationic hydrogen bond between the nicotinoid and Trp156, as indicated by the correlation between the α156-nicotinoid two body energy and the elongation of the C=O carbonyl in Trp156, the hydrogen bond acceptor (Figure 8.6). Deviations from this correlation occur naturally

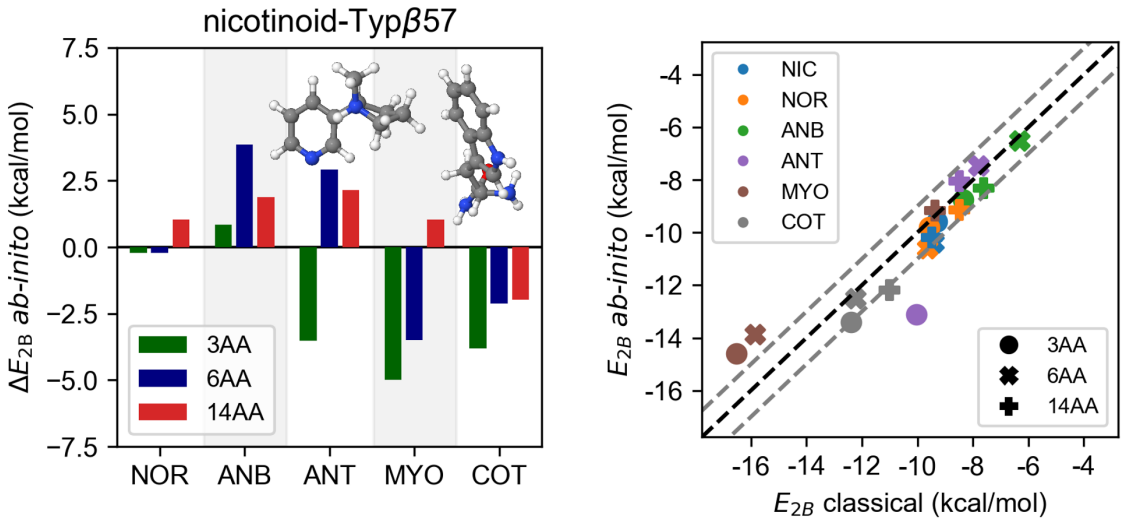

Figure 8.7: Left panel: the two body energies for the nicotinoid-Trpβ57 residue relative to that of NIC is plotted for the various nicotinoids. Right panel: Correlation between the ab initio and classical energies, indicating that these interactions are largely classical. Gray lines indicate an error of 1 kcal/mol.

through non-hydrogen bonding interactions. Unsurprisingly, COT forms the strongest hydrogen bond, as it is the most acidic. The hydrogen bonds are stronger in the 6AA model than the 3AA model, a direct consequence of the increased nucleophilicity of α156 chained and a tighter bound geometry to the Asp96 residue. This increased nucleophilicity manifests not only in the Trp156 residue, but all residues on this polypeptide, including Ser155.

Another conserved interaction is that between the nicotinoid and the Trpβ57 residue (Figure 8.7). It is strongly favorable in ANT, MYO and COT for the 3AA model, likely resulting from slight conformational changes decreasing H-π distances due to the missing steric hindrance. As additional amino acid residues are added in the larger pockets, the interaction becomes continually less favorable, and consistently less favorable than NIC for NOR, ANB, ANT, and MYO. The RMSD of the Trpβ57 residue shows that the closer geometry comes mostly from changes in the position of the residue.

As discussed earlier, the additional amino acid residues in the 6AA model change the way nicotiniods interact with the α156 residue and the β57 residue. These interactions do not, however, have

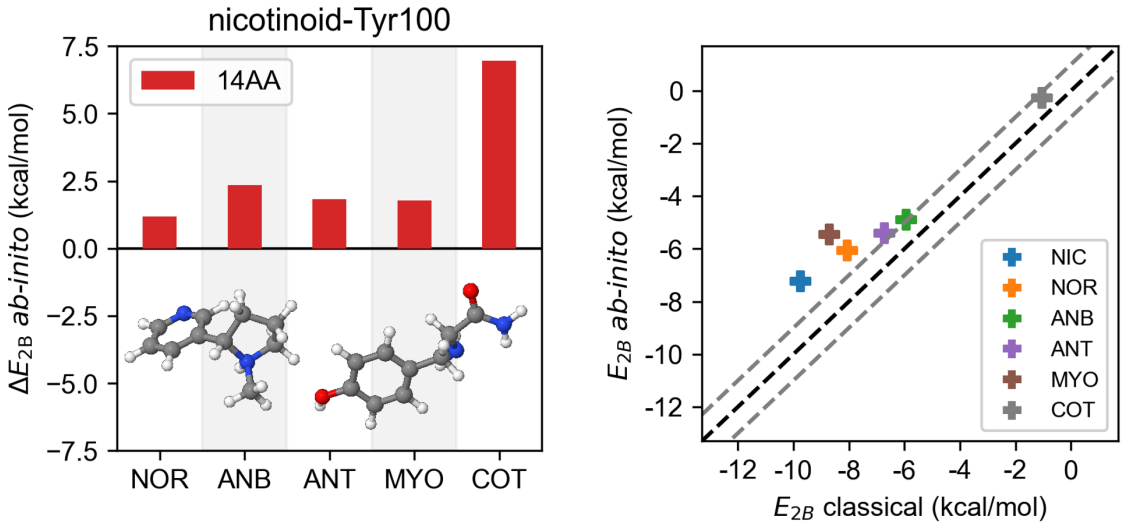

Figure 8.8: Left panel: the two body energy for the nicotinoid-Tyr100 residue relative to that of NIC shown for the various nicotinoids. Right panel: Correlation between the ab initio and classical energies, indicating that these interactions are largely classical. Gray lines indicate an error of 1 kcal/mol.

marked differences between nicotinoids. It becomes more important to discuss how the increasing nucleophilicity changes the structure of the nicotinoids in the pocket. The increased nucleophilicity of the pocket results in the change (bending) of the orientation of the nicotinoids, specifically moving the nicotinoid closer to the C-terminus of the α156 chain. This is a manifestation of the Trp156's increased nucleophilicity making the hydrogen bond to the carbonyl closer to 120° and stronger. This bending also adds a slightly favorable C-H hydrogen bond with the Ser155 residue and affects the ANB and ANT less favorably due to the larger 6-membered rings having larger steric interactions with neighboring residues. Steric interactions in the 14AA model return most orientations to those observed in the 3AA model. Therefore, the structures and subsequent energies computed in the 6AA model maybe an artifact of the geometries assumed in the smaller pockets compared to that of the full protein.

Amino acids in the 14AA pocket similarly change the way nicotinoids interact with the α156 chain as discussed earlier. Some have large contributions to the relative binding energy and will be discussed here. Nicotinoids form weak CH$\cdots$O interactions between the N-α carbon atoms and the

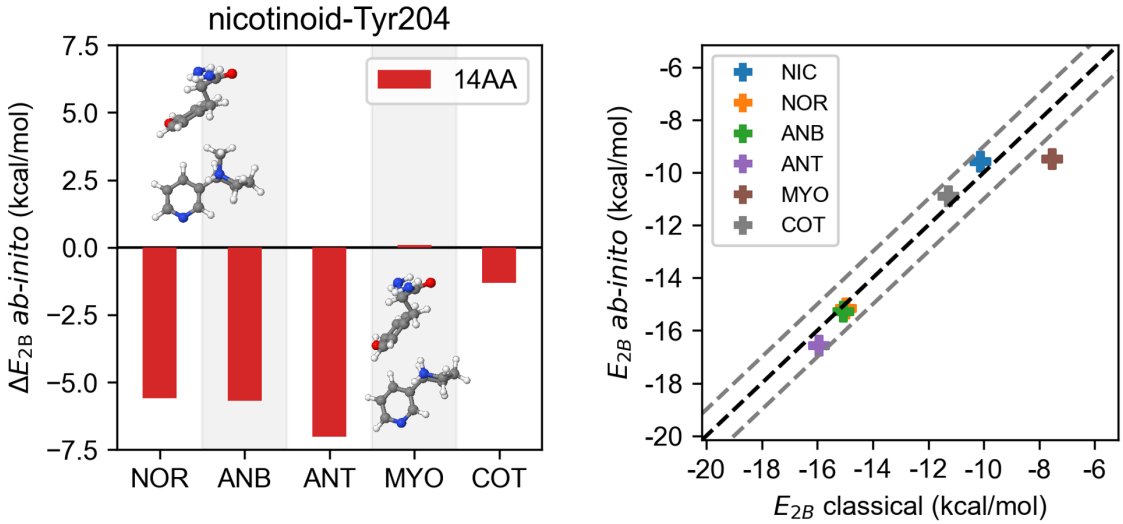

Figure 8.9: Left panel: the two body energy for the nicotinoid-Tyr204 residue relative to that of NIC shown for the various nicotinoids. Right panel: Correlation between the ab initio and classical energies, indicating that these interactions are largely classical. Gray lines indicate an error of 1 kcal/mol

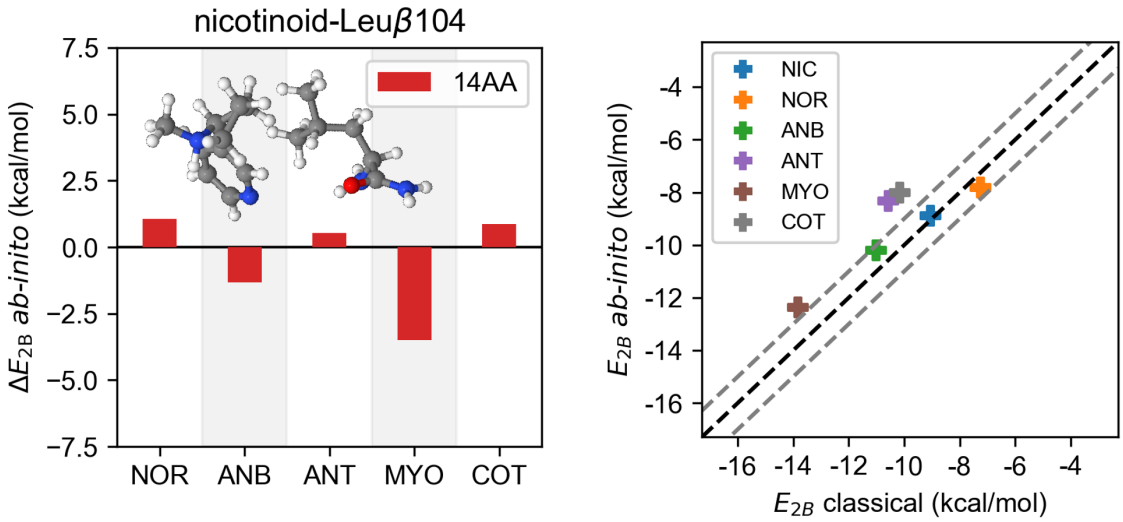

Figure 8.10: Left panel: the two body energy for the nicotinoid-Tyr204 residue relative to that of NIC shown for the various nicotinoids. Right panel: Correlation between the ab initio and classical energies, indicating that these interactions are largely classical. Gray lines indicate an error of 1 kcal/mol

hydroxide of the Tyr100 residue, see figure 8.8. This interaction is maximized in NIC by two N-α hydrogen atoms. In contrast, NOR, ANB, ANT, and MYO have only one N-α hydrogen atom so this interaction is consistently less favorable. On the other hand, the interactions of COT are distant and weak due to the additional carbonyl. Another residue, Tyr204, interacts with protonated secondary amines in favorable ways that strongly impact the relative binding energy (Figure 8.9). In NIC and COT, this residue accepts a weak CH$\cdots\pi$ bond between the methyl group and the phenol group. In NOR, ANB, and ANT, the methyl group is replaced by hydrogen, which forms a second protonated hydrogen bond. Leuβ121 has stronger interactions with MYO than with other nicotinoids (Figure 8.10). We attribute this result to the higher multipole moment on the pyridine ring of MYO due to larger partial positive and negative charges on the atoms, which comes from the conjugation of the pyridine ring to the protonated nitrogen on the five-membered ring. This increased multipole moment also appears as an increased electrostatic interaction when modeled via pairwise point charges. Leuβ121 has comparable dimer energies with the other nicotinoids and the remaining amino acid residues interact similarly with the set of nicotinoids.

The relative-to-NIC 2B energies are shown in the left panel of figure 8.11, whereas the relative-to-NIC 3B energies are plotted in the right panel of that figure. In general, two-body energies ($E_{2\mathrm{B}}$) between peptides are small and near equivalent regardless of the identity of the nicotinoid in the pocket. In addition, differences in peptide-peptide 2B energies in the presence of different nicotinoids are small, as shown by the visible blue sections centered around zero. The average value for these relative-to-NIC peptide-peptide 2B energies is 0.0094 kcal/mol with a standard deviation of 0.82 kcal/mol. A few terms are large in magnitude and will be discussed later.

Exceptions arise only for NOR and MYO, specifically for dimers Serβ108-α156, Serβ108-Leuβ104, Tyr197-Tyr204, Tyr204-α156, Tyr204-Asp96, Tyr204-α156, and Tyr204-Cys199/Cys200. Most 2B interactions pair up and cancel, such as the Serβ108-α156 and Serβ108-Leuβ121. The energetic differences can be attributed to a rotation of the hydroxyl group trading dipole-induced dipole interactions with the α156 chain for distant dipole-dipole interaction with the Leuβ121 carbonyl. A similar hydroxyl rotation pairs the Tyr204-Tyr197 and Tyr204-Asp96 residues. Interactions do not fully cancel for Tyr204-α156 and Tyr204-Cys199/Cys200 when MYO is bound to the pocket. Significant conformational changes in the Tyr204 residue increase favorable interactions with α156, while breaking favorable interactions with the C-Loop domain. This is of serious concern in the

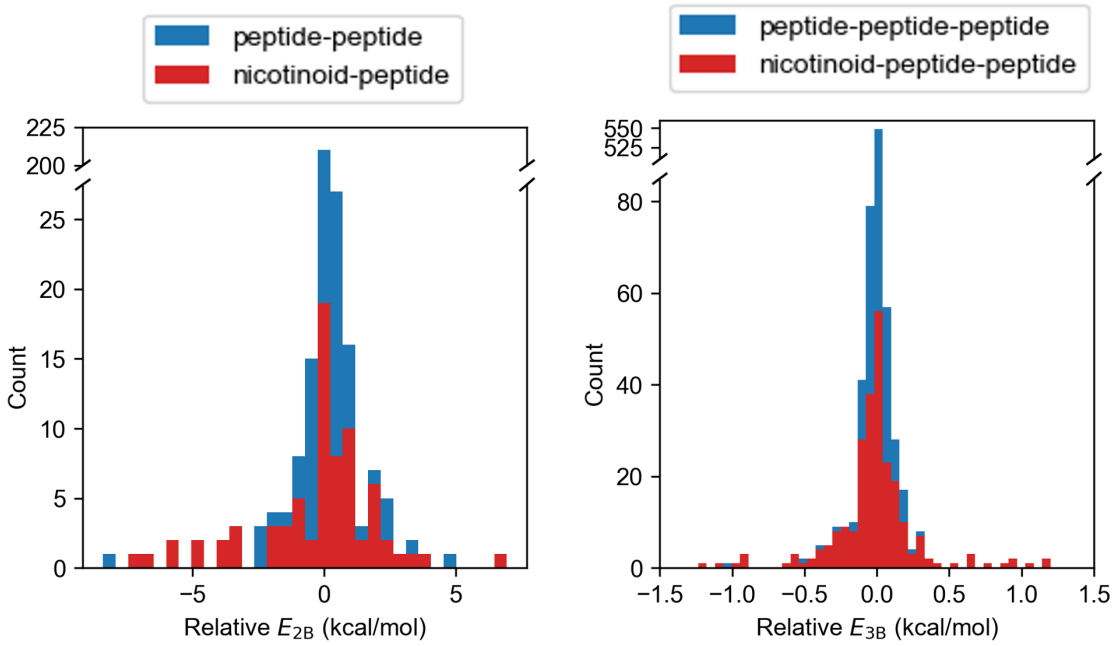

Figure 8.11: All two- and three-body terms, relative to those of nicotine, shown as histograms on the left and right panels, respectively. Red colors denote nicotinoid-peptide, whereas blue colors denote those between peptides and are stacked above the red ones. Note the breaks near the top of the y-axis.

closing of the pocket by the C-Loop.

The interactions of nicotinoids with Tyr204 are of interest because they were shown to interact differently with neighboring residues depending on the identity of the nicotinoid added to the pocket. This residue forms strong hydrogen bonding interactions with NOR, ANB, and ANT and the orientational changes alter the interactions of this residue with the cystines of the C-Loop domain. This interaction region is important for the activation process. Additionally, Tyr204 is suspected to interact with a water molecule that is present in the acetylcholine-binding protein (AChBP). The water molecule is not included in the pocket models in this chapter, which is intended to model the interaction of nicotinoids with the acetylcholine receptor (nAChR), where there is no water present. Although numerous past studies found a water molecule forming hydrogen bonds between Tyr204, Serβ108, and NIC in AChBPs and other variants of nAChRs, [373, 374, 376] the crystal structure by Morales et al. [93] and Walsh et al. [377] did not detect a water molecule

at this site. For this reason water was not added to the pocket. The changes made by NOR and MYO to the Tyr204 residue would disrupt the interactions with water. Further exploration, either computational or experimental, is warranted to confirm the absence or presence of water in the $\alpha_4\beta_2$ variant of the nAChR.

Three-body energies (E_{3B}) are consistently small and near equivalent. The relative-to-NIC 3B energies are plotted in the right panel of figure 8.11. None of these are large in magnitude and most are near zero; their average value of relative-to-NIC 3B energies is -0.0017 kcal/mol with a standard deviation of 0.17 kcal/mol. Even the absolute 3B energy terms are negligible and centered around zero; their average value is 0.011 kcal/mol with a standard deviation of 1.22 kcal/mol. The only large 3B terms belong to the nicotinoid-α156-Asp96 trimer. The corresponding dimers also contribute to large 2B terms.

8.3.4 Piecemeal Steps of Binding Provide a Predictive Model

The analysis of nicotinoid-peptide interactions with select amino acid residues provides a unique opportunity to develop a model for predicting addiction of nicotinoids. Trends in the proton exchange energy support the notion that simplistic, one-body effects can describe the ability for a nicotinoid to elicit changes in ICSS threshold and be addictive. Deviations from this trend are rectified via peptide-nicotinoid interactions, most importantly between the nicotinoid and the amino acid residues in its immediate vicinity. Figure 8.12 shows the trends when considering only a subset of interactions in each model. The upper left panel includes all one-body terms, viz. proton exchange, nicotinoid deformation, and peptide (pocket) deformation. The upper right panel incorporates all 2- and 3-body effects between the nicotinoid, the α156 chain, and the Trpβ57 residue. The lower left panel includes the 2- and 3-body interactions that include the nicotinoid, the α156 chain, the Trpβ57, and the Asp97 residues. Finally, the bottom right panel includes all prior and all 2-body interactions between the additional peptides in the 14AA model except for all Tyr204 interaction and the Leuβ121-nicotinoid interaction. The Tyr204 amino acid is excluded due to the deformations and interactions that have the potential to change the tertiary structure of the C-Loop region. The Leuβ121-nicotinoid was removed due to the strong quadrupole-quadrupole interaction between the Leuβ104 and MYO, which changes the orientation of the pyridine ring in

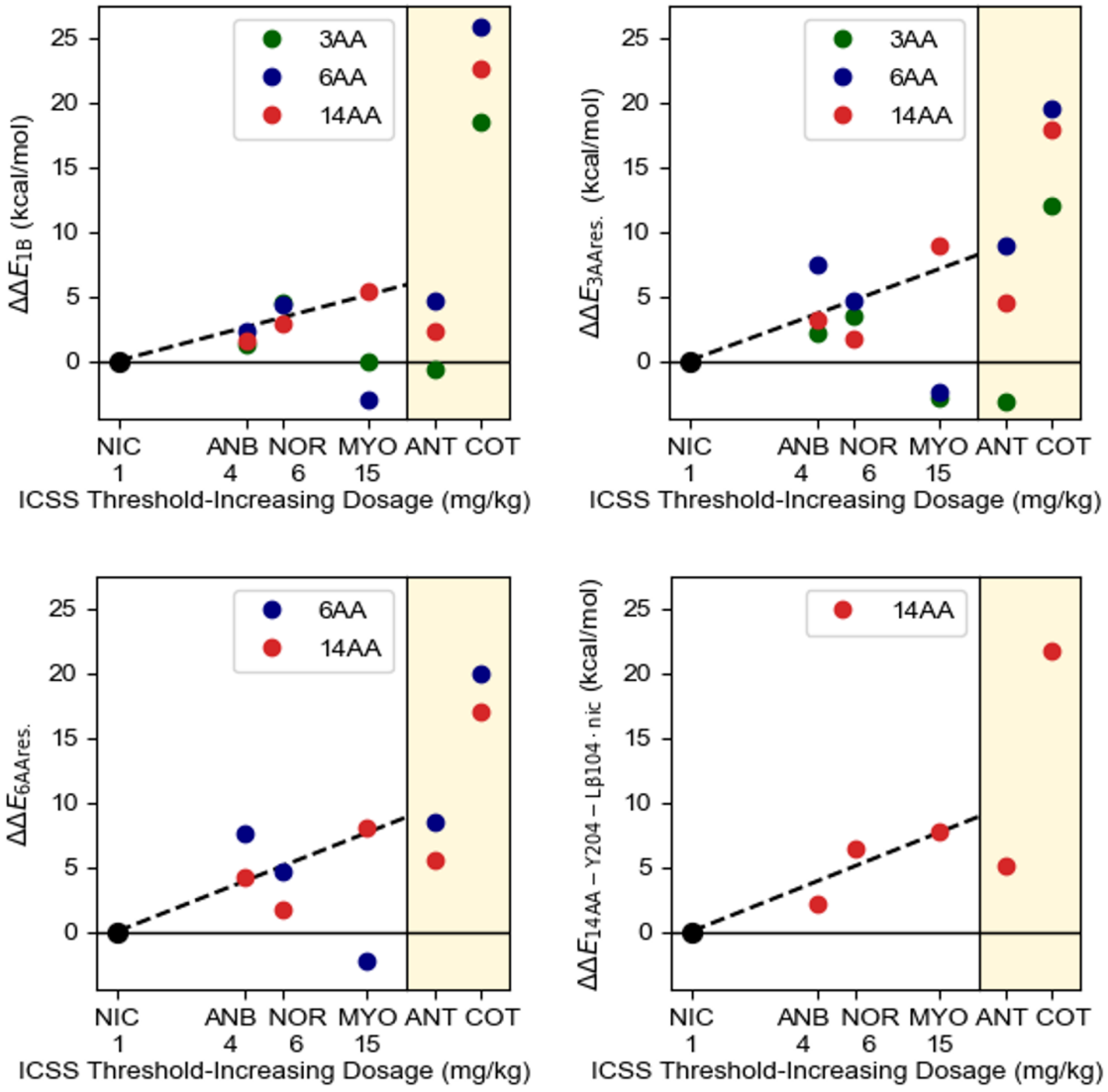

Figure 8.12: The ICSS threshold-increasing dosages are compared against the sum of certain binding terms, viz. one body terms (upper left), terms involving only fragments in the 3AA model (upper right), terms involving fragments in the 6AA model (lower left), and all terms except for those involving Tyr204 and Leuβ121-nicotinoid interactions (lower right).

MYO and may disrupt the speculative interaction with a water molecule (if present).

When taking into account only the nicotinoid and peptide deformations, a reasonable predictive relationship is established for the 14AA model. Steric type interactions, that restrict the ways MYO relaxes into the pocket appear to capture the relative unfavorable binding of MYO to the nAChR. The interactions of nicotinoids with local binding residues further diminish MYO's binding, therefore explaining its low addictive ability. In the simpler models (smaller pockets), NOR's quite relatively favorable binding to the α156 chain results in a lower binding energy than ANB. Incorporating the interactions of the nicotinoids and other amino acid residues prove to be important in predicting the order of ICSS threshold-increasing dosages. Interactions and their differences are small (<5 and <1 kcal/mol respectively), however numerous. The reduced summation of the 14AA model, in which the Tyr204 interactions and the Leuβ121-nicotinoid interaction were eliminated, proves to be sufficient in predicting the ICSS threshold-increasing doses. It serves as a good descriptor for predicting the psychoactive character and resulting addictive potential of nicotinoids. Caution should be taken, however, in ensuring that large changes are not observed with a test nicotinoid that is structurally different from nicotine and may change the structure of the nAChR in ways not observed in this study.

8.4 Conclusions

The binding energy of nicotinoids with the amino acid residues of the $\alpha_4\beta_2$ variant of the nAChR has proven to be an invaluable descriptor in predicting the addictive character of nicotine-like molecules. In general, nicotinoids that bind more strongly to the peptides are more addictive. The source of differences in binding spans intrinsic nicotinoid descriptors, such as proton affinity/proton exchange and to peptide-nicotinoid interactions.

The model pockets of increasing sizes, derived from the crystal structure of the nAChR, accurately describe the relative interactions of the guest nicotinoids to the host nAChR environment. The use of progressively larger model pockets provides insight into the nature of the interactions and differences thereof. Simplistic models were sufficient in describing methylated and de-methylated nicotinoids, whereas additional interactions were essential for understanding oxidized nicotinoids. Differences in the total interaction originate mainly from pairwise, two-body terms. Non-additive (3B) interactions are generally small, with the differences of these terms between nicotinoids be-

ing even smaller. Although many amino acids interact similarly with all nicotinoids, most of the differences arise in the restriction of the configuration space of the nicotinoid and therefore, its interactions with important residues such as Trp156 and Trpβ57. The pairwise interactions between the nicotinoids and distant amino acids (such as those present in the 14AA model) can be accurately described by simpler, lower scaling classical force fields. However, the interactions of nicotinoids with proximal amino acids (such as the α156 chain and Asp96 residues) should remain modeled by quantum mechanical methods. The current results also provide useful information about parallel spectroscopic infrared studies by other members of the group that target the interaction of nicotinoids with pieces of the pocket environment such as tryptophan. [399]

Evidence presented in this work supports future investigations into the possibility of the presence of water in the binding pocket. Although water was not detected in the X-ray structure by Morales and co-workers, [93, 377] there exist both hydrogen bond donors and acceptors spaced appropriately for water to bind in the pocket. The presence of a single water molecule would greatly impact the positioning of NOR and MYO in the pocket and thus affect the interactions of the nicotinoid with the Tyr204, Leuβ121, and Serβ108 residues. Simulations including a single water molecule will serve to modify these interactions and explain the potential role that water can have in the binding mechanism.

Chapter 9

THE GAS PHASE PROTONATION SITES OF SIX NATURALLY OCCURRING NICOTINOIDS [400]

9.1 Introduction

In the last chapter, we reported on a computational study in which we examined the relation between the binding energy of these six nicotinoids to model binding pockets of the $\alpha_4\beta_2$ nAChR and found similar trends in the binding affinity, with a surprisingly strong correlation to the addictive thresholds reported by Harris et al. [367] Through a fragmentation analysis, it was found that the main differences in the respective the binding energies can be attributed to the molecular properties of each nicotinoid, specifically the relative energy of the nicotinoid's *non*-Pyri-H^+ protomer. As a consequence of these past studies, it is believed that the *non*-Pyri-H^+ (Pyrro-H^+) protomer constitutes the bio-active form of NIC, and the availability of the *non*-Pyri-H^+ is essential for the addiction process to occur in the biological environment. Given the fact that NIC has also a higher addictive potential than NOR, it stands to reason to hypothesize that the *non*-Pyri-H^+/Pyri-H^+ ratio is related to its addictive potential. The work presented here aims towards identifying the preferred protonation sites of the nicotinoids in the gas phase and at the physiological temperature as a first step towards understanding their addictive character in the biological environment based on their molecular structure. Naturally, a direct structure/activity correlation will require additional simulations of their interaction with the binding pocket and the full consideration of how the steric effects from the rest of the receptor affect their binding motifs in that environment.

The discussion in chapter 8 suggests that molecular recognition is critical in determining the addictive potential of nicotinoids, with an important descriptor being the availability and stability of the *non*-Pyri-H^+ protomer. It is therefore important to determine the protonation site and subsequent population ratios of the two protomers of the various nicotinoids in different environments. In this chapter, we combined experimental techniques with high-level electronic structure theory as in prior studies for NIC and NOR [378, 380, 401] to determine the distribution of protomers for

the other four alkaloids (ANB, MYO, ANT, COT), which are naturally occurring in tobacco leaves. This will improve the characterization of one of the most important aspects of their binding affinity to the human receptor, namely the proton exchange energy, figure 8.5, and will enable comparisons with the addictive thresholds measured by Harris et al. [367] to further validate an important and simple molecular descriptor for ligand binding and addiction.

9.2 Experimental and Theoretical Methods

The experiments were performed by collaborators using a cryogenic ion trap setup combined with an electrospray ion source, cold ion trap infrared spectroscopy. [402–405] The experimental setup and the procedure are described in detail elsewhere. [380, 384, 406–409]. Briefly, a methanol solution containing a 10^{-5} M alkaloid with 0.5% of formic acidwas electrosprayed (ES) and the protonated alkaloid was introduced into the gas phase. The generated protonated alkaloid (alkaloid-H^+) was mass-selected by a quadruple mass filter (Q-MS) and introduced into a thermalizing trap. In the experiments, the thermalizing trap was maintained at 310 K, which is similar to the human body temperature. Subsequently, a He/H_2O gas mixture was introduced in the thermalizing trap. We have found previously that the introduced H_2O gas catalyzes the intra-molecular proton transfer for NIC, a necessary step for achieving a thermal equilibrium in the gas phase within a timescale of ion trapping (50 ms) [384]. The target molecules were subsequently mass-selected by the second Q-MS and trapped in a quadruple ion trap (QIT) maintained at 4 K. Here the experiment applied the hydrogen tagging method [405] to measure the infrared photo-dissociation (IRPD) spectra. He/H_2 gas is introduced to the QIT to allow H_2 to be attached to ions. A tunable IR laser (OPO/OPA: Laser Vision) was introduce to vibrationally excite the trapped alkaloid-H^+ in the QIT. H_2 was detached from the cluster when the IR laser resonances to the vibrational level of the tagged ions. The generated bare ions were monitored by a time-of-flight mass spectrometer (ToF-MS) so that the IRPD spectra could be obtained by the enhancement of the bare ion.

The geometries of the protonated nicotinoids were optimized at the Møller Plesset second-order perturbation theory (MP2) [144, 146] using the augmented correlation consistent basis sets of double and triple zeta quality (aug-cc-pVnZ, n =D, T) of Dunning and co-workers. [194, 195] Single point MP2/aug-cc-pVQZ energy calculations were performed at the MP2/aug-cc-pVTZ optimized geometries. The energies at the MP2/aug-cc-pv$n$Z (n =D, T, Q) levels of theory were used to

extrapolate the electronic energies to the complete basis set (CBS) limit via the heuristic 4-5 polynomial extrapolation [410, 411]

$$E(n) = E_{\mathrm{CBS}} + \frac{A}{(n+1)^4} + \frac{B}{(n+1)^5} \tag{9.1}$$

where $E(n)$ is the electronic energy with basis set n, E_{CBS} is the complete basis set limit, n is the cardinal number of the basis set ($n = 2$ for aug-cc-pVDZ, $n = 3$ for aug-cc-pVTZ, and $n = 4$ for aug-cc-pVQZ), and A and B are fitted parameters. The MP2 calculations were performed with the NWChem suite of electronic structure methods. [174]

Harmonic vibrations were calculated with the ωB97X-D [163] density functional and the aug-cc-pVDZ basis set [194] using Gaussian 16. [173] Thermal corrections were obtained by treating the vibrational degrees of freedom with harmonic partition functions and the rotational degrees of freedom with rigid rotors. The ratio of protomer populations (r_{calc}), theoretically estimated from thermodynamics, is

$$r_{\mathrm{calc}} = \frac{P_{\mathrm{Pyri-H^+}}}{P_{non-\mathrm{Pyri-H^+}}} = \frac{\sum_{i \in \mathrm{Pyri-H^+}} e^{-\Delta G_i/RT}}{\sum_{j \in non-\mathrm{Pyri-H^+}} e^{\Delta G_j/RT}} \tag{9.2}$$

where P is the population of each protomer obtained assuming a Boltzmann distribution, i and j index the individual isomer contributing to each protomer, ΔG_i is the free energy of isomer i, R is the universal gas constant, and T is temperature.

Fundamental vibrational transitions were calculated using second-order vibrational perturbation theory (VPT2) [269] at the ωB97X-D/aug-cc-pVDZ level of theory [163, 194] using Gaussian 16. [173] Only normal modes with frequencies above 1000 cm^{-1} were allowed to couple in the perturbation. Transitions were scaled by 0.985 to align the strong transition observed around 3390 cm^{-1}. The experimentally obtained ratio of protomer populations (r_{exp}) was estimated by combining the experimentally measured transition intensities with the calculated absorption coefficients through

$$r_{\mathrm{exp}} = \frac{P_{\mathrm{Pyri-H^+}}}{P_{non-\mathrm{Pyri-H^+}}} = \frac{I_{\mathrm{Pyri-H^+}} \times \varepsilon_{non-\mathrm{Pyri-H^+}}}{I_{non-\mathrm{Pyri-H^+}} \times \varepsilon_{\mathrm{Pyri-H^+}}} \tag{9.3}$$

where P is the population of each protomer, I is the integrated intensity from the experimental spectra, and ε is the absorption coefficient of the most energetically stable isomer of each protomer. It should be noted that ε did not vary much for different isomers of each protomer. The uncertainties (%) in the isomer populations determined from experiment are estimated from the

variance in the noise of the measured spectra. They range from 0.7%–2.3% and can be considered as representing the minimum detectable populations for each isomer. The theoretically and experimentally estimated protomer populations were compared against the ICSS thresholds [367] in order to investigate whether they can be used as a metric for relative addiction. These correlations were further quantitatively analyzed by regression analysis to obtain the corresponding coefficients of determination.

9.3 Results and Discussion

The collaborators first measured the IRPD spectra of alkaloids without using the thermalizing trap referred to as "without annealing" in figure 9.1. In these experiments, both the Pyri-H^+ and the non-Pyri-H^+ protomers can be observed because their populations are not annealed. Figure 9.1 shows a series of panels corresponding to the IR spectra of a) NIC-H^+, b) NOR-H^+, c) ANB-H^+, d) ANT-H^+, e) MYO-H^+, and f) COT-H^+. For each nicotinoid, panels (a)–(f) in figure 9.1, there are a series of panels displayed as follows: the top panel depicts the IRPD spectra without annealing, the second one to the spectra with annealing at T=310 K, the next two panels show the theoretically computed transitions using second order perturbation theory (VPT2) for the Pyri-H^+ (red) and *non*-Pyri-H^+ (blue) protomers and finally the bottom colored boxes trace the experimental and theoretically estimated populations of the two protomers for each nicotinoid. The experimentally observed bands are assigned by comparing them with the theoretical spectra.

The assignments for NIC-H^+ and NOR-H^+ are reported in our previous works [380, 384]. Briefly, the bands at 3394 cm^{-1} are assigned to the NH stretching vibrations of Pyri-H^+ in both molecules while other bands at lower frequencies are assigned to the NH stretching vibrations of the *non*-Pyri-H^+ protomers. The vibrational spectra of ANB-H^+ (panel c) and ANT-H^+ (panel d) are similar to that of NOR-H^+, and the bands at 3392 cm^{-1} and 3392 cm^{-1} correspond to the NH stretching vibrations of the Pyri-H^+ protomers of ANB-H^+ and ANT-H^+, respectively. The pairs of bands at lower frequencies are assigned to the NH stretching vibrations of the *non*-Pyri-H^+ protomers. The computed transitions for the lowest energy isomers of ANB, after scaling, are 3424 cm^{-1} with an absorption coefficient of 126.0 km/mol for ANB-Pyri-H^+, and 3276 and 3222 cm^{-1} with absorption coefficients of 32.1 and 32.1 km/mol for ANB-*non*-Pyri-H^+, respectively. For the lowest energy isomers of ANT, the computed transitions after scaling are 3413 cm^{-1} with an absorption

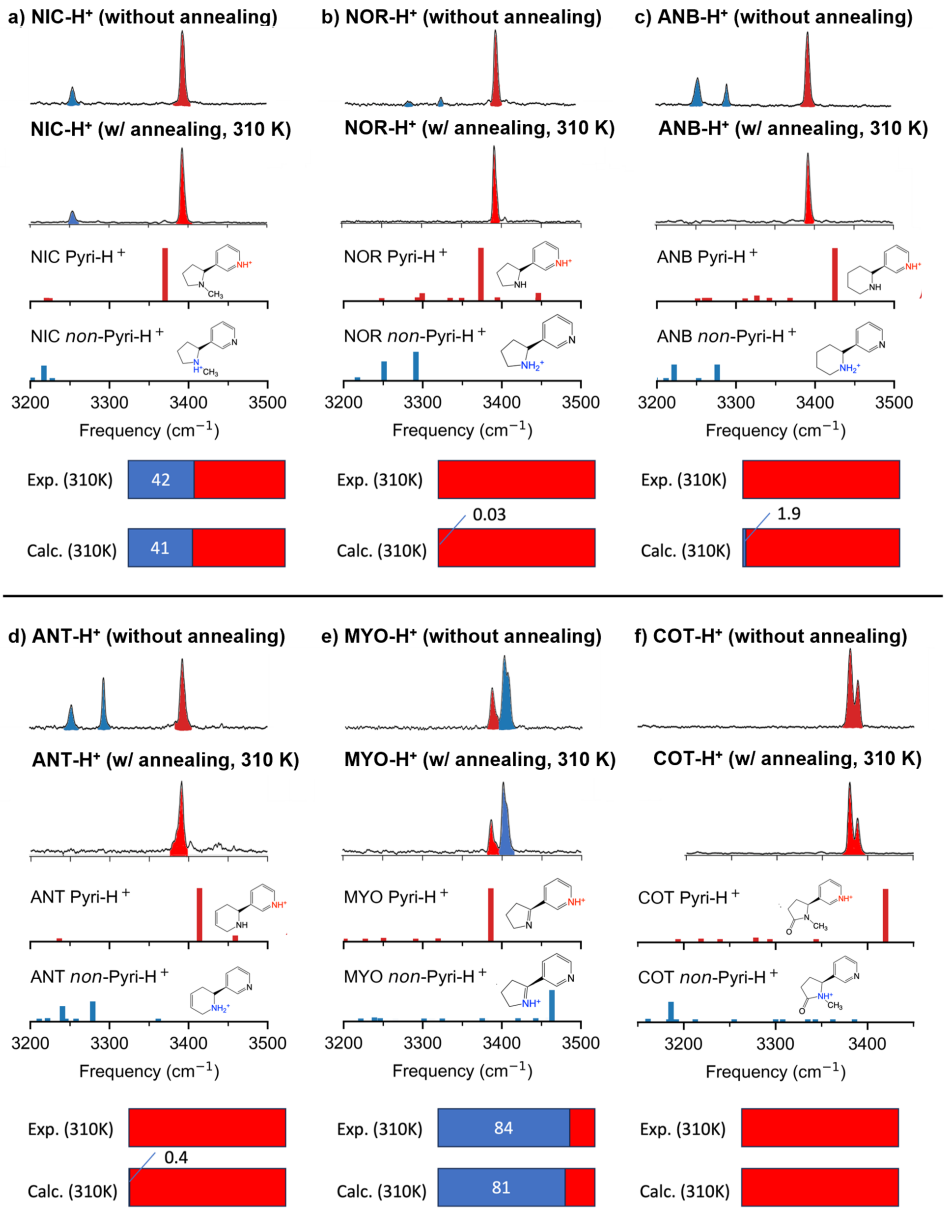

Figure 9.1: IR spectra of protonated nicotinoids, a) NIC-H^+, b) NOR-H^+, c) ANB-H^+, d) MYO-H^+, e) ANT-H^+ and f) COT-H^+. For each case the top panel traces the experimental IRPD spectra without annealing of the Pyri-H^+(red) and *non*-Pyri-H^+(blue) protomers. Right below are the experimental spectra that include thermalization at T=310 K ("with annealing"). The lowest two panels for each nicotinoid show the computed transitions using VPT2 (scaled by 0.985 to align the strong Pyri-H^+ transition). Colored bands at the bottom of each series of panels show the experimental and theoretically estimated populations of the the two protomers at T=310 K. The experimental uncertainties range from 0.7%–2.3%.

Table 9.1: Percentage (%) of the *non*-Pyri-H$^+$ protomers for all six alkaloids determined from the experimental spectra (Exp.), equation (9.3), and the theoretical populations (Calc.), equation (9.2). The calculated protomer populations are determined from the computed free energy differences between the Pyri-H$^+$ and *non*-Pyri-H$^+$ protomers assuming a Boltzmann distribution.

T	NIC-H$^+$	NOR-H$^+$	ANB-H$^+$	MYO-H$^+$	ANT-H$^+$	COT-H$^+$
180 K (Calc.)	14	0	0.08	90	0	0
310 K (Calc.)	41	0.03	1.9	81	0.4	0
310 K (Exp.)	42	0	0	84	0	0

coefficient of 154.3 km/mol for ANT Pyri-H$^+$, and 3279 and 3240 cm^{-1} with absorption coefficients of 51.2 and 35.15 km/mol, respectively, for ANT-*non*-Pyri-H$^+$.

However, MYO-H$^+$ and COT-H$^+$ exhibit different spectral features. For MYO-H$^+$ there are two bands mainly at 3387 cm^{-1} and 3402 cm^{-1}. These bands are assigned to the Pyri-H$^+$ and *non*-Pyri-H$^+$ protomers, respectively, by comparing them to the theoretical IR spectra. The computed MYO-Pyri-H$^+$ transition is at 3386 cm^{-1} with an absorption coefficient of 135.9 km/mol, and the computed MYO-*non*-Pyri-H$^+$ transition is at 3516 cm^{-1}, with an absorption coefficient of 73.9 km/mol. A shoulder at 3404 cm^{-1} is assigned to the NH stretching band of a rotational conformer of amine rings since both rotational conformers are isoenergetic ($\Delta G < 0.01$ kcal/mol). Similar assignments are obtained for COT-H$^+$, namely the band at 3380 cm^{-1} is attributed to the NH stretching vibration of the Pyri-H$^+$ protomer, while the other band at 3389 cm^{-1} is attributed to other Pyri-H$^+$ conformers of COT. No bands corresponding to the *non*-Pyri-H$^+$ protomer are observed, which is expected due to the large energetic difference between the Pyri-H$^+$ and *non*-Pyri-H$^+$ protomers of COT-H$^+$.

In the following we proceed with the population analysis of the non-Pyri-H$^+$ and Pyri-H$^+$ protomers of the six alkaloids at the physiological temperature (T=310 K) via both theory and experiment. Figure 9.1 shows the IRPD spectra at 310 K labeled "w/ annealing, 310K" for a) NIC-H$^+$, b) NOR-H$^+$, c) ANB-H$^+$, d) ANT-H$^+$, e) MYO-H$^+$, and f) COT-H$^+$ whereas the percentage of the *non*-Pyri-H$^+$ protomers determined from both experiment and theory is listed in table 9.1 for all alkaloids.

Nicotinoid	Global Minimum	Local Minimum	ΔE_e	ΔE_0	$\Delta G_{310\ K}$
NIC [a)]	Pyri-H$^+$ (S, anti)	*non*-Pyri-H$^+$ (N, syn)	0.26	1.14	0.37
NOR [a)]	Pyri-H$^+$ (S, anti)	*non*-Pyri-H+ (N, syn)	4.95	5.34	4.77
ANB	Pyri-H$^+$ (Equ, anti)	*non*-Pyri-H$^+$ (Equ, anti)	2.40	2.86	2.75
MYO	*non*-Pyri-H$^+$ (syn)	Pyri-H$^+$ (syn)	0.37	0.55	0.57
ANT	Pyri-H$^+$ (Ax, syn)	*non*-Pyri-H$^+$ (Equ, anti)	3.39	3.74	3.59
COT [b)]	Pyri-H$^+$ (S, anti)	*non*-Pyri-H$^+$ (N, syn)	20.84	20.32	20.32

[a)] Reference 6.

[b)] All energies at ωB97X-D/aVDZ.

Table 9.2: Relative electronic energies (ΔE_e, MP2/CBS), zero-point corrected energies (ΔE_0, ωB97X-D/aVDZ) and free energies ($\Delta G_{310\ K}$) between the global and the low lying local minima for each alkaloid. All energies are in kcal/mol.

We have previously reported the IRPD spectra of NIC-H$^+$ and NOR-H$^+$ at 300 K, [384] which is slightly lower than the average human body temperature of 310 K. The relative intensities of the NH stretching vibrations of the *non*-Pyri-H$^+$ and Pyri-H$^+$ protomers at 310 K are almost identical to the ones at 300 K for NIC-H$^+$. After taking into account the absorption coefficients of the *non*-Pyri-H$^+$ and Pyri-H$^+$ protomers, we obtained a population ratio of 42:58 at 310 K (40:60 at 300 K [384]), which is represented by a bar chart below the spectra in figure 9.1. Here, blue and red bars represent the relative populations of the *non*-Pyri-H$^+$ and Pyri-H$^+$ protomers, respectively. This result is consistent with the theoretical prediction of 41:59 for the ratio of the protomer populations at 310 K. The calculated electronic (MP2/CBS), zero-point corrected (ωB97X-D/aVDZ), and free energy differences between the global and the low-lying minima for each nicotinoid at 310 K are summarized in table 9.2. As reported in earlier studies, the small energy difference between NIC's two protomers results in their mixture in the gas-phase.

The population ratio between the *non*-Pyri-H$^+$ and Pyri-H$^+$ protomers clearly changed for NOR-H$^+$ upon raising the temperature from 300 K to 310 K. At 300 K, the non-Pyri-H$^+$ protomer was observed and the population ratio non-Pyri-H$^+$:Pyri-H$^+$ was reported as 8:92. [384] However,

at 310 K, we cannot detect the NH stretching bands of the non-Pyri-H^+ protomer at 3276 cm^{-1} and 3322 cm^{-1}, therefore we conclude that the population is 100% dominated by the Pyri-H^+ protomer at 310 K, a finding that is reproduced by the theoretical calculations, see the bar charts below the spectra in figure 9.1 and the entries in table 9.1.

The predominance of the Pyri-H^+ protomer is also observed for ANB, ANT, and COT. As shown in figure 9.1c, 9.1d and 9.1f, we observe only the NH stretching vibrational bands of the Pyri-H^+ protomers at around 3390 cm^{-1}. The NH bands of the non-Pyri-H^+ protomers should be at 3253 cm^{-1} and 3290 cm^{-1} for ANB-H^+, 3252.5 cm^{-1} and 3293 cm^{-1} for ANT-H^+, and around 3180 cm^{-1} for COT-H^+ , however no bands can be experimentally detected in that range. These are consistent to the theoretical calculations for the population ratio shown by the bar chart in figure 9.1. The large energetic difference between protomers in table 9.2 for NOR, ANB, ANT, and COT is consistent with the observation of a single protomer in the thermalized spectra shown in figure 9.1.

MYO is the only alkaloid in which the non-Pyri-H^+ protomer dominates the population. After annealing at 310 K, the observed two bands at 3387 cm^{-1} and 3402 cm^{-1} remain and the latter band is assigned to the *non*-Pyri-H^+ protomer. After the calibration by the theoretical absorption, we found that the *non*-Pyri-H^+ protomer is more than 80% of the population, a result that is consistent with the theoretical prediction of 81:19 for the population ratio (cf. table 9.1), resulting from the slight energetic bias for the *non*-Pyri-H^+ protomer reported in table 9.2.

Let us now discuss the previously stated hypothesis that the high addictive potential of a certain nicotinoid is related to the population ratio of the *non*-Pyri-H^+ to Pyri-H^+ protomers. The ICSS measurements, which measure the biological activity of the various nicotinoids, suggest that NIC is highly active, ANB and NOR are 4 and 6 times less active than NIC whereas ANT, COT and MYO do not show clear biological activities similar to NIC. This correlation has been recently established in the biological environment (binding pocket) via the molecular docking study in the prior chapter. ANB, NOR, ANT and COT do not show a measurable population of the *non*-Pyri-H^+ protomer from the IR spectra. This study provided a screening protocol that separates the highly addictive nicotinoids (NIC) from those that are less addictive (ANB and NOR). The current gas-phase measurements at this temperature cannot separate the non-addictive nicotinoids (COT) from mildly addictive ones (ANB and NOR). This corroborates the findings of the past chapter,

figure 8.5 and 8.12, with experimental validation. MYO remains as an exception. It does not have a strong biological activity like NIC, but 80% of its gas phase population is the *non*-Pyri-H^+ protomer, which is assumed to be the biologically active one. The case of MYO and the reason for this "exception" from the general trend will be attributed to its planar structure and the seric interactions identified in the chapter 8.

9.4 Conclusions

In conclusion, we determined the preferred gas phase protonation sites of six naturally occurring nicotinoids (NIC, NOR, ANB, ANT, MYO, COT) at room temperature via a joint experimental infrared (IR) and theoretical electronic structure study. We found that these are protonated as follows: NIC is a mixture of pyridine and *non*-pyridine (pyrrolidine) protomers, NOR, ANB, ANT and COT are pure pyridine protomers, whereas MYO is mostly a *non*-pyridine (pyroline) protomer. The experimental population ratios determined here serve as a first step in identifying a potential descriptor for highly addictive nicotinoids, which is in line with the previous energetic docking studies of the last chapter. This is certainly true for NIC (addictive, preference for the *non*-pyridine protomer), and the group of NOR, ANB, ANT, and COT (not addictive, preference for the pyridine protomer). The exception to this protomer-specific biological function relationship is MYO, which is not addictive with a preference for the *non*-pyridine protomer. However, this exception to the general trend is well attributed to MYO's distinct structural features arising from the conjugated π system. This highlights the careful consideration one must take connecting molecular properties to complicated biological processes. Finally, the completely non-addictive behavior of COT mostly stems from the large energetic bias towards its pyridine protomer.

This chapter shows that the conclusions of the molecular docking model presented in chapter 8 have experimental validation. The model, which predicted the addictive potential by considering the differences in the binding energy, was decomposed into fundamental processes including proton exchange, deformation, and interaction. The results presented in this chapter show that the molecular properties of the nicotinoids, which facilitate the proton exchange, are measurable through experimentation. The effect of deformation can be estimated through a steric comparison of the nicotinoids. Components of the interaction energy are currently being evaluated, with one recent report by Okura, Hallsten et al. highlighting the strong Trp156-NIC interaction. [399] Their report

finds that Trp156 stabilizes the bioactive protomer through cooperative effects of the indole side chain and the amide backbone. The minimum energy gas phase structure resembles that seen in the receptor, suggesting the ability for gas phase experiments to model the function of the receptor. This work bridges the gap between the purely computational docking models in chapter 8 and the experimentally verifiable screening techniques presented in this chapter.

Chapter 10

SWITCHING OF PROTONATION SITES IN HYDRATED NICOTINE VIA A GROTTHUSS MECHANISM [412]

10.1 Introduction

The last two chapters highlighted the importance of protonation in the molecular recognition of nicotine by nAChRs. One of NIC's protomers, Pyrro-H^+, binds favorably and strongly to the Trp156 residue in the nAChR, whereas the Pyri-H^+ protomer cannot bind due to steric concern. Chapters 8 and 9 highlighted the importance of the bioacitve protomer in the nicotinoid's addictive potential, and subsequent work has shown that the amino acids in the nAChR stabilize the Pyrro-H^+ protomer through cooperative effects. This chapter will discuss evidence for a low entropy mechanism for proton transfer in NIC-water clusters, and it's implication for proton transfer mechanisms in fully aqueous and confined but solvated NIC. In contrast to the past two chapters, this work will focus exclusively on NIC-H^+.

Recent joint experimental infrared (IR) and computational ab initio studies have shown that gas phase NIC is a mixture of the pyridine (Pyri) and Pyrro protomers, [384] whereas a single water molecule, present in the acetylcholine binding protein (AChBP), favors the Pyri- with respect to the Pyrro- protomer by 9:1. [380] The effect of the additional hydration (with more than a single water molecule) on NIC remains experimentally unexplored until this study. From a biological perspective, the position of the protonation site and the relative energy between the two protomers are crucial, as NIC must dehydrate in order to enter the anhydrous binding pocket of nAChR; the fewer water molecules required to stabilize the Pyrro- protomer, the easier it will be for this bioactive isomer to enter the binding pocket and be stabilized by hydrogen bonding to amino acid residues. Early investigations of NIC-water clusters have been performed by Gaigeot et al. [401] Although the energetics of that study do not align with those more recently reported at higher levels of theory, [380, 384] that early work provided strong insights into the subsequent hydration of NIC and the ability of the proton to shuttle between the nitrogen atoms of the pyridine and

pyrrolidine rings. Besides this single theoretical study, very little has been done to describe the hydration of NIC-H^+ as no experimental work to date has been reported for the NIC-$H^+(H_2O)_{n>1}$, clusters. The joint experimental-–theoretical results reported in this chapter represent the first experimental probe of the hydrated nicotine clusters revealing the critical role of water in inducing the shifting of the protonation site from the Pyri- to the Pyrro- nicotine protomer with increasing hydration via a Grotthuss mechanism. To the best of our knowledge, this is the first experimental probe of a Grotthuss mechanism in a finite aqueous cluster. [113, 413–415]

10.2 Methodology

Our collaborators applied cold ion trap infrared spectroscopy to determine the protonation sites of NIC-H^+-$(H_2O)_n$, NIC-H^+-$(D_2O)_n$. The procedure nicely follows the experiments performed in chapter 9, section 9.2, with minor modifications. The temperature of the termalizing trap was set to 130 K to allow larger NIC-H^+-$(H_2O)_n$ to form. In the specroscopic trap, Collision-Assisted Stripping Infrared (CAS-IR) spectroscopy, a protocol which had been developed by our collaborators and reported earlier, was performed. [406] Through this method, water molecules were removed from ions in the QIT by lowering offset voltage and increasing collision energy with buffer gas (He/H_2). H_2 tags were attached to the water-detached NIC-H^+ and removed through IRPD, producing pure NIC-H^+ monomer spectra in populations representative of NIC-H^+-$(H_2O)_n$. It should be emphasized that the IR spectrum of monohydrated NIC-H^+ without stripping a water molecule is already too complicated and the relative population of the two protomers is very difficult to be determined. This is the reason that the CAS-IR method was employed to measure the protomer populations of hydrated NIC-H^+. Additionally, the collisional stripping process has minimal effect on the population ratio of the protomers. [406]

The structures of the NIC-$H^+(H_2O)_{n=2-5}$ clusters were optimized at the DFT level of theory [154] with the ωB97X-D functional [163] and the aug-cc-pVDZ basis set [194] using the Gaussian 16 suite of electronic structure methods. [173] This level of theory has been shown to produce results in good agreement with the ones obtained at the second order Møller-Plesset (MP2) level of theory. [145, 147, 380, 384] We performed an extensive search for numerous isomers for each cluster with the lowest energy ones used in this report. Harmonic analysis of the vibrational frequencies augmented with rigid rotor evaluation of the partition function were used to obtain the

free energy (ΔG) of each isomer at T=130 K, and a Boltzmann population was used to predict the thermodynamic distribution between the Pyri-H^+ and Pyrro-H^+ protomers. In addition to the minima, the transition states for the Grotthuss mechanism were also identified at the ωB97X-D/aug-cc-pVDZ level of theory. The intrinsic reaction coordinate (IRC) paths were computed to ensure the connectivity between the respective Pyri-H^+ / Pyrro-H^+ minima and the computed transition states. [416] Free energies of the saddle points were computed using harmonic analysis and rigid rotor partition functions. These were used to compute rate constants via classical transition state theory (TST) [417–419] via

$$k_{\text{eff}} = \kappa \frac{k_B}{T} e^{-\Delta G^{\ddagger}/RT} \tag{10.1}$$

where κ is the recrossing term (assumed to be equal to one), k_B is Boltzmann's constant, T is temperature, h is Plank's constant, $\Delta G^{\ddagger}$ is the free energy difference between the transitions state and the reactants, and R is the universal gas constant. The rate-limiting step is assumed to be the proton transfer. All energies are reported with respect to the (global) Pyri-H^+ minimum for each cluster. The reported free energies are computed at T=130 K, which corresponds to the experimental conditions.

10.3 Results and Discussion

10.3.1 Thermodynamic Distribution of Protomers in NIC-H^+-$(H_2O)_n$ Clusters

Fig. 10.1 shows the CAS-IR spectra of NIC-H^+-$(H_2O)_n$, n=0–5, at T=130 K. The spectra trace the vibrational bands from the NIC-H^+ core after the water molecules have been collisionally removed. The bands at 3394 cm^{-1} and 3254 cm^{-1} have been previously assigned to the NH stretching vibrations of the Pyri-H^+ (red) and Pyrro-H^+ (blue) protomers, respectively. [384] The horizontal bars trace the protomer population ratio estimated using the calculated IR absorption cross sections. [380] Recently, we reported the CAS-IR spectra for the $n = 0$ and $n = 1$ protomers with the ratio of Pyrro-H^+:Pyri-H^+ of 2:3 at T=300 K for n=0 and 1:9 at T =220 K for n=1. [380, 384] The population ratios presented here are slightly different from the previous work, due to the different temperature (T=130 K in this study). The intensity of the NH stretching vibrations of Pyri-H^+ increases from n=0 to n=1, as reported previously. [380, 384] For the larger hydrated clusters, the N-H stretching vibration of Pyri-H^+ decreases with increasing number of water molecules, and

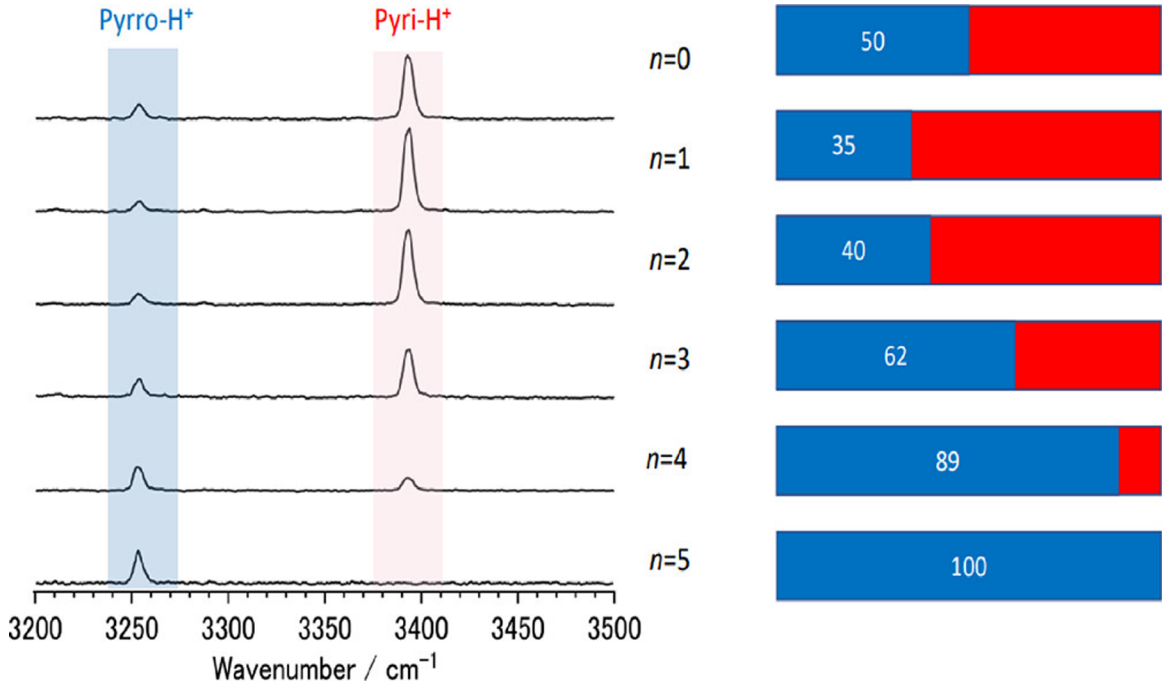

Figure 10.1: Experimental CAS-IR spectra at T=130 K of the Pyrro- (blue) and Pyri- (red) protomers of NIC-H^+-$(H_2O)_n$, n=0–5 and The CAS-IR spectra show a small shift to the Pyri-H^+ protomer from n=0 to n=1 and a gradual shift to the Pyrro-H^+ protomer for larger n.

finally vanishes at $n = 5$. The trend of the intensities of the N-H stretching vibration in the Pyrro-H^+ protomer behaves in an exactly opposite way: it decreases from $n = 0$ to $n = 1$, then increases from $n = 2$ to $n = 4$ and finally dominates at $n = 5$. The experimental spectra therefore suggest that hydration induces a shift of NIC's protonation site from the Pyri- to the Pyrro- protomer, the latter being the only protomer detected with 5 water molecules. In particular, the drastic decrease of the population of the Pyri-H^+ and the simultaneous rapid increase of the Pyrro-H^+ population from $n = 3$ to $n = 5$ suggests intra-cluster proton transfer. Also, it suggests the size dependence and threshold of the hydration-induced proton transfer.

The distributions of Pyri-H^+ and Pyrro-H^+ can be determined through theoretical calculations. The increased complexity in the sampling of the configuration space with increasing hydration produces an increased ensemble size of isomers available. The population ratios are the ratio of the sum of Boltzmann populations of each of the isomers. These population ratios can be converted

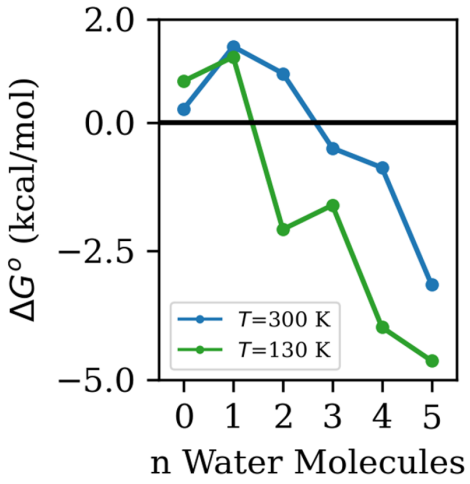

Figure 10.2: Theoretical free energy differences (ΔG°) between the two protomers at T=130 K and T=300 K with n water molecules.

into ensemble free energy differences ($\Delta G^\circ_{\text{eff}}$) through equation [420]

$$\Delta G^\circ_{\text{eff}} = -RT \ln \frac{P_{\text{Pyri}}}{P_{\text{Pyrro}}} \tag{10.2}$$

where P is the ensemble populations of each protomer, R is the universal gas constant, and T is the temperature. Populations of NIC-H^+-$(H_2O)_n$, n=0–1, were used from past work [380, 384] for completeness. The populations of NIC-H^+-$(H_2O)_n$, n=2–4, were obtained following a thorough search of all potentially relevant isomerization pathways. To ensure adequate sampling of the large ensemble of stationary points for NIC-H^+-$(H_2O)_5$, chemical intuition was applied to limit calculations to 127 energetically relevant and representative isomers for that cluster. The corresponding number of structures probed for the n=2–4 clusters were 9, 32 and 127, respectively.

The ensemble free energy differences between the Pyri-H^+ and Pyrro-H^+ protomers, plotted against the hydration number n in figure 10.2, confirm the experimental observations regarding the shift of the populations towards Pyri-H^+ with one water molecule and then gradually shift to Pyrro-H^+, which becomes the exclusive protomer observed for n=5. Depending on the temperature, the shift in the favored protonation site occurs between n=2 and n=3 with higher temperatures favoring larger n due to the stable high entropy branching isomers in the Pyri protomer of NIC-H^+-

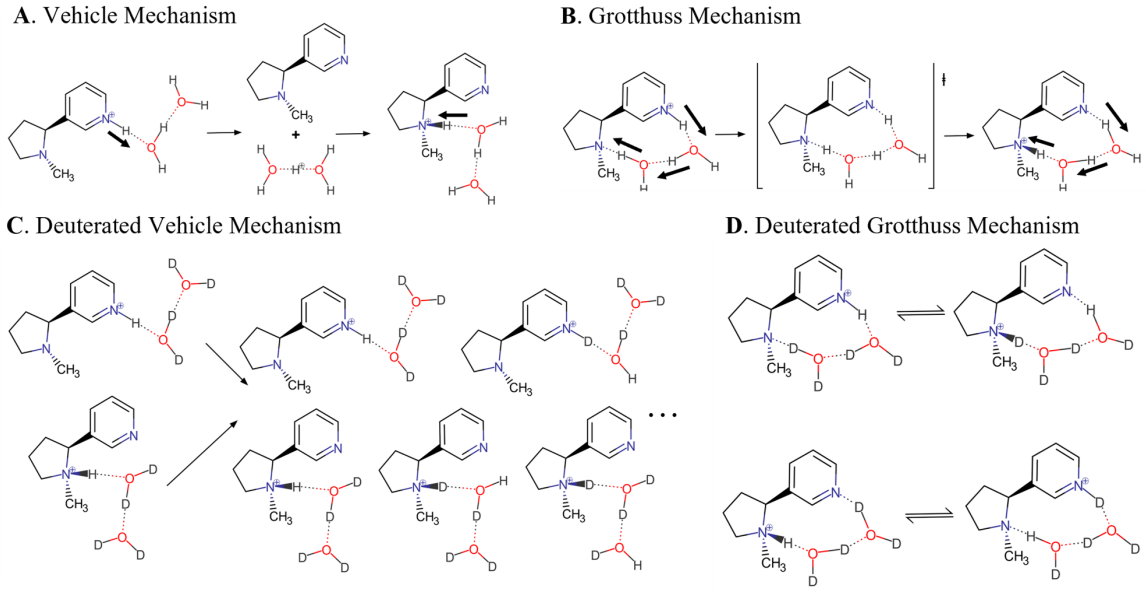

Figure 10.3: Proton transfer via the bimolecular vehicle (via translational diffusion, A) and Grotthuss (proton shuffling, B) mechanisms. Panels (C) and (D) depict the products of the vehicle (C) and Grotthuss (D) mechanisms when protonated NIC is solvated by deuterated water molecules.

$(H_2O)_2$. With two or more water molecules, bridge-like structures that connect the protonated acid to the base through hydrogen bonded networks become the most stable isomers. These structural motifs grow more complex with increased hydration to include diamond, pentagon, and cage like structures with increasing energy differences between Pyri-H^+ and Pyrro-H^+. Our energetic calculations (figure 10.2) support the changes in the experimentally observed intensities and subsequent protomer populations (figure 10.1) through the consideration of thermodynamics, however the observed and predicted populations are not directly comparable. Experimental populations likely represent a snapshot during the slow kinetic process of proton transfer. Therefore, understanding the mechanism of proton transfer remains important.

10.3.2 Mechanisms for Proton Transfer: Vehicle vs. Grotthuss

There are two possible mechanisms of solvent-induced proton transfer, namely the vehicle [421, 422] and Grotthuss [422, 423] mechanisms, which are shown in figure 10.3A,B for NIC solvated by two water molecules. In the vehicle mechanism (figure 10.3A), a water cluster picks up a proton

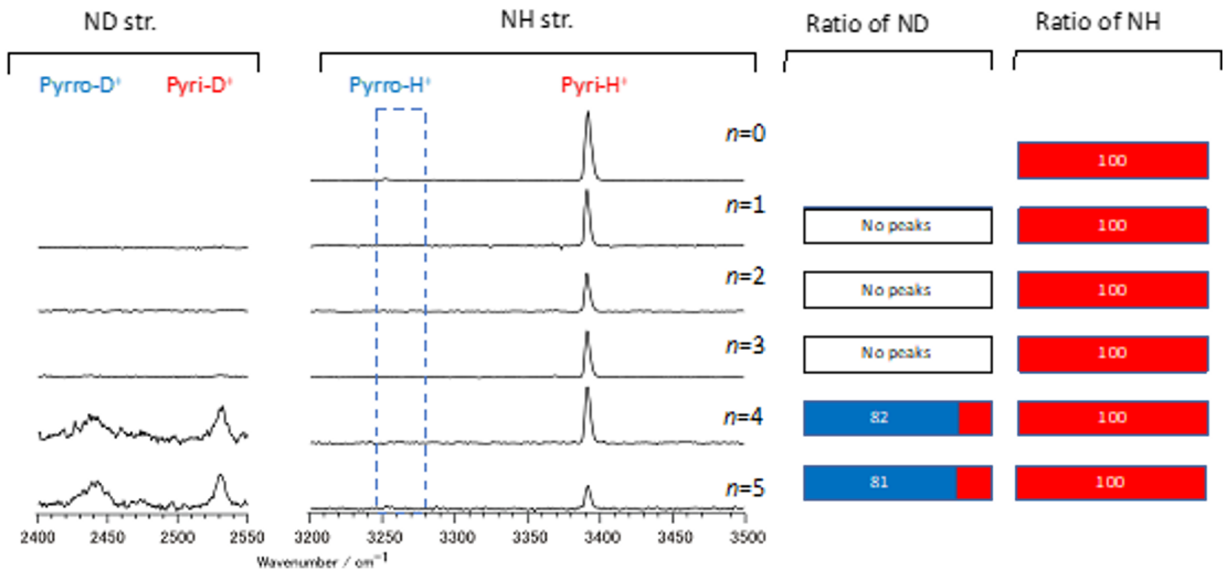

Figure 10.4: The CAS-IR spectra of NIC-H^+-$(D_2O)_n$, n =0–5, obtained with methanol/H_2O solvent. Initial population from ES ionization is only Pyri-H^+, and absence of Pyrro-H^+ represent Grotthuss mechanism. Similar ratio of Pyrro-D^+: Pyri-D^+ for all n=4 and 5 clusters predict that thermal equilibrium is archived between Pyrro-D^+ and Pyri-D^+.

from the NH of the Pyri- group and it subsequently dissociates from NIC-H^+. It then releases the proton to the N atom in Pyrro- group. In the Grotthuss mechanism (figure 10.3B), a water network forms a H-bonded bridge between the two protonation sites, and a proton is relayed along the water bridge via a concerted process. Since both mechanisms can produce the same Pyri- to Pyrro- proton transfer, it is difficult to determine the actual mechanism solely from the CAS-IR spectra of NIC-H^+-$(H_2O)_n$ shown in figure 10.1. To further reveal the actual mechanism, we resort to the experimental spectra with deuterated water coupled with theoretical analysis of the underlying potential energy surfaces.

The determination of the actual mechanism (vehicle vs. Grotthuss) responsible for the shifting of the protonation sites was achieved via isotopic substitution. In deuterated water clusters of the solvating water, the proton transfer from Pyri-H^+ by the Grotthuss mechanism (figure 10.3D) selectively produces Pyrro-D^+ because only the terminus D atom can be transferred to the N atom in the pyrrolidine group. In contrast, the vehicle mechanism (figure 10.3C) has more freedom by the

reorientation of the released water cluster, allowing for the possibility to produce both the Pyrro-D^+ and Pyrro-H^+ protomers. Similarly, the reverse proton transfer from Pyrro-H^+ to Pyri-H^+ in the deuterated water clusters generates only the Pyri-D^+ protomer by the Grotthuss mechanism (figure 10.3D), while both the Pyri-D^+ and Pyri-H^+ protomers can be populated via the vehicle type mechanism (figure 10.3C).

The IR transitions of Pyri-H^+, Pyrro-H^+, Pyri-D^+ and Pyrro-D^+ protomers of the NIC-H^+-$(D_2O)_n$ clusters can be easily distinguished by the NH and ND stretching bands, thus the CAS-IR spectroscopy has the potential to determine the type of the proton transfer mechanism. The problem with this approach is that the Pyri-H^+ and Pyrro-H^+ protomers co-exist at n=0, i.e., when no water is present. Their population ratio is about 1:1 when produced by ES ionization with methanol solvent, i.e., all possible species (Pyri-H^+, Pyrro-H^+, Pyri-D^+ and Pyrro-D^+) can be detected regardless of the type of mechanism. Therefore, it is necessary to bias the initial distribution towards one protomer. It has been reported that ionizing a sample with the ES of a solvent containing acetonitrile yields a protomer distribution different from that of a methanol solvent. [424]

To confirm the mechanism, NIC-H^+ populations were biased to Pyri-H^+ by electrospraying in methanol/H_2O solvent. The CAS-IR spectra of 100% Pyri-H^+ biased NIC-H^+-$(D_2O)_n$, are shown in figure 10.4. Here, Pyrro-H^+ which indicates the contribution of the vehicle mechanism cannot be detected with S/N=400. The band at 2439 cm^{-1} is assigned to the ND stretching vibration of the Pyrro-D^+ protomer, which is the result of the proton transfer in Pyri-H^+-$(D_2O)_{4,5}$, a finding that is consistent with the proton transfer occurring via the Grotthuss mechanism. The band at 2532 cm^{-1} corresponds to the ND stretching vibration of Pyri-D^+, which is the proton transfer product from Pyrro-$H^+(D_2O)_{4,5}$. Note band is not a direct product of the starting population (Pyri-H^+), but is instead a product of Pyrro-D^+ to Pyri-D^+ Grotthuss transfer after H/D exchange. The intensity of this Pyrro-D^+ band is comparable to that in the Pyri-D^+ protomer, and the relative Pyrro-D^+:Pyri-D^+ population is estimated at 85:15. The population ratio of Pyri-D^+ and Pyrro-D^+ is insensitive to the initial population ratio of Pyri-H^+:Pyrro-H^+, suggesting that the reaction products of the proton transfer are finally thermal equilibrium between Pyrro-D^+ and Pyri-D^+. It is reasonable because the reaction products could have excess energy after the proton transfer. We therefore conclude that the observed solvent-induced proton transfer is facilitated by the Grotthuss

mechanism, and the vehicle mechanism is a very minor process in these experimental conditions. To the best of our knowledge, this is the first experimental verification of a proton transfer mechanism on the ground state of a gas phase aqueous cluster.

Since the low energy structures for NIC-H^+-$(H_2O)_n$, n=2–5, consist of hydrogen bonded bridges, the Grotthuss mechanism displayed in figure 10.3C,D was investigated computationally to address the plausibility of the proton transfer under the experimental conditions. Figure 10.5A shows the free energy reaction profiles at $T = 130$ K for NIC-H^+-$(H_2O)_n$, n=2–5, demonstrating the lowering of the activation energy with increasing hydration.

The Grotthuss mechanism for the $n = 2$ cluster (see figure 10.5D,) occurs in one step via a shared-proton (Zundel-like [425]) transition state. The reaction for these single step reactions is described by equation (10.2). The transition state is 12.4 kcal/mol above the global Pyri-H^+ minimum, with a classical rate constant of 3.4×10^{-9} s^{-1}. The reverse classical rate constant is 2.9×10^{-11} s^{-1}. The lowest energy Grotthuss mechanism for NIC-H^+-$(H_2O)_3$ is similar, in that it also occurs in a single step as shown in Figure 10.5A. That transition state, shown in figure 10.5E, consists of a bridge with a hydronium ion in the center (Eigen-like [426]) and a free energy barrier of 9.0 kcal/mol above the global Pyri-H^+ minimum. There exists a second, higher energy Grotthuss-like path with a barrier of 9.8 kcal/mol, in which the hydrogen bond bridge bends in the opposite direction. This second higher path contributes about 5% to the computed total rate constant of 2.0×10^{-3} s^{-1}. The reverse classical rate constant is 1.4×10^{-6} s^{-1}.

The proton transfer via the Grotthuss mechanism for NIC-H^+-$(H_2O)_4$ also occurs through a hydrogen bond bridge with a three-water molecule motif, as in the $n = 3$ case. The additional water molecule accepts a hydrogen bond from the middle water to stabilize the positive charge in the proton transfer. Notably, this greatly stabilizes the hydronium structure, and the Eigen structure becomes a stable intermediate; the transition states resemble shared proton structures to either forward or backward directions of the stable hydronium intermediate. The lowest energy pathway has barriers of 5.2 and 5.4 kcal/mol (Figure 10.5A). Because both steps are nearly equal in energy, the reaction kinetics cannot be simplified into a single step. Instead, the rate constant is given via

$$k_{\text{rxn}} = \frac{k_1 k_2}{k_{-1} + k_2} = k_2 \cdot \frac{1}{1 + k_{-1}/k_2} \tag{10.3}$$

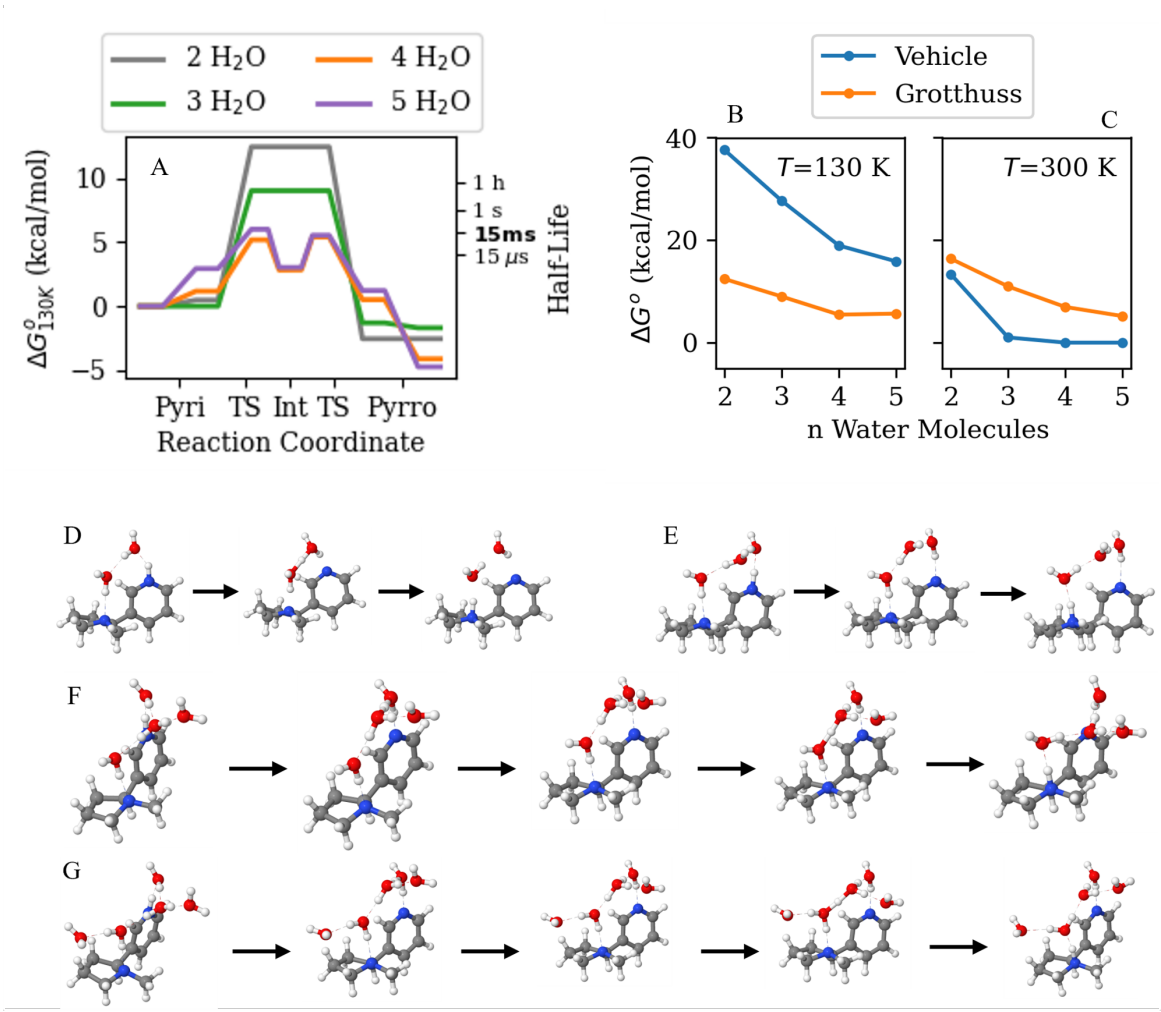

Figure 10.5: (A) The free energy for the Grotthuss mechanism at T=130 K for NIC-H$^+$-$(H_2O)_n$, n=2–5, showing the lowering of the activation energy with increasing hydration. The half-lives are derived from first-order kinetics, with the approximate time frame of the experiment (15 ms) in boldface. The barriers of the vehicle and Grotthuss mechanism are compared at 130 K (B) and 300 K (C), showing the low temperature preference for Grotthuss transfer, bur high temperature vehicle transfer. Pictorially, the structures for NIC-H$^+$-$(H_2O)_{n=2-5}$ are shown in panels D-G, respectively.

where k_{rxn} is the total rate of the reaction, k_1 is the forward rate of the first step forming the intermediate from the reactants k_{-1} is the reverse rate of the first step, and k_2 is the forward rate of the second step forming the products of the intermediates. Note that the expression can be simplified to create a rate constant equation that looks like

$$k_{\text{rxn}} = \kappa \frac{k_B}{T} h \frac{1}{e^{\Delta G^{\ddagger}_{\text{TS}_1}/RT} + e^{\Delta G^{\ddagger}_{\text{TS}_2}/RT}} \tag{10.4}$$

where $\Delta G^{\ddagger}_{\text{TS}_1}$ and $\Delta G^{\ddagger}_{\text{TS}_2}$ are the free energies of the transition states with respect to the global Pyri- minima and all other variables match definitions from equation (10.2). κ is assumed to be 1. This energy of the intermediate does not affect the rate constant under the steady-state approximation used in equation (10.3). Another low energy pathway exist through a similar 3-membered bridge. Using equation (10.4) and both pathways, the rate constant is estimated to be 1.8×10^{3} s^{-1}. The reverse classical rate constant is 1.6×10^{-2} s^{-1}. Proton transfer through the minimum energy structure, an antidromic water cycle, produced a high barrier and was not competitive with either discussed pathway.

The potential energy surface (PES) of NIC-H^+-$(H_2O)_5$ is associated with a large ensemble of bridge structures that hinders a full search of possible intermediates. However, chemical intuition gained from the PESs of the smaller clusters resulted in limiting the search for transition states to ∼30 bridge structures. Six important transition states become thermodynamically accessible, consisting of all modifications of the three-membered bridge found in the four-water cluster. The lowest energy path involves the extra water molecules accepting hydrogen bonds from the middle water molecule in the bridge and the water molecule near the Pyrro- nitrogen, shown in figure 10.5G. The reaction path again is stepwise, with free energy barriers of 6.0 and 5.5 kcal/mol (figure 10.5A). The additional 2 water molecules to the "core" 3-water bridge in NIC-H^+-$(H_2O)_5$ have the net effect of further reducing the barrier for the concerted proton transfer process. The cumulative rate constant for all six pathways is 4.8×10^{2} s^{-1}, slightly slower than the four-water case. The reverse classical rate constant is 4.5×10^{-6} s^{-1}. Evidence from the corresponding intrinsic reaction path suggests that tunneling could have an outsized role in the proton transfer in NIC-H^+-$(H_2O)_5$ resulting in the overall speed up of the reaction.The effect of tunneling for these processes is difficult to calculate due the high dimensionality of the cluster and the large curvature of the proton transfer pathway. Current advances in path integral molecular dynamics will allow for the computation of

tunneling effects in the Grotthuss mechanism, where high levels of theory are requied.

The barriers of the vehicle mechanism are approximated as either the energy of the minimum of the Pyri-H^+ protomer or the energy of the bimolecular species of NIC and a protonated water cluster with respect to the Pyri-H^+ protomer. It is acknowledged that these rate constants and barriers can be severely underestimated, and methods such as variational transition state theory [427] would serve as a better approximation for these rate constants. Nonetheless, the barriers of the vehicle mechanism are much higher than those of the Grotthuss mechanism at T=130 K (figure 10.5B) making the latter the preferred mechanism at the experimental conditions for all clusters with two or more water molecules. It is important to note that the vehicle mechanism has higher entropy due to the bimolecular character of the intermediates, a fact that allows it to be competitive to the Grotthuss mechanism at high temperatures (T=300 K, see figure 10.5C). Nevertheless, a more accurate kinetic theory would be needed to compare the rate constants at the physiological temperature (T=300 K).

10.4 Conclusions

This work probes the environmental stability of the bioactive and addictive form of NIC-H^+, namely the Pyrro- protomer, showing that it becomes the dominant one with only a few water molecules. Excess hydration of NIC-H^+ lowers the barriers for switching the protonation sites between the Pyri- and Pyrro- protomers. Experiments with deuterated water further suggest that the shifting of the protonation sites at the experimental conditions (T = 130 K) occurs via a Grotthuss rather than the bimolecular vehicle mechanism in the gas phase. This finding is corroborated by theoretical calculations, which also suggest that the vehicle mechanism is preferred in the gas phase at the room temperature (T = 300 K). As NIC-H^+ moves from the hydrous extracellular matrix to the pocket of nAChRs, it loses its aqueous solvation and enters a more anhydrous environment. The stability of the Pyrro-H^+ protomer with only a few water molecules supports the ability for this bioactive form to enter the binding pocket and remain available until stabilized by the binding residue. The current study unambiguously determines through experiments using deuterated water and theoretical modeling a detailed Grotthuss mechanism facilitating a concerted proton transfer in the ground state of a finite aqueous cluster.

Chapter 11

CONCLUSIONS

This thesis covered the physical and chemical applications of the hydrogen bond. The first half of this thesis established a physical relationship between the local properties of the hydrogen bond, namely energy, structure, and vibrational spectra. The decomposition of the hydrogen bond potential showed that the relationship is nearly linear for electrostatic hydrogen bonds, with initial slopes of 8.0 kcal/mol per pm and 4.5 kcal/mol per 100 cm^{-1}. The quantum mechanical nature of the hydrogen bond, namely exchange repulsion, adds a non-linear character to strong hydrogen bonds. This relationship was used to effectively employ the empirically fitted Iogansen relationship to a diverse set of hydrogen bonds, including intramolecular hydrogen bonds. Preliminary efforts at bridging the gap between the empirical universal relationship from Iogansen and the physical universal relationship show that additional parameterization of hydrogen bonded dimers is required to produce a smooth and universal relationship. None-the-less, there is now a deeper understanding of the interplay between energy, structure, and vibrational spectra in hydrogen bonds, with important applications to hydrogen bonds that are important to biochemistry and material science.

The second half of this thesis focused on specific studies of hydrogen bonds in aqueous and biological systems. Data science proved to be a useful tool in describing the hydrogen bond network in water clusters. Through machine learning techniques, models were developed that can identify structural similarities between clusters of different sizes. The models remained simple enough to connect back to the physical descriptors of the hydrogen bond, with adjacency, % HB, r_{OO}, θ_{OHO}, and q being the most important ones in describing the hydrogen bonded networks. Those descriptors were applied to the hydrogen bond networks of liquid water, where ensemble averaging effects perplex the structural descriptors of the hydrogen bond. Instead, the descriptors proved useful in describing the hydrogen bond breaking process, showing that hydrogen bond breaking is a continuous and smooth process, and the closely bound second solvation shell for low adjacency water molecules comes as an artifact in defining the hydrogen bond.

Finally a comparative analysis of the protonation sites of nicotine and similar nicotinoids highlighted the importance of hydrogen bonding in the molecular recognition of the $\alpha_4\beta_2$ nAChR. Models of the binding pockets, developed from the experimental X-ray data, suggested that the hydrogen bond interaction between the Trp156 residue and the nicotinoid is important in the selectivity of similarly structured analogs. Superfluous hydrogen bonding interactions with the C-loop region in demethylated nicotinoids seemed to inhibit active binding. Proton affinity also played an important role in predicting addiction. This most clearly appeared in the stability of the bioactive protomer over the gas-phase minimum protomer and became an experimentally obtainable descriptor for maximum addictive potential. The relative stability between the two protomers was found to be quite sensitive to the aqueous environment, where only a few water molecules were required to stabilize the bioactive protomer and facilitate a proton transfer. Although a bimolecular vehicle-like mechanism may be the preferred pathway for proton transfer at standard aqueous conditions, the Grotthuss mechanism is an accessible pathway for low entropy environments such as confined binding pockets or vesicles. The subtleties and importance of the hydrogen bond are becoming clearer every year through continuous and sustained research efforts. The work here has provided a small piece of the understanding of hydrogen bonding interactions and will hopefully guide future investigations into the role of hydrogen bonding in important chemical systems.

BIBLIOGRAPHY

(1) Buckingham, A. D.; Del Bene, J. E.; McDowell, S. A. C. The hydrogen bond. *Chem. Phys. Lett.* **2008**, *463*, 1–10.

(2) Pimentel, G. C.; McClellan, A. Hydrogen Bonding. *Annu. Rev. Phys. Chem.* **1971**, *22*, 347–385.

(3) Karas, L. J.; Wu, C.-H.; Das, R.; Wu, J. I.-C. Hydrogen bond design principles. *WIREs Comput Mol Sci.* **2020**, *10*, e1477.

(4) Nilsson, A.; Petterson, L. The structural origin of anomalous properties of liquid water. *Nat. Commun.* **2015**, *6*, 8998.

(5) Eaves, J. D.; Loparo, J. J.; Fecko, C. J.; Roberts, S. T.; Tokmakoff, A.; Geissler, P. L. Hydrogen bonds in liquid water are broken only fleetingly. *Proc. Natl. Acad. Sci.* **2005**, *102*, 13019–13022.

(6) Bordo, D.; Argos, P. The Role of Side-chain Hydrogen Bonds in the Formation and Stabilization of Secondary Structure in Soluble Proteins. *J. Mol. Biol.* **1994**, *243*, 504–519.

(7) Astbury, W. T. The hydrogen bond. The hydrogen bond in protein structure. *Trans. Faraday Soc.* **1940**, *36*, 871–880.

(8) Pauling, L.; Corey, R. B.; Branson, H. R. The structure of proteins: Two hydrogen-bonded helical configurations of the polypeptide chain. *Proc. Natl. Acad. Sci.* **1951**, *37*, 205–211.

(9) Pauling, L.; Corey, R. B. The Pleated Sheet A New Layer Configuration of Polypeptide Chains. *Proc. Natl. Acad. Sci.* **1951**, *37*, 251–256.

(10) Eberhardt, E. S.; Raines, R. T. Amide-Amide and Amide-Water Hydrogen Bonds: Implications for Protein Folding and Stability. *J. Am. Chem. Soc.* **1994**, *116*, 2149–2150.

(11) Appiling, D. R.; Anthony-Cahill, S. J.; Mathews, C. K. In *Biochemistry Concepts and Connections*, 2nd ed.; Pearson: 2019, pp 144–189.

(12) Watson, J.; Crick, F. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **1953**, *171*, 737–738.

(13) Jeffrey, G.; Maluszynska, H.; Mitra, J. Hydrogen bonding in nucleosides and nucleotides. *Int. J. Biol. Macromol.* **1985**, *7*, 336–348.

(14) Szatyłowicz, H.; Sadlej-Sosnowska, N. Characterizing the Strength of Individual Hydrogen Bonds in DNA Base Pairs. *J. Chem. Inf. Model.* **2010**, *50*, 2151–2161.

(15) Appiling, D. R.; Anthony-Cahill, S. J.; Mathews, C. K. In *Biochemistry Concepts and Connections*, 2nd ed.; Pearson: 2019, pp 72–107.

(16) Xie, Z.; Hu, B.-L.; Li, R.-W.; Zhang, Q. Hydrogen Bonding in Self-Healing Elastomers. *ACS Omega* **2021**, *6*, 9319–9333.

(17) Gadwal, I. A Brief Overview on Preparation of Self-Healing Polymers and Coatings via Hydrogen Bonding Interactions. *Macromol.* **2021**, *1*, 18–36.

(18) Song, T.; Jiang, B.; Li, Y.; Ji, Z.; Zhou, H.; Jiang, D.; Seok, I.; Murugadoss, V.; Wen, N.; Colorado, H. Self-healing Materials: A Review of Recent Developments. *ES Mater. Manuf.* **2021**, *14*, 1–19.

(19) Wang, K.; Wang, Q.; Wang, X.; Wang, M.; Wang, Q.; Shen, H.-M.; Yang, Y.-F.; She, Y. Intramolecular hydrogen bond-induced high chemical stability of metal–organic frameworks. *Inorg. Chem. Front.* **2020**, *7*, 3548–3554.

(20) Hall, E. A.; Redfern, L. R.; Wang, M. H.; Scheidt, K. A. Lewis Acid Activation of a Hydrogen Bond Donor Metal–Organic Framework for Catalysis. *ACS Catal.* **2016**, *6*, 3248–3252.

(21) Fumino, K.; Peppel, T.; Geppert-Rybczyńska, M.; Zaitsau, D. H.; Lehmann, J. K.; Verevkin, S. P.; Köckerling, M.; Ludwig, R. The influence of hydrogen bonding on the physical properties of ionic liquids. *Phys. Chem. Chem. Phys.* **2011**, *13*, 14064–14075.

(22) Hunger, L.; Al Sheakh, L.; Fritsch, S.; Villinger, A.; Ludwig, R.; Harville, P.; Moss, O.; Lachowicz, A.; Johnson, M. A. Spectroscopic Evidence for Doubly Hydrogen-Bonded Cationic Dimers in the Solid Liquid and Gaseous Phases of Carboxyl-Functionalized Ionic Liquids. *J. Phys. Chem. B* **2024**, *128*, 5463–5471.

(23) Glendening, E. D.; Streitwieser, A. Natural energy decomposition analysis: An energy partitioning procedure for molecular interactions with application to weak hydrogen bonding strong ionic and moderate donor–acceptor interactions. *J. Chem. Phys.* **1994**, *100*, 2900–2909.

(24) Mao, Y.; Horn, P. R.; Head-Gordon, M. Energy decomposition analysis in an adiabatic picture. *Phys. Chem. Chem. Phys.* **2017**, *19*, 5944–5958.

(25) Umeyama, H.; Morokuma, K. The origin of hydrogen bonding. An energy decomposition study. *J. Am. Chem. Soc.* **1977**, *99*, 1316–1332.

(26) Phipps, M. J. S.; Fox, T.; Tautermann, C. S.; Skylaris, C.-K. Energy decomposition analysis approaches and their evaluation on prototypical protein–drug interaction patterns. *Chem. Soc. Rev.* **2015**, *44*, 3177–3211.

(27) Kitaura, K.; Morokuma, K. A new energy decomposition scheme for molecular interactions within the Hartree-Fock approximation. *Int. J. Quantum Chem.* **1976**, *10*, 325–340.

(28) Horn, P. R.; Mao, Y.; Head-Gordon, M. Probing non-covalent interactions with a second generation energy decomposition analysis using absolutely localized molecular orbitals. *Phys. Chem. Chem. Phys.* **2016**, *18*, 23067–23079.

(29) Wu, Q.; Ayers, P. W.; Zhang, Y. Density-based energy decomposition analysis for intermolecular interactions with variationally determined intermediate state energies. *J. Chem. Phys.* **2009**, *131*, 164112.

(30) Morokuma, K. Why do molecules interact? The origin of electron donor-acceptor complexes hydrogen bonding and proton affinity. *Acc. Chem. Res.* **1977**, *10*, 294–300.

(31) Stone, A. J., *The Theory of Intermolecular Forces*; Oxford University Press: 2013.

(32) Hunt, P. A.; Ashworth, C. R.; Matthews, R. P. Hydrogen bonding in ionic liquids. *Chem. Soc. Rev.* **2015**, *44*, 1257–1288.

(33) Ford, T.; Glasser, L. Ab initio calculations of the structural energetic and vibrational properties of some hydrogen bonded and van der Waals dimers Part 3. The formaldehyde dimer. *J. Mol. Struct. THEOCHEM* **1997**, *398-399*, World Congress of Theoretically Oriented Chemists, 381–394.

(34) Vila, A.; Graña, A. M.; Mosquera, R. A. Electron density characterisation of intermolecular interactions in the formaldehyde dimer and trimer. *Chem. Phys.* **2002**, *281*, 11–22.

(35) H. J. C. Berendsen J. P. M. Postma W. F. von Gunstaren and J. Hermans in Intermolecular Forces edited by B. Pullman (Reidel Dordrecht Holland 1981) p. 331.

(36) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The missing term in effective pair potentials. *J. Phys. Chem.* **1987**, *91*, 6269–6271.

(37) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.

(38) Tian, C.; Kasavajhala, K.; Belfon, K. A. A.; Raguette, L.; Huang, H.; Migues, A. N.; Bickel, J.; Wang, Y.; Pincay, J.; Wu, Q.; Simmerling, C. ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *J. Chem. Theory Comput.* **2020**, *16*, 528–552.

(39) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A program for macromolecular energy minimization and dynamics calculations. *J. Comput. Chem* **1983**, *4*, 187–217.

(40) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.

(41) Roos, K.; Wu, C.; Damm, W.; Reboul, M.; Stevenson, J. M.; Lu, C.; Dahlgren, M. K.; Mondal, S.; Chen, W.; Wang, L.; Abel, R.; Friesner, R. A.; Harder, E. D. OPLS3e: Extending Force Field Coverage for Drug-Like Small Molecules. *J. Chem. Theory Comput.* **2019**, *15*, 1863–1874.

(42) Abascal, J. L. F.; Sanz, E.; García Fernández, R.; Vega, C. A potential model for the study of ices and amorphous water: TIP4P/Ice. *J. Chem. Phys.* **2005**, *122*, 234511.

(43) Abascal, J. L. F.; Vega, C. A general purpose model for the condensed phases of water: TIP4P/2005. *J. Chem. Phys.* **2005**, *123*, 234505.

(44) Mahoney, M. W.; Jorgensen, W. L. A five-site model for liquid water and the reproduction of the density anomaly by rigid nonpolarizable potential functions. *J. Chem. Phys.* **2000**, *112*, 8910–8922.

(45) Toukan, K.; Rahman, A. Molecular-dynamics study of atomic motions in water. *Phys. Rev. B* **1985**, *31*, 2643–2648.

(46) Habershon, S.; Markland, T. E.; Manolopoulos, D. E. Competing quantum effects in the dynamics of a flexible water model. *J. Chem. Phys.* **2009**, *131*, 024501.

(47) González, M. A.; Abascal, J. L. F. A flexible model for water based on TIP4P/2005. *J. Chem. Phys.* **2011**, *135*, 224516.

(48) Thole, B. Molecular polarizabilities calculated with a modified dipole interaction. *Chem. Phys.* **1981**, *59*, 341–350.

(49) Burnham, C. J.; Xantheas, S. S. Development of transferable interaction models for water. IV. A flexible all-atom polarizable potential (TTM2-F) based on geometry dependent charges derived from an ab initio monomer dipole moment surface. *J. Chem. Phys.* **2002**, *116*, 5115–5124.

(50) Fanourgakis, G. S.; Xantheas, S. S. The Flexible Polarizable Thole-Type Interaction Potential for Water (TTM2-F) Revisited. *J. Phys. Chem. A* **2006**, *110*, 4100–4106.

(51) Fanourgakis, G. S.; Xantheas, S. S. Development of transferable interaction potentials for water. V. Extension of the flexible polarizable Thole-type model potential (TTM3-F v. 3.0) to describe the vibrational spectra of water clusters and liquid water. *J. Chem. Phys.* **2008**, *128*, 074506.

(52) Misquitta A. J.; Stone A. J. CamCASP: a program for studying intermolecular interactions and for the calculation of molecular properties in distributed form. 2019.

(53) Ren, P.; Ponder, J. W. Polarizable Atomic Multipole Water Model for Molecular Mechanics Simulation. *J. Phys. Chem. B* **2003**, *107*, 5933–5947.

(54) Lennard-Jones, J. E. Cohesion. *Proc. Phys. Soc.* **1931**, *43*, 461.

(55) Schwerdtfeger, P.; Wales, D. J. 100 Years of the Lennard-Jones Potential. *J. Chem. Theory Comput.* **2024**, *20*, 3379–3405.

(56) Jones, J. E.; Chapman, S. On the determination of molecular fields. —II. From the equation of state of a gas. *Proc. R. Soc. Lond. A* **1924**, *106*, 463–477.

(57) Mie, G. Zur kinetischen Theorie der einatomigen Körper. *Ann. Phys.* **1903**, *316*, 657–697.

(58) Buckingham, R. A. The classical equation of state of gaseous helium neon and argon. *Proc. R. Soc. Lond. A* **1938**, *168*, 264–283.

(59) Werhahn, J. C.; Miliordos, E.; Xantheas, S. S. A new variation of the Buckingham exponential-6 potential with a tunable singularity-free short-range repulsion and an adjustable long-range attraction. *Chem. Phys. Lett.* **2015**, *619*, 133–138.

(60) Xantheas, S. S.; Werhahn, J. C. Universal scaling of potential energy functions describing intermolecular interactions. I. Foundations and scalable forms of new generalized Mie Lennard-Jones Morse and Buckingham exponential-6 potentials. *J. Chem. Phys.* **2014**, *141*, 064117.

(61) Babin, V.; Medders, G. R.; Paesani, F. Development of a "First Principles" Water Potential with Flexible Monomers. II: Trimer Potential Energy Surface Third Virial Coefficient and Small Clusters. *J. Theory Comput. Chem.* **2014**, *10*, 1599–1607.

(62) Yu, Q.; Qu, C.; Houston, P. L.; Conte, R.; Nandi, A.; Bowman, J. M. q-AQUA: A Many-Body CCSD(T) Water Potential Including Four-Body Interactions Demonstrates the Quantum Nature of Water from Clusters to the Liquid Phase. *J. Phys. Chem. Lett.* **2022**, *13*, 5068–5074.

(63) Kumar, R.; Schmidt, J. R.; Skinner, J. L. Hydrogen bonding definitions and dynamics in liquid water. *J. Chem. Phys.* **2007**, *126*, 204107.

(64) Jorgensen, W. L. Monte carlo results for hydrogen bond distributions in liquid water. *Chem. Phys. Lett* **1980**, *70*, 326–329.

(65) Luzar, A.; Chandler, D. Effect of Environment on Hydrogen Bond Dynamics in Liquid Water. *Phys. Rev. Lett.* **1996**, *76*, 928–931.

(66) Soper, A.; Phillips, M. A new determination of the structure of water at 25°C. *Chem. Phys.* **1986**, *107*, 47–60.

(67) Teixeira, J.; Bellissent-Funel, M. .-. Dynamics of water studied by neutron scattering. *J. Phys.: Condens. Matter* **1990**, *2*, SA105.

(68) Wernet, P.; Nordlund, D.; Bergmann, U.; Cavalleri, M.; Odelius, M.; Ogasawara, H.; Näslund, L. Å.; Hirsch, T. K.; Ojamäe, L.; Glatzel, P.; Pettersson, L. G. M.; Nilsson, A. The Structure of the First Coordination Shell in Liquid Water. *Science* **2004**, *304*, 995–999.

(69) Sprik, M.; Hutter, J.; Parrinello, M. Ab initio molecular dynamics simulation of liquid water: Comparison of three gradient-corrected density functionals. *J. Chem. Phys.* **1996**, *105*, 1142–1152.

(70) Kuo, I.-F. W.; Mundy, C. J. An ab Initio Molecular Dynamics Study of the Aqueous Liquid-Vapor Interface. *Science* **2004**, *303*, 658–660.

(71) Henchman, R. H.; Irudayam, S. J. Topological Hydrogen-Bond Definition to Characterize the Structure and Dynamics of Liquid Water. *J. Phys. Chem. B* **2010**, *114*, 16792–16810.

(72) Hammerich, A. D.; Buch, V. An alternative near-neighbor definition of hydrogen bonding in water. *J. Chem. Phys.* **2008**, *128*, 111101.

(73) Trivedi, R.; Nagarajaram, H. A. Intrinsically Disordered Proteins: An Overview. *Int. J. Mol. Sci.* **2022**, *23*, 14050.

(74) Merkel, J. S.; Sturtevant, J. M.; Regan, L. Sidechain interactions in parallel β sheets: the energetics of cross-strand pairings. *Structure* **1999**, *7*, 1333–1343.

(75) Lin, Z.; van Gunsteren, W. F. Exploring the Effect of Side-Chain Substitutions upon the Secondary Structure Preferences of β-Peptides. *J. Phys. Chem. B* **2011**, *115*, 12984–12992.

(76) Wang, M.; Song, X.; Chen, J.; Chen, X.; Zhang, X.; Yang, Y.; Liu, Z.; Yao, L. Intracellular environment can change protein conformational dynamics in cells through weak interactions. *Science Advances* **2023**, *9*, eadg9141.

(77) Devi, S.; Chaturvedi, M.; Fatima, S.; Priya, S. Environmental factors modulating protein conformations and their role in protein aggregation diseases. *Toxicology* **2022**, *465*, 153049.

(78) Macdonald, J. R.; Johnson JR., W. C. Environmental features are important in determining protein secondary structure. *Protein Science* **2001**, *10*, 1172–1177.

(79) Zarnecka, J.; Lukac, I.; Messham, S. J.; Hussin, A.; Coppola, F.; Enoch, S. J.; Dossetter, A. G.; Griffen, E. J.; Leach, A. G. Mapping Ligand-Shape Space for Protein–Ligand Systems: Distinguishing Key-in-Lock and Hand-in-Glove Proteins. *J. Chem. Inf. Model.* **2021**, *61*, 1859–1874.

(80) Bulusu, G.; Desiraju, G. R. Strong and Weak Hydrogen Bonds in Protein–Ligand Recognition. *J. Indian Inst. Sci.* **2020**, *100*, 31–41.

(81) Huang, Y.-m. M.; Kang, M.; Chang, C.-e. A. Switches of hydrogen bonds during ligand–protein association processes determine binding kinetics. *J. Mol. Recognit.* **2014**, *27*, 537–548.

(82) Puskar, N. L.; Xiu, X.; Lester, H. A.; Dougherty, D. A. Two Neuronal Nicotinic Acetylcholine Receptors $\alpha_4\beta_4$ and α_7 Show Differential Agonist Binding Modes. *J Biol Chem* **2011**, *286*, 14618–14627.

(83) Lu, Z.; Zhou, N.; Wu, Q.; Zhang, Y. Directional Dependence of Hydrogen Bonds: A Density-Based Energy Decomposition Analysis and Its Implications on Force Field Development. *J. Chem. Theory Comput.* **2011**, *7*, 4038–4049.

(84) Dani, J. A. In *Nicotine Use in Mental Illness and Neurological Disorders*, Biasi, M. D., Ed.; Int. Rev. Neurobiol. Vol. 124; Academic Press: 2015, pp 3–19.

(85) Unwin, N.; Fujiyoshi, Y. Gating Movement of Acetylcholine Receptor Caught by Plunge-Freezing. *J. Mol. Biol.* **2012**, *422*, 617–634.

(86) Hogg, R. C.; Rassenbass, M.; Bertrand, D. Nicotinic acetylcholine receptors: from structure to brain function. *Rev. Physiol. Biochem. Pharmacol.* **2003**, *147*, 1–46.

(87) Mike, A.; Castro, N. G.; Albuquerque, E. X. Choline and acetylcholine have similar kinetic properties of activation and desensitization on the α7 nicotinic receptors in rat hippocampal neurons. *Brain Research* **2000**, *882*, 155–168.

(88) Ma, K.-G.; Qian, Y.-H. Alpha 7 nicotinic acetylcholine receptor and its effects on Alzheimer's disease. *Neuropeptides* **2019**, *73*, 96–106.

(89) Tregellas, J. R.; Wylie, K. P. Alpha7 Nicotinic Receptors as Therapeutic Targets in Schizophrenia. *Nicotine Tob. Res.* **2018**, *21*, 349–356.

(90) Freedman, R. α7-Nicotinic Acetylcholine Receptor Agonists for Cognitive Enhancement in Schizophrenia. *Annu. Rev. Med.* **2014**, *65*, 245–261.

(91) Benowitz, N. L. Pharmacology of Nicotine: Addiction Smoking-Induced Disease and Therapeutics. *Annu. Rev. Pharmacol. Toxicol.* **2009**, *49*, 57–71.

(92) Papke, R. L.; Dwoskin, L. P.; Crooks, P. A. The pharmacological activity of nicotine and nornicotine on nAChRs subtypes: relevance to nicotine dependence and drug discovery. *J. Neurochem.* **2007**, *101*, 160–167.

(93) Morales-Perez, C. L.; Noviello, C. M.; Hibbs, R. E. X-ray structure of the human $\alpha 4\beta 2$ nicotinic receptor. *Nature* **2016**, *538*, 411–415.

(94) An original potentiating mechanism revealed by the cryo-EM structures of the human α7 nicotinic receptor in complex with nanobodies. *Nat Commun* **2023**, *14*, 5964.

(95) Blow, D. M. Structure and mechanism of chymotrypsin. *Acc. Chem. Res.* **1976**, *9*, 145–152.

(96) Appel, W. Chymotrypsin: Molecular and catalytic properties. *Clin Biochem* **1986**, *19*, 317–322.

(97) Xantheas, S. S.; Dunning, T. H. Ab initio studies of cyclic water clusters $(H_2O)_n$ n = 1-–6. I. Optimal structures and vibrational spectra. *J. Chem. Phys.* **1993**, *99*, 8774–8792.

(98) Mukhopadhyay, A.; Cole, W. T.; Saykally, R. J. The water dimer I: Experimental characterization. *Chem. Phys. Lett.* **2015**, *633*, 13–26.

(99) Mukhopadhyay, A.; Xantheas, S. S.; Saykally, R. J. The water dimer II: Theoretical investigations. *Chem. Phys. Lett.* **2018**, *700*, 163–175.

(100) Burnham, C. J.; Xantheas, S. S. Development of transferable interaction models for water. I. Prominent features of the water dimer potential energy surface. *J. Chem. Phys.* **2002**, *116*, 1479–1492.

(101) Rocher-Casterline, B. E.; Ch'ng, L. C.; Mollner, A. K.; Reisler, H. Communication: Determination of the bond dissociation energy (D) of the water dimer $(H_2O)_2$ by velocity map imaging. *J. Chem. Phys.* **2011**, *134*, 211101.

(102) Miliordos, E.; Aprà, E.; Xantheas, S. S. Optimal geometries and harmonic vibrational frequencies of the global minima of water clusters $(H_2O)_n$ n = 2-–6, and several hexamer local minima at the CCSD(T) level of theory. *J. Chem. Phys.* **2013**, *139*, 114302.

(103) Yang, D. C.; Kim, D. Y.; Kim, K. S. Quantum Monte Carlo Study of the Water Dimer Binding Energy and Halogen–π Interactions. *J. Phys. Chem. A* **2019**, *123*, 7785–7791.

(104) Shipman, L. L.; Owicki, J. C.; Scheraga, H. A. Structure energetics and dynamics of the water dimer. *J. Phys. Chem.* **1974**, *78*, 2055–2060.

(105) Dyke, T. R.; Muenter, J. S. Microwave spectrum and structure of hydrogen bonded water dimer. *J. Chem. Phys.* **1974**, *60*, 2929–2930.

(106) Dyke, T. R.; Mack, K. M.; Muenter, J. S. The structure of water dimer from molecular beam electric resonance spectroscopy. *J. Chem. Phys.* **1977**, *66*, 498–510.

(107) Fredin, L.; Nelander, B.; Ribbegrd, G. On the structure of the water dimer. *Chem. Phys. Lett.* **1975**, *36*, 375–376.

(108) Tursi, A. J.; Nixon, E. R. Matrix-Isolation Study of the Water Dimer in Solid Nitrogen. *J. Chem. Phys.* **1970**, *52*, 1521–1528.

(109) Van Thiel, M.; Becker, E. D.; Pimentel, G. C. Infrared Studies of Hydrogen Bonding of Water by the Matrix Isolation Technique. *J. Chem. Phys.* **1957**, *27*, 486–490.

(110) Heindel, J. P.; Xantheas, S. S. The Many-Body Expansion for Aqueous Systems Revisited: I. Water–Water Interactions. *J. Chem. Theory Comput.* **2020**, *16*, 6843–6855.

(111) Xantheas, S. S. Ab initio studies of cyclic water clusters $(H_2O)_n$ n = 1-–6. II. Analysis of many-body interactions. *J. Chem. Phys.* **1994**, *100*, 7523–7534.

(112) Xantheas, S. S. Cooperativity and hydrogen bonding network in water clusters. *Chem. Phys.* **2000**, *258*, 225–231.

(113) Wolke, C. T.; Fournier, J. A.; Dzugan, L. C.; Fagiani, M. R.; Odbadrakh, T. T.; Knorke, H.; Jordan, K. D.; McCoy, A. B.; Asmis, K. R.; Johnson, M. A. Spectroscopic snapshots of the proton-transfer mechanism in water. *Science* **2016**, *354*, 1131–1135.

(114) Boyer, M. A.; Marsalek, O.; Heindel, J. P.; Markland, T. E.; McCoy, A. B.; Xantheas, S. S. Beyond Badger's Rule: The Origins and Generality of the Structure–Spectra Relationship of Aqueous Hydrogen Bonds. *J. Phys. Chem. Lett.* **2019**, *10*, 918–924.

(115) Liu, J.; Yang, J.; Zeng, X. C.; Xantheas, S. S.; Yagi, K.; He, X. Towards complete assignment of the infrared spectrum of the protonated water cluster $H^+(H_2O)_{21}$. *Nat. Commun.* **2021**, *12*, 6141.

(116) Finney, J. M.; Choi, T. H.; Huchmala, R. M.; Heindel, J. P.; Xantheas, S. S.; Jordan, K. D.; McCoy, A. B. Isotope Effects in the Zundel-Eigen Isomerization of $H^+(H_2O)_6$. *J. Phys. Chem. Lett.* **2023**, *14*, 4666–4672.

(117) Hodges, M. P.; Stone, A. J.; Xantheas, S. S. Contribution of Many-Body Terms to the Energy for Small Water Clusters: A Comparison of ab Initio Calculations and Accurate Model Potentials. *J. Phys. Chem. A* **1997**, *101*, 9163–9168.

(118) Milet, A.; Moszynski, R.; Wormer, P. E. S.; van der Avoird, A. Hydrogen Bonding in Water Clusters: Pair and Many-Body Interactions from Symmetry-Adapted Perturbation Theory. *J. Phys. Chem. A* **1999**, *103*, 6811–6819.

(119) Herman, K. M.; Xantheas, S. S. An extensive assessment of the performance of pairwise and many-body interaction potentials in reproducing ab initio benchmark binding energies for water clusters *n*. *Phys. Chem. Chem. Phys.* **2023**, *25*, 7120–7143.

(120) Bernal, J. D.; Fowler, R. H. A Theory of Water and Ionic Solution with Particular Reference to Hydrogen and Hydroxyl Ions. *J. Chem. Phys.* **2004**, *1*, 515–548.

(121) Kuhs, W. F.; Lehmann, M. S. The structure of the ice Ih by neutron diffraction. *J. Phys. Chem.* **1983**, *87*, 4312–4313.

(122) Salzmann, C. G.; Radaelli, P. G.; Slater, B.; Finney, J. L. The polymorphism of ice: five unresolved questions. *Phys. Chem. Chem. Phys.* **2011**, *13*, 18468–18480.

(123) Haynes, W., *CRC Handbook of Chemistry and Physics*; CRC Press: 2016.

(124) Santra, B.; Klimeš, J.; Tkatchenko, A.; Alfè, D.; Slater, B.; Michaelides, A.; Car, R.; Scheffler, M. On the accuracy of van der Waals inclusive density-functional theory exchange-correlation functionals for ice at ambient and high pressures. *J. Chem. Phys.* **2013**, *139*, 154702.

(125) Herman, K. M.; Santis, S. S. Origins of Tetrahedral Order in Ice. *J. Am. Chem. Soc.* **2025**, under revision.

(126) Skinner, L. B.; Benmore, C. J.; Neuefeind, J. C.; Parise, J. B. The structure of water around the compressibility minimum. *J. Chem. Phys.* **2014**, *141*, 214507.

(127) Skinner, L. B.; Huang, C.; Schlesinger, D.; Pettersson, L. G. M.; Nilsson, A.; Benmore, C. J. Benchmark oxygen-oxygen pair-distribution function of ambient water from x-ray diffraction measurements with a wide Q-range. *J. Chem. Phys.* **2013**, *138*, 074506.

(128) Luck, W.; Ditter, W. Band-overlapping and water structure. *J. Mol. Struct.* **1968**, *1*, 339–348.

(129) Tanaka, M.; Girard, G.; Davis, R.; Peuto, A.; Bignell, N. Recommended table for the density of water between 0°C and 40°C based on recent experimental reports. *Metrologia* **2001**, *38*, 301.

(130) Wagner, W.; Pruß, A. The IAPWS Formulation 1995 for the Thermodynamic Properties of Ordinary Water Substance for General and Scientific Use. *J. Phys. Chem. Ref. Data* **2002**, *31*, 387–535.

(131) Heindel, J. P.; Xantheas, S. S. The Many-Body Expansion for Aqueous Systems Revisited: II. Alkali Metal and Halide Ion–Water Interactions. *J. Chem. Theory Comput.* **2021**, *17*, 2200–2216.

(132) Herman, K. M.; Heindel, J. P.; Xantheas, S. S. The many-body expansion for aqueous systems revisited: III. Hofmeister ion–water interactions. *Phys. Chem. Chem. Phys.* **2021**, *23*, 11196–11210.

(133) Luo, Y.; Jiang, W.; Yu, H.; MacKerell, A. D.; Roux, B. Simulation study of ion pairing in concentrated aqueous salt solutions with a polarizable force field. *Faraday Discuss.* **2013**, *160*, 135–149.

(134) Dinpajooh, M.; Intan, N. N.; Duignan, T. T.; Biasin, E.; Fulton, J. L.; Kathmann, S. M.; Schenter, G. K.; Mundy, C. J. Beyond the Debye–Hückel limit: Toward a general theory for concentrated electrolytes. *J. Chem. Phys.* **2024**, *161*, 230901.

(135) Kwak, K.; Jeon, J.; Chun, S. Y.; Cho, M. Ion Networks in Water-based Li-ion Battery Electrolytes. *Acc. Chem. Res.* **2025**, *58*, 199–207.

(136) Schrödinger, E. Quantisierung als Eigenwertproblem. *Annalen der Physik* **1926**, *385*, 437–490.

(137) Born, M.; Oppenheimer, R. Zur Quantentheorie der Molekeln. *Annalen der Physik* **1927**, *389*, 457–484.

(138) Hartree, D. R. The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part I. Theory and Methods. *Math. Proc. Camb. Philos. Soc.* **1928**, *24*, 89–110.

(139) Hartree, D. R. The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part II. Some Results and Discussion. *Math. Proc. Camb. Philos. Soc.* **1928**, *24*, 111–132.

(140) Slater, J. C. A Simplification of the Hartree-Fock Method. *Phys. Rev.* **1951**, *81*, 385–390.

(141) Roothaan, C. C. J. New Developments in Molecular Orbital Theory. *Rev. Mod. Phys.* **1951**, *23*, 69–89.

(142) Hall, G. G.; Lennard-Jones, J. E. The molecular orbital theory of chemical valency VIII. A method of calculating ionization potentials. *Proc. R. Soc. Lond. A* **1951**, *205*, 541–552.

(143) Møller, C.; Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.* **1934**, *46*, 618–622.

(144) Frisch, M. J.; Head-Gordon, M.; Pople, J. A. A Direct MP2 Gradient Method. *Chem. Phys. Lett.* **1990**, *166*, 275–280.

(145) Head-Gordon, M.; Pople, J. A.; Frisch, M. J. MP2 Energy Evaluation by Direct Methods. *Chem. Phys. Lett.* **1988**, *153*, 503–506.

(146) Frisch, M. J.; Head-Gordon, M.; Pople, J. A. Semi-direct Algorithms for the MP2 Energy and Gradient. *Chem. Phys. Lett.* **1990**, *166*, 281–289.

(147) Head-Gordon, M.; Head-Gordon, T. Analytic MP2 frequencies without fifth-order storage. Theory and application to bifurcated hydrogen bonds in the water hexamer. *Chem. Phys. Lett.* **1994**, *220*, 122–128.

(148) Řezáč, J. Non-Covalent Interactions Atlas Benchmark Data Sets: Hydrogen Bonding. *J. Chem. Theory Comput.* **2020**, *16*, 2355–2368.

(149) Johnson, E. R.; DiLabio, G. A. Structure and binding energies in van der Waals dimers: Comparison between density functional theory and correlated ab initio methods. *Chem. Phys. Lett.* **2006**, *419*, 333–339.

(150) Riley, K. E.; Platts, J. A.; Řezáč, J.; Hobza, P.; Hill, J. G. Assessment of the Performance of MP2 and MP2 Variants for the Treatment of Noncovalent Interactions. *J. Phys. Chem. A* **2012**, *116*, 4159–4169.

(151) Frisch, M. J.; Del Bene, J. E.; Binkley, J. S.; Schaefer, H. F. Extensive theoretical studies of the hydrogen-bonded complexes $(H_2O)_2$ $(H_2O)_2H^+$ $(HF)_2$ $(HF)_2H^+$ F_2H^- and $(NH_3)_2$. *J. Chem. Phys.* **1986**, *84*, 2279–2289.

(152) Frey, J. A.; Leutwyler, S. An ab Initio Benchmark Study of Hydrogen Bonded Formamide Dimers. *J. Phys. Chem. A* **2006**, *110*, 12512–12518.

(153) Su, P.; Li, H. Energy decomposition analysis of covalent bonds and intermolecular interactions. *J. Chem. Phys.* **2009**, *131*, 014102.

(154) Kohn, W.; Becke, A. D.; Parr, R. G. Density Functional Theory of Electronic Structure. *J. Chem. Phys.* **1996**, *100*, 12974–12980.

(155) Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140*, A1133–A1138.

(156) Perdew, J. P.; Schmidt, K. Jacob's ladder of density functional approximations for the exchange-correlation energy. *AIP Conf. Proc* **2001**, *577*, 1–20.

(157) Dirac, P. A. M. Note on Exchange Phenomena in the Thomas Atom. *Math. Proc. Camb. Philos. Soc.* **1930**, *26*, 376–385.

(158) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

(159) Grimme, S.; Neese, F. Double-hybrid density functional theory for excited electronic states of molecules. *J. Chem. Phys.* **2007**, *127*, 154116.

(160) Baer, R.; Livshits, E.; Salzner, U. Tuned Range-Separated Hybrids in Density Functional Theory. *Annu. Rev. Phys. Chem.* **2010**, *61*, 85–109.

(161) Grimme, S. Accurate description of van der Waals complexes by density functional theory including empirical corrections. *J. Comput. Chem.* **2004**, *25*, 1463–1473.

(162) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.

(163) Chai, J.-D.; Head-Gordon, M. Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615–6620.

(164) Hamprecht, F. A.; Cohen, A. J.; Tozer, D. J.; Handy, N. C. Development and assessment of new exchange-correlation functionals. *J. Chem. Phys.* **1998**, *109*, 6264–6271.

(165) Chai, J.-D.; Head-Gordon, M. Systematic optimization of long-range corrected hybrid density functionals. *J. Chem. Phys.* **2008**, *128*, 084106.

(166) Morse, P. M. Diatomic Molecules According to the Wave Mechanics. II. Vibrational Levels. *Phys. Rev.* **1929**, *34*, 57–64.

(167) Qadeer, S.; Santis, G. D.; Stinis, P.; Xantheas, S. S. Vibrational levels of a generalized Morse potential. *The Journal of Chemical Physics* **2022**, *157*, 144104.

(168) Scott, A. P.; Radom, L. Harmonic Vibrational Frequencies: An Evaluation of HartreeFock, MøllerPlesset, Quadratic Configuration Interaction, Density Functional Theory, and Semiempirical Scale Factors. *J. Phys. Chem.* **1996**, *100*, 16502–16513.

(169) Boyer, M. A.; McCoy, A. B. A flexible approach to vibrational perturbation theory using sparse matrix methods. *J. Chem. Phys.* **2022**, *156*, 054107.

(170) Barone, V. Anharmonic vibrational properties by a fully automated second-order perturbative approach. *J. Chem. Phys.* **2004**, *122*, 014108.

(171) and, J. K. G. W. Simplification of the molecular vibration-rotation hamiltonian. *Mol. Phys.* **2002**, *100*, 47–54.

(172) Piccardo, M.; Bloino, J.; Barone, V. Generalized vibrational perturbation theory for rotovibrational energies of linear, symmetric and asymmetric tops: Theory, approximations, and automated approaches to deal with medium-to-large molecular systems. *Int. J. Quantum Chem.* **2015**, *115*, 948–982.

(173) Frisch, M. J. et al. Gaussian˜16 Revision C.01, Gaussian Inc. Wallingford CT, 2016.

(174) Aprà, E. et al. NWChem: Past Present and Future. *J. Chem. Phys.* **2020**, *152*, 184102.

(175) Martin, J. M. L.; Lee, T. J.; Taylor, P. R.; François, J.-P. The anharmonic force field of ethylene, C2H4, by means of accurate ab initio calculations. *J. Chem. Phys.* **1995**, *103*, 2589–2602.

(176) Boyer, M. A.; McCoy, A. B. A wave function correction-based approach to the identification of resonances for vibrational perturbation theory. *J. Chem. Phys* **2022**, *157*, 164113.

(177) Reproduced from Santis, G. D.; Xantheas, S. S. Extending Badger's rule I: The relationship between energy and structure in hydrogen bonds. *J. Chem. Phys.* **2025**, *162*, 044106. with the permission of AIP Publishing.

(178) Mollner, A. K.; Casterline, B. E.; Ch'ng, L. C.; Reisler, H. Imaging the State-Specific Vibrational Predissociation of the Ammonia-Water Hydrogen-Bonded Dimer. *J. Phys. Chem. A* **2009**, *113*, 10174–10183.

(179) Case, A. S.; Heid, C. G.; Kable, S. H.; Crim, F. F. Dissociation energy and vibrational predissociation dynamics of the ammonia dimer. *J. Chem. Phys.* **2011**, *135*, 084312.

(180) Bohac, E. J.; Marshall, M. D.; Miller, R. E. Initial state effects in the vibrational predissociation of hydrogen fluoride dimer. *J. Chem. Phys.* **1992**, *96*, 6681–6695.

(181) Kemeys, T. R. Hydrogen bonding in the vapour phase between water and hydrogen fluoride: the infrared spectrum of the 1:1 complex. *Proc. R. Soc. Lond. A* **1975**, *344*, 579–592.

(182) Gaw, J. F.; Yamaguchi, Y.; Vincent, M. A.; Schaefer, H. F. Vibrational frequency shifts in hydrogen-bonded systems: the hydrogen fluoride dimer and trimer. *J. Am. Chem. Soc.* **1984**, *106*, 3133–3138.

(183) Badger, R. M. A Relation Between Internuclear Distances and Bond Force Constants. *J. Chem. Phys.* **1934**, *2*, 128–131.

(184) Badger, R. M. The Relation Between the Internuclear Distances and Force Constants of Molecules and Its Application to Polyatomic Molecules. *J. Chem. Phys.* **1935**, *3*, 710–714.

(185) Badger, R. M.; Bauer, S. H. Spectroscopic Studies of the Hydrogen Bond. II. The Shift of the O–H Vibrational Frequency in the Formation of the Hydrogen Bond. *J. Chem. Phys.* **1937**, *5*, 839–851.

(186) Bellamy, L.; Owen, A. A simple relationship between the infra-red stretching frequencies and the hydrogen bond distances in crystals. *Spectrochim. Acta - A: Mol. Biomol. Spectrosc.* **1969**, *25*, 329–333.

(187) Green, M. T. Application of Badger's Rule to Heme and Non-Heme Iron–Oxygen Bonds: An Examination of Ferryl Protonation States. *J. Am. Chem. Soc.* **2006**, *128*, 1902–1906.

(188) Spaeth, A. D.; Gagnon, N. L.; Dhar, D.; Yee, G. M.; Tolman, W. B. Determination of the Cu(III)–OH Bond Distance by Resonance Raman Spectroscopy Using a Normalized Version of Badger's Rule. *J. Am. Chem. Soc.* **2017**, *139*, 4477–4485.

(189) Iogansen, A. V.; Rassadin, B. V. Dependence of intensity and shift in the infrared bands of the hydroxyl group on hydrogen bonding energy. *J. Appl. Spectosc.* **1969**, *11*, 1318–1325.

(190) Bhatta, R. S.; Iyer, P. P.; Dhinojwala, A.; Tsige, M. A brief review of Badger–Bauer rule and its validation from a first-principles approach. *Mod. Phys. Lett. B.* **2014**, *28*, 1430014.

(191) Hermansson, K. Electric-field effects on the OH vibrational frequency and infrared absorption intensity for water. *J. Chem. Phys.* **1993**, *99*, 861–868.

(192) Virtanen, P.; Gommers, R.; Oliphant, T. E. a. SciPy 1.0: Fundamental Algorithms for Scientific. *Nat. Methods* **2020**, *17*, 261–272.

(193) Wales, D. J.; Doye, J. P. K. Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters Containing up to 110 Atoms. *J. Phys. Chem. A* **1997**, *101*, 5111–5116.

(194) Dunning, T. H. Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron Through Neon and Hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007–1023.

(195) Kendall, R. A.; Dunning, T. H.; Harrison, R. J. Electron Affinities of the First-row Atoms Revisited. Systematic Basis Sets and Wave Functions. *J. Chem. Phys.* **1992**, *96*, 6796–6806.

(196) Rakshit, A.; Bandyopadhyay, P.; Heindel, J. P.; Xantheas, S. S. Atlas of putative minima and low-lying energy networks of water clusters n = 3-25. *J. Chem. Phys.* **2019**, *151*, 214307.

(197) Santis, G. D.; Herman, K. M.; Heindel, J. P.; Xantheas, S. S. Descriptors of water aggregation. *J. Chem. Phys.* **2024**, *160*, 054306.

(198) Reproduced from Santis, G. D.; Xantheas, S. S. Extending Badger's rule: II. The relationship between energy and spectra in hydrogen bonds. *J. Chem. Phys.* **2025**, accepted. with the permission of AIP Publishing.

(199) Koné, M.; Illien, B.; Laurence, C.; Graton, J. Can Quantum-Mechanical Calculations Yield Reasonable Estimates of Hydrogen-Bonding Acceptor Strength? The Case of Hydrogen-Bonded Complexes of Methanol. *J. Phys. Chem. A* **2011**, *115*, 13975–13985.

(200) Drago, R. S.; O'Bryan, N.; Vogel, G. C. A frequency shift-enthalpy correlation for a given donor with various hydrogen-bonding acids. *J. Am. Chem. Soc.* **1970**, *92*, 3924–3929.

(201) Epley, T. D.; Drago, R. S. Calorimetric studies on some hydrogen-bonded adducts. *J. Am. Chem. Soc.* **1967**, *89*, 5770–5773.

(202) Drago, R. S.; Epley, T. D. Enthalpies of hydrogen bonding and changes in hydroxy frequency shifts for a series of adducts with substituted phenols. *J. Am. Chem. Soc.* **1969**, *91*, 2883–2890.

(203) Góralski, P. Hydrogen bonds between cholesterol and oxygen bases: a thermodynamic study. *J. Chem. Soc. Faraday Trans.* **1993**, *89*, 2433–2435.

(204) Joesten, M. D.; Drago, R. S. The Validity of Frequency Shift-Enthalpy Correlations. I. Adducts of Phenol with Nitrogen and Oxygen Donors. *J. Am. Chem. Soc.* **1962**, *84*, 3817–3821.

(205) Rao, C. N. R.; Dwivedi, P. C.; Ratajczak, H.; Orville-Thomas, W. J. Relation between O—H stretching frequency and hydrogen bond energy: re-examination of the Badger–Bauer rule. *J. Chem. Soc. Faraday Trans. 2* **1975**, *71*, 955–966.

(206) Sherry, A. D.; Purcell, K. F. Linear enthalpy-spectral shift correlation for perfluoro-tert-butyl alcohol. *J. Am. Chem. Soc.* **1972**, *94*, 1853–1857.

(207) Iogansen, A. Direct proportionality of the hydrogen bonding energy and the intensification of the stretching ν(XH) vibration in infrared spectra. *Spectrochim. Acta A Mol. Biomol. Spectrosc* **1999**, *55*, 1585–1612.

(208) Jabłoński, M. A Critical Overview of Current Theoretical Methods of Estimating the Energy of Intramolecular Interactions. *Molecules* **2020**, *25*, 5512.

(209) Nikolaienko, T. Y.; Bulavin, L. A.; Hovorun, D. M. Bridging QTAIM with vibrational spectroscopy: the energy of intramolecular hydrogen bonds in DNA-related biomolecules. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7441–7447.

(210) Santis, G. D.; Xantheas, S. S. Extending Badger's rule: I. The relationship between energy and structure in hydrogen bonds. *J. Chem. Phys.* **2025**, *162*, 044106.

(211) Santis, G. D.; Xantheas, S. S. Extending Badger's rule: II. The relationship between energy and spectra in hydrogen bonds. *J. Chem. Phys.* **2025**, accepted.

(212) Zhang, B.; Yang, S.; Huang, Q.-R.; Jiang, S.; Chen, R.; Yang, X.; Zhang, D. H.; Zhang, Z.; Kuo, J.-L.; Jiang, L. Deconstructing Vibrational Motions on the Potential Energy Surfaces of Hydrogen-Bonded Complexes. *CCS Chem.* **2021**, *3*, 829–835.

(213) Ch'ng, L. C.; Samanta, A. K.; Czakó, G.; Bowman, J. M.; Reisler, H. Experimental and Theoretical Investigations of Energy Transfer and Hydrogen-Bond Breaking in the Water Dimer. *J. Am. Chem. Soc.* **2012**, *134*, 15430–15435.

(214) Johnson, G.; Andrews, L. Matrix infrared spectrum of the ammonia-hydrogen fluoride hydrogen-bonded complex. *J. Am. Chem. Soc.* **1982**, *104*, 3043–3047.

(215) Slipchenko, M. N.; Sartakov, B. G.; Vilesov, A. F.; Xantheas, S. S. Study of NH Stretching Vibrations in Small Ammonia Clusters by Infrared Spectroscopy in He Droplets and ab Initio Calculations. *J. Phys. Chem. A* **2007**, *111*, 7460–7471.

(216) Nauta, K.; Miller, R. E. The hydrogen fluoride dimer in liquid helium: A prototype system for studying solvent effects on hydrogen bonding. *J. Chem. Phys.* **2000**, *113*, 10158–10168.

(217) Nelander, B.; Nord, L. Complex between water and ammonia. *J. Phys. Chem.* **1982**, *86*, 4375–4379.

(218) Huang, X.; Braams, B. J.; Bowman, J. M. Ab Initio Potential Energy and Dipole Moment Surfaces of $(H_2O)_2$. *J. Phys. Chem. A* **2006**, *110*, 445–451.

(219) Wang, Y.; Carter, S.; Braams, B. J.; Bowman, J. M. MULTIMODE quantum calculations of intramolecular vibrational energies of the water dimer and trimer using ab initio-based potential energy surfaces. *J. Chem. Phys.* **2008**, *128*, 071101.

(220) Howard, J. C.; Gray, J. L.; Hardwick, A. J.; Nguyen, L. T.; Tschumper, G. S. Getting down to the Fundamentals of Hydrogen Bonding: Anharmonic Vibrational Frequencies of $(HF)_2$ and $(H_2O)_2$ from Ab Initio Electronic Structure Computations. *J. Chem. Theory Comput.* **2014**, *10*, 5426–5435.

(221) Sälli, E.; Salmi, T.; Halonen, L. Computational High-Frequency Overtone Spectra of the Water–Ammonia Complex. *J. Phys. Chem. A* **2011**, *115*, 11594–11605.

(222) Sexton, T. Dissociation Energy of the $H_2O\cdots HF$ Dimer. *J. Phys. Chem. A* **2018**, *122*, 4902–4908.

(223) Brindle, C. A.; Chaban, G. M.; Gerber, R. B.; Janda, K. C. Anharmonic vibrational spectroscopy calculations for $(NH_3)(HF)$ and $(NH_3)(DF)$: fundamental overtone and combination transitions. *Phys. Chem. Chem. Phys.* **2005**, *7*, 945–954.

(224) Lane, J. R.; Vaida, V.; Kjaergaard, H. G. Calculated electronic transitions of the water ammonia complex. *J. Chem. Phys.* **2008**, *128*, 034302.

(225) Viglaska, D.; Wang, X.-G.; Carrington, T.; Tew, D. P. Computational study of the rovibrational spectrum of H_2O-HF. *J. Mol. Spectrosc.* **2022**, *384*, 111587.

(226) Lomboy, A. J. V.; Topper, R. Q. Nonuniform Proton Transfer and Strong Hydrogen Bonding within Cation Anion and Neutral Clusters of Ammonia and Hydrogen Fluoride. *J. Phys. Chem. A* **2021**, *125*, 2546–2557.

(227) Ovsyannikov, R. I.; Makhnev, V. Y.; Zobov, N. F.; Koput, J.; Tennyson, J.; Polyansky, O. L. Highly accurate HF dimer ab initio potential energy surface. *J. Chem. Phys.* **2022**, *156*, 164305.

(228) Legon, A.; Millen, D.; North, H. M. Experimental determination of the dissociation energies D_0 and D_e of $H_2O\ldots HF$. *Chem. Phys. Lett.* **1987**, *135*, 303–306.

(229) Aling, J.; Szalewicz, K.; van der Avoird, A. Ammonia dimer: extremely fluxional but still hydrogen bonded. *Nat. Commun* **2022**, *13*, 1470.

(230) Miliordos, E.; Xantheas, S. S. An accurate and efficient computational protocol for obtaining the complete basis set limits of the binding energies of water clusters at the MP2 and CCSD(T) levels of theory: Application to $(H_2O)_m$ m = 2–6, 8, 11, 16, and 17. *J. Chem. Phys.* **2015**, *142*, 234303.

(231) Curotto, E.; Mella, M. Quantum Monte Carlo simulations of selected ammonia clusters (n = 2-–5): Isotope effects on the ground state of typical hydrogen bonded systems. *J. Chem. Phys.* **2010**, *133*, 214301.

(232) Ho, K.-L.; Lee, L.-Y.; Katada, M.; Fujii, A.; Kuo, J.-L. An ab initio anharmonic approach to study vibrational spectra of small ammonia clusters. *Phys. Chem. Chem. Phys.* **2016**, *18*, 30498–30506.

(233) Dixon, R. N. Anomalous Band Intensity in Fermi Resonance. *J. Chem. Phys.* **1959**, *31*, 258–260.

(234) Pankewitz, T.; Lagutschenkov, A.; Niedner-Schatteburg, G.; Xantheas, S. S.; Lee, Y.-T. Infrared spectrum of $NH_4^+(H_2O)$: Evidence for mode specific fragmentation. *J. Chem. Phys.* **2007**, *126*, 074307.

(235) Chaban, G. M.; Xantheas, S. S.; Gerber, R. B. Anharmonic Vibrational Spectroscopy of the $F^-(H_2O)_n$ Complexes n. *J. Phys. Chem. A* **2003**, *107*, 4952–4956.

(236) Weis, P.; Kemper, P. R.; Bowers, M. T.; Xantheas, S. S. A New Determination of the Fluoride Ion - Water Bond Energy. *J. Am. Chem. Soc.* **1999**, *121*, 3531–3532.

(237) Chaplin, M. Anomalous Properties of Water. *Angew. Chem. Int. Ed.* **2001**, *40*, 1808–1827.

(238) Maibaum, L.; Dinner, A. R.; Chandler, D. Micelle Formation and the Hydrophobic Effect. *J. Phys. Chem. B* **2004**, *108*, 6778–6781.

(239) Kuhn, B.; Mohr, P.; Stahl, M. Intramolecular Hydrogen Bonding in Medicinal Chemistry. *J. Med. Chem.* **2010**, *53*, 2601–2611.

(240) Tsemekhman, K.; Goldschmidt, L.; Eisenberg, D.; Baker, D. Cooperative hydrogen bonding in amyloid formation. *Protein Sci.* **2007**, *16*, 761–764.

(241) Harreld, J. H.; Wong, M. S.; Hansma, P. K.; Morse, D. E.; Stucky, G. D. Self-healing Organosiloxane Materials Containing Reversible and Energy-dispersive Crosslinking Domains, US 6783709B2, 2004.

(242) Wu, D. Y.; Meure, S.; Solomon, D. Self-healing polymeric materials: A review of recent developments. *Prog. Polym Sci.* **2008**, *33*, 479–522.

(243) Buemi, G.; Zuccarello, F. Is the intramolecular hydrogen bond energy valuable from internal rotation barriers? *J. Mol. Struct. THEOCHEM* **2002**, *581*, 71–85.

(244) Bader, R. F. W. A quantum theory of molecular structure and its applications. *Chem. Rev.* **1991**, *91*, 893–928.

(245) Espinosa, E.; Molins, E.; Lecomte, C. Hydrogen bond strengths revealed by topological analyses of experimentally observed electron densities. *Chem. Phys. Lett.* **1998**, *285*, 170–173.

(246) Guevara-Vela, J. M.; Francisco, E.; Rocha-Rinza, T.; Martín Pendás, Á. Interacting Quantum Atoms—A Review. *Molecules* **2020**, *25*, 4028.

(247) Denisov, G. S.; Sheikh-Zade, M. I.; Eskina, M. V. Determination of the energy of an intramolecular hydrogen bond by means of competing equilibria. *J. Appl. Spectrosc.* **1977**, *27*, 1574–1578.

(248) West, A. C.; Schmidt, M. W.; Gordon, M. S.; Ruedenberg, K. A Comprehensive Analysis in Terms of Molecule-Intrinsic Quasi-Atomic Orbitals. II. Strongly Correlated MCSCF Wave Functions. *J. Phys. Chem. A* **2015**, *119*, 10360–10367.

(249) West, A. C.; Schmidt, M. W.; Gordon, M. S.; Ruedenberg, K. A Comprehensive Analysis in Terms of Molecule-Intrinsic Quasi-Atomic Orbitals. III. The Covalent Bonding Structure of Urea. *J. Phys. Chem. A* **2015**, *119*, 10368–10375.

(250) West, A. C.; Schmidt, M. W.; Gordon, M. S.; Ruedenberg, K. A Comprehensive Analysis in Terms of Molecule-Intrinsic Quasi-Atomic Orbitals. IV. Bond Breaking and Bond Forming along the Dissociative Reaction Path of Dioxetane. *J. Phys. Chem. A* **2015**, *119*, 10376–10389.

(251) Galvez Vallejo, J. L.; Heredia, J. D.; Gordon, M. S. Bonding analysis of water clusters using quasi-atomic orbitals. *Phys. Chem. Chem. Phys.* **2021**, *23*, 18734–18743.

(252) Harville, T.; Gordon, M. S. Intramolecular hydrogen bonding analysis. *J. Chem. Phys.* **2022**, *156*, 174302.

(253) Deshmukh, M. M.; Gadre, S. R.; Bartolotti, L. J. Estimation of Intramolecular Hydrogen Bond Energy via Molecular Tailoring Approach. *J. Phys. Chem. A* **2006**, *110*, 12519–12523.

(254) Deshmukh, M. M.; Gadre, S. R. Estimation of N–H$\cdots$O. *J. Phys. Chem. A* **2009**, *113*, 7927–7932.

(255) Korth, H.-G.; de Heer, M. I.; Mulder, P. A DFT Study on Intramolecular Hydrogen Bonding in 2-Substituted Phenols: Conformations Enthalpies and Correlation with Solute Parameters. *J. Phys. Chem. A* **2002**, *106*, 8779–8789.

(256) Lampert, H.; Mikenda, W.; Karpfen, A. Intramolecular Hydrogen Bonding in 2-Hydroxybenzoyl Compounds: Infrared Spectra and Quantum Chemical Calculations. *J. Phys. Chem.* **1996**, *100*, 7418–7425.

(257) Filgueiras, C. A.; Rocha, O. G. An infrared study of hydrogen-bonding between phenols and sulphoxides. *Tetrahedron* **1982**, *38*, 1213–1216.

(258) Abraham, M. H.; Prior, D. V.; Schulz, R. A.; Morris, J. J.; Taylor, P. J. Hydrogen bonding Part 44 Thermodynamics of complexation of 35-dichlorophenol with ketones and ethers in cyclohexane: the Badger–Bauer relationship. *J. Chem. Soc. Faraday Trans.* **1998**, *94*, 879–885.

(259) Sherry, A. D.; Purcell, K. F. Hydrogen bond interactions with sulfur donors. *J. Am. Chem. Soc.* **1972**, *94*, 1848–1853.

(260) Purcell, K. F.; Drago, R. S. Theoretical aspects of the linear enthalpy wavenumber shift relation for hydrogen-bonded phenols. *J. Am. Chem. Soc.* **1967**, *89*, 2874–2879.

(261) Purcell, K. F.; Stikeleather, J. A.; Brunk, S. D. Linear enthalpy-spectral shift correlations for 1,1,1,3,3,3-hexafluoro-2-propanol. *J. Am. Chem. Soc.* **1969**, *91*, 4019–4027.

(262) Ouvrard, C.; Berthelot, M.; Laurence, C. An enthalpic scale of hydrogen-bond basicity part 1: halogenoalkanes. *J. Phys. Org. Chem.* **2001**, *14*, 804–810.

(263) Yurenko, Y. P.; Zhurakivsky, R. O.; Ghomi, M.; Samijlenko, S. P.; Hovorun, D. M. Comprehensive Conformational Analysis of the Nucleoside Analogue 2'-β-Deoxy-6-azacytidine by DFT and MP2 Calculations. *J. Phys. Chem. B* **2007**, *111*, 6263–6271.

(264) Yurenko, Y. P.; Zhurakivsky, R. O.; Ghomi, M.; Samijlenko, S. P.; Hovorun, D. M. How Many Conformers Determine the Thymidine Low-Temperature Matrix Infrared Spectrum? DFT and MP2 Quantum Chemical Study. *The J. Phys. Chem. B* **2007**, *111*, 9655–9663.

(265) Yurenko, Y. P.; Zhurakivsky, R. O.; Ghomi, M.; Samijlenko, S. P.; Hovorun, D. M. Ab Initio Comprehensive Conformational Analysis of 2'-Deoxyuridine the Biologically Significant DNA Minor Nucleoside and Reconstruction of Its Low-Temperature Matrix Infrared Spectrum. *The J. Phys. Chem. B* **2008**, *112*, 1240–1250.

(266) Baker, A. W.; Shulgin, A. T. Intramolecular Hydrogen Bonding. II. The Determination of Hammett Sigma Constants by Intramolecular Hydrogen Bonding in Schiff's Bases. *J. Am. Chem. Soc.* **1959**, *81*, 1523–1529.

(267) Hammett, L. P. The Effect of Structure upon the Reactions of Organic Compounds. Benzene Derivatives. *J. Am. Chem. Soc.* **1937**, *59*, 96–103.

(268) Gu, Q.; Trindle, C.; Knee, J. L. Communication: Frequency shifts of an intramolecular hydrogen bond as a measure of intermolecular hydrogen bond strengths. *J. Chem. Phys.* **2012**, *137*, 091101.

(269) Mills, I. Vibrational Perturbation Theory. *J. Mol. Spectrosc.* **1961**, *5*, 334–340.

(270) Barone, V. Anharmonic vibrational properties by a fully automated second-order perturbative approach. *J. Chem. Phys.* **2005**, *122*, 014108.

(271) Howard, D. L.; Kjaergaard, H. G. Influence of Intramolecular Hydrogen Bond Strength on OH-Stretching Overtones. *J. Phys. Chem. A* **2006**, *110*, 10245–10250.

(272) Lane, J. R.; Schrøder, S. D.; Saunders, G. C.; Kjaergaard, H. G. Intramolecular Hydrogen Bonding in Substituted Aminoalcohols. *J. Phys. Chem. A* **2016**, *120*, 6371–6378.

(273) Boyarkin, O. V. Cold ion spectroscopy for structural identifications of biomolecules. *Int. Rev. Phys. Chem.* **2018**, *37*, 559–606.

(274) Stearns, J. A.; Boyarkin, O. V.; Rizzo, T. R. Spectroscopic Signatures of Gas-Phase Helices: Ac-Phe-$(Ala)_5$-Lys-H_+ and Ac-Phe-$(Ala)_{10}$-Lys-H^+. *J. Am. Chem. Soc.* **2007**, *129*, 13820–13821.

(275) Gerhards, M.; Unterberg, C. Structures of the protected amino acid Ac–Phe–OMe and its dimer: A β-sheet model system in the gas phase. *Phys. Chem. Chem. Phys.* **2002**, *4*, 1760–1765.

(276) Stearns, J. A.; Boyarkin, O. V.; Rizzo, T. R. Effects of N-Terminus Substitution on the Structure and Spectroscopy of Gas-Phase Helices. *CHIMIA* **2008**, *62*, 240.

(277) Klotz, I. M. Solvent water and protein behavior: View through a retroscope. *Protein Sci.* **1993**, *2*, 1992–1999.

(278) Abo-Riziq, A.; Grace, L.; Nir, E.; Kabelac, M.; Hobza, P.; de Vries, M. S. Photochemical selectivity in guanine–cytosine base-pair structures. *Proc. Natl. Acad. Sci.* **2005**, *102*, 20–23.

(279) Nosenko, Y.; Kunitski, M.; Stark, T.; Göbel, M.; Tarakeshwar, P.; Brutschy, B. Vibrational signatures of Watson–Crick base pairing in adenine–thymine mimics. *Phys. Chem. Chem. Phys.* **2013**, *15*, 11520–11530.

(280) Williams, L. D.; Chawla, B.; Shaw, B. R. The hydrogen bonding of cytosine with guanine: Calorimetric and ^{1}H-NMR analysis of the molecular interactions of nucleic acid bases. *Biopolymers* **1987**, *26*, 591–603.

(281) Yanson, I. K.; Teplitsky, A. B.; Sukhodub, L. F. Experimental studies of molecular interactions between nitrogen bases of nucleic acids. *Biopolymers* **1979**, *18*, 1149–1170.

(282) Plützer, C.; Hünig, I.; Kleinermanns, K.; Nir, E.; de Vries, M. S. Pairing of Isolated Nucleobases: Double Resonance Laser Spectroscopy of Adenine–Thymine. *Chem. Phys. Chem.* **2003**, *4*, 838–842.

(283) Alonso, D.; Mondragón, A. Mechanisms of catalytic RNA molecules. *Biochemical Society Transactions* **2021**, *49*, 1529–1535.

(284) Kelleher, P. J.; Johnson, C. J.; Fournier, J. A.; Johnson, M. A.; McCoy, A. B. Persistence of Dual Free Internal Rotation in $NH_4^+(H_2O)\cdot He_n$. *J. Phys. Chem. A* **2015**, *119*, 4170–4176.

(285) Leavitt, C. M.; Wolk, A. B.; Kamrath, M. Z.; Garand, E.; Van Stipdonk, M. J.; Johnson, M. A. Characterizing the Intramolecular H-bond and Secondary Structure in Methylated $GlyGlyH^+$ with H_2 Predissociation Spectroscopy. *J. Am. Soc. Mass Spectrom.* **2011**, *22*, 1941–1952.

(286) Wolke, C. T.; DeBlase, A. F.; Leavitt, C. M.; McCoy, A. B.; Johnson, M. A. Diffuse Vibrational Signature of a Single Proton Embedded in the Oxalate Scaffold $HO_2CCO_2^-$. *J. Phys. Chem. A* **2015**, *119*, 13018–13024.

(287) McCoy, A. B.; Guasco, T. L.; Leavitt, C. M.; Olesen, S. G.; Johnson, M. A. Vibrational manifestations of strong non-Condon effects in the $H_3O^+\cdot X_3$ (X = Ar, N_2, CH_4, H_2O) complexes: A possible explanation for the intensity in the "association band" in the vibrational spectrum of water. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7205–7214.

(288) Santis, G. D.; Herman, K. M.; Heindel, J. P.; Xantheas, S. S. Descriptors of water aggregation, Reproduced from Santis, G. D.; Herman, K. M.; Heindel, J. P.; Xantheas, S. S. Descriptors of water aggregation. *J. Chem. Phys.*, **2024**, *160*, 054306. with the permission of AIP Publishing., 2024.

(289) Steinbach, C.; Andersson, P.; Melzer, M.; Kazimirski, J. K.; Buck, U.; Buch, V. Detection of the book isomer from the OH-stretch spectroscopy of size selected water hexamers. *Phys. Chem. Chem. Phys.* **2004**, *6*, 3320–3324.

(290) Combariza, J. E.; Kestner, N. R.; Jortner, J. Energy- structure relationships for microscopic solvation of anions in water clusters. *J. Chem. Phys.* **1994**, *100*, 2851–2864.

(291) Xantheas, S. S. Quantitative Description of Hydrogen Bonding in Chloride-Water Clusters. *J. Phys. Chem.* **1996**, *100*, 9703–9713.

(292) Wasserman, E.; Rustad, J. R.; Xantheas, S. S. Interaction potential of Al^{3+} in water from first principles calculations. *J. Chem. Phys.* **1997**, *106*, 9769–9780.

(293) Heindel, J. P.; Hao, H.; LaCour, R. A.; Head-Gordon, T. Spontaneous Formation of Hydrogen Peroxide in Water Microdroplets. *J. Phys. Chem. Lett.* **2022**, *13*, 10035–10041.

(294) Yoo, S.; Xantheas, S. S. In *Handbook of Computational Chemistry*, Leszczynski, J., Kaczmarek-Kedziera, A., Puzyn, T., G. Papadopoulos, M., Reis, H., K. Shukla, M., Eds.; Springer International Publishing: 2017, pp 1139–1173.

(295) Lagutschenkov, A.; Fanourgakis, G. S.; Niedner-Schatteburg, G.; Xantheas, S. S. The spectroscopic signature of the all-surface to internally solvated structural transition in water clusters in the *n*. *J. Chem. Phys.* **2005**, *122*, 194310.

(296) Pradzynski, C. C.; Dierking, C. W.; Zurheide, F.; Forck, R. M.; Buck, U.; Zeuch, T.; Xantheas, S. S. Infrared detection of $(H_2O)_{20}$ isomers of exceptional stability: a drop-like and a face-sharing pentagonal prism cluster. *Phys. Chem. Chem. Phys.* **2014**, *16*, 26691–26696.

(297) Bilbrey, J. A.; Heindel, J. P.; Schram, M.; Bandyopadhyay, P.; Xantheas, S. S.; Choudhury, S. A look inside the black box: Using graph-theoretical descriptors to interpret a Continuous-Filter Convolutional Neural Network (CF-CNN) trained on the global and local minimum energy structures of neutral water clusters. *J. Chem. Phys.* **2020**, *153*, 024302.

(298) Doi, H.; Takahashi, K. Z.; Aoyagi, T. Mining of Effective Local Order Parameters to Classify Ice Polymorphs. *J. Phys. Chem. A* **2021**, *125*, 9518–9526.

(299) Geiger, P.; Dellago, C. Neural networks for local structure detection in polymorphic systems. *J. Chem. Phys.* **2013**, *139*, 164105.

(300) Fulford, M.; Salvalaglio, M.; Molteni, C. DeepIce: A Deep Neural Network Approach To Identify Ice and Water Molecules. *J. Chem. Inf. Model.* **2019**, *59*, 2141–2149.

(301) Shi, R.; Tanaka, H. Microscopic structural descriptor of liquid water. *J. Chem. Phys.* **2018**, *148*, 124503.

(302) Faccio, C.; Benzi, M.; Zanetti-Polzi, L.; Daidone, I. Low- and high-density forms of liquid water revealed by a new medium-range order descriptor. *J. Mol. Liq.* **2022**, *355*, 118922.

(303) Kazachenko, S.; Thakkar, A. J. Are there any magic numbers for water nanodroplets $(H_2O)_n$ in the range $36 \leq n \leq 50$? *Mol. Phys.* **2010**, *108*, 2187–2193.

(304) Chau, P.; Hardwick, A. J. A new order parameter for tetrahedral configurations. *Mol. Phys.* **1998**, *93*, 511–518.

(305) Erringon, J.; Debenedetti, P. Relationship between structural order and the anomalies of liquid water. *Nature* **2001**, *406*, 318–321.

(306) Steinhardt, P.; Nelson, D.; Ronchetti, M. Bond-Orientational Order in Liquids and Glasses. *Phys. Rev. B Condens. Matter Mater. Phys.* **1983**, *28*, 784–805.

(307) Ten Wolde, P.-R.; Ruiz-Montero, M. J.; Frenkel, D. Simulation of Homogeneous Crystal Nucleation Close to Coexistence. *Faraday Discuss.* **1996**, *104*, 93–110.

(308) Auer, S.; Frenkel, D. Numerical Simulation of Crystal Nucleation in Colloids. *Adv. Polym. Sci.* **2005**, *173*, 149.

(309) Lechner, W.; Dellago, C. Accurate Determination of Crystal Structures Based on Averaged Local Bond Order Parameters. *J. Chem. Phys.* **2008**, *129*, 114707.

(310) Shiratani, E.; Sasai, M. Molecular scale precursor of the liquid–liquid phase transition of water. *J. Chem. Phys.* **1998**, *108*, 3264–3276.

(311) Yan, Z.; Buldyrev, S. V.; Kumar, P.; Giovambattista, N.; Debenedetti, P. G.; Stanley, H. E. Structure of the first- and second-neighbor shells of simulated water: Quantitative relation to translational and orientational order. *Phys. Rev. E* **2007**, *76*, 051201.

(312) Russo, J.; Tanaka, H. Understanding water's anomalies with locally favoured structures. *Nat. Commun.* **2014**, *5*, 3556.

(313) Monserrat, B.; Brandenburg, J. G.; Engel, E. A.; Cheng, B. Liquid water contains the building blocks of diverse ice phases. *Nat. Commun.* **2022**, *11*, 5757.

(314) Donkor, E. D.; Laio, A.; Hassanali, A. Do Machine-Learning Atomic Descriptors and Order Parameters Tell the Same Story? The Case of Liquid Water. *J. Chem. Theory Comput.* **2023**, *19*, 4596–4605.

(315) Wendler, K.; Thar, J.; Zahn, S.; Kirchner, B. Estimating the Hydrogen Bond Energy. *J. Phys. Chem. A* **2010**, *114*, 9529–9536.

(316) Steiner, T. The Hydrogen Bond in the Solid State. *Angew. Chem. Int. Ed.* **2002**, *41*, 48–76.

(317) Kirov, M. V.; Fanourgakis, G. S.; Xantheas, S. S. Identifying the most stable networks in polyhedral water clusters. *Chem. Phys. Lett.* **2008**, *461*, 180–188.

(318) Morrison, I.; Li, J.-C.; Jenkins, S.; Xantheas, S. S.; Payne, M. C. Ab-Initio Total Energy Studies of the Static and Dynamical Properties of Ice Ih. *J. Phys. Chem. B* **1997**, *101*, 6146–6150.

(319) Duboué-Dijon, E.; Laage, D. Characterization of the Local Structure in Liquid Water by Various Order Parameters. *J. Phys. Chem. B* **2015**, *119*, 8406–8418.

(320) Yang, N.; Duong, C. H.; Kelleher, P. J.; McCoy, A. B.; Johnson, M. A. Deconstructing water's diffuse OH stretching vibrational spectrum with cold clusters. *Science* **2019**, *364*, 275–278.

(321) Wiener, H. Structural Determination of Paraffin Boiling Points. *J. Am. Chem. Soc.* **1947**, *69*, 17–20.

(322) Rouvray, D. H. In Rouvray, D. H., King, R. B., Eds.; Woodhead Publishing: 2002; Chapter The Rich Legacy of Half a Century of the Wiener Index, pp 16–37.

(323) Rahman, A.; Stillinger, F. H. Hydrogen-bond patterns in liquid water. *J. Am. Chem. Soc.* **1973**, *95*, 7943–7948.

(324) King, S. V. Ring Configurations in a Random Network Model of Vitreous Silica. *Nature* **1967**, *213*, 1112–1113.

(325) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.

(326) Katritzky, A. R.; Kuanar, M.; Slavov, S.; Hall, C. D.; Karelson, M.; Kahn, I.; Dobchev, D. A. Quantitative Correlation of Physical and Chemical Properties with Chemical Structure: Utility for Prediction. *Chem. Rev.* **2010**, *110*, 5714–5789.

(327) Karthikeyan, M.; Glen, R. C.; Bender, A. General Melting Point Prediction Based on a Diverse Compound Data Set and Artificial Neural Networks. *J. Chem. Inf. Model.* **2005**, *45*, 581–590.

(328) Gilbert, N.; Mewis, R. E.; Sutcliffe, O. B. Classification of fentanyl analogues through principal component analysis (PCA) and hierarchical clustering of GC–MS data. *Forensic Chem.* **2020**, *21*, 100287.

(329) Chanana, S.; Thomas, C. S.; Zhang, F.; Rajski, S. R.; Bugni, T. S. *hcapca*: Automated Hierarchical Clustering and Principal Component Analysis of Large Metabolomic Datasets in R. *Metabolites* **2020**, *10*, 297.

(330) Blaziak, K.; Tzeli, D.; Xantheas, S. S.; Uggerud, E. The activation of carbon dioxide by first row transition metals (Sc–Zn). *Phys. Chem. Chem. Phys.* **2018**, *20*, 25495–25505.

(331) Benedict, W. S.; N.Gailar; E.K.Plyler Rotation-vibration spectra of deuterated water vapor. *J. Chem. Phys.* **1956**, *24*, 1139.

(332) Baianu, I. C.; Boden, N.; Lightowlers, D.; Mortimer, M. A new approach to the structure of concentrated aqueous electrolyte solutions using pulsed NMR methods. *Chem. Phys. Lett.* **1978**, *54*, 169.

(333) Floriano, M. A.; Klug, D. D.; Whalley, E.; Svensson, E. C.; Sears, V. F.; Hallman, E. D. Direct determination of the intramolecular O–D distance in ice Ih and Ic by neutron diffraction. *Nature* **1987**, *329*, 821.

(334) Mantz, Y. A.; Geiger, F. M.; Molina, L. T.; Molina, M. J.; Trout, B. L. First-principles molecular-dynamics study of surface disordering of the (0001) face of hexagonal ice. *J. Chem. Phys.* **2000**, *113*, 10733.

(335) Jeon, J.; Lefohn, A. E.; Voth, G. A. An improved Polarflex water model. *J. Chem. Phys.* **2003**, *118*, 7504.

(336) Fanourgakis, G. S.; Xantheas, S. S. The bend angle of water in ice Ih and liquid water: The significance of implementing the nonlinear monomer dipole moment surface in classical interaction potentials. *J. Chem. Phys.* **2006**, *124*, 174504.

(337) Imoto, S.; Xantheas, S. S.; Saito, S. Molecular origin of the difference in the HOH bend of the IR spectra between liquid water and ice. *J. Chem. Phys.* **2013**, *138*, 054506.

(338) Kumar, P.; Buldyrev, S. V.; Stanley, H. E. A tetrahedral entropy for water. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 22130–22134.

(339) Bandyopadhyay, D.; Mohan, S.; Ghosh, S. K.; Choudhury, N. Correlation of Structural Order Anomalous Density and Hydrogen Bonding Network of Liquid Water. *J. Phys. Chem. B* **2013**, *117*, 8831–8843.

(340) Gartner III, T. E.; Hunter, K. M.; Lambros, E.; Caruso, A.; Riera, M.; Medders, G. R.; Panagiotopoulos, A. Z.; Debenedetti, P. G.; Paesani, F. Anomalies and Local Structure of Liquid Water from Boiling to the Supercooled Regime as Predicted by the Many-Body MB-pol Model. *J. Phys. Chem. Lett.* **2022**, *13*, 3652–3658.

(341) Keutsch, F. N.; Cruzan, J. D.; Saykally, R. J. The Water Trimer. *Chem. Rev.* **2003**, *103*, 2533–2578.

(342) Wang, Y.; Bowman, J. M. Ab initio potential and dipole moment surfaces for water. II. Local-monomer calculations of the infrared spectra of water clusters. *J. Chem. Phys.* **2011**, *134*, 154510.

(343) Liu, K.; Brown, M. G.; Carter, C.; Saykally, R. J.; Gregory J. K.and Clary, D. C. Characterization of a cage form of the water hexamer. *Nature* **1996**, *381*, 501–503.

(344) Wang, Y.; Babin, V.; Bowman, J. M.; Paesani, F. The Water Hexamer: Cage Prism or Both. Full Dimensional Quantum Simulations Say Both. *J. Am. Chem. Soc.* **2012**, *134*, 11116–11119.

(345) Whalley, E. Energies of the phases of ice at zero temperature and pressure. *J. Chem. Phys.* **1984**, *81*, 4087–4092.

(346) Huš, M.; Urbic, T. Strength of hydrogen bonds of water depends on local environment. *J. Chem. Phys.* **2012**, *136*, 144305.

(347) Yoo, S.; Kirov, M. V.; Xantheas, S. S. Low-Energy Networks of the T-Cage $(H_2O)_24$ Cluster and Their Use in Constructing Periodic Unit Cells of the Structure I (sI) Hydrate Lattice. *J. Am. Chem. Soci.* **2009**, *131*, 7564–7566.

(348) Heindel, J. P.; Kirov, M. V.; Xantheas, S. S. Hydrogen bond arrangements in $(H_2O)_{20\ 24\ 28}$ clathrate hydrate cages: Optimization and many-body analysis. *J. Chem. Phys.* **2022**, *157*, 094301.

(349) Chmiela, S.; Tkatchenko, A.; Sauceda, H. E.; Poltavsky, I.; Schütt, K. T.; Müller, K.-R. Machine learning of accurate energy-conserving molecular force fields. *Sci. Adv.* **2017**, *3*, e1603015.

(350) Schütt, K. T.; Sauceda, H. E.; Kindermans, P.-J.; Tkatchenko, A.; Müller, K.-R. SchNet – A deep learning architecture for molecules and materials. *J. Chem. Phys.* **2018**, *148*, 241722.

(351) Schran, C.; Thiemann, F. L.; Rowe, P.; Müller, E. A.; Marsalek, O.; Michaelides, A. Machine learning potentials for complex aqueous systems made simple. *Proc. Natl. Acad. Sci.* **2021**, *118*, e2110077118.

(352) Yao, S.; Van, R.; Pan, X.; Park, J. H.; Mao, Y.; Pu, J.; Mei, Y.; Shao, Y. Machine learning based implicit solvent model for aqueous-solution alanine dipeptide molecular dynamics simulations. *RSC Adv.* **2023**, *13*, 4565–4577.

(353) Panapitiya, G.; Girard, M.; Hollas, A.; Sepulveda, J.; Murugesan, V.; Wang, W.; Saldanha, E. Evaluation of Deep Learning Architectures for Aqueous Solubility Prediction. *ACS Omega* **2022**, *7*, 15695–15710.

(354) Schroeter, T.; Schwaighofer, A.; Mika, S.; Laak, A. t.; Sulzle, D.; Ganzer, U.; Heinrich, N.; Mller, K.-R. Estimating the Domain of Applicability for Machine Learning QSAR RModels: A Study on Aqueous Solubility of Drug Discovery Molecules. *J. Comput. Aided Mol. Des.* **2007**, *21*, 651–664.

(355) Lishchuk, S. V.; Malomuzh, N. P.; Makhlaichuk, P. V. Contribution of H-bond vibrations to heat capacity of water. *Phys. Lett. A* **2011**, *375*, 2656–2660.

(356) Kell, G. S. Isothermal compressibility of liquid water at 1 atm. *J. Chem. Eng. Data* **1970**, *15*, 119–122.

(357) Brites, C. D. S.; Zhuang, B.; Debasu, M. L.; Ding, D.; Qin, X.; Maturi, F. E.; Lim, W. W. Y.; Soh, D. W.; Rocha, J.; Yi, Z.; Liu, X.; Carlos, L. D. Decoding a Percolation Phase Transition of Water at ∼330 K with a Nanoparticle Ruler. *J. Phys. Chem. Lett.* **2020**, *11*, 6704–6711.

(358) Suresh, S. J.; Naik, V. M. Hydrogen bond thermodynamic properties of water from dielectric constant data. *J. Chem. Phys.* **2000**, *113*, 9727–9732.

(359) Luzar, A.; Stefan, J. Dielectric behaviour of DMSO-water mixtures. A hydrogen-bonding model. *J. Mol. Liq.* **1990**, *46*, 221–238.

(360) Zhao, L.; Ma, K.; Yang, Z. Changes of Water Hydrogen Bond Network with Different Externalities. *Int. J. Mol. Sci.* **2015**, *16*, 8454–8489.

(361) Reddy, S. K.; Straight, S. C.; Bajaj, P.; Huy Pham, C.; Riera, M.; Moberg, D. R.; Morales, M. A.; Knight, C.; Götz, A. W.; Paesani, F. On the accuracy of the MB-pol many-body potential for water: Interaction energies vibrational frequencies and classical thermodynamic and dynamical properties from clusters to liquid water and ice. *J. Chem. Phys.* **2016**, *145*, 194504.

(362) Vega, C.; Conde, M. M.; McBride, C.; Abascal, J. L. F.; Noya, E. G.; Ramirez, R.; Sesé, L. M. Heat capacity of water: A signature of nuclear quantum effects. *J. Chem. Phys.* **2010**, *132*, 046101.

(363) Shiga, M.; Shinoda, W. Calculation of heat capacities of light and heavy water by path-integral molecular dynamics. *J. Chem. Phys.* **2005**, *123*, 134502.

(364) Gasparotto, P.; Hassanali, A. A.; Ceriotti, M. Probing Defects and Correlations in the Hydrogen-Bond Network of ab Initio Water. *J. Chem. Theory Comput.* **2016**, *12*, 1953–1964.

(365) Reprinted with permission from Santis, G. D.; Okura, Y.; Hirata, K.; Ishiuchi, S.; Fujii, M.; Xantheas S. S. Affinity of Nicotinoids to a Model Nicotinic Acetylcholine Receptor. *J. Phys. Chem. B.* **2024**, *128*, 4577-4789. Copyright 2024 American Chemical Society.

(366) Griffiths, R. R.; Bigelow, G. E.; Liebson, I. Facilitation of Human Tobacco Self-Administration by Ethanol: A Behavioral Analysis. *J. Exp. Anal. Behav.* **1976**, *25*, 279–292.

(367) Harris, A. C.; Tally, L.; Muelken, P.; Banal, A.; Schmidt, C. E.; Cao, Q.; LeSage, M. G. Effects of nicotine and minor tobacco alkaloids on intracranial-self-stimulation in rats. *Drug Alcohol Depend.* **2015**, *153*, 330–334.

(368) US Food and Drug Administation. Results from the Annual National Youth Tobacco Survey `https://www.fda.gov/tobacco-products/public-health-education/youth-and-tobaccwebo` (accessed Oct 19, 2023).

(369) U.S. Department of Health and Human Services. Patterns of E-Cigarette Use Among U.S. Youth and Young Adults. In *E-Cigarette Use Among Youth and Young Adults: A Report of the Surgeon General.* U.S. Department of Health and Human Services.

(370) Federal Register. Product Standard for Nicotine Level of Combusted Cigarettes. `https://www.federalregister.gov/documents/2018/03/16/2018-05345/tobacco- product-standard-for-nicotine-level-of-combusted-cigarettes` (accessed Oct 19, 2023).

(371) US Food and Drug Administation. Announces Plans for Proposed Rule to Reduce Addictiveness of Cigarettes and Other Combusted Tobacco Products. `https://www.fda.gov/news-events/press-announcements/fda-announces-plans-proposed-rule- reduce-addictiveness-cigarettes-and-other-combusted-tobacco` (accessed Oct 19, 2023).

(372) Dani, J. A.; Biasi, M. D. Cellular mechanisms of nicotine addiction. *Pharmacol. Biochem. Behav.* **2001**, *70*, 439–446.

(373) Amiri, S.; Sansom, M. S.; Biggin, P. C. Molecular dynamics studies of AChBP with nicotine and carbamylcholine: the role of water in the binding pocket. *Protein Eng. Des. Sel.* **2007**, *20*, 353–359.

(374) Brejc, K.; van Dijk, W. J.; Klaassen, R. V.; Schuurmans, M.; van der Oost, J.; B., S. A.; Sixma, T. K. Crystal structure of an ACh-binding protein reveals the ligand-binding domain of nicotinic receptors. *Nature* **2001**, *411*, 269–276.

(375) Zhao, Y.; Liu, S.; Zhou, Y.; Zhang, M.; Chen, H.; Xu, H. E.; Sun, D.; Liu, L.; Tian, C. Structural basis of human α7 nicotinic acetylcholine receptor activation. *Cell Res.* **2021**, *31*, 713–716.

(376) Gharpure, A.; Teng, J.; Zhuang, Y.; Noviello, C. M.; Walsh, R. M.; Cabuco, R.; Howard, R. J.; Zaveri, N. T.; Lindahl, E.; Hibbs, R. E. Agonist Selectivity and Ion Permeation in the $\alpha3\beta4$ Ganglionic Nicotinic Receptor. *Neuron* **2019**, *104*, 501–511.e6.

(377) Walsh Jr., R. M.; Roh, S.-H.; Gharpure, A.; Morales-Perez, C. L.; Ryan E. Hibbs, J. T. Structural principles of distinct assemblies of the human $\alpha4\beta2$ nicotinic receptor. *Nature* **2018**, *557*, 261–265.

(378) Graton, J.; Berthelot, M.; Gal, J.-F.; Girard, S.; Laurence, C.; Lebreton, J.; Le Questel, J.-Y.; Maria, P.-C.; Nauš, P. Site of Protonation of Nicotine and Nornicotine in the Gas Phase: Pyridine or Pyrrolidine Nitrogen? *J. Am. Chem. Soc.* **2002**, *124*, 10552–10562.

(379) Yoshida, T.; Farone, W. A.; Xantheas, S. S. Isomers and Conformational Barriers of Gas-Phase Nicotine Nornicotine and Their Protonated Forms. *J. Chem. Phys. B* **2014**, *118*, 8273–8285.

(380) Santis, G. D.; Takeda, N.; Hirata, K.; Tsuruta, K.; Ishiuchi, S.; Xantheas, S. S.; Fujii, M. Structure of Gas Phase Monohydrated Nicotine: Implications for Nicotine's Native Structure in the Acetylcholine Binding Protein. *J. Am. Chem. Soc.* **2022**, *144*, 16698–16702.

(381) Clayton, P. M.; Vas, C. A.; Bui, T. T. T.; Drake, A. F.; McAdam, K. Spectroscopic investigations into the acid–base properties of nicotine at different temperatures. *Anal. Methods* **2013**, *5*, 81–88.

(382) Linnell, R. H. Anabasine. *J. Am. Chem. Soc.* **1954**, *76*, 1391–1393.

(383) Chen, B.; Xu, Q. Protonating and determining myosmine intactly by association with citrate anion. *Analyst* **2011**, *136*, 4846–4854.

(384) Takeda, N.; Hirata, K.; Tsuruta, K.; Santis, G. D.; Xantheas, S. S.; Ishiuchi, S.; Fujii, M. Gas phase protonated nicotine is a mixture of pyridine- and pyrrolidine-protonated conformers: implications for its native structure in the nicotinic acetylcholine receptor. *Phys. Chem. Chem. Phys.* **2022**, *24*, 5786–5793.

(385) Zoltewicz, J. A.; Bloom, L. B.; Kem, W. R. Quantitative determination of the ring-chain hydrolysis equilibrium constant for anabaseine and related tobacco alkaloids. *J. Org. Chem.* **1989**, *54*, 4462–4468.

(386) Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; DeFrees, D. J.; Pople, J. A. Self-consistent molecular orbital methods. XXIII. A polarization-type basis set for second-row elements. *J. Chem. Phys.* **1982**, *77*, 3654–3665.

(387) Hariharan, P. C.; Pople, J. A. The influence of polarization functions on molecular orbital hydrogenation energies. *Theoret. Chim. Acta* **1973**, *28*, 213–222.

(388) Gordon, M. S.; Binkley, J. S.; Pople, J. A.; Pietro, W. J.; Hehre, W. J. Self-consistent molecular-orbital methods. 22. Small split-valence basis sets for second-row elements. *J. Am. Chem. Soc.* **1982**, *104*, 2797–2803.

(389) Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. V. R. Efficient diffuse function-augmented basis sets for anion calculations. III. The 3-21+G basis set for first-row elements Li–F. *J. Comput. Chem.* **1983**, *4*, 294–301.

(390) Tomasi, J.; Mennucci, B.; Cammi, R. Quantum Mechanical Continuum Solvation Models. *Chem. Rev.* **2005**, *105*, 2999–3094.

(391) Remya, K.; Suresh, C. H. Which density functional is close to CCSD accuracy to describe geometry and interaction energy of small noncovalent dimers? A benchmark study using Gaussian09. *J Comput Chem* **2013**, *34*, 1341–1353.

(392) Herman, K. M.; Aprà, E.; Xantheas, S. S. A critical comparison of CH$\cdots\pi$ versus $\pi\cdots\pi$ interactions in the benzene dimer: obtaining benchmarks at the CCSD(T) level and assessing the accuracy of lower scaling methods. *Phys. Chem. Chem. Phys.* **2023**, *25*, 4824–4838.

(393) Morales-Perez C. L.; Noviello C.M.; Hibbs R. E. X-ray structure of the human $\alpha4\beta2$ nicotnic receptor (2016) https://doi.org/10.2210/pdb5KXI/pdb (accessed Aug 17, 2023).

(394) Negus, S. S.; Miller, L. L.; Nader, M. A. Intracranial Self-Stimulation to Evaluate Abuse Potential of Drugs. *Pharmacol. Rev.* **2014**, *66*, 869–917.

(395) Roberts, F.; Tesman, B. U., *Applied Combinatorics*, 2nd; Taylor & Francis: 2009.

(396) Hankins, D.; Moskowitz, J. W.; Stillinger, F. H. Water Molecule Interactions. *J. Chem. Phys.* **1970**, *53*, 4544–4554.

(397) Reed, A. E.; Weinstock, R. B.; Weinhold, F. Natural Population Analysis. *J. Chem. Phys.* **1985**, *83*, 735–746.

(398) Mautner, M. The ionic hydrogen bond and ion solvation. 1. $NH^+\cdots O$ $NH^+\cdots N$ and $OH^+\cdots O$ bonds. Correlations with proton affinity. Deviations due to structural effects. *J. Am. Chem. Soc.* **1984**, *106*, 1257–1264.

(399) Okura, Y.; Hallsten, M. N.; Santis, G. D.; Hirata, K.; Fujii, M.; Ishiuchi, S.; Xantheas, S. S. The Nicotine–Tryptophan Dimer: Probing the Principal Interaction of Nicotine with the Nicotinic Acetylcholine Receptor (nAChR) the Binding Pocket in the Human Brain. *J. Am. Chem. Soc.* **2025**, *147*, 14043–14047.

(400) Reprinted with permission from Okura, Y; Santis, G. D.; Hirata, K.; Xantheas, S. S.; Fujii, M.; Ishiuchi, S. The Gas Phase Protonation Sites of Six Naturally Occuring Nicotinoids. *J. Phys. Chem. Lett.* **2024**, *15*, 6966-6973. Copyright 2024 American Chemical Society.

(401) Gaigeot, M.-P.; Cimas, A.; Seydou, M.; Kim, J.-Y.; Lee, S.; Schermann, J.-P. Proton Transfer from the Inactive Gas-Phase Nicotine Structure to the Bioactive Aqueous-Phase Structure. *J. Am. Chem. Soc.* **2010**, *132*, 18067–18077.

(402) Garand, E. Spectroscopy of Reactive Complexes and Solvated Clusters: A Bottom-Up Approach Using Cryogenic Ion Traps. *J. PHys. Chem. A* **2018**, *122*, 6479–6490.

(403) Redwine, J. G.; Davis, Z. A.; Burke, N. L.; Oglesbee, R. A.; McLuckey, S. A.; Zwier, T. S. A Novel Ion Trap Based Tandem Mass Spectrometer for the Spectroscopic Study of Cold Gas Phase Polyatomic Ions. *Int. J. Mass Spectrom.* **2013**, *348*, 9–14.

(404) Stearns, J. A.; Mercier, S.; Seaiby, C.; Guidi, M.; Boyarkin, O. V.; Rizzo, T. R. Conformation-Specific Spectroscopy and Photodissociation of Cold Protonated Tyrosine and Phenylalanine. *J. Am. Chem. Soc.* **2007**, *129*, 11814–11820.

(405) Kamrath, M. Z.; Relph, R. A.; Guasco, T. L.; Leavitt, C. M.; Johnson, M. A. Vibrational Predissociation Spectroscopy of the H_2-tagged Mono- and Dicarboxylate Anions of Dodecanedioic Acid. *Int. J. Mass Spectrom.* **2011**, *300*, 91–98.

(406) Hirata, K.; Fuad, H.; Otto, D.; Ishiuchi, S.; Fujii, M. Collision-assisted Stripping for Determination of Microsolvation-dependent Protonation Sites in Hydrated Clusters by Cryogenic Ion Trap Infrared Spectroscopy: the Case of Benzocaine$H^+(H_2O)_n$. *Phys. Chem. Chem. Phys.* **2022**, *24*, 5774–5779.

(407) Ishiuchi, S.; Wako, H.; Kato, D.; Fujii, M. High-cooling-efficiency Cryogenic Quadrupole Ion Trap and UV-UV Hole Burning Spectroscopy of Protonated Tyrosine. *J. Mol. Spectrosc.* **2017**, *332*, Molecular Spectroscopy in Traps, 45–51.

(408) Otsuka, R.; Hirata, K.; Sasaki, Y.; Lisy, J. M.; Ishiuchi, S.; Fujii, M. Alkali and Alkaline Earth Metal Ions Complexes with a Partial Peptide of the Selectivity Filter in K^+ Channels Studied by a Cold Ion Trap Infrared Spectroscopy. *Chem. Phys. Chem.* **2020**, *21*, 712–724.

(409) Tamura, M.; Sekiguchi, T.; Ishiuchi, S.; Zehnacker-Rentien, A.; Fujii, M. Can the Partial Peptide SIVSF of the β2-Adrenergic Receptor Recognize Chirality of the Epinephrine Neurotransmitter? *J. Phys. Chem. Lett.* **2019**, *10*, 2470–2474.

(410) Klopper, W.; Noga, J. An Explicitly Correlated Coupled Cluster Calculation of the Helium–Helium Interatomic Potential. *J. Chem. Phys.* **1995**, *103*, 6127–6132.

(411) Xantheas, S. S. On the Importance of the Fragment Relaxation Energy Terms in the Estimation of the Basis Set Superposition Error Correction to the Intermolecular Interaction Energy. *J. Chem. Phys.* **1996**, *104*, 8821–8824.

(412) Reprinted (adapted) with permission from Okura, Y.; Santis, G. D.; Hirata, K.; Melissas, V. S.; Ishiuchi, S.; Fujii, M.; Xantheas S. S. Switching of Protonation Sites in Hydrated Nicotine via a Grotthuss Mechanism. *J. Am. Chem. Soc.*, **2024**, *146*, 3023-3030. Copyright 2024 American Chemical Society.

(413) Popov, I.; Zhu, Z.; Young-Gonzales, A. R.; Sacci, R. L.; Mamontov, E.; Gainaru, C.; J., P. S.; Sokolov, A. P. Search for a Grotthuss Mechanism through the Observation of Proton Transfer. *Commun. Chem.* **2023**, *6*, 1–10.

(414) Cukierman, S. Et tu Grotthuss! and other unfinished stories. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **2006**, *1757*, Proton Transfer Reactions in Biological Systems, 876–885.

(415) Zeng, H. J.; Johnson, M. A. Demystifying the Diffuse Vibrational Spectrum of Aqueous Protons Through Cold Cluster Spectroscopy. *Annu. Rev. Phys. Chem.* **2021**, *72*, 667–691.

(416) Fukui, K. The path of chemical reactions - the IRC approach. *Acc. Chem. Res.* **1981**, *14*, 363–368.

(417) Atkins, P.; de Paula, J. In *Physical Chemistry: Thermodynamcis Structure and Change*; Oxford University Press: 2014; Chapter Transtion-State Theory, pp 895–903.

(418) Chorkendorf, I.; Niemantsverdriet, J. W. In *Concepts of Modern Catalysis and Kinetics*; Wiley-VCH: 2017; Chapter Transition State Theory, pp 107–113.

(419) Truhlar, D. G.; Garrett, B. C.; Klippenstein, S. J. Current Status of Transition-State Theory. *J. Phys. Chem.* **1996**, *100*, 12771–12800.

(420) McQuarrie, D. A. In *Statistical Mechanics*; University Science Books: 2000; Chapter Chemical Equilibrium, pp 142–159.

(421) Kreuer, K.-D.; Rabenau, A.; Weppner, W. Vehicle Mechanism A New Model for the Interpretation of the Conductivity of Fast Proton Conductors. *Angew. Chem. Int. Ed. Engl.* **1982**, *21*, 208–209.

(422) Ludueña, G. A.; K, T. D.; Sebastiani, D. Mixed Grotthuss and Vehicle Transport Mechanism in Proton Conducting Polymers from Ab initio Molecular Dynamics Simulations. *Chemistry of Materials* **2011**, *23*, 1424–1429.

(423) De Grotthuss, C. Sur la décomposition de l'eau et des corps qu'elle tient en dissolution à l'aide de l'électricité galvanique. *Ann. Chim* **1806**, *58*, 54.

(424) Marlton, S. J. P.; McKinnon, B. I.; Ucur, B.; Maccarone, A. T.; Donald, W. A.; Blanksby, S. J.; Trevitt, A. J. Selecting and identifying gas-phase protonation isomers of nicotineH^+ using combined laser ion mobility and mass spectrometry techniques. *Faraday Discuss.* **2019**, *217*, 453–475.

(425) Zundel, G.; Metzger, H. Energiebänder der tunnelnden Überschuß-Protonen in flüssigen Säuren. Eine IR-spektroskopische Untersuchung der Natur der Gruppierungen $H_5O_2^+$. *Z. Phys. Chem.* **1968**, *58*, 225–245.

(426) Eigen, M. Proton Transfer Acid-Base Catalysis and Enzymatic Hydrolysis. Part I: ELEMENTARY PROCESSES. *Angew. Chem. Int. Ed. Engl.* **1964**, *3*, 1–19.

(427) Bao, J. L.; Truhlar, D. G. Variational transition state theory: theoretical framework and recent developments. *Chem. Soc. Rev.* **2017**, *46*, 7548–7596.